Equating Forms E and F of the P&P-GATB

Daniel O. Segall
Reynaldo I. Monzon

Navy Personnel Research and Development Center

December 8, 1995

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.

Contents

List of Tables

Chapter 1
INTRODUCTION

In response to a review of the General Aptitude Test Battery (GATB) by the National Research Council of the National Academy of Sciences (Hartigan & Wigdor, 1989) two new forms are under development. Changes in the new forms are intended to address concerns raised by the National Academy of Sciences. These changes address such issues as test security, test aesthetics, test speededness, and susceptibility to coaching. Specific changes to the battery include: (1) an increase in testing time (and a reduction of the number of items) for three power tests; (2) reducing the number of items for some tests (without changing the time-limits); (3) inclusion of a correction for guessing for the four speeded tests; (4) changes in test instructions; (5) a change in test administration sequence; (6) modifications to question format for some tests; (7) modifications to overall test appearance; (8) a change in answer sheet format, (9) the deletion of the Form Matching subtest, and renaming the GATB.

In response to these changes, there are several concerns about the use and meaning of the scores obtained from these new forms. One concern is that the distribution of scores obtained from the new forms will differ systematically from the distribution of scores obtained from the old forms. For example, changes in test time-limits and test-lengths are likely to affect the mean and standard deviation (and possibly higher moments) of the test score distributions. Decisions and interpretations made relative to the scale of the old forms will almost surely not be appropriate for individuals taking the new forms. For example scores may be systematically higher or lower on the new forms relative to the old forms. These expected differences caused by the modifications to the GATB support the need for an equating of the old and new forms to help provide score exchangeability.

A second issue in the evaluation of the new forms concerns precision. Several of the new test versions have fewer numbers of items than their original counterparts. Although fewer items may be offset by an increase in testing time, it is important to show that the new forms have sufficiently high levels of reliability relative to the old GATB forms. Lower reliability would lead to lower levels of validity.

The third issue in the evaluation of the new forms concerns construct validity. It is highly desirable for the new and old GATB forms to measure identical or highly correlated constructs. The measurement of similar constructs would enable the validity of the new forms to be inferred from the large body of existing validity research conducted on the old forms of the GATB. The deletion of the Form Matching subtest (from the new Forms E and F) raises an additional issue concerning the equivalency of the constructs measured by the new and old batteries.

This report provides a description of the data collection and analyses for the equating study of the new paper-and-pencil (P&P) GATB forms. The report addresses issues and analyses covering equating, reliability, and construct validity.

Chapter 2
GATB COMPOSITION

Two new forms of the P&P-GATB (designated as Forms E and F) will be developed in two phases. In the first phase, the six non-psychomotor subtests of Forms E and F will be developed, equated, and evaluated. This report addresses the equating data collection and analyses associated with this phase of the non-psychomotor development and evaluation.

In the second phase of the form development, new versions of the five psychomotor subtests will be developed for Forms E and F. These new psychomotor subtests will be equated and evaluated from a separate data collection and analysis effort which is not addressed in this report. It should be noted that equating the psychomotor and non-psychomotor tests in two separate data collection studies places several constraints on the equating and evaluation effort. These constraints limit the inferences that can be made concerning both, the construct validity, and the distribution of composite scores (for those composites including psychomotor subtests). A separate sample consisting of 538 examinees was collected to provide partial answers to these issues. This sample combined the new Form F (non-psychomotor subtests) with Form A (psychomotor and non-psychomotor subtests).

2.1 Forms A–D

The subtests of GATB Forms A–D, and their order of administration are provided in Table 2.1. Also provided in Table 2.1 are test-lengths, time-limits, and an indication of each test's speededness.

Table 2.1 GATB Composition (Forms A–D)

Subtest	Order	Time-Limit	Test-Length	Power/Speeded
Name Comparison (NC)	1	6 min.	150	Speeded
Computation (CO)	2	6 min.	50	Speeded
Three-Dimensional Space (3D)	3	6 min.	40	Power
Vocabulary (VO)	4	6 min.	60	Power
Tool Matching (TM)	5	5 min.	49	Speeded
Arithmetic Reasoning (AR)	6	7 min.	25	Power
Form Matching (FM)	7	6 min.	60	Speeded

Table 2.2 GATB Composition (Forms E and F)

Subtest	Order	Time-Limit	Test-Length	Power/Speeded
Arithmetic Reasoning (AR)	1	20 min.	18	Power
Vocabulary (VO)	2	8 min.	19	Power
Three-Dimensional Space (3D)	3	8 min.	20	Power
Computation (CO)	4	6 min.	40	Speeded
Name Comparison (NC)	5	6 min.	90	Speeded
Object (Tool) Matching (OM)	6	5 min.	42	Speeded

2.2 Forms E and F

The subtests of GATB Forms E and F, and their order of administration are provided in Table 2.2. Additional information about test-lengths and time-limits is also provided.

2.3 Summary of Changes

To address concerns raised by the National Academy of Sciences, several changes were made to the new Forms E and F. These changes address criticisms of the old Forms A–D in the areas of test security, test aesthetics, test speededness, and susceptibility to coaching. Major changes are outlined below.

2.3.1 Power Subtests

In Forms A–D, these subtests have tended to possess a speeded component. In Forms E and F, an effort has been made to make the power tests less speeded by increasing the time-limits and decreasing the test-length:

- *Arithmetic Reasoning.* Testing time increased from 7 minutes to 18 minutes. Number of items decreased from 25 to 18.
- *Three Dimensional Space.* Testing time increased from 6 minutes to 8 minutes. Number of items decreased from 40 to 20.
- *Vocabulary.* Testing time increased from 6 minutes to 8 minutes. Number of items decreased from 60 to 19.

2.3.2 Speeded Subtests

The number of items included in speeded subtests have been slightly reduced in Forms E and F (compared to Forms A–D). The total number of items for each of these subtests has been set equal to the number reached by examinees scoring in the 99th percentile of the GATB (based on a sample of over 18,000 respondents). No change in testing times have been made:

- *Name Comparison.* Number of items were reduced from 150 to 90.
- *Computation.* Number of items were reduced from 50 to 40.

- *Tool Matching.* Name changed to Object Matching. Number of items were reduced from 49 to 42.
- *Form Matching.* Dropped from Forms E and F. This subtests was dropped because of concerns over maintaining equivalence with the CAT-GATB. (The item types of this subtest are incompatible with computer display.)

2.3.3 Speeded Subtest Scoring

To help reduce the influence of coaching and random guessing, a correction for guessing will be used when scoring the three GATB speeded subtests of Forms E and F.

2.3.4 Instructions

Instructions for Forms E and F have been modified to ensure that examinees receive information regarding the optimum test taking strategy. The instructions for each subtest include information about guessing, and the appropriate pace that should be followed when answering questions. Speeded and power tests have different sets of instructions.

2.3.5 Subtest Administration Order

In older GATB Forms A–D, power tests were mixed among the speeded tests (see Table 2.1). In the new Forms E and F, the three power tests are administered at the beginning of the battery, followed by the three speeded subtests (see Table 2.2).

2.3.6 Question Format

Slight modifications were made to question formats. These include presenting math problems in a column format, placing mathematical symbols in the actual items rather than above them, and updating the terminology used in some test items.

2.3.7 Test Appearance

The overall appearance of the GATB subtests has been improved in Forms E and F. Efforts were made to make the test items easier to read and follow. Changes include larger fonts, cleaner printing of items, more liberal use of white space within and between items, and presenting items in columns down the page, rather than in rows across the page.

2.3.8 Answer Sheet Format

The format of the answer sheet has been changed for Forms E and F. The primary changes include making answer bubbles oval rather than round, and starting answer bubbles in the same location in each of the answer sheet sections.

Chapter 3
DATA COLLECTION DESIGN AND PROCEDURES

The data collection design is presented in three sections, each section corresponding to one of the three primary samples included in the GATB equating study. Section 3.1 provides a description of the data-collection design for the *independent-groups* sample. This sample was used to equate the new and old GATB forms. Section 3.2 provides a description of the data-collection design for the *repeated-measures* sample. This sample was used primarily for comparing the reliability and construct validity of the new and old forms. However, a portion of this sample was used as supplemental data for the equating analysis. Section 3.3 provides a description of the data-collection design for the *psychomotor* sample. This sample was used to examine the need for composite equatings, and to examine construct validity issues involving psychomotor subtests.

3.1 Independent-Groups Sample

Examinees contained in the independent-groups sample were randomly assigned to one of three Forms: A, E, and F. As indicated in Table 3.1, a total of 5892 examinees were tested. Approximately equal numbers of examinees were tested on each of the three forms ($N \approx 1964$). Table 3.1 also displays the numbers of examinees tested on each form at each of the five ARDC's.

Across each of the five ARDC's, there were a total of approximately 40 testing sites. At each site, examinees were randomly assigned to test form (A, E, or F). The assignment of test-forms to examinees was complicated by the fact that the old (A) and new (E and F) forms of the GATB possess different subtest ordering, time-limits, and instructions. Consequently these versions can not be administered to a single group simultaneously. They must be administered in different testing sessions, where the sessions are separated physically by either location (testing room) or by time. Consequently, at a given testing

Table 3.1 Independent-Groups Sample Sizes

ARDC	Form A	Form E	Form F	Total
EARDC	447	370	389	1206
NARDC	436	370	401	1207
SARDC	301	330	334	965
PARDC	402	372	392	1166
WARDC	455	456	437	1348
Total	2041	1898	1953	5892

Table 3.2 Repeated Measures Design (and Sample Sizes)

	Second Test			
First Test	A	B	E	F
A		1 (411)		3 (236)
B	2 (432)		5 (209)	
E		6 (215)		7 (446)
F	4 (216)		8 (446)	

site, one of two methods of assignment was used, depending on whether one or two testing rooms were available.

At two-room sites, examinees were randomly assigned to Forms A, E, and F upon arrival. Those examinees assigned to Form A were tested in one room. Those examinees assigned to either of Forms E or F were tested in a second room.

At one-room sites, some sessions were dedicated to Form A, and other sessions were dedicated to the new Forms E and F. All examinees at one-room sites were scheduled for testing prior to their arrival at the test-site. At the time of scheduling, each examinee was randomly assigned to one of the three forms (A, E, or F). Once assigned to a specific form, then the examinee was given a choice of several test-dates which had been dedicated to the assigned form.

3.2 Repeated-Measures Sample

Examinees in the repeated-measures sample were administered two forms of the GATB. These data were used primarily for examining the reliability and construct validity of the GATB. However, a portion of these data were also used to supplement the equating data. These data were used to perform a detailed comparison of measurement properties between the old and new forms.

Each examinee participating in the repeated-measures portion of the study was randomly assigned to one of eight conditions. These conditions are listed in Table 3.2. The numbers of examinees in each condition is also presented in Table 3.2. Note that conditions 1, 2, 7, and 8 consist of samples of approximately equal size ($N \approx 430$). The remaining conditions listed in Table 3.2 also consist of approximately equal sizes samples ($N \approx 218$). The rationale for the sample size requirement is outlined in the data analysis section. The numbers of examinees tested in each condition at each site is provided in Table 3.3.

At two-room sites, examinees were randomly assigned to the eight conditions upon arrival. At one-room sites, all examinees were scheduled for testing prior to arrival. This procedure ensured random assignment of examinee to condition.

3.3 Psychomotor Sample

This sample of $N = 538$ was administered the five psychomotor tests along with the non-psychomotor portions of Forms A and F. The design is presented in Table 3.4. Examinees were randomly assigned to one of two groups. Each group received three sections: (1) Form A (non-psychomotor), (2) Form A (psychomotor), and (3) Form F (non-psychomotor), with

Table 3.3 Repeated-Measures Sample Sizes (by test-site)

ARDC	Condition 1	2	3	4	5	6	7	8	Total
EARDC	43	46	27	22	24	25	53	48	288
NARDC	81	82	41	45	40	40	88	93	510
SARDC	94	100	52	51	47	45	76	77	542
PARDC	107	98	63	49	53	49	103	108	630
WARDC	86	106	53	49	45	56	126	120	641
Total	411	432	236	216	209	215	446	446	2611

Table 3.4 Psychomotor Data Collection Design

	Group 1 ($N = 265$)	Group 2 ($N = 273$)
Morning		
	1. Form A (non-pmotor)	1. Form F (non-pmotor)
	2. Form A (pmotor)	2. Form A (pmotor)
Afternoon		
	3. Form F (non-pmotor)	3. Form A (non-pmotor)

the order of presentation counterbalanced across the two groups. As indicated in Table 3.4, Group 1 received Form A (non-psychomotor) and Form A (psychomotor) portions in the morning session, and Form F (non-psychomotor) in the afternoon. Group 2 received the same battery of tests with the order of the non-psychomotor sections of Forms A and F reversed.

At two-room sites, examinees were randomly assigned to the two conditions upon arrival. At one-room sites, all examinees were scheduled for testing prior to arrival. This procedure ensured random assignment of examinee to condition.

Chapter 4
SAMPLE CHARACTERISTICS

This section provides an evaluation of the demographic characteristics of each of the three samples included in this study: (1) the independent-groups sample, (2) the repeated-measures sample, and (3) the psychomotor sample. An evaluation of the random equivalence of selected groups within each of the three samples is also provided, since the random equivalence of these groups is a key assumption made in the equating, reliability, and validity analyses.

This section also provides a description of the data editing procedure used to remove unmotivated examinees and other highly influential cases from each of the three samples. The editing procedure is described, along with the numbers of cases removed from each sample.

4.1 Demographics and Group Equivalence

For each of the three primary samples (independent-groups, repeated-measures, and psychomotor), descriptive statistics are provided for sex, race, age, and education. Within each sample, these demographic variables are displayed separately for randomly equivalent groups. Significance tests of the difference across randomly equivalent groups, for each of these demographic variables, are also provided. Non-significant results would be consistent with the expectation based on random assignment of examinees to condition, and would support the assumption of equivalent groups made in the equating, reliability, and validity analyses.

4.1.1 Independent-Groups (IG) Sample

Table 4.1 displays the percentages of males and females for the independent-groups sample. About 53-percent of the sample is male. As indicated by the non-significant χ^2-value, the small difference in the percentages of males (and females) across the three groups (defined by test form) is about equal to that expected by sampling error.

Table 4.2 displays the distribution of racial composition for the independent-groups sample. Nearly half the sample is White, with about 35-percent Blacks, and about 12-percent Hispanic. The distributions of race across the three-independent groups do not differ statistically, as indicated by the non-significant χ^2-value.

Tables 4.3 and 4.4 display the means and SD's for age and education, respectively. The mean age (in years) was about 36, and did not differ significantly (as indicated by the non-significant F-value) across the three randomly equivalent groups. The average number of years of education was 12.7, and also did not differ significantly across the three groups.

Table 4.1 IG Sample: Percentages of Males and Females

	Form			
Sex	A	E	F	Total
Male	54	53	51	53
Female	46	47	49	47
Total	100	100	100	100
N	2041	1897	1951	5889

Note. $\chi^2 = 2.57$ ($df = 2$, $p = .28$)

Table 4.2 IG Sample: Racial Composition

	Form			
Race	A	E	F	Total
White	48.1	46.9	48.4	47.8
Black	34.3	36.1	34.7	35.0
Hispanic	11.7	11.9	12.2	11.9
Amer. Ind.	1.3	1.6	1.2	1.4
Asian	3.0	2.6	2.0	2.5
Other	1.6	.8	1.5	1.3
Total	100	100	100	100
N	2041	1892	1945	5878

Note. $\chi^2 = 13.20$ ($df = 10$, $p = .21$)

Table 4.3 IG Sample: Descriptive Statistics for Age (Years)

	Form			
Statistic	A	E	F	Total
N	2029	1895	1949	5873
Mean	35.3	35.7	35.7	35.6
SD	12.04	12.33	12.23	12.20

Note. $F = .56$ ($df = 2, 5870$; $p = .57$)

Table 4.4 IG Sample: Descriptive Statistics for Education (Years)

	Form			
Statistic	A	E	F	Total
N	2040	1895	1946	5881
Mean	12.7	12.7	12.7	12.7
SD	2.17	2.08	2.09	2.11

Note. $F = .36$ ($df = 2, 5878$; $p = .70$)

Table 4.5 RM Sample: Percentages of Males and Females

Sex	Condition 1	2	3	4	5	6	7	8	Total
Male	59	59	57	58	59	58	59	55	58
Female	41	41	43	42	41	42	41	45	42
Total	100	100	100	100	100	100	100	100	100
N	411	432	234	216	208	215	446	446	2608

Note. $\chi^2 = 2.06$ ($df = 7$, $p = .96$)

Table 4.6 RM Sample: Racial Composition

Race	Condition 1	2	3	4	5	6	7	8	Total
White	30.4	37.0	35.6	33.3	32.7	35.8	34.3	32.5	33.9
Black	48.7	42.8	46.4	50.0	51.0	42.3	42.8	46.0	45.8
Hispanic	13.4	12.3	13.7	10.2	13.5	14.4	15.9	14.8	13.7
Amer. Ind.	2.2	2.1	2.1	3.2	1.9	.5	.7	2.2	1.8
Asian	3.6	4.6	1.7	1.4	.5	3.3	3.6	3.4	3.1
Other	1.7	1.2	.4	1.9	.5	3.7	2.7	1.1	1.6
Total	100	100	100	100	100	100	100	100	100
N	411	432	233	216	208	215	446	446	2607

Note. $\chi^2 = 48.02$ ($df = 35$, $p = .07$)

4.1.2 Repeated-Measures (RM) Sample

Table 4.5 displays the percentages of males and females for the repeated-measures sample. About 58-percent of the sample is male. As indicated by the non-significant χ^2-value, the small difference in the percentages of males (and females) across the eight groups is about equal to that expected by sampling error.

Table 4.6 displays the distribution of racial composition for the repeated-measures sample. Nearly 46-percent of the sample is Black, with about 34-percent White, and about 14-percent Hispanic. The distributions of race across the eight groups do not differ statistically, as indicated by the non-significant χ^2-value.

Tables 4.7 and 4.8 display the means and SD's for age and education, respectively for the repeated-measures sample. The mean age (in years) was about 35, and did not differ significantly (as indicated by the non-significant F-value) across the eight randomly equivalent groups. The average number of years of education was 12.6, and also did not differ significantly across the eight groups.

4.1.3 Psychomotor (PM) Sample

Table 4.9 displays the percentages of males and females for the psychomotor sample. About 59-percent of the sample is male. As indicated by the non-significant χ^2-value, the difference in the percentages of males (and females) across the two groups is about equal to that expected by sampling error.

Table 4.10 displays the distribution of racial composition for the psychomotor sam-

Table 4.7 RM Sample: Descriptive Statistics for Age (Years)

Statistic	Condition 1	2	3	4	5	6	7	8	Total
N	411	432	232	216	207	215	444	445	2602
Mean	34.9	35.8	35.8	36.6	36.4	34.1	34.6	34.7	35.2
SD	11.54	11.21	12.20	11.99	12.50	11.42	12.26	12.06	11.88

Note. $F = 1.56$ ($df = 7, 2594$, $p = .14$)

Table 4.8 RM Sample: Descriptive Statistics for Education (Years)

Statistic	Condition 1	2	3	4	5	6	7	8	Total
N	409	432	233	216	208	215	445	445	2603
Mean	12.8	12.6	12.6	12.9	12.4	12.6	12.7	12.5	12.6
SD	1.99	2.13	2.00	2.03	1.80	1.96	2.05	1.90	2.00

Note. $F = 1.57$ ($df = 7, 2595$, $p = .14$)

Table 4.9 PM Sample: Percentages of Males and Females

Sex	Form: Group 1	Group 2	Total
Male	56	63	59
Female	44	37	41
Total	100	100	100
N	265	273	538

Note. $\chi^2 = 2.86$ ($df = 1$, $p = .09$)

Table 4.10 PM Sample: Racial Composition

Race	Form Group 1	Form Group 2	Total
White	49.1	44.3	46.7
Black	39.2	44.3	41.8
Hispanic	6.4	7.3	6.9
Amer. Ind.	1.9	2.6	2.2
Asian	1.1	.4	.7
Other	2.3	1.1	1.7
Total	100	100	100
N	265	273	538

Note. $\chi^2 = 4.07$ ($df = 5$, $p = .54$)

Table 4.11 PM Sample: Descriptive Statistics for Age (Years)

Statistic	Form Group 1	Form Group 2	Total
N	265	273	538
Mean	36.4	36.1	36.2
SD	10.79	10.95	10.87

Note. $F = .11$ ($df = 1,536$; $p = .74$)

ple. Nearly 47-percent of the sample is White, with about 42-percent Black, and about 7-percent Hispanic. The distributions of race across the two groups do not differ statistically, as indicated by the non-significant χ^2-value.

Tables 4.11 and 4.12 display the means and SD's for age and education, respectively for the psychomotor sample. The mean age (in years) was about 36, and did not differ significantly (as indicated by the non-significant F-value) across the two randomly equivalent groups. The average number of years of education was 12.8, and also did not differ significantly across the two groups.

4.1.4 Summary

The demographic summaries indicate diverse samples with respect to gender, race, age, and education. Furthermore, the significance tests performed on the three samples (independent-

Table 4.12 PM Sample: Descriptive Statistics for Education (Years)

Statistic	Form Group 1	Form Group 2	Total
N	265	273	538
Mean	12.7	12.8	12.8
SD	1.95	1.96	1.95

Note. $F = .0041$ ($df = 1,536$; $p = .95$)

groups, repeated-measures, and psychomotor) provide some reassurance that the assignment procedures worked as intended—producing groups that are randomly equivalent with respect demographic characteristics. Although the equivalence of the groups with respect to cognitive and psychomotor abilities is not testable with existing data, the results based on the demographic variables provide additional confidence in this assumption, since in some instances demographic and cognitive/psychomotor variables tend to be correlated.

4.2 Outlier Analysis

Prior to data analysis, a small number of cases with unlikely scores were deleted from the database. These were deleted using a procedure suggested by Hotelling (1931) which identifies cases that are unlikely given that the observations are sampled from multivariate elliptical-shaped distribution.

Separate outlier analyses were performed for the three samples (independent-groups, repeated-measures, and psychomotor) described in Section 3. Furthermore, separate analyses were performed for each group, within each sample. The specific variables used in each analysis is presented in Table 4.13. As indicated across the top of Table 4.13, there were a total of 13 groups, spanning the independent-groups (IG), repeated-measures (RM), and psychomotor (PM) samples. The main body of Table 4.13 illustrates the number and type of variables used to edit the data. The bottom row displays the total number of variables (denoted as q) used in editing each group. For example, for examinees administered Form A in the IG-sample, the seven subtests of Form-A were used in the editing (second column of Table 4.13). For Group-1 of the RM-sample (5th column), the 14-variables of the two Forms A and B were used in the editing.

For simplicity, the description that follows is provided in terms of $\mathbf{X}$, a $n \times q$ matrix of scores, where n is the number of examinees in the group, and q is the number of variables used in the data editing (bottom row of Table 4.13). The analysis described below was repeated for the 13 groups listed across the top of Table 4.13.

Let $\mathbf{X}$ be an $n \times q$ matrix of test-scores (number-right) for n subjects. We can define a $n \times q$ matrix of difference scores by

$$\mathbf{D} = \mathbf{X} - \mathbf{1}\mathbf{m}_x' ,$$

where $\mathbf{1}$ is an $n \times 1$ column vector of 1's, and $\mathbf{m}_x'$ is a row vector of means of the q variables. In addition we let $\boldsymbol{\Sigma}$ denote the covariance matrix of the q variables. The vector of indices can be computed from

$$\mathbf{w} = \text{diag}\{\mathbf{D}\boldsymbol{\Sigma}^{-1}\mathbf{D}'\} .$$

The distribution of w is directly related to the F-distribution by the relationship

$$w = \frac{q(n-1)}{n-q} F_{q,n-q} ,$$

where q is the number of variables used in the analysis. Cases were deleted if w was sufficiently large (i.e. if $p < .001$).

Note that it is possible for two types of patterns to be flagged and deleted by the above procedure. One type occurs when an examinee receives extreme scores on many

Table 4.13 Data Editing Design

	Sample												
	IG			RM								PM	
Variable	A	E	F	1	2	3	4	5	6	7	8	1	2
New Form E/F													
AR		1	1			1	1	1	1	2	2	1	1
VO		1	1			1	1	1	1	2	2	1	1
3D		1	1			1	1	1	1	2	2	1	1
CO		1	1			1	1	1	1	2	2	1	1
NC		1	1			1	1	1	1	2	2	1	1
OM		1	1			1	1	1	1	2	2	1	1
Old Form A/B													
AR	1			2	2	1	1	1	1			1	1
VO	1			2	2	1	1	1	1			1	1
3D	1			2	2	1	1	1	1			1	1
CO	1			2	2	1	1	1	1			1	1
NC	1			2	2	1	1	1	1			1	1
TM	1			2	2	1	1	1	1			1	1
FM	1			2	2	1	1	1	1			1	1
PM Form A													
MM												1	1
PL												1	1
TU												1	1
AS												1	1
DI												1	1
N Vars (q)	7	6	6	14	14	13	13	13	13	12	12	18	18

Table 4.14 Group Sizes for Edited IG/Equating Samples

Sample	Form A	E	F
IG	2041	1898	1953
RM-1 (A/B)	411		
RM-3 (A/F)	236		
RM-6 (E/B)		215	
RM-7 (E/F)		446	
RM-4 (F/A)			216
RM-8 (F/E)			446
Total N	2688	2559	2615
No. Deletes	34	14	18
Final N	2654	2545	2597

Table 4.15 Group Sizes for Edited RM Sample

Frequency	Sample 1	2	3	4	5	6	7	8	Total
Total N	411	432	236	216	209	215	446	446	2611
No. Deletes	13	10	4	3	3	5	12	10	60
Final N	398	422	232	213	206	210	434	436	2551

subtests (i.e. all low scores). Another type of unlikely pattern occurs when an examinee scores high on one subtest and low on a second which is highly correlated with the first (i.e. alternate forms of the same subtest). A small number of additional cases were deleted because of zero number-right scores on one or more subtests. Since DOL policy dictates that scores should not be provided to such examinees, these cases were excluded.

4.2.1 Independent-Groups Sample

Table 4.14 provides the editing results for the independent-groups sample. Here, selected cases from the RM sample were combined with the IG sample. Data from the first test administered (Forms A, E, or F) of Conditions 1, 3, 4, 6, 7, and 8, were combined with the IG data to increase the sample sizes for the equating study. This resulted in the group sizes listed at the bottom of Table 4.14. Data editing was performed in each of the three IG/RM groups: "A," "E," and "F." A small number of cases were removed from each group—ranging from 14–34 examinees. The final group sizes used in the equating analysis are listed last row of Table 4.14.

4.2.2 Repeated-Measures Sample

Table 4.15 displays the editing results for the repeated-measures sample. Between 3 and 13 cases were deleted from each of the eight groups. The final group sizes after editing are displayed on the bottom row of Table 4.15.

Table 4.16 Group Sizes for Edited Psychomotor Sample

	Form		
Frequency	1	2	Total
Total N	265	273	538
No. Deletes	4	0	4
Final N	261	273	534

4.2.3 Psychomotor Sample

Table 4.16 displays the editing results for the repeated-measures sample. Four cases were deleted from Group-1, and zero were deleted from Group-2. The final group sizes after editing are displayed on the bottom row of Table 4.16.

Chapter 5

SCORING THE GATB

A number of GATB scores are routinely produced and used. Scoring for the new Forms E and F is complicated by the absence of the Form Matching subtest, and by the use of formula scores for the speeded subtests. The method of score computation for the old and new GATB forms is detailed below. This description includes the computation of both subtest, and composite scores.

5.1 Subtest Scoring

For the purpose of this study, all scores for Forms A and B were computed according to conventions specified in the "Manual for the USES general Aptitude Test Battery" (Section I).

5.1.1 Forms A and B

The set of raw scores for Forms A and B, denoted by

$$\{X_{ar}(A), X_{vo}(A), ..., X_{di}(A); X_{ar}(B), X_{vo}(B), ..., X_{di}(B)\} ,$$

are taken as the simple sum of the number of correct responses for the power and speeded subtests. The raw scores for the five psychomotor subtests are obtained according to the operational procedures. Standard-scores S are obtained from a look-up table (pp. 77–92). These standard scores are summed in various combinations to form the nine aptitude-scores $\{A_g, A_v, A_n, A_s, A_p, A_q, A_k, A_f, A_m\}$ displayed in Table 5.1. The conversion of raw-score to standard score depends on which aptitude the subtest scores will be used for. For subtests which enter into two different aptitude scores, there are two conversion tables. For subtests which enter into only one aptitude score, there is a single conversion table. Table 5.1 provides the notation for the subtest standard score and the aptitude score composition. Aptitude scores are formed from the simple sum (down each column of Table 5.1) of subtest standard scores.

5.1.2 Forms E and F

The set of raw scores for forms E and F, denoted by

$$\{X_{ar}(E), X_{vo}(E), ..., X_{om}(E); X_{ar}(F), X_{vo}(F), ..., X_{om}(F)\} ,$$

are taken as either: (a) the simple sum of the number of correct responses (for the three power subtests AR, VO, and 3D), or (b) the chance corrected formula score (for the three speeded subtests CO, NC, and OM). The formula scores for each of the speeded tests are

Table 5.1 Aptitude Score Composition

Subtest	Aptitude Score								
	A_g	A_v	A_n	A_s	A_p	A_q	A_k	A_f	A_m
1. Arithmetic Reasoning	$S_{ar}^{(g)}$		$S_{ar}^{(n)}$						
2. Vocabulary	$S_{vo}^{(g)}$	$S_{vo}^{(v)}$							
3. 3D Space	$S_{3d}^{(g)}$			$S_{3d}^{(s)}$					
4. Computation			$S_{co}^{(n)}$						
5. Name Comparison						$S_{nc}^{(q)}$			
6. Object (Tool) Matching					$S_{ob}^{(p)}$				
7. Form Matching					$S_{fm}^{(p)}$				
8. Mark Making							$S_{mm}^{(k)}$		
9. Place									$S_{pl}^{(m)}$
10. Turn									$S_{tu}^{(m)}$
11. Assemble								$S_{as}^{(f)}$	
12. Disassemble								$S_{di}^{(f)}$	

given by the general formula

$$X = NC - W/(A-1) \; , \tag{5.1}$$

where NC is the number of correct response, W is the number of wrong answers (items answered incorrectly; does not include omits or not-reached), and A is the number of response-options associated with the subtest items. For the three speeded subtests (5.1) simplifies to

$$\begin{aligned} X_{co}(f) &= NC - W/4 \\ X_{nc}(f) &= NC - W \\ X_{om}(f) &= NC - W/3 \, , \end{aligned}$$

where f equals Form "E" or "F."

Standard-scores S for Forms E and F are obtained from look-up tables produced from the equating described in the following chapter. Thus the process of computing aptitude-scores is identical to that described above for Forms A and B with one exception. Since Form Matching (FM) was dropped from Forms E and F, the aptitude score A_p is set equivalent to $S_{ob}^{(p)}$, rather than computed as $A_p = S_{ob}^{(p)} + S_{fm}^{(p)}$ (as in Forms A and B). Note that the distribution of A_p across the new and old forms is ensured to be equal through the appropriate specification of the equating transformation. This point is described in more detail in the following chapter.

5.2 Composite Scoring

In addition to the aptitude-score composites described above, there were two other sets of composites studied. These include three *component* composites and five *job-family* composites.

Each of the three component composites were computed from the sum of selected aptitude scores:

$$\begin{array}{lrcl} \text{Cognitive:} & C_{gvn} & = & A_g + A_v + A_n \\ \text{Perceptual:} & C_{spq} & = & A_s + A_p + A_q \\ \text{Psychomotor:} & C_{kfm} & = & A_k + A_f + A_m \,. \end{array}$$

Scores on the cognitive and perceptual composites were calculated for all examinees in the equating, reliability, and validity analyses. These composites were included in all key analyses. However, scores on the psychomotor composites were calculated only for examinees in the psychomotor sample, which was used to address selected composite equating and validity issues.

Also to address these same composite equating and validity issues, five job-family composites were computed from the weighted linear combinations:

$$\begin{array}{rcl} J_1 & = & .59 \times C_{gvn} + .30 \times C_{spq} + .11 \times C_{kfm} \\ J_2 & = & .13 \times C_{gvn} + .87 \times C_{kfm} \\ J_3 & = & C_{gvn} \\ J_4 & = & .73 \times C_{gvn} + .27 \times C_{kfm} \\ J_5 & = & .44 \times C_{gvn} + .56 \times C_{kfm} \,. \end{array}$$

Since all composites but one (which is redundant with C_{gvn}) are a function of one or more psychomotor subtests, these composites were computed only for the psychomotor sample.

Chapter 6
SMOOTHING AND EQUATING

The objective of equipercentile equating is to provide a transformation that will match score distributions of the new forms with the distribution of scores from the reference Form A. This transformation, which will be applied to the new Forms E and F, will allow scores on the new versions to be interpreted relative to the old scale represented by Form A.

One method for estimating this transformation involves the use of the two empirical cumulative distribution functions (CDF's). For example, scores on Forms A and E could be equated by matching the empirical proportion scoring at or below observed score levels. However, this transformation is subject to random sampling errors contained in the CDF's. It is commonly believed that the precision of the equating transformation can be improved by smoothing either: (a) the equating transformation, or (b) the two empirical distributions which form the equating transformation. A number of alternative procedures exist however for choosing the amount or degree of smoothing.

The precision of any estimated equating transformation can be decomposed into a *bias* component and a *variance* component. Smoothing procedures that attempt to eliminate the bias will increase the random variance of the transformation. A high order polynomial provides one example. The polynomial may track the data closely, but capitalize on chance errors and replicate poorly in a new sample. On the other hand, smoothing procedures that attempt to eliminate the random variance do so at the expense of introducing systematic error, or bias, into the transformation. Linear equating methods often replicate well but may display marked departure from the population transformation. It should be noted that whatever equating method is being used, the choice of method, either implicitly or explicitly involves a trade-off between random and systematic error.

One primary objective of the method of equating proposed here is to use smoothing procedures that provide an acceptable trade-off between random and systematic error. In this study smoothing was performed on each distribution (of Forms A, E, and F) separately. These smoothed distributions were then used to specify the equipercentile transformation.

6.1 Smoothing Procedures

Here polynomial log-linear smoothing was performed on each subtest of Forms A, E, and F. The number of terms used in each smoothing was based on the χ^2 difference test. This procedure began by fitting a 12-th order log-linear polynomial. The number of terms was decreased by one, and the likelihood ratio χ^2-difference test ($df = 1$) was performed to examine the significance of the difference between the existing n-order model and one that included all terms up to and including the $(n-1)$-order. If non-significant, then the number of terms was again decremented by one, and another significance test was performed. This process was repeated until the χ^2-difference test indicated that eliminating the $(n-1)$-th term resulted in a significant degradation in model fit. The deleted term (providing the significant degradation) was then included in the final n-term model.

The data used in this analysis were provided from two samples—the independent-groups sample, and the repeated-measures sample. Data collected on the first administered test from selected groups of the repeated-measures sample were combined with same-form data of the independent groups sample. The sample sizes used to estimate the Forms A, E, and F distributions are provided in Table 4.14.

6.2 Zero Cells

The log-linear procedure is undefined for bins (score-levels) which have a frequency of zero. For these analyses, a two step procedure was used for specifying zero-cell entries. First the maximum score was identified for each subtest. Then the distribution smoothing was performed for the raw-score range: "1–max," where "max" is the maximum score for the subtest. Then any zero-bins falling within the "1–max" range were changed to .5 for log-linear estimation. For the speeded-test distributions of Forms E and F, the low-range was set equal to the observed minimum in the sample, since for formula-scores, the minimum can fall below 1.

6.3 P-Aptitude

The new Forms E and F do not contain the Form Matching subtest. In the old Forms A–D this subtest was used in the P-Aptitude Score (see Table 5.1). The P-Aptitude score (Forms A–D) was computed from the sum of Form Matching and Tool (Object) Matching. Rather than equating Object Matching across the new and old forms, Object Matching (Forms E and F) was equated directly to the P-Aptitude distribution (Form A). This direct equating will allow scores on Object Matching (Forms E and F) to be transformed to a P-Aptitude score which has the same distribution as Form A, even though the Form Matching subtest has been omitted from the new forms.

This matching involved several steps. First the distribution of the P-Aptitude score was computed from Form A data. This was accomplished by summing the standard scores of Form Matching and Tool Matching. Next the distribution of P-Aptitude scores was smoothed by applying log-linear smoothing. The final equating was obtained by matching this smoothed distribution with the smoothed Object Matching distributions of Forms E and F, using the equipercentile procedure.

6.4 Polynomial Extrapolation

Several of the new and old subtests differ substantially in their length, which leads to a significant difference in their maximum attainable scores (see Tables 2.1 and 2.2). For example, Vocabulary had 60 items in Form A and 19 items in Forms E and F. A maximum score on the Form A version represented several standard deviations above the mean, while a maximum score on Form E/F represents less than two standard deviations above the mean. Consequently, matching the maximum scores on the new and old versions (which is a consequence of equipercentile equating) does not appear to be appropriate.

Although the test-lengths for the three speeded tests were shortened, the test-content and time-limits remained virtually unchanged across the new and old GATB versions. Since the shortening of these tests did not cause a noticeable ceiling effect, the maximum score on the new versions were set equal to their equated scores on Form A.

For the three power tests (AR, VO, and 3D), a polynomial extrapolation was used to specify the maximum equated score. The highest five points of the equipercentile transformation were approximated by a second-order polynomial fitted with least-squares. The equated value for the maximum score-level of Forms E and F was set equal to the predicted value of the resulting polynomial.

6.5 Equating Transformations

A total of 18 smoothings were performed—six smoothings for each of the three GATB forms. These smoothed distributions were used to compute 12 equating transformation—six transformations equating Form A and E subtests, and another six transformations equating Forms A and F.

The tables below provide a translation between Form E/F raw-scores x and Form A standard scores. Linear interpolation was used in conjunction with equipercentile equating to specify appropriate standard scores. The equated standard scores were obtained from

$$S(x) = S_L + \frac{F_{E/F}(x) - C_L}{C_U - C_L}\left[S_U - S_L\right] , \tag{6.1}$$

where $F_{E/F}(x)$ is the cumulative distribution function on the new form evaluated at raw-score level x, (C_L, C_U) are the lower and upper values of the Form A cumulative distribution interval which contains $F_{E/F}(x)$, i.e.

$$C_L \leq F_{E/F}(x) \leq C_U ,$$

and (S_L, S_U) are the Form A standard-scores corresponding to the interval defined by (C_L, C_U). Formula 6.1 was used to specify raw to standard score conversions for each raw-score level of Forms E and F. The resulting tables are listed below.

Table 6.1 Aptitude Score Composition: Arithmetic Reasoning Form E

Raw	Aptitude G	Aptitude N
1	2	2
2	5	4
3	6	5
4	9	6
5	11	8
6	13	10
7	15	11
8	17	12
9	19	14
10	21	15
11	23	17
12	24	18
13	26	19
14	29	21
15	31	23
16	35	26
17	38	28
18	42	31

Table 6.2 Aptitude Score Composition: Vocabulary Form E

Raw	Aptitude G	Aptitude V
1	44	64
2	46	66
3	47	69
4	48	72
5	49	75
6	50	78
7	51	81
8	51	83
9	53	85
10	54	88
11	55	91
12	57	96
13	58	100
14	60	105
15	61	108
16	63	112
17	65	117
18	68	125
19	71	134

Table 6.3 Aptitude Score Composition: 3D Space Form E

Raw	Aptitude G	Aptitude S
1	2	51
2	3	56
3	4	60
4	4	63
5	5	66
6	6	68
7	7	71
8	8	74
9	9	78
10	10	82
11	11	86
12	13	90
13	14	93
14	15	97
15	16	102
16	17	107
17	18	112
18	20	118
19	23	126
20	26	134

Table 6.4 Aptitude Score Coversion: Computation Form E

Raw	Aptitude N	Raw	Aptitude N
-4	37	19	79
-3	39	20	81
-2	41	21	83
-1	42	22	85
0	44	23	87
1	45	24	89
2	47	25	92
3	48	26	94
4	50	27	96
5	52	28	98
6	54	29	99
7	56	30	101
8	58	31	102
9	60	32	103
10	61	33	104
11	63	34	105
12	65	35	106
13	67	36	106
14	68	37	108
15	70	38	110
16	74	39	112
17	76	40	114
18	77		

Table 6.5 Aptitude Score Coversion: Name Comparison Form E

Raw	Aptitude Q	Raw	Aptitude Q	Raw	Aptitude Q
-8	53	25	81	58	111
-7	56	26	82	59	112
-6	58	27	82	60	113
-5	60	28	83	61	114
-4	63	29	84	62	116
-3	65	30	85	63	117
-2	66	31	85	64	119
-1	67	32	86	65	120
0	68	33	86	66	121
1	69	34	87	67	123
2	70	35	88	68	124
3	70	36	89	69	125
4	71	37	90	70	126
5	71	38	90	71	127
6	72	39	91	72	128
7	73	40	92	73	129
8	74	41	93	74	132
9	74	42	94	75	133
10	75	43	95	76	134
11	76	44	96	77	136
12	76	45	96	78	137
13	77	46	97	79	138
14	77	47	98	80	140
15	77	48	99	81	142
16	78	49	100	82	145
17	78	50	102	83	147
18	78	51	104	84	150
19	79	52	105	85	152
20	79	53	105	86	154
21	80	54	106	87	156
22	80	55	107	88	157
23	80	56	108	89	160
24	81	57	109	90	162

Table 6.6 Aptitude Score Coversion: Object Matching Form E

Raw	Aptitude P	Raw	Aptitude P
-2	33	21	92
-1	38	22	95
0	41	23	98
1	43	24	102
2	44	25	105
3	46	26	108
4	48	27	111
5	49	28	114
6	51	29	117
7	53	30	120
8	55	31	124
9	58	32	127
10	60	33	131
11	63	34	134
12	65	35	137
13	68	36	140
14	71	37	143
15	74	38	145
16	77	39	146
17	80	40	147
18	83	41	148
19	86	42	148
20	89		

Table 6.7 Aptitude Score Composition: Arithmetic Reasoning Form F

Raw	Aptitude G	Aptitude N
1	2	2
2	4	3
3	6	4
4	8	5
5	10	7
6	11	8
7	12	9
8	14	11
9	16	12
10	18	13
11	20	15
12	22	16
13	24	18
14	26	19
15	29	21
16	32	23
17	37	27
18	42	31

Table 6.8 Aptitude Score Composition: Vocabulary Form F

Raw	Aptitude G	Aptitude V
1	44	63
2	45	65
3	46	66
4	47	69
5	48	72
6	49	75
7	50	77
8	51	80
9	51	82
10	52	85
11	54	88
12	55	92
13	57	97
14	59	102
15	61	107
16	63	111
17	65	116
18	67	124
19	69	131

Table 6.9 Aptitude Score Composition: 3D Space Form F

Raw	Aptitude G	Aptitude S
1	2	51
2	2	51
3	4	58
4	4	62
5	5	65
6	6	67
7	6	69
8	7	72
9	8	75
10	9	79
11	10	82
12	11	85
13	13	90
14	14	93
15	15	97
16	16	102
17	17	107
18	19	115
19	22	124
20	25	135

Table 6.10 Aptitude Score Coversion: Computation Form F

Raw	Aptitude N	Raw	Aptitude N
-2	38	20	77
-1	40	21	79
0	41	22	82
1	42	23	84
2	42	24	87
3	44	25	89
4	45	26	91
5	47	27	93
6	48	28	95
7	50	29	97
8	52	30	99
9	54	31	101
10	56	32	104
11	57	33	105
12	59	34	105
13	61	35	106
14	63	36	107
15	65	37	108
16	67	38	110
17	69	39	112
18	72	40	114
19	75		

Table 6.11 Aptitude Score Composition: Name Comparison Form F

Raw	Aptitude Q	Raw	Aptitude Q	Raw	Aptitude Q
-10	53	24	79	58	108
-9	53	25	79	59	109
-8	55	26	80	60	111
-7	56	27	80	61	112
-6	57	28	81	62	113
-5	58	29	81	63	114
-4	59	30	82	64	115
-3	60	31	83	65	117
-2	62	32	84	66	118
-1	63	33	85	67	119
0	65	34	85	68	120
1	66	35	86	69	122
2	66	36	86	70	123
3	67	37	87	71	123
4	68	38	88	72	124
5	68	39	89	73	125
6	69	40	90	74	126
7	69	41	91	75	127
8	70	42	92	76	127
9	70	43	93	77	128
10	71	44	94	78	129
11	71	45	95	79	132
12	72	46	96	80	133
13	72	47	96	81	134
14	73	48	97	82	135
15	74	49	98	83	137
16	74	50	99	84	138
17	75	51	100	85	141
18	76	52	102	86	144
19	76	53	104	87	148
20	77	54	104	88	153
21	77	55	105	89	156
22	78	56	106	90	162
23	78	57	107		

Table 6.12 Aptitude Score Coversion: Object Matching Form F

Raw	Aptitude P	Raw	Aptitude P
-5	27	19	82
-4	31	20	85
-3	33	21	88
-2	35	22	91
-1	37	23	94
0	39	24	97
1	41	25	100
2	42	26	104
3	44	27	107
4	46	28	111
5	48	29	114
6	49	30	118
7	51	31	122
8	53	32	126
9	55	33	130
10	58	34	133
11	60	35	137
12	62	36	140
13	65	37	143
14	68	38	145
15	70	39	146
16	73	40	147
17	76	41	148
18	79	42	148

Chapter 7
COMPOSITE EQUATING

Equating the new and old P&P-GATB forms involves matching subtest distributions using an equipercentile method. This distribution matching provides a transformation of Forms E and F to standard-score equivalents on the reference form (Form A) scale. Once this transformation is specified for each subtest, standard-score equivalents can be computed. These standard-score equivalents provided the basis for the computation of GATB composites. The same formulas used to compute composites from standard scores on Form A can be used to compute composite scores from standard-score equivalents on the new forms.

One concern is that the distribution of composite scores from the new forms will differ systematically from the corresponding distributions of the old forms. This difference could result from differences in subtest intercorrelations between the old and new forms. Different subtest intercorrelations may result from one or more revisions made to the new P&P-GATB (change in test-lengths, time-limits, instructions, etc.). Since the variance of a composite is partially affected by the correlations among subtests, differences in composite variances could result as a consequence. Higher order moments of the composite distributions could be affected in a similar manner.

This section presents an analysis of the composite distributions across new and old GATB forms. This analysis is presented in two sections. The first section presents the analysis of the non-psychomotor composites. The second section provides the analysis of composites containing both psychomotor and non-psychomotor subtests.

7.1 Non-Psychomotor Composites

The data used in this analysis were the same data used to estimated the equating transformations. These data were provided from two samples—the independent-groups sample, and the repeated-measures sample. Data collected on the first administered test from selected groups of the repeated-measures sample were combined with same-form data of the independent-groups sample. The sample sizes used to examine composite distributions across Forms A, E, and F are provided in Table 4.14.

The distributions of several composites were examined. A number of steps were involved in comparing composite score distributions. First, scores on Forms E and F were transformed to standard-score equivalents using the transformation estimated from the equating. Next for each composite, scores were computed for three groups:

1. Composite scores were obtained for those examinees taking Form A by applying the composite-formulas to the standard-scores. A total of four composites were examined:

A_g and A_n (aptitude scores, Table 5.1); and C_{gvn} and C_{spq} (Cognitive and Perceptual composites, Section 5.2). All other aptitude scores are a function of a single subtest—thus the agreement of their distributions across the new and old forms will be guaranteed by the equipercentile transformation. Since the A_p aptitude score distribution of Form A was equated directly to the A_p score distribution of Forms E and F (which is a nonmonotonic function of Object Matching) these distributions are also matched through the equating transformation, and thus no confirmation is necessary. Composites which contain psychomotor subtests were analyzed separately (see section below).

2. Composite scores were obtained for those examinees taking Form E. Composite scores were obtained for A_g, A_n, C_{gvn}, and C_{spq} by applying the composite formulas to the standard-score equivalent scores.

3. Composite scores were obtained for those examinees taking Form F. Composite scores were obtained for A_g, A_n, C_{gvn}, and C_{spq} by applying the composite formulas to the standard-score equivalent scores.

The distributions of scores for each of the new forms (Forms E and F) were compared to the corresponding composite distribution of the reference form (Form A). Four cut-points were used to divide the distribution into five groups. Cut scores were based on the area under a normal density function. The z-values (computed from Form A means and standard deviations) which divided the distribution into groups having the expected proportions

$$\{.10, .25, .30, .25, .10\}$$

were applied to the composite distributions to produce the observed proportions displayed in Tables 7.1–7.4. The proportion of examinees falling in each group was compared across the two new versions (E and F) and the single old version (A). The significance of the difference in these proportions was examined using a 3×5 contingency table analysis. The Pearson χ^2 statistic was used to test the null hypothesis of no difference among distributions. The proportion of each composite distribution falling in each of the five groups are displayed for each of the three GATB forms in Tables 7.1–7.4.

Although two composites (A_g and C_{spq}) have marginally significant differences, an examination of the distributions indicate the four composites are very similar across the new and old forms. None of the composites were significant at the .01 level. These results suggest that the standard GATB composite formulas can be applied to the equated standard-scores of the new forms—and that these composite scores will have similar distributions across the new and old versions. These results indicate that separate composite equating tables for the non-psychomotor composites are unnecessary for the new Forms E and F.

7.2 Psychomotor Composites

In this study the evaluation of composite equating is complicated by the absence of the psychomotor tests from the new Forms E and F. New versions of the psychomotor tests are being developed under a separate data collection and analysis effort. Since some of the composites computed under the Job-Family System include a combination of psychomotor

Table 7.1 Distributions for Aptitude Score A_g

Group	Form A	Form E	Form F
1	.09	.10	.10
2	.28	.25	.27
3	.29	.29	.26
4	.23	.23	.24
5	.11	.13	.13
Total	1.00	1.00	1.00
N	2654	2545	2597

$\chi^2 = 16.14$ ($df = 8$, $p = .04$)

Table 7.2 Distributions for Aptutude Score A_n

Group	Form A	Form E	Form F
1	.10	.10	.10
2	.26	.25	.25
3	.29	.29	.29
4	.24	.25	.26
5	.11	.10	.11
Total	1.00	1.00	1.00
N	2654	2545	2597

$\chi^2 = 3.40$ ($df = 8$, $p = .91$)

Table 7.3 Distributions for Composite C_{gvn}

Group	Form A	Form E	Form F
1	.09	.09	.09
2	.28	.26	.26
3	.28	.30	.29
4	.23	.23	.24
5	.11	.12	.12
Total	1.00	1.00	1.00
N	2654	2545	2597

$\chi^2 = 7.87$ ($df = 8$, $p = .45$)

Table 7.4 Distributions for Composite C_{spq}

	Form		
Group	A	E	F
1	.10	.10	.09
2	.26	.25	.25
3	.30	.29	.29
4	.23	.26	.27
5	.11	.10	.10
Total	1.00	1.00	1.00
N	2654	2545	2597

$\chi^2 = 16.94$ ($df = 8$, $p = .03$)

and non-psychomotor subtests, a complete evaluation should examine the similarity of these distributions (across old and new forms). For this analysis we use data collected from the psychomotor sample (see Section 3.3) This sample was administered one old Form (A), one new Form (F), and one form of the psychomotor portion of the battery (from Form A).

Distributions of composite scores were compared across the two groups using data collected in the morning session only (Table 3.4). For Group 1 (which was administered Form A psychomotor and non-psychomotor subtests) four composites (J_1, J_2, J_4, J_5) were computed according to the procedures described in Section 5.2. These composites consist of the four Job-family composites that contain a mixture of psychomotor and non-psychomotor subtests. Group 2 scores for the corresponding composites were computed in a similar manner from standard-score equivalents using the Form F (non-psychomotor) and Form A (psychomotor) portions of the battery.

Four cut-points were used to divide the distributions into five groups as in the analysis above. Cut scores were based on the area under a normal density function. The z-values (computed from Group 1 means and standard deviations) which divided the distribution into groups having the expected proportions $\{.10, .25, .30, .25, .10\}$ were applied to the composite distributions to produce the observed proportions displayed in Tables 7.5–7.8. The proportion of examinees falling in each group was compared across the new (Form F) and old (Form A) versions. The significance of the difference in these proportions was examined using a 2×5 contingency table analysis. The Pearson χ^2 statistic was used to test the null hypothesis of no difference among distributions. The proportion of each composite distribution falling in each of the five groups are displayed for Forms A and F in Tables 7.5 through 7.8.

Note that these proportions are very similar across new and old GATB forms, and do not differ significantly from what would be expected from sampling error. These results indicate that the Job-Family composite formulas can be applied to the equated standard-scores of the new forms—and that these composite scores will have similar distributions across the new and old versions. These results suggest that separate composite equating tables for the psychomotor composites are unnecessary for the new Forms E and F.

Note that one key assumption of this analysis is that the new forms of the psychomotor subtests (currently under development for Forms E and F) will be parallel to the form

Table 7.5 Distributions for Aptitude Score J_1

Group	Form A	Form F
1	.11	.12
2	.26	.21
3	.27	.30
4	.25	.29
5	.12	.09
Total	1.00	1.00
N	261	273

$\chi^2 = 3.88$ ($df = 4$, $p = .42$)

Table 7.6 Distributions for Aptitude Score J_2

Group	Form A	Form F
1	.11	.09
2	.23	.23
3	.28	.27
4	.29	.28
5	.09	.12
Total	1.00	1.00
N	261	273

$\chi^2 = 2.08$ ($df = 4$, $p = .72$)

Table 7.7 Distributions for Aptitude Score J_4

Group	Form A	Form F
1	.10	.10
2	.25	.22
3	.28	.28
4	.28	.31
5	.09	.08
Total	1.00	1.00
N	261	273

$\chi^2 = 1.09$ ($df = 4$, $p = .90$)

Table 7.8 Distributions for Aptitude Score J_5

	Form	
Group	A	F
1	.11	.09
2	.21	.23
3	.31	.26
4	.26	.31
5	.10	.10
Total	1.00	1.00
N	261	273

$\chi^2 = 2.63$ ($df = 4$, $p = .62$)

used in this study. If they are not parallel, then the covariances between these tests and the non-psychomotor subtests may be poorly represented by those obtained in the current study. That is, if the new and old forms of the psychomotor tests are not parallel, then the results obtained in this study may not generalize to the new psychomotor tests.

Chapter 8
SUBGROUP COMPARISONS

Although equipercentile equating matches subtest distributions for the total sample, it does not necessarily guarantee a match for distributions of subgroups contained in the sample. This result follows from the fact that the new and old versions of the P&P-GATB are not strictly parallel. Although we might expect to observe small differences in subgroup performance across the new and old versions as a result of differences in measurement precision, many of the other revisions made to the new forms could also cause group differences. It is therefore instructive to examine the performance of subgroups to determine if any are placed at a substantial disadvantage by the new forms, relative to their level of performance on the old GATB forms.

Two analyses examining subgroup performance were conducted. The first set of analyses examined the level of performance of each subgroup across the new and old forms. The second set of analyses examined adverse impact for selected subgroups, also broken out separately for new and old forms. Each set of analyses is described in separate sections below.

8.0.1 Subgroup Performance Across Forms

In the analyses described below, four subgroups were examined: (1) Blacks, (2) Hispanics, (3) Females, and (4) examinees 41 years of age or older. The equating transformation based on the total sample was applied to subgroup members who had taken the new GATB forms (E and F). For each subgroup, mean performance levels were compared across new and old forms. Six subtest variables were examined: $S_{ar}^{(g)}$, $S_{vo}^{(g)}$, $S_{3d}^{(g)}$, $S_{co}^{(n)}$, $S_{nc}^{(q)}$, and A_p. (See Chapter 5 for variable definitions.) These variables are monotonic functions of the six subtests scores. (For Form A, the variable A_p is a function of both Form Matching and Object Matching.)

Tables 8.1–8.4 display the ANOVA results for each of the four subgroups. Within each table the significance of the difference among means is examined for Forms A, E, and F. Among the 24 comparisons, only one was significant at the .01 level. A significant difference was observed across A, E, and F for Blacks on the Vocabulary subtest. As indicated from Table 8.1, Blacks administered Forms E and F tended to score slightly higher than Blacks administered Form A. In general, the results indicate similar average performance levels across new and old versions for each of the four subgroups examined.

8.0.2 Adverse Impact

Adverse impact analyses were conducted for three subgroups: (1) Blacks, (2) Hispanics, and (3) females. The results of these analyses are presented in Tables 8.5, 8.6, and 8.7, respectively. Differences in mean levels between majority and minority groups are reported

Table 8.1 Significance Tests for Mean Differences Across Forms: Blacks

Variable	Mean			SD			ANOVA	
	A	E	F	A	E	F	F-ratio	p
$S_{ar}^{(g)}$	18.35	18.78	18.74	6.59	7.27	7.38	1.09	0.34
$S_{vo}^{(g)}$	54.05	54.90	54.79	4.54	5.08	5.18	8.63	0.00
$S_{3d}^{(g)}$	11.98	12.46	12.13	4.61	4.89	4.90	2.60	0.07
$S_{co}^{(n)}$	71.17	71.51	71.64	11.40	11.75	11.99	0.38	0.68
$S_{nc}^{(q)}$	98.88	100.17	99.87	13.62	15.59	14.37	2.11	0.12
A_p	93.79	94.65	94.76	18.68	18.93	18.87	0.76	0.47
N	996	960	974					

Table 8.2 Significance Tests for Mean Differences Across Forms: Hispanics

Variable	Mean			SD			ANOVA	
	A	E	F	A	E	F	F-ratio	p
$S_{ar}^{(g)}$	21.23	21.01	22.09	6.74	7.27	7.85	2.01	0.13
$S_{vo}^{(g)}$	55.26	55.59	56.15	4.94	5.53	5.39	2.31	0.10
$S_{3d}^{(g)}$	15.13	14.95	15.34	5.14	5.21	5.31	0.45	0.64
$S_{co}^{(n)}$	73.79	73.35	74.49	10.76	10.88	11.23	0.89	0.41
$S_{nc}^{(q)}$	101.29	102.98	103.30	14.07	15.20	15.40	1.68	0.19
A_p	101.30	101.92	103.33	19.24	19.38	19.32	0.93	0.39
N	320	326	325					

Table 8.3 Significance Tests for Mean Differences Across Forms: Females

Variable	Mean			SD			ANOVA	
	A	E	F	A	E	F	F-ratio	p
$S_{ar}^{(g)}$	22.43	22.49	22.84	8.41	8.42	8.88	0.82	0.44
$S_{vo}^{(g)}$	57.56	57.38	57.44	6.04	6.12	5.88	0.30	0.74
$S_{3d}^{(g)}$	14.10	14.25	14.27	5.34	5.50	5.69	0.32	0.72
$S_{co}^{(n)}$	76.59	76.62	76.78	11.75	11.91	11.77	0.09	0.91
$S_{nc}^{(q)}$	108.31	108.27	108.21	16.00	16.38	16.49	0.01	0.99
A_p	103.33	104.17	104.63	19.80	19.74	19.65	1.35	0.26
N	1192	1164	1226					

Table 8.4 Significance Tests for Mean Differences Across Forms: Age > 40

Variable	Mean			SD			ANOVA	
	A	E	F	A	E	F	F-ratio	p
$S_{ar}^{(g)}$	24.22	24.11	24.85	8.96	9.05	9.23	1.63	0.20
$S_{vo}^{(g)}$	58.76	59.28	59.21	6.43	6.66	6.28	1.59	0.20
$S_{3d}^{(g)}$	14.09	14.10	14.75	5.38	5.70	5.84	3.87	0.02
$S_{co}^{(n)}$	76.36	75.73	76.29	12.79	12.40	12.26	0.66	0.52
$S_{nc}^{(q)}$	103.88	102.18	102.77	16.13	15.59	16.05	2.47	0.08
A_p	93.42	93.64	94.37	18.67	18.01	18.01	0.64	0.53
N	850	835	865					

Table 8.5 Adverse Inpact: Black – White

	Δ		
Score	Form A	Form E	Form F
$S_{ar}^{(g)}$	−0.99	−0.92	−0.89
$S_{vo}^{(g)}$	−1.02	−0.87	−0.89
$S_{3d}^{(g)}$	−0.84	−0.77	−0.79
$S_{co}^{(n)}$	−0.70	−0.65	−0.61
$S_{nc}^{(q)}$	−0.70	−0.52	−0.52
A_p	−0.59	−0.59	−0.57

for new and old GATB forms. Six subtest variables were examined: $S_{ar}^{(g)}$, $S_{vo}^{(g)}$, $S_{3d}^{(g)}$, $S_{co}^{(n)}$, $S_{nc}^{(q)}$, and A_p. (See Chapter 5 for variable definitions.) These variables are monotonic functions of the six subtests scores. (For Form A, the variable A_p is a function of both Form Matching and Object Matching.)

Each numerical entry in Tables 8.5–8.7 represents

$$\Delta = \frac{\bar{X}_m - \bar{X}_M}{s_T},$$

where $\bar{X}_m$ = the minority group mean, $\bar{X}_M$ = the majority group mean, and s_T = the total group standard deviation. The individual sample statistics used to compute the adverse impact values are provided in Appendix A. For example, the '−.99' in Table 8.5 indicates that Blacks scored .99 standard deviation units lower than whites on Form A subtest $S_{ar}^{(g)}$, and was computed from means and sd values provided in Table 11.1 (Appendix A):

$$\Delta = \frac{18.35 - 26.79}{8.55} = -.99.$$

All adverse impact values can be recreated from values displayed in Appendix A, Tables 11.1–11.6.

In general, levels of adverse impact tended to be similar across the new and old forms—although some minor trends were evident. The new GATB Forms E and F tended to display slightly lower levels of adverse impact for Blacks. For Hispanics, the adverse impact statistics tended to possess greater variability than for other minority subgroups—probably because these statistics were a function of small samples ($N \approx 325$).

Table 8.6 Adverse Inpact: Hispanic – White

	Δ		
Score	Form A	Form E	Form F
$S_{ar}^{(g)}$	−0.65	−0.66	−0.51
$S_{vo}^{(g)}$	−0.82	−0.76	−0.66
$S_{3d}^{(g)}$	−0.29	−0.33	−0.23
$S_{co}^{(n)}$	−0.49	−0.50	−0.38
$S_{nc}^{(q)}$	−0.55	−0.35	−0.31
A_p	−0.21	−0.22	−0.13

Table 8.7 Adverse Inpact: Female – Male

	Δ		
Score	Form A	Form E	Form F
$S_{ar}^{(g)}$	−0.06	−0.07	−0.03
$S_{vo}^{(g)}$	0.17	−0.01	0.00
$S_{3d}^{(g)}$	−0.19	−0.17	−0.15
$S_{co}^{(n)}$	0.15	0.19	0.19
$S_{nc}^{(q)}$	0.43	0.44	0.45
A_p	0.26	0.30	0.33

Chapter 9
RELIABILITY ANALYSIS

A primary issue in the investigation of new GATB forms is that of precision. Several of the new test versions have fewer numbers of items than their original counterparts. Although fewer items may be offset by an increase in testing time, it is important to show that the new forms have sufficiently high levels of reliability relative to the old GATB forms. Lower reliability would lead to lower levels of validity.

9.1 Method

Four groups of the repeated-measures sample were used in this analysis. (See Table 3.2.) Data from Groups 1 and 2 were combined to form a sample of $N = 820$, and were used to compute the alternate form correlations between the old Forms A and B. These correlations are displayed in Table 9.1 for subtest, aptitude-score, and composite variables. (See Chapter 5 for a listing of variable definitions.) Groups 7 and 8 were combined to form a sample of $N = 870$, and were used to compute the alternate form correlations between the two new Forms E and F. These alternate form correlations are also displayed in Table 9.1.

Fisher's z-transformation was used to test the significance of the difference between the alternate form correlations of the new and old GATB forms. As described in Cohen and Cohen (1975, p. 50–51), the significance of the difference between two correlation coefficients obtained from two different random samples can be evaluated from the normal curve deviate

$$z = \frac{z'_n - z'_o}{\sqrt{\frac{1}{n_n-3} + \frac{1}{n_o-3}}} ,$$

where

$$z'_n = \frac{1}{2}\left[\ln(1 + r_n) - \ln(1 - r_n)\right] ,$$

$$z'_o = \frac{1}{2}\left[\ln(1 + r_o) - \ln(1 - r_o)\right] ,$$

and r_n is the alternate form correlation for the new test-variable (based on Forms E and F), r_o is the alternate form correlation for the old-test variable (based on Forms A and B), and n_n and n_o are the sample sizes for the groups used to compute the alternate form correlations ($n_n = 870$, $n_o = 820$). Normal deviates z were computed for each subtest, aptitude-score, and composite variable. The results are displayed in Table 9.1. Also displayed in Table 9.1 are the probability values associated with these normal deviates: $1 - \Phi(|z|)$, where Φ is the normal cumulative distribution function.

Table 9.1 Alternate Form Reliability Estimates

	Correlations		Significance Test	
Variable	r_n	r_o	z	p
$S_{ar}^{(g)}$	.800	.803	-.179	.429
$S_{ar}^{(n)}$	.800	.802	-.127	.450
$S_{vo}^{(g)}$	.846	.859	-.925	.177
$S_{vo}^{(v)}$	.850	.858	-.580	.281
$S_{3d}^{(g)}$	.829	.805	1.529	.063
$S_{3d}^{(s)}$	.832	.805	1.648	.050
$S_{co}^{(n)}$	.818	.846	-1.873	.031
$S_{nc}^{(q)}$	.778	.755	1.145	.126
$S_{ob}^{(p)}$	.823	.770	2.996	.001
A_g	.908	.886	2.284	.011
A_v	.850	.858	-.580	.281
A_n	.876	.884	-.704	.241
A_s	.832	.805	1.648	.050
A_p	.823	.824	-.049	.480
A_q	.778	.755	1.145	.126
C_{gvn}	.919	.913	.772	.220
C_{spq}	.893	.849	3.694	.000

Note. r_n = new form reliability
Note. r_o = old form reliability
Note. $p = 1 - \Phi(|z|)$

9.2 Results

The results indicate that the alternate form reliabilities of the new GATB forms are generally as high, or higher than those of the old GATB Forms A and B. This is very encouraging, since for the three power tests, subtest lengths were decreased. However, the increase in testing time may have added to the reliability of these power tests, offsetting the detrimental effects of shortening test-lengths. Only one of the comparisons displayed a significantly lower alternate form correlation for the new form; this was $S_{co}^{(n)}$ (Computation). However, the magnitude of the difference is small, and none of the composites which use Computation display a significantly lower new-form reliability estimate.

Chapter 10
VALIDITY ANALYSIS

This chapter address the third primary issue in the evaluation of the new forms, that of construct validity. It is highly desirable for the new and old GATB forms to measure identical or highly correlated constructs. The measurement of similar constructs would enable the validity of the new forms to be inferred from the large body of existing validity research conducted on the old forms of the GATB.

The construct-validity analysis is presented in two sections. The first section describes an analysis of the non-psychomotor subtest and composite variables based on the repeated-measures data. The second section addresses the construct validity of variables which enter into the job-family composites, specifically: the cognitive, perceptual, and psychomotor composites. These analyses are based on the psychomotor sample.

10.1 Non-Psychomotor Construct Validity

For this analysis, all eight groups of the repeated-measures sample were used. (See Table 3.2.) Data from Groups 3, 4, 5, and 6 were combined to form a sample of $N = 861$. For the purpose of these analyses, scores on Forms A and B were treated as an "Old-Test" variable—scores on Forms E and F were treated as a "New-Test" variable. Correlations between the old and new batteries (denoted by $r_{n,o}$) were obtained for subtest, aptitude, and composite variables. These correlations are displayed in Table 10.1. (Variable definitions are provided in Chapter 5.)

The alternate-form reliability estimates computed from Groups 1 and 2 (old forms) and Groups 3 and 4 (new forms) were used to obtain the disattenuated correlations between the new and old forms. These alternate form correlations were computed as described in Chapter 9. The disattenuated correlations were computed from the Classical Test Theory expression

$$\rho(\tau_n, \tau_o) = \frac{r_{n,o}}{\sqrt{r_n \times r_o}} .$$

These values are displayed in Table 10.1.

As indicated, all disattenuated correlations between old and new forms were extremely high, ranging from .905 to .982. Although changes were made in test format, time-limits, test-lengths, deletion of the Form Matching Subtest, change in scoring formulas, etc., these changes do not appear to have significantly altered the dimensionality of the battery. Because of the high correlations between the dimensions measured by the new and old forms, the large number of validity studies conducted on the old GATB forms can continue to provide useful data for inferring the validity of the new GATB forms.

Table 10.1 Disattenuated Correlations Between New and Old GATB Forms

Variable	$r_{n,o}$	$\rho(\tau_n, \tau_o)$
$S_{ar}^{(g)}$	.762	.951
$S_{ar}^{(n)}$	.765	.955
$S_{vo}^{(g)}$	.802	.940
$S_{vo}^{(v)}$	.801	.938
$S_{3d}^{(g)}$	.743	.910
$S_{3d}^{(s)}$	.743	.908
$S_{co}^{(n)}$	.806	.969
$S_{nc}^{(q)}$	.730	.952
$S_{ob}^{(p)}$	.721	.905
A_g	.857	.955
A_v	.801	.938
A_n	.864	.982
A_s	.743	.908
A_p	.748	.907
A_q	.730	.952
C_{gvn}	.887	.968
C_{spq}	.817	.938

10.2 Psychomotor Construct Validity

The five Job Family Composites are expected to play a central role in the future use of the GATB. Many or most examinees will be counselled and referred on a basis of their scores on these composites. Since four of the five composites are computed, in part, on a basis of psychomotor subtests, a complete evaluation of the new GATB forms should include these psychomotor subtests. If for example, the covariance between psychomotor and non-psychomotor subtests differed across the old and new forms, then the validity of the five Job-Family composites might be affected. That is, these five composites computed from the old forms might measure different traits than those computed from the new forms.

For this analysis, 534 subjects of the psychomotor-sample were used. As indicated in Table 3.4, each examinee was administered Form A (non-psychomotor), Form A (psychomotor), and Form F (non-psychomotor) portions of the GATB. For each examinee five scores were computed: (1) $C_{gvn}(A)$, (2) $C_{spq}(A)$, (3) $C_{gvn}(F)$, (4) $C_{spq}(F)$, (5) $C_{kfm}(A)$; where $C_{gvn}(A)$ denotes the cognitive composite computed from Form A subtests; $C_{gvn}(F)$ denotes the same cognitive composite computed from Form F subtests, etc. To compute the Form F composites, scores were transformed to Form A "standard-score equivalents" using the equating transformation (Chapter 6). The formulas given in Chapter 5 were applied to the aptitude scores to compute the five composites. Table 10.2 displays the correlations among these five variables. As indicated in the last row, the pattern of correlations between the psychomotor composite C_{kfm} and the cognitive and perceptual composites (C_{gvn} and

Table 10.2 Correlations among Cognitive, Perceptual, and Psychomotor Composites

Composite	$C_{gvn}(A)$	$C_{spq}(A)$	$C_{gvn}(F)$	$C_{spq}(F)$	$C_{kfm}(A)$
$C_{gvn}(A)$	1.0				
$C_{spq}(A)$	.73	1.0			
$C_{gvn}(F)$	.89	.61	1.0		
$C_{spq}(F)$	.71	.84	.69	1.0	
$C_{kfm}(A)$	.27	.44	.23	.42	1.0

C_{spq}) appear to be similar* across Forms A and F (.27 vs .23, .44 vs .42). These similar patterns provide some assurance that the same relations among cognitive, perceptual, and psychomotor composites hold for both the new and old GATB forms. These results taken in conjunction with the high disattenuated correlations between the new and old cognitive and perceptual composites (.97 and .94, respectively; Table 10.1) suggest that the dimensions measured by the new and old job-family composites (which are linear combinations of C_{gvn}, C_{spq}, and C_{kfm}) will also be very highly correlated.

*The hypothesis that these pairs of correlations were significantly different was tested using a confirmatory factor model where Φ was set equal to the correlation matrix of the observed variables, and by constraining $\phi_{51} = \phi_{53}; \phi_{52} = \phi_{54}$. Based on a chi-square difference test, these pairs of correlations did not differ significantly ($\chi^2 = 4.07$, $df = 2$ $p = .13$).

Chapter 11
APPENDIX A: DESCRIPTIVE SUBGROUP STATISTICS

Table 11.1 Descriptive Statistics by Subgroup: $S_{ar}^{(g)}$

Subgroup	Form A			Form E			Form F		
	N	X	s	N	X	s	N	X	s
White	1178	26.79	8.62	1112	26.77	8.46	1154	26.67	8.80
Black	996	18.35	6.59	960	18.78	7.27	974	18.74	7.38
Hispanic	320	21.23	6.73	326	21.01	7.26	325	22.09	7.83
Other	158	22.63	7.78	146	23.09	8.13	142	24.02	8.73
Male	1460	22.93	8.66	1380	23.07	8.86	1369	23.09	8.97
Female	1192	22.43	8.41	1164	22.49	8.42	1226	22.84	8.88
Total	2652	22.70	8.55	2544	22.81	8.67	2595	22.97	8.93

Table 11.2 Descriptive Statistics by Subgroup: $S_{vo}^{(g)}$

Subgroup	Form A			Form E			Form F		
	N	X	s	N	X	s	N	X	s
White	1178	60.12	5.77	1112	60.20	5.92	1154	60.10	5.61
Black	996	54.05	4.54	960	54.90	5.07	974	54.79	5.18
Hispanic	320	55.26	4.93	326	55.59	5.52	325	56.15	5.39
Other	158	56.08	5.29	146	56.68	5.75	142	56.68	5.67
Male	1460	56.56	5.80	1380	57.43	6.07	1369	57.41	6.03
Female	1192	57.56	6.03	1164	57.38	6.12	1226	57.44	5.88
Total	2652	57.01	5.93	2544	57.41	6.09	2595	57.42	5.96

Table 11.3 Descriptive Statistics by Subgroup: $S_{3d}^{(g)}$

Subgroup	Form A			Form E			Form F		
	N	X	s	N	X	s	N	X	s
White	1178	16.76	5.75	1112	16.85	5.75	1154	16.64	5.60
Black	996	11.98	4.61	960	12.46	4.89	974	12.13	4.90
Hispanic	320	15.13	5.13	326	14.95	5.20	325	15.34	5.30
Other	158	15.30	5.44	146	13.90	5.82	142	15.13	5.58
Male	1460	15.15	5.92	1380	15.23	5.88	1369	15.10	5.66
Female	1192	14.10	5.33	1164	14.25	5.50	1226	14.27	5.69
Total	2652	14.68	5.69	2544	14.78	5.73	2595	14.70	5.69

Table 11.4 Descriptive Statistics by Subgroup: $S_{co}^{(n)}$

	Form A			Form E			Form F		
Subgroup	N	$\bar{X}$	s	N	$\bar{X}$	s	N	$\bar{X}$	s
White	1178	79.69	11.69	1112	79.35	11.51	1154	79.07	11.44
Black	996	71.17	11.39	960	71.51	11.75	974	71.64	11.99
Hispanic	320	73.79	10.74	326	73.35	10.86	325	74.49	11.21
Other	158	76.16	12.16	146	75.41	12.17	142	76.67	11.67
Male	1460	74.73	12.39	1380	74.37	12.18	1369	74.49	12.33
Female	1192	76.59	11.75	1164	76.62	11.91	1226	76.78	11.76
Total	2652	75.57	12.14	2544	75.40	12.11	2595	75.58	12.12

Table 11.5 Descriptive Statistics by Subgroup: $S_{nc}^{(q)}$

	Form A			Form E			Form F		
Subgroup	N	$\bar{X}$	s	N	$\bar{X}$	s	N	$\bar{X}$	s
White	1178	109.99	16.18	1112	108.51	15.65	1154	108.25	16.39
Black	996	98.88	13.61	960	100.17	15.58	974	99.87	14.37
Hispanic	320	101.29	14.05	326	102.98	15.17	325	103.30	15.38
Other	158	105.75	16.40	146	104.05	16.08	142	107.08	16.31
Male	1460	101.42	15.10	1380	101.13	15.00	1369	101.03	14.76
Female	1192	108.31	15.99	1164	108.27	16.37	1226	108.21	16.48
Total	2652	104.52	15.88	2544	104.40	16.04	2595	104.42	16.00

Table 11.6 Descriptive Statistics by Subgroup: A_p

	Form A			Form E			Form F		
Subgroup	N	$\bar{X}$	s	N	$\bar{X}$	s	N	$\bar{X}$	s
White	1178	105.59	19.89	1112	106.28	18.96	1154	105.94	19.00
Black	996	93.79	18.67	960	94.65	18.92	974	94.76	18.87
Hispanic	320	101.30	19.21	326	101.92	19.35	325	103.33	19.29
Other	158	103.01	20.51	146	99.24	20.84	142	102.35	20.14
Male	1460	98.17	20.10	1380	98.20	19.48	1369	98.16	19.30
Female	1192	103.33	19.80	1164	104.17	19.73	1226	104.63	19.64
Total	2652	100.49	20.12	2544	100.93	19.82	2595	101.22	19.72