Technical Manual:
2002 Series GED Tests

GENERAL EDUCATIONAL DEVELOPMENT
TESTING SERVICE

American Council on Education

GED Testing Service
One Dupont Circle NW, Suite 250
Washington, DC 20036-1163
(202) 939-9490
Fax: (202) 659-8875
www.GEDtest.org

The *Technical Manual: 2002 Series GED Tests* is a publication of the GED Testing Service, a national program of the American Council on Education.

Additional copies of this publication are available for purchase online at www.acenet.edu/bookstore for $45.00 per copy, plus shipping and handling. Copies may also be purchased by contacting:

ACE Fulfillment Service
Department 191
Washington, DC 20055-0191
Phone: (301) 632-6757
Fax: (301) 843-0159

When ordering, please specify Item #311925.

Table of Contents

Acknowledgments ..... **v**

Preface ..... **vi**

Chapter 1: Introduction **1**
ABOUT THE GED TESTING PROGRAM 1
GEDTS Vision 1
GEDTS Mission 1
GEDTS Values 1
PURPOSE OF THE GED TESTS 1
Proper Uses of GED Test Scores and Credential 2
HISTORY OF THE GED TESTS 3
OVERVIEW OF THE 2002 SERIES GED TESTS 5
Item Format 6
Accommodations 6
Test Administration 7
Time Limits 7
Scoring Procedures 7
The 2002 Series GED Writing Test Official Essay Scoring Guide: The Four-Point Scale 8
Explanation of Standard Scores 10
Passing Standard 11
The GED Tests and the Compensatory Model 12
Meaning of the Average GED Test Battery Standard Score 13
Test Security 14

Chapter 2: GED Tests Specifications and Forms Development **15**
DESCRIPTION OF THE GED TESTS 15
Testing Contexts 15
Readability 16
Cognitive Levels 16
SPECIFICATIONS FOR THE 2002 SERIES TESTS 17
Language Arts, Writing Test, Part I 17
Language Arts, Writing Test, Part II 19
Social Studies Test 20
Science Test 23
Language Arts, Reading Test 26

Mathematics Test 28
DEVELOPMENT AND SELECTION OF MULTIPLE-CHOICE ITEMS 31
Item Writer Representation 31
Item Development Procedures 31
Final Forms Development 32
DEVELOPMENT OF PART II (ESSAY) OF LANGUAGE ARTS, WRITING TEST 34
Overview 34
Development of Essay Topics 35
Standards for Scoring Essays 35
CANADIAN VERSION 36
SPANISH-LANGUAGE GED TESTS 36
FRENCH-LANGUAGE GED TESTS 36

Chapter 3: Standardization and Norming, Scaling, and Equating 37
STANDARDIZATION AND NORMING 37
Sample 38
SCALING OF MULTIPLE-CHOICE TESTS 40
EQUATING OF MULTIPLE-CHOICE TESTS 41
2001 Standardization Study 41
Equating in Subsequent Years 42
SCALING AND EQUATING OF THE LANGUAGE ARTS, WRITING TEST 42
Deriving and Scaling the Writing Composite Score 42
Equating Forms of the Language Arts, Writing Test 44

Chapter 4: Reliability 45
RELIABILITY OF GED TEST SCORES 45
Internal Consistency Reliability 45
Standard Error of Measurement 46
K-R 20 and SEM Results for the GED Tests 47
Conditional Standard Errors of Measurement 50
Alternate-Form Reliability 52
RELIABILITY OF ESSAY SCORES ON THE LANGUAGE ARTS, WRITING TEST 56
Site Monitoring and Scoring Stability 56
Essay Score Inter-rater Reliability 58
RELIABILITY OF THE LANGUAGE ARTS, WRITING TEST COMPOSITE SCORE 58
Weighting the Essay and Multiple-Choice Scores 59
Nunnally's Formula 59
DECISION CONSISTENCY 60
Decision Consistency Based on Content Area Tests 60
Decision Consistency Based on Entire GED Test Battery 62

Chapter 5: Validity 63
PURPOSE OF THE GED TESTS 63
EVIDENCE BASED ON TEST CONTENT 64
Development of GED Tests Specifications 64
Content Review of GED Tests and Items 65
GED Tests Content and Workplace Skills 65
EVIDENCE BASED ON INTERNAL STRUCTURE 67
Factor Analyses 67
Differential Item Functioning 70
EVIDENCE BASED ON RELATIONS WITH OTHER VARIABLES 75
Correlations Among Content Area GED Tests 75
The Relationship Between GED Test Scores and High School Grades 75
The Relationship Between GED Test Scores and Prior Instruction 78
Predictive Utility of the GED Credential 82
Comparisons Between GED Credential Recipients and Graduating High School Seniors 82
VALIDITY ISSUES SPECIFIC TO THE LANGUAGE ARTS, WRITING TEST 83
Essay Scoring Sessions 84
How the Scoring Standards Are Defined 84
Selecting and Training Chief Readers and Essay Readers 84
Site Certification 85
Site Monitoring 85

Chapter 6: Accommodations for GED Examinees with Disabilities 87
AVAILABLE ACCOMMODATIONS FOR GED EXAMINEES 87
EFFECT OF TEST ACCOMMODATIONS ON GED TEST PERFORMANCE 88
Overview 88
Method 89
Results 91
Discussion 94

Chapter 7: English-Language Canadian GED Tests 95
OVERVIEW 95
HISTORY OF THE ENGLISH-LANGUAGE CANADIAN GED TESTS 95
TEST SPECIFICATIONS AND DEVELOPMENT 95
STANDARDIZATION AND NORMING 96
SCALING AND EQUATING 97
RELIABILITY 97
K-R 20 and SEM Results 97
Conditional Standard Errors of Measurement 99
Reliability of Essay Scores on the Language Arts, Writing Test 99
Decision Consistency 99

VALIDITY 101
Evidence Based on Test Content 101
Evidence Based on Internal Structure 101
Evidence Based on Relations with Other Variables 103

Chapter 8: French-Language GED Tests 109
OVERVIEW 109
HISTORY OF THE FRENCH-LANGUAGE GED TESTS 109
TEST SPECIFICATIONS AND DEVELOPMENT 109
STANDARDIZATION AND NORMING 110
SCALING AND EQUATING 110
RELIABILITY 111
K-R 20 and SEM Results 111
Conditional Standard Errors of Measurement 112
Reliability of Essay Scores on the Language Arts, Writing Test 112
Decision Consistency 114
VALIDITY 114
Evidence Based on Test Content 115
Evidence Based on Relations with Other Variables 115

Chapter 9: Spanish-Language GED Tests 121
OVERVIEW 121
HISTORY OF THE SPANISH-LANGUAGE GED TESTS 121
TEST SPECIFICATIONS AND DEVELOPMENT 121
STANDARDIZATION AND NORMING 122
SCALING AND EQUATING 123
RELIABILITY 123
K-R 20 and SEM Results 123
Conditional Standard Errors of Measurement 124
Reliability of Essay Scores on the Language Arts, Writing Test 125
Decision Consistency 127
VALIDITY 128
Evidence Based on Test Content 128
Evidence Based on Relations with Other Variables 128

References 135

Appendices 138

Acknowledgments

The GED Testing Service® (GEDTS), a program of the American Council on Education®, is pleased to issue the *Technical Manual: 2002 Series GED Tests*. We wish to acknowledge the work of psychometricians Carol Ezzelle and J. Carl Setzer, who authored this technical manual. GED Testing Service colleagues who provided comments that improved this publication include Margaret Patterson, Martin Kehe, Ruth Mensel, Yung-chen Hsu, Wei Song, Chhaya Rao, Anne Guison-Dowdy, Jizhi Zhang, Jeanna Grimes, and Usha Paulson. We would also like to thank American Council on Education staff who edited and designed this publication.

Preface

This manual was written to provide technical information regarding the 2002 Series GED Tests. Throughout this manual, documentation is provided regarding the development of the GED Tests, data collection activities, as well as reliability and validity evidence. The purpose of this manual is to provide evidence that the GED Tests are technically sound.

This manual is made up of nine chapters, which include the following information:

Chapter 1: Introduction to the GED Tests and an overview of the GED testing program, including the purposes of the tests and proper uses of test scores.

Chapter 2: Test specifications and development of the GED Tests.

Chapter 3: The standardization process, including the norming, scaling, and equating processes.

Chapter 4: The reliability of the English-language U.S. GED test scores.

Chapter 5: The validity of English-language U.S. GED test scores.

Chapter 6: Information and research on accommodations.

Chapter 7: The development of the English-language Canadian GED Tests, with reliability and validity evidence.

Chapter 8: The development of the French-language GED Tests, with reliability and validity evidence.

Chapter 9: The development of the Spanish-language GED Tests, with reliability and validity evidence.

This manual was written for anyone who is interested in (a) the background of the GED testing program, (b) understanding how the GED Tests are developed and scored, (c) the statistical characteristics of the GED Tests, and (d) knowing more, in general, about the GED testing program. Individuals interested in additional information are encouraged to contact the GED Testing Service (GEDTS) at www.GEDtest.org.

Chapter 1: Introduction

ABOUT THE GED TESTING PROGRAM

The General Educational Development Testing Service (GEDTS) is a program of the American Council on Education (ACE). As such, our mission, vision, and values are tied to those of ACE, and we share ACE's core values of inclusiveness and diversity. We recognize the responsibility of those in the educational community to contribute to society, and we embrace the belief that widespread access to postsecondary education, particularly for those adult learners who seek lifelong learning, is the cornerstone of a democratic society.

GEDTS Vision

In an ideal society, everyone would graduate from high school. Until that becomes a reality, we, the General Educational Development Testing Service, will offer the opportunity to earn a high school equivalency diploma so that individuals can have a second chance to advance their educational, personal, and professional aspirations.

GEDTS Mission

As a nonprofit program of ACE, GEDTS stands as the only legitimate and time-honored architect of the Tests of General Educational Development (GED Tests) that certify the high school–level academic achievement of national and international non–high school graduates. In collaboration with key partners, we develop, deliver, and safeguard our tests; we analyze the testing program and its participants; and we develop policies, procedures, and programs to ensure equal access to our tests.

GEDTS Values

The integrity of GEDTS and its products (GED Tests) rests on our commitment to excellence, diversity, inclusiveness, educational opportunities, and lifelong learning, as reflected in our proactive approach to developing collaborative solutions, our research-based decision making, and our timely support to the people we serve.

PURPOSE OF THE GED TESTS

The GED Tests began as a way for military personnel returning from World War II to demonstrate that they had the knowledge and skills necessary for employment and higher education. Since its beginning in 1942, the GED testing program has grown and evolved. There have been three previous series of the GED Tests: 1942, 1978, and 1988. Changes made in each series were the result of the identification of specific areas of need or assessment that would strengthen the tests and provide evidence of test score validity and credibility in a changing world.

The current GED Tests measure academic skills and knowledge associated with a high school program of study with an increased emphasis on the workplace and higher education. The GED test battery comprises five content area tests:

- Language Arts, Reading
- Language Arts, Writing
- Mathematics
- Science
- Social Studies

The GED Tests have been designed to provide an opportunity for adults who did not complete a formal high school program to certify their attainment of high school–level academic knowledge and skills and earn their jurisdictions' high school–level equivalency credential, diploma, or certificate. Thus, the intended use of the GED credential is similar to that of a high school diploma—to qualify for jobs and job promotions, to enable further education and training, and to enhance an adult's personal satisfaction.

Upon taking the GED test battery, standard scores for each test are provided to each GED examinee. These standard scores are used to compare an examinee's performance on a test to the performance of graduating high school seniors who took the test. Percentile ranks, which are also reported to each examinee, indicate the percentage of graduating high school seniors who earned scores at or below that test score. Test scores are also compared to a passing standard. It is inferred that those examinees who meet the GED test battery passing standard perform as well as approximately 60 percent of graduating high school seniors. The details of GED test scores and the passing standard are provided below.

The acceptance of the GED test scores as a valid measure for awarding a high school equivalency credential is fundamental to the success of the GED testing program. All 50 states, the District of Columbia, eight insular areas, 13 Canadian provinces and territories, and various federal institutions such as U.S. military bases, the Federal Bureau of Prisons, Michigan prisons, and Veterans Affairs hospitals use scores earned on the GED Tests as a basis for awarding high school equivalency credentials. The testing program serves more than 700,000 examinees annually at more than 3,400 Official GED Testing Centers. A recent national survey confirms that most U.S. employers and training programs (96 percent) consider applicants who hold a GED credential in the same manner as those who hold traditional high school diplomas (Society for Human Resource Management, 2002). In addition, for admissions purposes, almost all U.S. colleges and universities (98 percent) accept GED score reports as being equivalent to high school transcripts (Annual Survey of Colleges, 2007).

The GED Tests have been designed to measure only the content and cognitive aspects related to a high school curriculum. The test specifications are thus centered on content and cognitive facets that are reflected within a traditional high school program of study. Therefore, GED test scores should not be used to make inferences regarding any non-cognitive aspects often developed by attending high school, such as creativity, team work, planning and organization, ethics, leadership, self-discipline, and socialization. In addition, ACE policy clearly states that the GED Tests should not be used to validate high school diplomas and does not permit the tests to be administered to high school students still enrolled in school or high school graduates, except under special circumstances. Employers and postsecondary institutions are explicitly forbidden to use the GED Tests to verify the achievement level of high school graduates.

Proper Uses of GED Test Scores and Credential

The GED Tests were developed to provide adults with an opportunity to obtain a high school equivalency credential. As such, the intended uses of the test scores and credential are similar to those for which a high school diploma is appropriate. More specifically, the GED Tests were designed to measure how much an examinee knows in relation to a population of graduating high school seniors regarding national and state high school curricula standards. Thus, the information in this manual directly relates to providing evidence regarding the appropriateness of using the GED Tests for this purpose. Information provided in this manual should not be extended to other purposes beyond that which is stated by GEDTS. For example, GED test scores should not be interpreted as grade point average equivalents, nor should they be used to establish concordance with extant data sources such as the Scholastic Aptitude Test (SAT), the ACT Assessment, or the Comprehensive Adult Student Assessment System (CASAS).

HISTORY OF THE GED TESTS

The first GED Tests were developed in 1942 to measure the major outcomes and concepts generally associated with four years of high school education. Initiated by the United States Armed Forces Institute (USAFI), the original tests were administered only to military personnel so that returning World War II veterans could more easily pursue their educational, vocational, and personal goals.

The USAFI examination staff, composed of civilian testing experts, worked with an advisory committee established with the support and cooperation of ACE, the National Association of Secondary School Principals, and regional U.S. accrediting associations. Lindquist (1944) paved the way for the GED Tests by establishing a philosophical and technical basis for exam-based equivalency. In 1945, ACE established the Veterans' Testing Service (VTS; predecessor of today's GED Testing Service). The VTS took over the development and administration of the GED Tests and focused on helping World War II veterans pursue educational and vocational goals without returning them to the classroom.

The opportunity to document the attainment of high school–level academic skills served as a significant aid to the many service members whose academic careers had been disrupted during the war. During the 1950s, it became apparent that civilians could also benefit from the program—a need that ACE undertook to fulfill. New York was the first state to allow nonveteran adults to take the GED Tests in late 1947. In 1955, policies regarding administration of the tests in federal correctional and health institutions were modified. By 1959, more nonveterans than veterans were taking the GED Tests. With the growth of the high school equivalency program, ACE made the decision in 1960 to transfer the college-level GED Tests to the Educational Testing Service (ETS). Those tests that ETS developed are known today as part of the College Board's College-Level Examination Program® (CLEP).

From 1945 to 1963, the program was administered by the VTS. In 1958, a policy to allow overseas testing U.S. civilians and foreign nationals was approved. In 1963, in recognition of the transition to a program chiefly for nonveteran adults, the name was changed to GED Testing Service. To serve all qualified examinees equally, the Commission on Accreditation of Service Experiences approved English-language versions in audio, Braille, and large-print formats in 1964. In addition, Nova Scotia became the first Canadian province to offer GED testing to civilians in 1969, and in 1970, the first English-language Canadian version of the GED Tests was published. In 1973, GEDTS reached a milestone when California became the last state to adopt a uniform acceptance of the GED Tests. (For more history on the GED Tests, see Mullane [2001].)

The five original GED Tests in use from 1942 to 1978 were titled:

Test 1: Correctness and Effectiveness of Expression
Test 2: Interpretation of Reading Materials in the Social Studies
Test 3: Interpretation of Reading Materials in the Natural Sciences
Test 4: Interpretation of Literacy Materials
Test 5: General Mathematical Ability

The entire battery took 10 hours to administer. (For more information about the content of the first generation of GED Tests, see American Council on Education [1964].) In the 1970s it became apparent that the effects of changed secondary curricula and, perhaps, changed attitudes toward education among the general public necessitated a review of the specifications of the GED Tests. This review resulted in a thorough revision of the first series of GED Tests.

The second series of GED Tests, introduced in 1978, was based on test specifications defined in the mid-1970s by committees of high school curriculum specialists. Among the major changes were the development of a Reading Skills test to replace Test 4: Interpretation of Literary Materials, and the reduction of the reading load in the Science and Social Studies tests. In addition, "concept" items were developed to make up one-third of the Science and Social Studies Tests. These items required much less reading than the reading comprehension items, which dominated previous tests, but assumed that examinees had some prior knowledge in science and social studies. In addition, Test 1: Correctness and Effectiveness of Expression was replaced by a Writing Skills test. The Mathematics test included more practically-oriented test items.

The second series of GED Tests originally required six hours of test administration time. On the basis of research (Whitney & Patience, 1981), the Commission on Accreditation of Service Experiences in 1981 increased the time limits for the Writing Skills Test from 60 minutes to 75 minutes and for the Mathematics

Test from 60 minutes to 90 minutes. This series of tests, used from 1979 through 1987, consisted of the following titles:

Test 1: The Writing Skills Test
Test 2: The Social Studies Test
Test 3: The Science Test
Test 4: The Reading Skills Test
Test 5: The Mathematics Test

These tests retained the emphasis on demonstrating the designated high school outcomes but introduced "real-life" contexts into many of the test items. They also introduced many reading materials likely to be encountered in an adult's daily life (such as schedules and newspaper articles). (For further information about the development of the second-generation tests, including detailed descriptions, see Patience and Whitney [1982].)

The development of the third series of GED Tests, used from 1988 through 2001, began in November 1982 in order to ensure that the GED Tests addressed and measured the educational academic outcomes expected of graduating high school seniors during the late 1980s and early 1990s. The Tests Specifications Committee offered recommendations for the entire GED test battery centered around five major themes. First, the new GED Tests should require examinees to demonstrate their higher-level thinking abilities and problem-solving skills. Second, the new GED Tests should include a clear emphasis on the relationship of the skills tested to aspects of the world of work. Third, awareness of the role and impact of computer technology should be represented in the new GED Tests. Fourth, certain consumer skills should be addressed in the context of many of the new GED Tests. Fifth, the new GED Tests should use contexts that adult examinees would recognize and include stimulus materials should relate to aspects of everyday life in all five tests. The GED Tests introduced in 1988:

- Required a direct writing sample.
- Demanded more highly developed levels of critical thinking.
- Reflected many roles of adults.
- Acknowledged the sources of change affecting individuals and society.
- Contained contexts that adult examinees would recognize as relevant to daily life.

This third series of GED Tests required seven hours and 45 minutes of test administration time. The official titles of the five separate subject tests were:

Test 1: Writing Skills
Test 2: Social Studies
Test 3: Science
Test 4: Interpreting Literature and the Arts
Test 5: Mathematics

OVERVIEW OF THE 2002 SERIES GED TESTS

In the late 1990s, GEDTS undertook a study comparing national and state standards in English language arts, mathematics, science, and the disciplines within social studies to survey what graduating students should know and be able to do (American Council on Education, 1999). The major purpose of the study was to inform the education community and the public of the development of the fourth series of GED Tests. By identifying the common elements among national and state standards and aligning the test specifications to these standards, GEDTS provided support for the claim that the GED test scores can be used to measure what adult examinees know and are able to do relative to high school seniors who are completing a traditional four-year program of study.

This fourth and current series of GED Tests (English-language U.S. and Canadian versions) was released in 2002.[1] The official titles of the five separate subject tests are:

Language Arts, Writing
Social Studies
Science
Language Arts, Reading
Mathematics

During the first year of administration, three different test forms of the English-language U.S. GED Tests were available for each of the content area tests. These forms were labeled as IA, IB, and IC. During subsequent years, eight new forms (ID through IK) were introduced, for a total of 11 operational forms. The purposes of introducing new forms include increasing test security and maintaining an alignment with curricula and standards. Each new form was created using the same test specifications developed for the 2002 Series GED Tests.

The development of the 2002 series began with an extensive review of the test specifications and realignment with the current national curriculum and standards. Next, item-writing procedures were performed until an adequate item pool was available for forms development. Item tryout studies were performed using nationally representative samples of high school seniors. In addition, graduating U.S. high school seniors were administered the GED Tests in 2001 (using forms IA, IB, and IC) for the purpose of developing national norms—to which future GED examinees' scores would be compared—and to establish the passing standard. One of each of these three forms in each content area was used as an anchor form for each subsequent year. Once the national norms and passing standard were set, the 2002 series became operational and GED examinees began taking the new test forms.

Because the bulk of the jurisdictions are within the United States, the majority of those who take the GED Tests are administered the English-language U.S. version.[2] However, several jurisdictions have invested in the development of additional versions of the test, including an English-language Canadian version, a Spanish-language version (based on the U.S. version), as well as a French-language version (based on the Canadian version).[3] The purpose of these versions of the GED Tests is to provide the same opportunity for obtaining a high school equivalency credential, certificate, or diploma to adults in Canada and those who speak French or Spanish as a primary language. The development of these tests was very similar to the development of the English-language U.S. GED Tests. However, these tests serve different populations of GED examinees and are normed on different groups of graduating high school seniors (see Chapters 3, 7, 8, and 9).

GEDTS also offers an English as a Second Language (ESL) Test to jurisdictions that wish to test the English-language reading comprehension proficiency of an examinee. The ESL test is sometimes administered by jurisdictions to those examinees who take the Spanish- or French-language GED Tests.

---

[1] In addition to the standard print editions, large-print, audiocassette, and Braille editions were introduced for the 2002 Series at the same time. These editions were developed for those in need of special accommodations. Further details on accommodations are provided in Chapter 6.

[2] Throughout the remainder of this manual, the term *jurisdiction* is used to refer to an entity such as a U.S. state, U.S. insular area, Canadian province or territory, U.S. military facility, correctional institution, or VA hospital that administered a GED testing program.

[3] The French- and Spanish-language editions were first introduced in 2004. In both cases, only two forms were administered during the first year (IA and IC). Two additional forms were subsequently introduced for each of these two editions (ID and IE).

Details of the ESL Test are not provided in this technical manual, but will be addressed in a forthcoming publication.

In addition to the operational test versions and editions mentioned above, GEDTS develops Official GED Practice Tests (OPT) for each of the five content area tests, available for purchase through Steck-Vaughn. Uses of the OPT vary by jurisdiction. For example, some jurisdictions require a GED examinee to take and meet specified score requirements on the OPT prior to attempting the operational GED Tests. Other jurisdictions use the OPT only for training purposes. Details of the OPT, including reliability and validity evidence, are provided in the Official GED Practice Tests Administrator's Manual (GED Testing Service, 2002b).

Item Format

Each of the content area tests contains multiple-choice items and each multiple-choice item contains five response options. The Language Arts, Writing Test also includes an essay section. The Mathematics Test includes alternate format items that require a grid to capture some answers (20 percent of items are of this type), which are ultimately scored as right or wrong. The multiple-choice and alternate format items are scored electronically; essays are hand-scored by expert readers. Both types of items are scored at various locations that undergo a strict certification process. Table 1.1 displays the number of multiple-choice and essay items on each test.

TABLE 1.1
NUMBER OF MULTIPLE-CHOICE AND ESSAY/ALTERNATE FORMAT ITEMS PER TEST

	Multiple-Choice Items	Essay/Alternate Format Items
Language Arts, Writing	50	1 essay
Social Studies	50	0
Science	50	0
Language Arts, Reading	40	0
Mathematics	40	10 alternate format

Accommodations

GED Testing Service has established procedures for adults with documented disabilities to obtain accommodations for the GED Tests. GEDTS encourages individuals who may benefit from accommodations to take advantage of the opportunities available to them (e.g., large print, audiocassette, Braille editions) through the GED testing program. Accommodations are provided for adults with physical, learning, and psychological disabilities, as well as those with attention-deficit/hyperactivity disorder.

Individuals with disabilities must be able to provide adequate documentation and must request accommodations through their local Official GED Testing Center. They are required to submit appropriate forms (based on the type of disability) that have been completed by certified professionals. Individuals with disabilities may qualify for one or more of the following accommodations based on documentation and recommendation from certified professionals:

- Audiocassette edition.
- Braille edition.
- Large-print edition (no documentation required).
- Vision-enhancing technologies.
- Use of video equipment for examinees who are deaf or hard-of-hearing in composing the Language Arts, Writing essay.
- Use of a talking calculator or abacus.
- Certified sign-language interpreter; use of a scribe.
- Extended time; supervised extra breaks.
- Use of a private room.
- One-on-one testing at a health facility.
- Other accommodations as warranted, based on individual needs.

Approval for accommodations and use of special editions for adults with disabilities must be obtained through an accommodations request process. However, any adult may request to take the large-print edition of the tests under normal time limits. The English-, Spanish-, and French-language GED Tests are available in large-print and audiocassette editions. The English- and Spanish-language GED Tests are available in Braille editions. Additional information on accommodations is provided in Chapter 6.

Test Administration

Like many other testing programs, the GED Tests are administered at local testing centers. However, the GED Tests are unique in that the GED credential is awarded by the participating jurisdiction. Therefore, the administration of the GED Tests is a shared responsibility between participating jurisdictions and GEDTS. Standards and policies for the GED testing program have been established by the GED Advisory Committee and are detailed in the *GED Examiner's Manual for the Tests of General Educational Development* (GEDTS, 2002a) and the *GED Testing Service Policies and Procedures Manual* (GEDTS, 2008b). This manual states that

> [t]he proper administration, supervision, and the integrity of the GED testing program are joint responsibilities of participating jurisdictional departments or ministries of education, the contracting agencies, and the GED Testing Service. In the case of U.S. federal correctional facilities and military installations, the GED testing program is the joint responsibility of the federal agency and the GED Testing Service. (p. 1)

Time Limits

The time limits for each of the test versions are provided in Table 1.2.

TABLE 1.2
TIME LIMITS APPLIED TO THE GED TESTS

	English	Spanish/French
Language Arts, Writing, Part I (multiple-choice)	75 minutes	80 minutes
Language Arts, Writing, Part II (essay)	45 minutes	45 minutes
Language Arts, Reading	65 minutes	70 minutes
Social Studies	70 minutes	75 minutes
Science	80 minutes	85 minutes
Mathematics: Part I	45 minutes	50 minutes
Mathematics: Part II	45 minutes	50 minutes

Time limits for the operational forms were determined based on item tryout studies (see Chapter 3). The time limits were set such that 90 percent of examinees could complete the test in the allotted time.

When accommodations are provided to examinees with proper documentation, time limits may be modified as necessary. The *GED Examiner's Manual* (GEDTS, 2002a) provides details for test administrators regarding modifications to time limits. For example, those persons using either the Braille or audiocassette edition are generally provided twice as much time for completion. In general, those who take the large-print edition are given the standard amount of completion time, unless proper documentation provides sufficient evidence that additional time is warranted.

Scoring Procedures

Historically, GEDTS has used classical test theory in the development and scoring of the GED Tests. In January 2000, the testing service convened a psychometric panel to discuss the possibility of implementing item response theory for the 2002 test series. In addition, data from a study designed to determine the potential difference in the numbers of examinees passing the GED test battery under classical test theory scoring (using raw scores) and item response theory scoring (pattern scoring) were presented. Following a lengthy and comprehensive discussion, the psychometric panel concluded that, while item response theory could provide more information on examinees' ability levels, the cost-benefit analysis of implementing item

response theory did not outweigh the benefits of continuing to employ classical test theory at the current time. Item response theory will be further considered during the development of the 2012 Series GED Tests.

The scoring of the GED Tests is decentralized. That is, each jurisdiction is responsible for accurate scoring of the tests. However, all Official Scoring Sites adhere to the same scoring standards developed by GEDTS. Scoring sites must undergo a strict certification procedure prior to becoming operational and additional site monitoring occurs at various times.

The specific instructions for scoring the GED Tests are listed in the *GED Examiner's Manual* (GEDTS, 2002a). The multiple-choice and essay sections of the GED Tests must be scored by a GEDTS-certified Official Scoring Site. The multiple-choice sections are electronically scored by one of several Official Scoring Sites; in 2008, there were 19 sites that were certified by GED Testing Service to score multiple-choice sections of the GED Tests. The essay section is scored holistically by expert readers; in 2008, 18 sites were certified by GED Testing Service to perform essay scoring.

Essays are scored on a four-point scale, with 1 being the lowest possible score. An examinee who earns a score of less than 2 on the essay must retake both parts of the Language Arts, Writing Test. If an examinee leaves the essay blank, writes to a topic different from the assigned topic, or produces illegible handwriting, a score for the essay will not be generated, and the examinee must then retake both parts of the test. The essay accounts for 35 percent of the Language Arts, Writing Test standard score (see Chapter 3 for details on the weighting of the Language Arts, Writing Test).

Each essay is scored by two GED Testing Service-certified readers. Readers score essays holistically, a process by which each essay is evaluated on the basis of its overall effectiveness. The readers consider all of the elements of the essay and do not weigh specific features in order to arrive at a score for the essay. Errors that affect the overall effectiveness of the essay will influence the score an essay receives. For example, a well-written essay that establishes clear organization and achieves coherent development, with specific and relevant details and appropriate word choices, would remain effective despite an occasional spelling or punctuation error. However, numerous sentence or spelling errors in an essay could make it difficult for the reader to follow or understand the writer's ideas. This would result in a less effective essay and, thus, a lower score.

The 2002 Series GED Writing Test Official Essay Scoring Guide: The Four-Point Scale

The 2002 Series GED Writing Test Official Essay Scoring Guide describes the general features of essays at the different points on a four-point scale. The score scale used by each reader ranges from 1 (low) to 4 (high).

When holistically scoring essays, each reader assigns a score to an essay. The two readers' scores are then added, resulting in a range from two to eight, and then divided by two. If the two readers differ by more than one point, a Chief Reader also reads and scores the essay.[4] The Chief Reader's score is averaged with the original score he or she feels is more appropriate. Because the four-point scale is an even-numbered scale, there is no midpoint. The lack of a midpoint forces readers away from a natural tendency to drift toward the middle.

Score Definitions

Under each score point on the essay scoring guide, a statement describes the type of writing found at that level. These statements are directed toward the reader of the essay and reinforce that writing is part of a two-way communication process.

4 – Effective - Reader understands and easily follows the writer's expression of ideas.
3 – Adequate - Reader understands the writer's ideas.
2 – Marginal - Reader occasionally has difficulty understanding or following the writer's ideas.
1 – Inadequate - Reader has difficulty identifying or following the writer's ideas.

[4] The Chief Reader supervises essay scoring operations for a jurisdiction in consultation with GEDTS. He or she must be trained and certified by GEDTS (see Chapter 5).

Descriptors

There are five major areas or descriptors that are used in evaluating an essay. They are as follows:

- Response to the prompt
- Organization
- Development and details
- Conventions of Edited American English (EAE)
- Word choice

Response to the prompt refers to how well the examinee responded to the topic, including whether the focus of the response shifted or whether the focus was maintained.

Organization refers to whether there is evidence that the examinee had a clear idea about what he or she would write and was able to establish a definable plan for writing the essay.

Development and details refers to the examinee's ability to expand on initial concepts through the use of examples and specific details rather than simply using lists or reiterating information.

Conventions of Edited American English refers to the examinee's ability to use appropriate, edited, written English, including the application of the basic rules of grammar, such as sentence structure, mechanics, usage, and so forth.

Word choice refers to the use of appropriate words to express an idea.

Explanation of Standard Scores

GED test standard score scales were developed through norming studies (described in Chapter 3) involving high school seniors who are about to graduate. GED standard scores thus provide a standard against which an adult's test performance can be evaluated. This standard involves an external yardstick based on the achievement levels of contemporary high school seniors.

Standard scores permit an examinee's results to be reported on a consistent scale for all five tests, although the tests in the GED test battery contain different numbers of items. Standard scores represent the most frequently used method for establishing a common basis for comparing an examinee's achievement on different content area tests.

The process GED Testing Service uses to establish standard scores helps ensure that minor differences in difficulty among the various forms of the GED Tests will neither help nor hinder an examinee's completion of a particular form. That is, standard scores are used to make appropriate adjustments for the fact that the items on some test forms may be slightly easier or slightly more difficult than those on another form. The use of standard scores ensures that an examinee can expect to earn about the same score regardless of test form.

The standard scores are used to compare the achievement of GED examinees directly with the demonstrated achievement of recent high school graduates. To qualify for a credential, a GED examinee must perform at least as well as a certain percentage of graduating high school seniors (see Passing Standard, on the next page).

In reporting scores earned on the GED Tests, GEDTS uses standard scores and percentile ranks (the percentage of the high school senior norm group that scored at or below that standard score). Both score scales involve transforming the examinee's raw score (number of items correctly answered) to new numerical scales. Higher raw scores are associated with higher standard scores and percentile ranks. The standard score scale used to report results for each of the five U.S. 2002 Series English-Language GED Tests has the following properties:

- The scale ranges from a minimum of 200 to a maximum of 800.[5]
- The scale has a mean of 500 and a standard deviation of 100.[6]
- About two-thirds of all U.S. high school seniors earn standard scores between 400 and 600. Standard scores lower than 300 or higher than 700 are earned by only about 2 percent of graduating high school seniors.

The relationship between GED standard scores and associated percentile ranks for the English- and Spanish-language U.S. GED Tests is presented in Table 1.3. (Appendix A lists the GED standard scores and associated percentile ranks for the Spanish-language tests administered in Puerto Rico, the French-language tests, and the Canadian version tests.) A standard score of 500 represents the average performance of high school seniors. The percentile rank of 50 associated with the standard score of 500 indicates that half of the high school seniors scored at or below this level. Eighteen percent of high school seniors scored at or below the standard score of 410 on each test; 31 percent scored at or below the standard score of 450 on each test; 69 percent scored at or below the standard score of 550 on each test, and so on. These percentile ranks associated with standard scores represent performance of the norm group on any one GED content area test. Any adjustments necessitated by shifts in norm group performance affect the conversion of raw scores to standard scores (see Chapter 3 for more information on norming processes).

[5] The scale for the English as a Second Language Test ranges from 20 to 80, with mean 50 and standard deviation 10.

[6] The Canadian version, as well as the French- and Spanish-language GED Tests, is also on a scale ranging from 200 to 800. However, the scale means and standard deviations may not be equal to those from the U.S. English-language version.

TABLE 1.3
GED TESTS STANDARD SCORES AND PERCENTILE RANKS FOR THE ENGLISH- AND SPANISH-LANGUAGE U.S. GED TESTS

Standard Score	Percentile Rank	Standard Score	Percentile Rank
800	99	500	50
790	99	490	46
780	99	480	42
770	99	470	38
760	99	460	34
750	99	450	31
740	99	440	27
730	99	430	24
720	99	420	21
710	98	410	18
700	98	400	16
690	97	390	14
680	96	380	12
670	96	370	10
660	95	360	8
650	93	350	7
640	92	340	5
630	90	330	4
620	88	320	4
610	86	310	3
600	84	300	2
590	82	290	1
580	79	280	1
570	76	270	1
560	73	260	1
550	69	250	1
540	66	240	1
530	62	230	1
520	58	220	1
510	54	210	1
		200	1

Passing Standard

To identify and recommend a passing standard for the 2002 Series GED Tests, a Psychometric Expert Panel, the GED Advisory Committee, and members of the Commission on Adult Learning and Educational Credentials compared the failure rates of high school seniors on each of the five content area GED Tests using data from the 2001 norming study. They examined the current failure rates on the 2002 Series GED Tests with different passing standards across the five tests. The data revealed that a standard score of 410 on each content area test would result in similar failure rates as the 1988 Series GED Tests for the Language Arts, Writing and Social Studies Tests, slightly higher failure rates for the Language Arts, Reading and Science Tests, and a decidedly higher failure rate increase of 6 percent for the Mathematics Test. However, this 6 percent increase in the failure rate brought the Mathematics Test failure rate into alignment with the other four tests.

The Psychometric Panel, the GED Advisory Committee, and the Commission thus recommended setting the passing standard at 410 for each of the five tests, and an average of 450 for the battery, which required increased performance on some of the tests.

This requirement, which took effect January 1, 2002, with the introduction of the new series, represents the reasoned judgment by ACE that such requirements should be neither so high as to represent levels of achievement far above that demonstrated by recent high school graduates (and, as such, arbitrarily unfair to adult examinees) nor so low as to threaten the credibility of the high school equivalency credential. It was estimated that the new passing standard was met by 58 percent of graduating high school seniors in the norm group.

Each jurisdiction that contracts to use the GED Tests establishes its own minimum score requirements for issuance of the GED credential. However, ACE requires that such score requirements be set at a standard no lower than that which would result from requiring the following: an average standard score of at least 450 on the five tests in the battery and a minimum standard score of 410 on each test in the battery.

Minimum score requirements in jurisdictional GED testing programs are usually stated in terms of the battery average of the five test scores, the minimum score on each of the five tests in the battery, or a combination of average and minimum standard scores. Requirements are always specified as standard scores on a range of 200 to 800. The minimum score requirements for credentials for each jurisdiction are listed in Appendix B.

A standard score on any one of the GED Tests represents a level of achievement attained or exceeded by a certain percent (100 percent minus the percentile rank of the standard score) of the reference group of U.S. high school seniors. Standard scores of 350, 450, and 500 are met or exceeded by about 93 percent, 69 percent, and 50 percent of this group, respectively, on any one test. The meaning of a minimum score requirement for any one test is that the criterion represents a level of performance met or exceeded by a particular percentage of the sample of graduating seniors in the norm group.

To understand existing jurisdictional passing requirements, it may be helpful to inspect the percentage of the 2001 high school seniors who met several passing requirements. From the norming sample data, it is possible to estimate the percentage of all similar high school seniors who would have met a given requirement if all seniors in the nation had been tested. Table 1.4 shows the estimated percentage of the 2001 U.S. norm group who meet various score requirements for the GED test battery.

TABLE 1.4
ESTIMATED PERCENTAGES OF 2001 NORM GROUP MEETING VARIOUS PASSING REQUIREMENTS

Passing Requirement	Estimated Percentage of High School Seniors Meeting Requirement
Average 450 and minimum score 400	60
Average 450 and minimum score 410	58
Average 420 and minimum score 420	59
Average 450 and minimum score 420	55

The GED Tests and the Compensatory Model

In the compensatory model, a minimum overall or average performance level/score must be met as well as minimum performance levels/scores on various tests that contribute to the overall or average score. Such is the case with the GED Tests. In order to pass the GED test battery, an examinee *must* have an average of the five individual subject area test scores of 450 or greater; in addition, each individual subject area test score *must* be 410 or greater. Only by achieving these standards does passing the GED test battery indicate that an examinee has scored better than at least 40 percent of the graduating high school seniors in the norm group.

This model allows examinees to "compensate" for performance in one subject area by stronger performance in another; that is, a score below 450 (but greater than 410) on one test can be compensated by a score greater than 450 on another test and result in passing the GED test battery. The model advocates that many skills make important contributions to achievement and that it is possible for most examinees to compensate for weaknesses in one area using strengths in other areas. The model also carries some psychometric benefits, such as an increase in the reliability of the battery scores and a reduction of the effects of error of measurement.

GED examinees who do not meet the minimum score for a content area test(s) have the option to retake the test(s) at a later date. Policies for retaking the test vary by jurisdiction (see GEDTS, 2008). However, no examinee can take the test more times than the number of operational forms in a given year. In any case, the examinee needs to retake only that test in an attempt to achieve both the minimum

average battery score and minimum score on the individual content area test, as determined by each jurisdiction.[7] GED examinees who do not meet the average passing standard, but earn the minimum score on each of the individual content area tests, may choose to retake a test or several tests. This option allows an examinee the opportunity to improve his or her individual test score(s) in order to raise the average score to 450.

Meaning of the Average GED Test Battery Standard Score

Most participating jurisdictions set minimum performance levels for the average GED test battery standard score and each GED content area test standard score when defining requirements for their high school equivalency credential. (See Appendix B for a list of these jurisdictions and their minimum standard score requirements for 2007. Note that the standards set by individual jurisdictions are subject to change; see the Annual GED Testing Program Statistical Reports [GEDTS, 2008a, 2007, 2006a, 2006b, 2005b, 2004].) For this reason, it is helpful to understand some of the properties of average battery scores. Average battery standard scores must not be interpreted using the figures in Table 1.3 because those figures apply only to individual tests. The battery averages have unique percentiles that are slightly different from those for each individual content area test. This occurs because the same individuals who, for example, earn standard scores of 410 and below on one of the tests do not necessarily earn average scores of 410 and below on the battery. Percentiles for battery averages are not included in the standard score report received by examinees, but the percentile rankings are available on the GEDTS web site (www.GEDtest.org). Estimates of percentiles of battery averages can give the examinees a general idea of where their composite achievement level is in comparison with that of current high school graduates. Thus, percentile ranks associated with the average GED test battery standard score can loosely be thought of as a GED examinee's approximate class rank with respect to the population of graduating high school seniors across the United States (Table 1.5).

TABLE 1.5
GED AVERAGE STANDARD SCORE AND ESTIMATED NATIONAL (U.S.) CLASS RANK OF GRADUATING HIGH SCHOOL SENIORS

Battery Average Standard Score	Estimated National Class Rank
700	Top 1%
670	Top 2%
660	Top 3%
640	Top 5%
610	Top 10%
580	Top 15%
570	Top 20%
550	Top 25%
530	Top 33%
520	Top 40%
500	Top 50%
460	Top 55%
450	Top 60%

Source: College Admissions and Candidates with GED High School Credential, GEDTS brochure, 2003.

[7] Minimum scores vary by jurisdiction. See Appendix B for details.

Test Security

The nature of the GED Tests requires that test security be maintained throughout the development and administration of each test form. The validity of the test score interpretations relies on keeping the test items free from security violations. GEDTS provides a number of safeguards regarding test security, all of which depend on those people who develop, administer, score, and take the GED Tests. Although the details of GEDTS test security measures are beyond the scope of this manual, some noteworthy issues are highlighted below.

- Although GEDTS develops Official GED Practice Tests, those items are not used on the operational versions of the GED Tests. The operational forms are not available for individual sale and thus are not available to the general public outside of official testing conditions.
- During the test development stage, item writers see only the items they write and thus have limited exposure to the item pool. All item writers, reviewers, and external contractors must sign a confidentiality agreement.
- The administration of the GED Tests is maintained within each jurisdiction and is therefore decentralized. All GED Administrators, as well as the Chief Examiners within each jurisdiction, are required to follow all procedures detailed in the *GED Examiner's Manual* (GEDTS, 2002a). The reader is referred to this document for details on these and other security procedures maintained within each jurisdiction.
- Likely the biggest threat to test security for the GED Tests is item exposure. Because the number of people who take the GED Tests each year is considerable, and given the number of operational forms available, item exposure is high. One preventative measure used by GEDTS is to cycle test forms both within and across years. Specifically, a test form is never used in consecutive years. In addition, examinees who retake any GED content area test within the same calendar year are not exposed to the same items.

Chapter 2: GED Tests Specifications and Forms Development

The content of the five English-language U.S. and Canadian versions of the GED Tests corresponds to material that graduating high school students in the United States and Canada, respectively, are required to know and demonstrate. GED Testing Service (GEDTS) staff members rely heavily on the experience of the GED Tests Specifications Committee, which comprises secondary school educators from various academic disciplines, to develop specifications for the tests that seek to synthesize the academic curricula of high school programs throughout the United States and Canada.

In preparation for the 2002 Series GED Tests, a nationwide selection process produced a select group of educators to serve on a new GED Tests Specifications Committee. This committee met in January 1997 to draft recommendations for the content and format of the five tests. Committee members represented four content areas: language arts (reading and writing), social studies, science, and mathematics.

By 2001, GEDTS staff members and item writers had compiled a database of several hundred test items for all five tests. The majority of these items were newly written, with a select few revised from the 1988 series. To ensure that the content of the new GED Tests would reflect contemporary high school curricula, content specialists with backgrounds in secondary or adult education, from a variety of ethnic groups and geographic areas, participated in the review process. In preparation for the release of the new series of GED Tests in January 2002, these items underwent initial field testing in high schools in a number of states and Canadian provinces and territories.

The process of developing the GED test forms is described in detail in this chapter. After development of the first year's test forms, each form was first standardized with a national sample of graduating high school students. The obtained test scores from each new form were subsequently equated prior to becoming fully operational. (For a detailed description of the standardization process, see Chapter 3.)

DESCRIPTION OF THE GED TESTS

Testing Contexts

The contexts used in the GED Tests are designed to be relevant to adults, to be as practical and realistic as possible, and to reinforce key themes of global awareness and the impact of technology.

Settings Relevant to Adults

Because the GED Tests are taken by a diverse adult population, they must be carefully developed with this audience in mind. The context of the passages and items in each of the five tests reflects the following themes:

- Contain settings that adults will recognize as relevant to their daily lives.
- Reflect the many roles of the individual (i.e., as worker, family member, consumer, and citizen).
- Acknowledge the sources of change affecting individuals and society.

Global Awareness

The contexts in which items are presented are also intended to reinforce key themes of which an educated adult should be aware, such as the global nature of society. Many items, particularly in the Social Studies Test, use contexts or situations that refer to areas outside North America and that recognize and address the global nature of our culture.

For example, an economics item addressing the laws of supply and demand might portray the political, economic, or geographic interdependence of world oil suppliers and consumers. Similarly, a science context might emphasize the global nature of ecological issues, and the selections in the Language Arts, Reading Test might include international authorship and themes.

Impact of Technology

Another important theme highlighted in test item contexts is the role and impact of technology—especially computer technology in modern society. Though the tests do not include items that directly address or require computer literacy, test item stimuli materials in all five tests occasionally incorporate information or situations that refer to computer technology and its impact.

Readability

Items used in the GED Tests are written or selected by practitioners: teachers and content experts current in the academic disciplines represented on the tests. All items are then screened by at least three independent teachers or content experts and the GEDTS content area test specialist (who is a professional educator certified in that discipline). These reviewers determine whether the difficulty levels of the reading selections and items are appropriate for a high school graduate. Thus, the readability of all the GED Tests is monitored early in the test development process through the judgment of experienced educators.

Those items that are not eliminated during the first stage of screening are field-tested by administering them to graduating high school seniors. A review of examinees' performance on field-tested items represents a second check of the readability and difficulty of reading selections on the various tests in the GED test battery. An item may be rejected by either the test specialist or other reviewers' estimates of the difficulty level.

Cognitive Levels

The Language Arts, Reading, and Social Studies Tests are classified solely according to an adaptation of Bloom's taxonomy (Bloom, 1956; see Appendix C). Although specific cognitive levels are not designated for the Science Test, higher levels of Bloom's taxonomy are emphasized for the items on the Science Test. Items on the Mathematics Test are classified using a system recommended by the GED Tests Specifications Committee. The Language Arts, Writing Test multiple-choice items are based on a classification system that is similar to Bloom's taxonomy.

SPECIFICATIONS FOR THE 2002 SERIES TESTS

The following sections detail the major content areas within each of the five GED Tests. Unless stated otherwise, the specifications listed are relevant to each test version.

Language Arts, Writing Test, Part I

The Language Arts, Writing Test has two parts. The first part has 50 multiple-choice items and a time limit of 75 minutes. Part I requires examinees to demonstrate the ability to revise and edit workplace and informational documents by answering multiple-choice items. The second part assesses the examinee's ability to write an essay. The scores earned on both parts are combined and reported as a single standard score.

Content

The content areas in the first part of the Language Arts, Writing Test include the following:

Organization (15%): Organization items require the examinee to edit and revise the document by adding, removing, or repositioning sentences. Organizational skills include effective text divisions (within or among paragraphs, forming new paragraphs within multi-paragraph documents, and combining paragraphs to form a more effective document), topic sentences, and unity/coherence.

Sentence Structure (30%): Sentence structure items involve sentence fragments, run-on sentences, comma splices, improper coordination and subordination, modification, and parallelism.

Usage (30%): Usage errors may include subject-verb agreement (including agreement in number, interrupting phrases, and inverted structure), verb tense errors (including sequence of tenses, word clues to tense in sentences, word clues to tense in paragraphs, and verb form), and pronoun reference errors (including incorrect relative pronouns, pronoun shift, vague or ambiguous references, and agreement with antecedents).

Mechanics (25%): Mechanics problems may include capitalization (including proper names and adjectives, titles, and months/seasons), punctuation (including commas in a series, commas between independent clauses joined by a conjunction, introductory elements, appositives, and the overuse of commas), and spelling (restricted to errors related to possessives, contractions, and homophones).

Context

The subject matter for Language Arts, Writing Test, Part I includes those topics with which the examinee is likely to be familiar. Part I passages are all approximately 200-300 words (12 to 22 sentences) and are based on the following types of documents:

Workplace and community documents: Workplace and community documents are those common to an adult's everyday environment. These documents are letters, memos, meeting notes, reports, executive summaries, applications, and similar correspondence.

"How to" texts: "How to" texts are documents that provide instructions or directions. These documents focus on such topics as securing a job, writing a resume, dressing for success, leasing a car, getting to a specific location, and so forth.

Informational texts: Informational texts provide an analysis of a particular topic. These texts include position papers, critical evaluations, support papers, and the like. Sample topics include: why the rainforests should be saved, the growing popularity of mega-malls, the building of a monument, and examining the transit needs of a local community.

Format

Part I directly measures proofreading and editing skills based on the three document types described above. Each document, when corrected, is an example of good writing. The errors to be corrected are those most likely to hinder the writer's ability to communicate effectively to various audiences.

Several of the sentences within each paragraph contain errors that the examinee must locate and correct, such as a faulty verb or misplaced comma. In some cases, however, the sentences do not contain specific errors but require revision for clarity or logic. This revision may entail restructuring the sentence, moving the sentence to another position in the document, or occasionally removing the sentence from the document altogether. Other revisions may require a paragraph to be divided or two paragraphs to be joined.

In each item, the sentence to be corrected or revised is presented, and the five possible alternatives (or answers) are presented in the order in which they occur in the sentence. On occasion, the fifth alternative is "no correction is necessary" or "no revision is necessary." In other types of items, the first alternative will be the same as an underlined word or phrase in the sentence, requiring the examinee to recognize that no revision is necessary. Alternatives for each item often come from any of the four Language Arts, Writing Test content areas to create a realistic editing or proofreading situation.

Cognitive Levels

The Language Arts, Writing Test, Part I items are classified as being Correction, Revision, or Construction Shift item types. These three terms apply to the cognitive skills necessary to draft, edit, and revise written documents. Correction items can be considered similar to Bloom's comprehension/analyzing category; Revision items are similar to Bloom's analyzing/synthesis category; and Construction Shift items are similar to Bloom's synthesis category. The cognitive skills measured by the Language Arts, Writing Test, Part I are described next.

Correction items (45%) test skills in the following content areas:

- Organization
- Sentence structure
- Usage
- Mechanics

These items may involve one sentence, a number of sentences, a complete paragraph, or the text as a whole. This item type provides a series of choices and asks what correction should be made.

Revision items (35%) test skills in the following content areas:

- Organization
- Sentence structure
- Usage
- Mechanics

The revision item presents a sentence with an underlined portion that may or may not contain an error. The answers present five possible corrections or revisions to the underlined section of the sentence. In these items, the first alternative always matches the original sentence. This requires the examinees to recognize that no correction or revision is necessary.

Construction shift items (20%) test skills in the following content areas:

- Organization
- Sentence structure
- Usage

The construction shift item presents a sentence that must be rewritten by revising the sentence structure. The resulting sentence must be correct and clearly stated. Original construction shift sentences do not contain errors. These items test an examinee's ability to manipulate sentence structures to create a better sentence.

Construction shift items require the examinee to use logic to think through the process of changing a sentence, or in the case of organization, change the structure of a document. Construction shift also tests an examinee's understanding of the sequence of events.

Organization construction shift items may require the examinee to combine paragraphs, separate paragraphs, or insert a new sentence within a paragraph.

Specifications for the items on Language Arts, Writing Test, Part I are shown below in Table 2.1.

TABLE 2.1
SPECIFICATIONS FOR THE LANGUAGE ARTS, WRITING TEST, PART I: NUMBERS OF ITEMS, BY ITEM CONTENT AND ITEM TYPE

	COGNITIVE SKILLS		
ITEM CONTENT	Correction 45% (25 items)	Revision 35% (17-18 items)	Construction Shift 20% (7-8 items)
Organization 15% (7 items)	3-4	2-3	1
Sentence Structure 30% (15 items)	7-8	4-5	3-4
Usage 30% (15 items)	7-8	4-5	3-4
Mechanics 25% (13 items)	7-8	5-6	0

Language Arts, Writing Test, Part II

Part II of the Language Arts, Writing Test measures the examinee's ability to write, and carries a time limit of 45 minutes. Examinees are required to compose an essay using personal observations and experiences to support their ideas.

Content

In the Language Arts, Writing Test, Part II, examinees receive a single expository topic and are asked to present an opinion or an explanation regarding a situation about which adults would be expected to have some general knowledge.

Context

The topics are specifically chosen by GED Testing Service because they are found to be potentially interesting and meaningful to examinees as well as to the readers who will score them. Essay topics do not require specialized content knowledge.

Format

The test directions encourage examinees to plan their essays, make notes before writing, and revise and edit their final products.

Examinees must write their essays only on two lined pages in the answer booklet, with scratch paper for pre-writing and drafting. Only the writing on the two pages in the answer booklet will be scored. Examinees are not required to fill both pages, nor is there a minimum word count to attain.

Examinees may use either cursive or print when writing their essay. Handwriting does not factor into the evaluation process because it is not a part of the content or substance of the essay. Two papers differing only in the neatness or style of handwriting should receive the same score. However, readability is important.

Social Studies Test

The Social Studies Test measures an examinee's skill in understanding, interpreting, and applying key history, geography, economics, and civics and government concepts and principles. Source materials and items on the Social Studies Test address the experiences of citizens, consumers, and workers in the United States, Canada, and the rest of the world. The test items are based on written and visual texts drawn from a variety of sources including academic and workplace texts, as well as primary and secondary sources. The Social Studies Test has 50 multiple-choice items and a time limit of 70 minutes.

Content

The U.S. version of the Social Studies Test includes items in each of the following areas:

U.S. History (25%): Beginnings to 1820 (Native Peoples, Colonization, Settlement, Revolution, the New Nation); 1801–1900 (Expansion, Reform, Civil War, Reconstruction, Industrial Development); and 1890–present (Emergence of Modern America, Great Depression, World War II, Postwar United States, Contemporary United States)

World History (15%): Beginnings–1000 B.C. (Beginnings and Early Civilizations); 1000 B.C.–300 B.C. (Classical Traditions, Empires, Religions); 300 B.C.–A.D. 1770 (Growing Trade, Hemispheric Interactions, First Global Age); 1750–1914 (Age of Revolutions); and 1900–present (Urbanization; World Wars; Global Depression; Advances in Science and Technology; New Democracies of Africa, Asia, South America; the Cold War; "Global Culture")

Civics and Government (25%): Civic Life, Politics, Government; Foundations of the American Political System; American Government; Relationship of United States to Other Nations; and the Roles of Citizens in American Democracy

Geography (15%): World in Spatial Terms; Places and Regions; Physical Systems; Human Systems; Environment and the Society; and Uses of Geography

Economics (20%): Economic Reasoning and Choice; Comparison of Economic Systems, Business in a Free Enterprise System, Production, Consumers; Financial Institutions; and Government's Role in the Economy, Labor and the Economy, Global Markets, and Foreign Trade

The Canadian version of the Social Studies Test includes the same items as the U.S. version that relate to World History, Geography, and Economics. It also includes Canadian-specific government, civics, and history.

Canadian Government and Civics (25%): National Unity; Canada in the World; Canadian Governance; Civil and Social Responsibility; Residual Powers to Federal Government; Provincial vs. Federal Relations; Canada-U.S. Relations; Canadian Demography; and Economic Issues

Canadian History (25%): Canada's Aboriginal Peoples; European Exploration and Colonization; Growth and Change; Growing Frontier Community; Political Reform and Confederation; the Age of Macdonald (1867–1891); Canada in the 20th Century; and Facing the Challenges of the Modern World

Context

Approximately 60 percent of the Social Studies Test items relate to concepts and issues taken from a global or international perspective, and 40 percent address a specific national setting (either United States or Canada).

Historical Documents: Each form of the Social Studies Test includes an excerpt from at least one of the following fundamental historical documents of the United States and Canada:

- Declaration of Independence (U.S. version only)
- U.S. Constitution (U.S. version only)
- Landmark Supreme Court cases (U.S. version only)
- The Charter of Rights and Freedoms (Canada version only)

Practical Documents: Each form includes one practical document (a source of information used by most adults in their roles as citizens, consumers, and workers) such as:

- Consumer information
- Voters' guides
- Atlases
- Tax forms
- Budget graphs
- Political speeches
- Almanacs
- Statistical abstracts
- Advertisements

Format

All Social Studies Test items are multiple-choice items based on one of the following types of source materials:

- Prose (40%): narratives, high school textbooks and resources, editorials, speeches, newspapers, news magazines, historical documents
- Visual text (40%): maps, graphs, charts, diagrams, political cartoons, photographs, lithographs, works of art
- Written and visual text (20%): a combination of narrative and graphic stimuli

Prose sources are no longer than 150 words, with text for a single item ranging 50–60 words. Approximately 60 percent of the items are grouped into sets that share stimulus material (e.g., two to five items based on an excerpt or a set of items based on visual text).

Cognitive Levels

The Social Studies Test requires that examinees use higher-level thinking skills, as defined by Bloom's taxonomy (see Appendix C). These skills often require prior knowledge of important social studies concepts, principles, events, and skills.

Comprehension items (20%) measure the examinee's understanding of the meaning and intent of text and/or visual material. These items measure the examinee's ability to:

- Restate information.
- Summarize ideas.
- Identify implications and make inferences.

Application items (20%) measure the examinee's ability to use information and ideas in a situation different from that provided by the item stimulus. These items measure the examinee's ability to:

- Identify an illustration of a generalization, principle, or strategy.
- Apply the appropriate abstraction to a new problem without prompting or instruction.

Analysis items (40%) measure the examinee's ability to break down information and explore the examinee's understanding of the relationship between component ideas. These items measure the examinee's ability to:

- Distinguish facts from opinions and hypotheses.
- Distinguish conclusions from supporting statements.
- Recognize information that is designed to persuade an audience.
- Recognize unstated assumptions.
- Recognize fallacies in logic in arguments or conclusions.
- Identify cause and effect relationships and distinguish them from other sequential relationships.
- Recognize the point of view of a writer in a historical account.
- Recognize the historical context of the text, avoiding "present-mindedness."
- Identify comparisons and contrasts among points of view and interpretations of issues.
- Determine implications, effects, and the value of presenting visual data in different ways.

Evaluation items (20%) measure the examinee's ability to use criteria provided to make judgments about the validity or accuracy of information. These items measure the examinee's ability to:

- Assess the appropriateness of information to substantiate conclusions, hypotheses, and generalizations (using such criteria as source, objectivity, technical correctness, and currency).
- Assess the accuracy of facts.
- Compare and contrast differing accounts of the same event.
- Recognize the role that values, beliefs, and convictions play in decision making.

The content and cognitive specifications for the 50 items in the Social Studies Test are presented in Table 2.2.

TABLE 2.2
SPECIFICATIONS FOR THE SOCIAL STUDIES TEST: NUMBERS OF ITEMS, BY ITEM CONTENT AND COGNITIVE LEVEL

	COGNITIVE LEVELS			
ITEM CONTENT	Comprehension 20% (10 items)	Application 20% (10 items)	Analysis 40% (20 items)	Evaluation 20% (10 items)
History: National 26% (13 items)	3	2	6	2
History: World 14% (7 items)	1	2	3	1
Geography 16% (8 items)	1/1	1/1	2/0	1/1
Economics 20% (10 items)	1/1	1/0	4/1	1/1
Civics & Government 24% (12 items)	1/1	2/1	3/1	2/1

Note: The number to the left of the slash (/) indicates the number of items in an operational form that must represent a global perspective. The number on the right represents the number of items that must represent a specific U.S. (or Canadian) perspective.

Science Test

The Science Test items are designed to measure an examinee's skills and knowledge in the content areas of life science, physical science (physics and chemistry), and Earth and space science. The test items are based on written and visual texts from academic and workplace contexts. Even though no specific cognitive levels are designated, upper levels of Bloom's taxonomy (comprehension, application, analysis, and evaluation) are emphasized in the Science Test. The items reflect the many roles of individuals (for example, worker, family member, consumer, and citizen). The Science Test has 50 multiple-choice items and a time limit of 80 minutes.

The Science Test measures the major and lasting expected outcomes of a sound, well-rounded high school science education. These outcomes include the acquisition of a broad knowledge base and the ability to use a range of reasoning skills. Test items focus on the comprehensive, integrated skills typical of what the examinee must know, understand, and be able to perform in order to be scientifically literate.

Content

The GED Tests Specifications Committee recommended that the Science Test items be based on the eight content standards for grades 9–12 as outlined by the National Science Education Standards (NSES). According to the committee's recommendations, 60 percent of the Science Test items measure an examinee's fundamental understanding of basic knowledge, principles, concepts, and vocabulary associated with physical science, life science, and Earth and space science. The three main content areas are provided below.

Physical Science (35%). Items are drawn from the following subsets:

- Structure of atoms
- Structure and properties of matter
- Chemical reactions
- Motions and forces
- Conservation of energy and increase in disorder
- Interactions of energy and matter

Life Science (45%). Items are drawn from the following subsets:

- The cell
- Molecular basis of heredity
- Biological evolution
- Interdependence of organisms
- Matter, energy, and organization in living systems
- Behavior of organisms

Earth and Space Science (20%). Items are drawn from the following subsets:

- Energy in the Earth system
- Geochemical cycles
- Origin and evolution of the Earth system
- Origin and evolution of the universe

Standards/Context

The remaining five NSES content standards are used in context of the three core content standards for the Science Test. These areas comprise 40 percent of the test items and are described as follows.

Unifying Concepts and Processes outlines the broad concepts and processes that need to be developed over an examinee's entire education and that transcend disciplinary boundaries. Test items in this category are drawn from the following concepts and processes:

- Systems, order, and organization
- Evidence, models, and explanations
- Change, constancy, and measurement
- Evolution and equilibrium

Science as Inquiry advances the examinee toward higher-level content knowledge and cognitive skills by helping him or her develop questioning and reasoning abilities. Items under this standard come from the following specific processes associated with scientific inquiry:

- Identifying questions and concepts that guide scientific investigations.
- Designing and conducting scientific investigations.
- Using appropriate tools and techniques to gather data.
- Thinking critically and logically about relationships between evidence and explanations.
- Analyzing alternative explanations.
- Communicating scientific arguments.
- Understanding scientific inquiry.

The remaining content standard categories build on the examinee's knowledge and understanding of physical science, life science, and Earth and space science.

Science and Technology focuses on an examinee's ability to identify, change, or improve a piece of technology or technique and understand the links between science and technology. Specific foci might include an examinee's decision-making abilities in identifying and stating new problems or needs and designing, implementing, and evaluating a solution.

Science in Social and Personal Perspectives addresses the scientific foundation an examinee needs to evaluate in order to make decisions about personal and social issues that he or she may encounter. Items under this standard are drawn from the following subsets:

- Personal and community health
- Population growth
- Natural resources
- Environmental quality
- Natural and human-induced hazards
- Science and technology in local, national, and global challenges

History and Nature of Science addresses an examinee's understanding of the nature of science and science in different historical and cultural perspectives. Items under this standard are drawn from the following subsets:

- Science as human endeavor
- Nature of scientific knowledge
- Historical perspectives

Format

The Science Test includes items based on both text passages and visual text (e.g., graphs, tables, charts, diagrams). Up to 60 percent of the items are presented with visual text, which reduces the amount of written explanatory text on the test. Examinees must demonstrate that they can interpret and analyze different types of visual text.

Written text ranges in length from text included in a single item to a short article followed by one or more items. Articles are written at a reading level that does not interfere with the assessment of the examinee's knowledge and application of science principles.

Approximately 25 percent of the items are grouped into sets that share stimulus material (e.g., two to five items based on an excerpt or a set of items based on visual text). Passages and visual text represent realistic situations.

Table 2.3 presents specifications for the 50 items on the Science Test.

TABLE 2.3
SPECIFICATIONS FOR THE SCIENCE TEST: NUMBERS OF ITEMS, BY STANDARD AND CORE CONTENT

	CORE CONTENT		
STANDARD	Life Science 45% (23 items)	Earth & Space Science 20% (11 items)	Physical Science 35% (16 items)
Fundamental Understanding 60% (30 items)	14	6	10
Unifying Concepts & Processes 4% (2 items)	1	0	1
Science as Inquiry 8% (4 items)	2	1	1
Science & Technology 4% (2 items)	1	1	0
Science in Personal & Social Perspectives 16% (8 items)	4	2	2
History & Nature of Science 8% (4 items)	1	1	2

Note: The distribution of the number of items for each of the five standards across the core content standards (life, physical, and earth and space) is not fixed. For example, Unifying Concepts & Processes has two items (4%), which could fall under any one of the three core content categories, as long as not all are under one category.

Language Arts, Reading Test

The Language Arts, Reading Test is a passage-based, multiple-choice test that measures an examinee's ability to comprehend and interpret literary and workplace reading selections and to apply those interpretations to new contexts. The Language Arts, Reading Test has 40 items and a time limit of 65 minutes.

Content

The content of the Language Arts, Reading Test reflects the variety of texts a high school student encounters. On each test, 75 percent of the excerpts are from literary texts, and 25 percent are from nonfiction texts. Texts and authors that could be expected to appear in a high school classroom for examination and critical review appear on the Language Arts, Reading Test. Sources for the literary texts reflect a commitment to quality writing from writers of recognized stature.

Literary texts (75%) include at least one selection from each of the following areas:

- Poetry
- Drama
- Prose fiction before 1920
- Prose fiction between 1920 and 1960
- Prose fiction after 1960

Nonfiction texts (25%) include two selections of nonfiction prose representing two of the three following areas on a rotating basis:

- Nonfiction prose
- Critical reviews of visual or performing arts
- Workplace and community documents, such as mission and goal statements, rules for employee behavior, legal documents, memos, letters, excerpts from manuals, etc.

Context

The subject matter chosen for the Language Arts, Reading Test reflects the multicultural backgrounds and diverse age groups of GED examinees. Texts are examined carefully to ensure that no particular group is presented in a discriminatory manner. At the same time, texts are also chosen to reflect the variety of experiences of the general population without giving undue attention to any particular group's experiences. Each test is constructed with this diversity in mind so that no examinee feels either excluded or advantaged by the set of texts in any given reading test.

Format

The selections in the Language Arts, Reading Test are coherent excerpts with a beginning, middle, and end. Excerpts range from 200 to 400 words, with poetry running from 8 to 25 lines. Each selection is followed by four to eight items that measure reading skills at several cognitive levels.

Each selection is preceded by a purpose question. This question is designed to focus the examinee and provide a purpose for reading the text. In an unnatural reading situation such as a standardized test, the focus question efficiently provides the examinee with an orientation to the text that, in a natural setting, would spring from the reader's ability to survey a selection before reading.

Cognitive Levels

The multiple-choice items on the Language Arts, Reading Test are constructed on four cognitive levels based on Bloom's taxonomy. As would be expected in high school instruction, higher cognitive levels receive relatively more emphasis.

Comprehension items (20%) measure the examinee's ability to extract the basic meaning and intent of the text. This item type can refer to specific parts of the text or to the text as a whole. Comprehension items measure the ability to:

- Restate or paraphrase information.
- Summarize main ideas.
- Explain the primary implications of the text.

Application items (15%) measure the examinee's ability to transfer concepts and principles from the reading to a new context.

Analysis items (30–35%) measure the examinee's ability to break down information into basic elements; these items can require multiple or complex references. Analysis items generally refer to specific parts of a passage and also measure the examinee's ability to:

- Draw conclusions, understand consequences, and make inferences.
- Identify elements of style and structure (by concept, not by literary term), and identify the use of different techniques (e.g., tone, word usage, characterization, use of detail and example, and figurative language).
- Identify cause and effect relationships.
- Distinguish conclusions from supporting statements and recognize unstated assumptions.

Synthesis items (30-35%) measure the examinee's ability to put elements together to form a whole. Synthesis items require multiple inferences that draw on many parts of the text. Although *synthesis* often implies the integration of information from multiple sources into a new whole, for the purpose of the Language Arts, Reading Test, synthesis also refers to integrating information from many parts of a single selection. Synthesis items measure the examinee's ability to:

- Interpret the organizational structure or pattern of a text.
- Interpret the overall tone, point of view, style, or purpose of a work.
- Make connections among parts of the text.
- Compare and contrast.
- Integrate information from outside the passage with elements within the passage.

The last synthesis subskill listed above appears on the test as a multiple-choice item, in which additional information about the text or author is given in the item stem. The item then asks the examinee to synthesize this new information with information obtained from the selection itself to form a new understanding of the text. For example, a reading selection may be provided from a piece of fiction, such as a Chekhov short story. A synthesis item of the last type might include in the item stem a quote from the author about the nature of the human struggle. The item then might ask the examinee to identify an element in the reading passage that illustrates the author's stated philosophy.

Table 2.4 presents current specifications for the 40 items on the Language Arts, Reading Test.

TABLE 2.4
SPECIFICATIONS FOR THE LANGUAGE ARTS, READING TEST: NUMBERS OF ITEMS, BY ITEM CONTENT AND COGNITIVE LEVEL

	COGNITIVE LEVEL			
ITEM CONTENT	Comprehension 20% (8 items)	Application 15% (5-7 items)	Analysis 30-35% (12-14 items)	Synthesis 30-35% (12-14 items)
Literary Text 75% (30 items)	6	4-5	9-10	9-10
Nonfiction/Prose Text 25% (10 items)	2	1-2	3-4	3-4

Mathematics Test

The Mathematics Test is divided into two equally weighted halves, each consisting of 25 items. A page of formulas is provided as a reference for the examinees in each of the test halves.

On Part I of the test, a Casio *fx-260* calculator is provided for each examinee at the Official GED Testing Center. Directions for using the calculator are found in the test booklet. The calculator is not permitted on Part II of the test, in which estimation and mental math are critical skills.

A total of 90 minutes is allotted for completing the entire test. Part I is issued first. After 45 minutes, the Part I booklet and calculator are collected and the Part II booklet issued. If an examinee completes Part II before the remaining time has expired, it is permissible to turn in the booklet and return to Part I with the calculator. Once Part I has been returned and the calculator reissued, an examinee may not return to Part II.

The Mathematics Test assesses an understanding of mathematical concepts and the application of those concepts to various situations. Specifically, the test:

- Measures problem-solving, analytical, and reasoning skills.
- Determines whether an examinee can interpret information from both word problems and graphic formats, including charts, tables, graphs, and diagrams.
- Presents problems in real-life contexts.

Content

Four major areas are tested on the Mathematics Test. The content areas are:

Number Operations and Number Sense (20-30%): The skills tested include the ability to:

- Represent and use numbers in a variety of equivalent forms (integer, fraction, decimal, percent, exponential, and scientific) in real-world and mathematical problem situations.
- Represent, analyze, and apply whole numbers, decimals, fractions, percents, ratios, proportions, exponents, roots, and scientific notation in a wide variety of situations.
- Recognize equivalencies and order relations for whole numbers, fractions, decimals, integers, and rational numbers.
- Select the appropriate operations to solve problems (for example, When should I divide?).
- Relate basic arithmetic operations to one another.
- Calculate mentally, with pencil and paper, and with a scientific calculator using whole numbers, fractions, decimals, and integers.
- Use estimation to solve problems and assess the reasonableness of an answer.

Measurement and Geometry (20-30%): The skills tested include the ability to:

- Model and solve problems using the concepts of perpendicularity, parallelism, congruence, and similarity of geometric figures.
- Use spatial visualization skills to describe and analyze geometric figures and translations/rotations/dilations of geometric figures.
- Use the Pythagorean theorem to model and solve problems.
- Find, use, and interpret the slope of a line, the *y*-intercept of a line, and the intersection of two lines.
- Use coordinates to design and describe geometric figures.
- Identify and select appropriate units of metric and customary measures.
- Convert and estimate units of metric and customary measure (all conversions within systems).
- Solve and estimate solutions to problems involving length, perimeter, area, surface area, volume, angle measurement, capacity, weight, and mass.
- Use uniform rates (e.g., miles per hour, bushels per acre) in problem situations.
- Read and interpret scales, meters, and gauges.
- Predict the impact of changes in linear dimension on the perimeter, area, and volume of figures.

Data Analysis, Statistics, and Probability (20-30%): The skills tested include the ability to:

- Construct, interpret, and draw inferences from tables, charts, and graphs.
- Make inferences and convincing arguments based on data analysis.
- Evaluate arguments based on data analysis, including distinguishing between correlation and causation.
- Represent data graphically in ways that make sense and are appropriate to the context.
- Apply measures of central tendency (mean, median, mode) and analyze the effect of changes in data on these measures.
- Use an informal line of best fit to make predictions from data.
- Apply and recognize sampling and bias in statistical claims.
- Make predictions based on experimental or theoretical probabilities, including listing possible outcomes.
- Compare and contrast different sets of data on the basis of measures of central tendency and dispersion (range, standard deviation).

Algebra, Functions, and Patterns (20-30%): The skills tested include the ability to:

- Analyze and represent situations involving variable quantities with tables, graphs, verbal descriptions, and equations.
- Recognize that a variety of problem situations may be modeled by the same function or type of function (e.g., $y = mx + b$, $y = ax^2$, $y = ax$, $y = 1/x$).
- Convert between different representations, such as tables, graphs, verbal descriptions, and equations.
- Create and use algebraic expressions and equations to model situations and solve problems.
- Evaluate formulas.
- Solve equations, including first degree, quadratic, power, and systems of linear equations.
- Recognize and use direct and indirect variation.
- Analyze tables and graphs to identify and generalize patterns and relationships.
- Analyze and use functional relationships to explain how a change in one quantity results in a change in another quantity, including linear, quadratic, and exponential functions.

Context

The context of items on the Mathematics Test incorporates tasks with which the examinee has had considerable experience. Context situations are natural, rather than contrived, and deal with such topics as the world of work, the consumer, technology, and family experiences and situations.

Format

Eighty percent of the mathematics items are multiple-choice, leaving 20 percent of the items to require examinees to construct an answer of their own. In these alternate format items, rather than selecting from five choices, the examinees must record answers on either standard or coordinate plane grids. Both Parts I and II of the Mathematics Test have multiple-choice, standard grid, and coordinate plane grid items, and directions for completing the items are found in both test booklets.

Seventy-five percent of the items are individual mathematical problems. The other 25 percent are sets of two to four items centered on a common stimulus. Visual text formats are used in approximately 50 percent of the items.

Cognitive Levels

The Mathematics Test assesses different ways of applying math skills through the use of three different item types. Cognitive skills are tested through the use of items at the following levels:

Procedural items (20%) require an examinee to select and apply the appropriate process for solving a problem. Procedural items test the examinee's ability to:

- Select and apply the correct operation or procedure to solve a problem.
- Verify and justify the correctness of a procedure using concrete models or symbolic methods.
- Modify procedures to deal with factors inherent in problem settings.
- Use numerical algorithms.
- Read and interpret graphs, charts, and tables.
- Execute geometric constructions.
- Round, estimate, and order numbers as needed in a given situation.

Conceptual items (30%) require an examinee to demonstrate knowledge of how basic math concepts and principles work. In some conceptual problems, examinees will be required to identify how to solve a problem, but they will not be required to actually compute the answer. Examinees who have a clear understanding of math concepts and principles know how, when, and why to use a particular mathematical concept. These items assess the examinee's ability to:

- Recognize and label basic mathematical concepts.
- Generate examples and counter-examples of concepts.
- Interrelate models, diagrams, and representatives of math concepts.
- Identify and apply concepts and principles of mathematics.
- Know and apply facts and definitions.
- Compare, contrast, and integrate related concepts and principles.
- Recognize, interpret, and apply signs, symbols, and mathematical terms.
- Interpret assumptions and relationships.

Application/Modeling/Problem-Solving items (50%) assess the ability to apply mathematical principles and problem-solving strategies. These items assess the examinee's ability to:

- Recognize and identify the type of problem that is represented.
- Decide whether or not there is sufficient information provided to solve a problem.
- Select only the information that is necessary to solve a given problem.
- Apply the appropriate problem-solving strategy to compute an answer.
- Adapt strategies or procedures to solve a problem.
- Determine whether an answer is reasonable and correct.

Table 2.5 presents current specifications for the 50 items on the Mathematics Test.

TABLE 2.5
SPECIFICATIONS FOR THE MATHEMATICS TEST: NUMBERS OF ITEMS, BY ITEM CONTENT AND COGNITIVE LEVEL

	COGNITIVE LEVEL		
ITEM CONTENT	Procedural 20%	Conceptual 30%	Application/Modeling/Problem Solving 50%
Number Operations (20-30%)	2-3	2-3	6-8
Measurement/Geometry (20-30%)	2-3	2-3	6-8
Data Analysis/Statistics (20-30%)	2-3	2-3	6-8
Algebra (20-30%)	2-3	2-3	6-8

DEVELOPMENT AND SELECTION OF MULTIPLE-CHOICE ITEMS

Item Writer Representation

Across multiple years, the GED Testing Service staff contracted with professional educators to select or write stimulus material and to write items for new operational forms of the GED Tests. A large number of items were needed because, even with good item writers, many items fail to meet the high judgmental and psychometric standards established for usable test items. For the 2002 Series GED Tests, item writers were content specialists with secondary teaching experience in the academic disciplines for which they were contracted to write items. Every attempt was made to contract with a representative cross-section of North American educators who represent the diversity of the population, in terms of ethnic background, gender, and geographic location.

Recruitment of item writers and reviewers was an ongoing process. Test specialists recruited at various national conferences, including the National Council of Teachers of Mathematics (NCTM), National Council of Teachers of English (NCTE), National Science Teachers Association (NSTA), National Council on Social Studies (NCSS), National Testing Network on Writing (NTNW), National Council on Measurement in Education (NCME), and American Association of Adult and Continuing Education (AAACE). Recruitment also occurred at regional and state meetings of adult and secondary teachers and administrators, as well as in association publications. Each potential item writer was given test specifications and a set of content and style guidelines, and he or she was required to submit a specified set of sample items for evaluation.

Item Development Procedures

Each potential item was subjected to a multi-step review process before being included in a test form for field-testing. These steps are represented in Figure 2.1. First, an item was reviewed by the test specialist for its content accuracy, representation of context, appropriateness for high school–level work, fairness, and general quality. The item was then either accepted as is for further review, edited and submitted for further review, or rejected.

Each item that passed the initial in-house review and revision by the test specialist was then submitted to three external content reviewers who are content specialists in the appropriate discipline. GEDTS recruited a cross-section of educators to ensure multicultural, multiethnic, and geographically diverse representation among external content reviewers. For the development of the operational test forms, contracts were signed with over 100 reviewers. These 100 reviewers represented 20 states and all U.S. regions and seven Canadian provinces and territories.

Concurrently, the director of the GED Testing Service Test Development Unit conducted an independent content review of each item. The content reviewers and director of the Test Development Unit independently judged the accuracy, clarity, suitability, cognitive skills level, and fairness of each item.

Following this external content review, each test specialist studied the ratings of the external content reviewers and director of the Test Development Unit and revised or rejected each item. Items that passed content review then underwent a measurement and fairness review, which was conducted by three external sensitivity reviewers.

After this second round of external review, the test specialist again either revised or rejected the items. The remaining items were then reviewed by a professional editor/proofreader for grammar, spelling, vocabulary, format, and surface errors. Finally, the test specialist revised the items based on comments by the editor/proofreader.

Items surviving this rigorous screening were then administered to graduating high school seniors in a field test or item tryout study. Data from field testing allowed for the examination of each item's psychometric properties, including item difficulty and item discrimination indices.

Based on the results of the item analyses of field-tested items, GED Testing Service psychometricians and test specialists screened items for potential use in operational test forms. Any item with a difficulty level less than 0.40 (fewer than 40 percent of examinees answering an item correctly) and/or a point biserial correlation (item discrimination index) less than 0.20 was eliminated from the pool of eligible items. Each test form in the battery was constructed to yield an average item difficulty of 0.70 and an average discrimination index (or average point biserial correlation) of 0.40.

Final Forms Development

Specific guidelines were followed when new operational forms of the GED Tests were assembled. In previous GED Tests series, a limited number of passages could appear in more than one operational test form, although an item related to the passage could not appear in more than one test form. Since 1989, it has been testing service policy to restrict passage use to a single test form. Once a single item or item set is selected for an operational test form, all related items are locked out of the pool of eligible items for other test forms.

Items included in operational forms of the 2002 Series GED Tests were to some extent ordered by degree of difficulty, from least difficult to most difficult. Passage or graphic item sets were ordered chronologically, in order of text reference, to avoid confusion on the examinee's part and to minimize time spent skipping at random through the printed text of a passage.

When items and item sets were selected to meet specifications of the content grid, test specialists sought a balanced range and variety of context and topics within a specific test form. All multiple-choice items on the 2002 Series GED Tests have five answer choices, numbered one through five. During assembly, test specialists ensured that there was a balance among the correct responses and that no systematic pattern of correct responses can be discerned from one test form to the next.

After a test specialist completed assembly of a form, the form was reviewed by five reviewers: three external and two internal. The internal reviewers were the test specialist for the content area and the director of the Test Development Unit. The five reviewers comprised the Final Form Review Committee. At least two weeks prior to an on-site group review of a test form, the Final Form Review Sheet (see Appendix D) was sent to the Final Form Review Committee. The director of the Test Development Unit served as the final form measurement reviewer. All five reviewers were required to complete independent written reviews, which they brought to the on-site review. Once there, content reviewers compared their independent reviews and prepared a committee consensus report. The Final Form Review Committee either accepted the preliminary form or recommended changes or substitutions for individual items or sets. The content area test specialist received the Final Form Review Committee's recommended changes. After reviewing each comment from the committee, the test specialist met with the director of the Test Development Unit to review all revisions based on comments by the final form reviewers. The test specialist then documented all responses to the committee's suggestions and adjusted the test form accordingly.

Once a test form was approved and printed, a standardization or equating study was performed by administering the test forms to stratified random samples of U.S. and Canadian graduating high school seniors, during the spring of their graduation year. Score scales were equated to the appropriate norming sample (see Chapter 3). The tests were then ready to be administered to any GED examinee seeking to qualify for a GED high school equivalency credential.

FIGURE 2.1. GEDTS TESTS DEVELOPMENT FLOWCHART.

Sensitivity Review

Analyses of differential item functioning (DIF) is a process of examining test items to determine whether groups of examinees perform similarly on the items. DIF analyses are generally a part of the item development process to ensure that membership in any particular group will not make an examinee more likely to answer an item correctly. In DIF analyses, examinees belonging to different groups—for example, males and females—are "matched" according to some criterion representative of ability, usually the total test score. The performance of individuals in each matched group should be similar, given equal instruction and opportunity to learn the test material.

Data obtained via the item tryout and equating studies (see Chapter 3) for the 2002 series typically did not have sample sizes large enough to perform adequate DIF analyses prior to making items operational. Therefore, no DIF results from high school data are reported here. It should be noted, however, that during the item development process, each item was scrutinized for potential bias on multiple occasions by expert item reviewers. Because sample sizes obtained via the operational GED test forms are much larger, more adequate DIF analyses were performed post-data collection. Details on this process are described in Chapter 5.

GED Testing Service used a judgmental sensitivity review of item content throughout the item development process. The GEDTS sensitivity review process was included in the content review stages of both the items and the tests. GEDTS test specialists and reviewers of content and final form all reviewed the test items and passages for material that might be construed as offensive, advantageous, or disadvantageous to any particular group of examinees. In particular, reviewers were asked to evaluate items and passages to determine whether they contained any material that may portray any group unfavorably (or favorably) or in a stereotypical fashion. Reviewers were asked to pay particular attention to material that may advantage or disadvantage examinees based on an examinee's gender, age, race/ethnicity, religion, disabilities, lifestyle, or community type. The director of the Test Development Unit also reviewed all tryout items and final form items for potentially sensitive or offensive material.

All GEDTS staff members involved in test development were trained in the process of sensitivity review. In addition to removing items to which different groups of people may be sensitive or disadvantaged/advantaged, GEDTS attempted to ensure that the tests included content familiar to all groups of examinees.

DEVELOPMENT OF PART II (ESSAY) OF LANGUAGE ARTS, WRITING TEST

Overview

The 1982 GED Tests Specifications Committee recommended adding an essay to the GED test battery, believing that no one should receive credit for high school equivalency without being asked directly, as well as indirectly, to demonstrate writing ability. The 1997–1998 Specifications Committee recommended that an examinee's writing skills should continue to be measured, both directly and indirectly, on the 2002 series Language Arts, Writing Test. As in the previous test series, examinees are asked to compose an essay and to support the essay through personal observations and experience. Part II thus remains an essay-writing exercise.

Acting on this recommendation, GED Testing Service established a permanent Writing Advisory Committee to oversee the development and maintenance of Part II of the Language Arts, Writing Test. The Writing Advisory Committee not only provides expert judgment in a variety of aspects for the essay section, but it also provides continuity across administrations, topics, scoring sites, and testing sessions. To qualify for the committee, nominees must be able to meet all Chief Reader specifications and must pass the Chief Reader training (see Chapter 5). A new member of the committee must attend all meetings and matriculate gradually into the decision-making process. The current Writing Advisory Committee members include three university professors with a background in national writing assessment, an English language arts content specialist for an urban public school system, and a high school English department chair who participates in national writing projects.

Because Writing Advisory Committee members are or have been directly involved in writing instruction on a regular basis, they are familiar with instructional expectations for GED examinees at different grade levels and levels of performance. The Writing Advisory Committee helps stabilize the essay scoring scale by assisting with the decision on the type of essay required, the choice of the scoring rubric, the development

of the 2002 Series GED Writing Test Official Essay Scoring Guide (which provides the scoring criteria), and the selection of the reader training sets (described in Chapter 5).

In Part II of the Language Arts, Writing Test, examinees are given a single expository topic and are directed to write an original essay. (A list of sample GED essay topics is provided in Appendix E.) The following sections describe the development and maintenance of the essay test itself. Issues relating to reliability and validity of the writing assessment are discussed in Chapters 4 and 5, respectively.

Development of Essay Topics

Acting on the recommendation of the GED Tests Specifications Committee, the Writing Advisory Committee decided to limit essay topics to one type: expository. This immediately helped minimize the problem of controlling topic variability. Judgment criteria were developed to help the Writing Advisory Committee evaluate potential topics (see Appendix F for a partial list of these judgmental criteria, excerpted from GEDTS, 1993). All topics developed for use in the GED Tests are of the same rhetorical type and are of equal length and format.

Developing operational topics was a multi-step process. After trial essay topics were reviewed and edited by the Writing Advisory Committee, each topic was field-tested. The Writing Advisory Committee read and evaluated essays written on the potential topics from these field tests, rejecting any topics for which the essays did not meet specified criteria (see Appendix F). During the field testing, anchor essay topics were also administered, which assisted in evaluating the trial topics.

Statistical evaluations were then performed on the surviving topics. A Kruskal-Wallis test was conducted to determine the distribution of scores on each topic. The Kruskal-Wallis test compared the 2001 norming sample score distribution of the anchor topic with the score distribution of the trial topic. For a topic to be acceptable, the two distributions were not allowed to differ significantly ($p > 0.10$). In addition, a statistical test was performed to determine the equality of correlations between multiple-choice scores and essay scores. This test evaluated the equality of the correlation between the anchor topic essay and multiple-choice scores, and the correlation between the trial topic essay scores and multiple-choice scores. For the trial topic to be accepted, the difference between these two correlations could *not* differ significantly ($p > 0.10$). Only trial topics that yielded non-significant results on both statistical tests were retained for operational use.

Standards for Scoring Essays

To ensure clarity for evaluating the consistency of essay scoring across topics, as well as within and across scoring sites and sessions, the Official Essay Scoring Guide was developed.[8] The design of the 2002 Series GED Writing Test Official Essay Scoring Guide (see Appendix G) had to be descriptive in nature, rather than prescriptive, because the direct assessment of writing involved in Part II of the Language Arts, Writing Test is norm-referenced. In order to define the scoring guide used to evaluate the writing of GED examinees, GEDTS staff members procured a large, stratified, and random national sample of direct writing from graduating high school seniors in 1987. After the 1997–1998 Tests Specifications Committee changed the original six-point scale to a four-point scale, the Writing Advisory Committee was asked to rank-order the essays in four categories, based on the quality of the writing. The 2002 Series GED Writing Test Official Essay Scoring Guide defines the general characteristics of the essays at each of the four scoring points. The committee's standards are invariant but not prompt-specific, and they are intended to apply across topics. These standards will remain in effect until a new norming standard is conducted.

[8] The Official GED Essay Scoring Guide was originally developed in 1997. However, see GEDTS (2005a) for full details.

CANADIAN VERSION

The English-language Canadian version of the GED Tests follows the same test specifications as those for the English-language U.S. version, with the exception of the Social Studies Test (see Social Studies Test specifications section above). The GED Tests Specifications Committee did contain Canadian members, who represented the Canadian educational curriculum. Thus, although the test specifications were developed to primarily represent the U.S. national and state curricula, they are also representative of Canadian education standards.

SPANISH-LANGUAGE GED TESTS

The Spanish-language GED Tests follow most of the same specifications as those for the English-language U.S. version. More specifically, the Spanish-language version of the Social Studies; Science; Language Arts, Reading; and Mathematics Tests is a direct translation of the English-language U.S. version. Almost 90 percent of the Language Arts, Writing Test items were also direct translations. A select few (less than 10 percent) of the English-language U.S. version of the Language Arts, Writing Test items were replaced altogether to avoid translation issues. The testing times for each Spanish-language GED content area test are listed in Chapter 1.

FRENCH-LANGUAGE GED TESTS

The French-language GED Tests have specifications that are similar to the English-language Canadian version. The French-language version of the Social Studies; Science; Language Arts, Reading; and Mathematics Tests are direct translations of the respective English-language Canadian version. The content and cognitive specifications for these tests are identical to the English-language Canadian version. The Language Arts, Writing Test was developed independently by the Quebec Ministry of Education and has somewhat different content and cognitive test specifications. The Language Arts, Writing Test is comprised of Spelling and Grammar (50 percent), Syntax and Punctuation (35 percent), and Organization of Text and Ideas (15 percent). The testing times for each of the French-language GED Tests are listed in Chapter 1 and are similar to those for the Spanish-language GED Tests.

Chapter 3: Standardization and Norming, Scaling, and Equating

This chapter describes the processes of norming, scaling, and equating the GED Tests. *Standardization* and *norming* refer to the process of administering the GED Tests to a nationally representative sample of graduating high school seniors (the norm group) to establish typical scores for that norm group. *Scaling* refers to the process of transforming raw GED test scores (e.g., number of items correctly answered) to *scaled scores* that possess desirable qualities useful for comparing scores on the same content area tests across different forms. *Equating* refers to the statistical process of adjusting test scores so that the level of performance indicated by a particular scaled score is consistent from form to form. These three processes are described more thoroughly in this chapter.

STANDARDIZATION AND NORMING

As stated in Chapter 1, the purpose of the GED Tests is to provide an opportunity for adults who did not complete a formal high school program to certify their attainment of high school–level academic knowledge and skills and earn their jurisdictions' high school equivalency credential. In order to allow adults the opportunity to demonstrate that their knowledge and skills are comparable to that of high school graduates, the score scales for the GED Tests are referenced to the performance of graduating high school seniors on these same tests. This referencing of the GED Tests score scales to a nationally representative group of graduating high school seniors is called *norming*. The 2002 Series GED Tests were standardized and normed using a nationally representative sample of graduating seniors who took the GED Tests during March, April, and May 2001. The standards for score scales for the test forms developed after 2001 have been based on the performance of this norm group.

Periodically, changes in national curricular trends dictate changes in the content of the GED Tests. When these changes occur, the "new" forms cannot be equated to the "older" forms, and a new standardization and norming study must be performed. Norming studies are also conducted whenever it is suspected that changes in achievement levels may have occurred in the norm group (i.e., graduating high school seniors). In 1967, 1977, 1987, and 2001, norming studies on the GED Tests were conducted because of changes in test content. In 1955, 1980, and 1996, norming studies were conducted because of perceived changes in the achievement levels of graduating high school seniors. In all cases, the new norms reflected a new set of performance standards for obtaining a GED credential.

Sample

The sample for the 2001 standardization and norming was obtained through a two-stage stratified random sampling process. The two stages consisted of sampling schools and then sampling students within schools. This procedure ensured that the makeup of the norm group sufficiently represented graduating high school seniors in the United States. The population of eligible U.S. schools included all public and private schools located in the 50 states and District of Columbia that enrolled students at grade 12, except those that (a) did not graduate a senior class; (b) did not enroll students of their own, but rather only received students enrolled in other schools; or (c) were schools such as university lab schools, schools for the deaf or blind, reservation schools, Montessori, special education, vocational, ungraded, alternative, Department of Defense, or other nontraditional schools.

Sampling of schools. Eligible schools were stratified according to school type (public or private) and four geographic regions (Northeast, Midwest, South, and West). The public school sample was stratified proportionally based on 12th grade enrollment data from the National Center for Education Statistics (NCES) Common Core of Data. The stratification of the private school sample was based on data from the 1997–98 Private School Universe Survey conducted by NCES.

Public schools were also stratified according to community type (urban, suburban, and rural) and free lunch eligibility (eligible or not eligible). For the 2001 standardization and norming, there were a total of 28 strata: 24 strata (four regions by three community type levels by two free lunch eligibility levels) for public schools and four strata for non-public schools (four regions). The sample of schools was determined by randomly selecting schools in each stratum in proportion to the stratification variables.

Sampling of students. Within each school, personnel staff compiled a sequentially numbered list of eligible students. Eligible students were grade 12 students expected to graduate by fall 2001. Students were excluded who (a) would not be awarded a traditional high school diploma by the following September; or (b) would require a special edition (Braille, audiocassette, or large print) or special administration (e.g., individual rather than group administration) of the test. The number of seniors selected typically ranged from 30 to 40 per school. If the total number of seniors was equal to or less than 30, then the entire group of eligible seniors was tested. If the total number of eligible seniors was greater than 30, a random sample of students was selected from the population of eligible students using a computerized random number generator. The list of random numbers generated for each school was specific to each school.

Because of the length of time required to take the entire GED test battery, high school students in the standardization and norming study were administered only one to three of the five content area GED Tests. Sample sizes for each test ranged from 300 to 700 students. Sample sizes for the Language Arts, Writing Test were somewhat larger than those for the other tests. This larger sample was used in order to increase the number of matched essay and multiple-choice records; matched essay and multiple-choice records can be reduced due to (a) essay topics found to be unacceptable for operational use, or (b) students not obtaining the minimum valid essay score (a score of 2 or greater) in order to generate a valid total Language Arts, Writing Test standard score.

The number of schools participating in the 2001 standardization and norming, and 2002, 2003, and 2005 equating studies (see Equating of Multiple-Choice Tests section) is presented in Table 3.1.

TABLE 3.1
NUMBER OF SCHOOLS PARTICIPATING IN THE 2001 U.S. STANDARDIZATION AND NORMING, AND 2002, 2003, AND 2005 EQUATING STUDIES FOR THE ENGLISH-LANGUAGE GED TESTS

	2001		2002		2003		2005	
	N	%	N	%	N	%	N	%
Public Schools								
Northeast	41	11.4	9	6.8	12	8.2	8	6.5
Urban	29	8.1	6	4.5	7	4.8	5	4.0
Free Lunch Eligible	3	0.6	1	0.8	1	0.7	0	0.0
Free Lunch Ineligible	26	7.2	5	3.8	6	4.1	5	4.0
Suburban	4	1.1	2	1.5	3	2.1	2	1.6
Free Lunch Eligible	1	0.3	0	0.0	1	0.7	1	0.8
Free Lunch Ineligible	3	0.8	2	1.5	2	1.4	1	0.8
Rural	8	2.2	1	0.8	2	1.4	1	0.8
Free Lunch Eligible	0	0.0	0	0.0	0	0.0	0	0.0
Free Lunch Ineligible	8	2.2	1	0.8	2	1.4	1	0.8
Midwest	101	28.1	40	30.3	49	33.6	38	30.6
Urban	54	15.0	13	9.8	14	9.6	10	8.1
Free Lunch Eligible	8	2.2	1	0.8	3	2.1	1	0.8
Free Lunch Ineligible	46	12.5	12	9.1	11	6.8	9	7.3
Suburban	21	5.8	7	5.3	15	10.3	10	8.1
Free Lunch Eligible	2	0.6	0	0.0	6	4.1	5	4.0
Free Lunch Ineligible	19	5.3	7	5.3	9	6.2	5	4.0
Rural	26	7.2	20	15.2	20	13.7	18	14.5
Free Lunch Eligible	6	1.7	3	2.3	3	2.1	3	2.4
Free Lunch Ineligible	20	5.6	17	12.9	17	11.6	15	12.1
South	142	39.6	65	49.2	64	43.8	38	30.6
Urban	81	22.6	31	23.5	31	21.2	16	12.9
Free Lunch Eligible	18	4.7	9	6.8	9	6.2	6	4.8
Free Lunch Ineligible	63	16.2	22	16.7	22	15.1	10	8.1
Suburban	30	8.4	17	12.9	13	8.9	11	8.9
Free Lunch Eligible	16	4.5	9	6.8	7	4.8	7	5.6
Free Lunch Ineligible	14	3.9	8	6.1	6	4.1	4	3.2
Rural	31	8.6	17	12.9	20	13.7	11	8.9
Free Lunch Eligible	16	4.5	6	4.5	7	4.8	5	4.0
Free Lunch Ineligible	15	4.2	11	8.3	13	8.9	6	4.8
West	63	17.5	18	13.6	21	14.4	14	11.3
Urban	44	12.3	10	7.6	11	7.5	7	5.6
Free Lunch Eligible	11	3.1	3	2.3	5	3.4	3	2.4
Free Lunch Ineligible	33	9.2	7	5.3	6	4.1	4	3.2
Suburban	12	3.3	5	3.8	6	4.1	4	3.2
Free Lunch Eligible	1	0.3	2	1.5	1	0.7	1	0.8
Free Lunch Ineligible	11	3.1	3	2.3	5	3.4	3	2.4
Rural	7	1.9	3	2.3	4	2.7	3	2.4
Free Lunch Eligible	3	0.6	0	0.0	2	1.4	2	1.6
Free Lunch Ineligible	4	1.1	3	2.3	2	1.4	1	0.8
Total Public Schools	347	96.7	132	89.2	146	93.0	98	79.0
Private Schools								
Northeast	2	16.7	3	18.8	2	18.2	11	42.3
Midwest	6	50.0	9	56.3	8	72.7	4	15.3
South	3	25.0	1	6.3	0	0.0	5	19.2
West	1	8.3	3	18.8	1	9.1	6	23.1
Total Private Schools	12	3.3	16	10.8	11	7.0	26	20.9
Total Schools	359		148		157		124	
Total Students	10,160		4,718		5,145		5,356	

An incentive was provided to each participating school in an effort to maintain an adequate sample size. However, incentives were not necessarily provided directly to the high school seniors who participated in either the standardization study or the subsequent equating studies (described in the Equating of Multiple-Choice Tests section) or item-tryout studies.[9] As is the case with many other large-scale testing programs that use similar norming processes, the low stakes associated with the administration of these tests may have affected the quality and integrity of the reported data. No measure of motivation or effort was obtained directly from the student during these studies. However, data records in which more than one-third of the item responses were missing were excluded from analyses.

Adequacy of the Sampling Procedure. Benners and George-Ezzelle (2006) compared the ethnicity distributions of the high school senior samples from the 2001 standardization and subsequent equating studies to national estimates. They concluded that the samples were representative of the U.S. population of high school seniors.

SCALING OF MULTIPLE-CHOICE TESTS

Prior to the 2002 Series GED Tests, the GED standard scores were scaled to a mean of 50, standard deviation of 10, and ranged from 20 to 80. A new standard score scale was constructed for the 2002 test series. For each of the five tests in the GED test battery, a standard score scale was constructed with a mean of 500, standard deviation of 100, and a range of 200 to 800. Each operational and practice test form was placed on this scale in order to permit comparisons of examinee scores across test forms. The procedure used to establish the standard score scale for the test forms (IA, IB, and IC) administered in the 2001 standardization and norming study is described next.

For each test in the battery, cumulative proportion distributions of scores were pre-smoothed using the log-linear method.[10] These smoothed distributions were independently normalized by converting the midpoint of each interval (i.e., raw score unit) in the smoothed distribution to a standard normal deviate, or *z*-score. These scores were transformed linearly to produce a distribution of standard scores with a mean of 500 and a standard deviation of 100. Scores more than three standard deviations from the mean were truncated to conform to the 200 to 800 range.[11]

[9] Participating schools may have provided incentives to students.

[10] The implementation of this method varied by year, using computer software programs developed by Hanson (1992) or Cui and Chien (2004).

[11] Additional information related to scale score stability can be found in Benners and George-Ezzelle (2006).

EQUATING OF MULTIPLE-CHOICE TESTS

2001 Standardization Study

As previously noted, three test forms were administered during the 2001 standardization and norming study, namely, Forms IA, IB, and IC. These three forms were spiraled within the standardization sample. Thus, the groups taking each of these test forms were assumed to be approximately randomly equivalent. The test forms were constructed to be similar to one another: All test forms were produced from the same content specifications and all items were matched as closely as possible on psychometric characteristics. Form IA served as the anchor form for the Language Arts, Reading; Social Studies; and Mathematics Tests; and Form IC served as the anchor form for the Language Arts, Writing and Science Tests for the 2002 series forms.[12] In order to control for any minor, unintended differences in test forms, raw scores on the two additional forms were equated to those for the anchor form using the equipercentile method within the random groups design.

The purpose of equating is to produce a relationship of equivalence between raw scores on two or more tests. Several methods for equating have been documented and reviewed within the literature (e.g., Petersen, Kolen, & Hoover, 1989; Kolen & Brennan, 2004). Although several methods for equating the GED Tests have been examined (e.g., Kolen & Whitney, 1982), the equipercentile method of equating test scores has been used since the GED testing program's inception.

For the two additional forms that were equated during the 2001 standardization and norming study, the five tests in each form were equated separately to their counterparts on the anchor form. To do this, cumulative proportion distributions were created from the raw score distributions. The cumulative distributions were then pre-smoothed using the log-linear method. Using the midpoint of each raw score interval in the distribution, a percentile rank was determined for each possible raw score. The raw score on the anchor form that corresponded to the same percentile rank on the two additional forms was considered to be equivalent to the related raw score on the new operational form.

Once the raw score equivalents on the two additional forms were established, standard score conversions for those forms were determined. The standard score associated with the anchor form raw score was linked to the corresponding raw score on the new form. If linear interpolation was needed to obtain the equivalent raw score, then it was also used to obtain the standard score.

The equipercentile method of equating used to equate test forms may also be illustrated by the following example. Suppose scores on a new hypothetical form "XX" are to be equated to an anchor test form "AA." The raw scores on form XX are equated to the standard scores on form AA by identifying the standard scores on form AA that have percentile ranks equivalent to the percentile ranks observed for the form XX raw scores. For example, if a score of 17 on form XX has a percentile rank of 12 (i.e., 12 percent of the students tested with form XX scored at or below a raw score of 17), the standard score on form AA that corresponded to a percentile rank of 12 on form AA is identified. For the purposes of this example, assume that this form AA standard score is 340. A raw score of 17 on form XX would then be assigned a standard score of 340. The standard score of 340 represents the same level of achievement signified by the particular raw score on either test form.

[12] For foreign-language GED Tests, in which some tests are not direct translations of the English-language version, Form IA served as the anchor form for the Language Arts, Writing Test for both French- and Spanish-language GED Tests; Form IC served as the anchor form for the French-language Language Arts, Reading Test.

Equating in Subsequent Years

Eight additional operational forms were developed to be equated in years subsequent to 2001 (ID, IE, and IF in 2002; IG, IH, and II in 2003; and IJ and IK in 2005). To maintain the score scale established in the 2001 standardization and norming study, the new forms were equated to the anchor forms. To make the equating possible, the new forms were spiraled with one of the anchor forms and administered to a nationally representative sample of graduating high school seniors.[13] The sampling procedure was similar to that used in the 2001 standardization and norming study (see Table 3.1 for participating school information). Equating the new forms to the anchor form was accomplished using the equipercentile procedure as described above, with an additional step carried out to ensure that the equated scores maintained the performance levels reflected by the 2001 norm group.

Because the standards of performance required to pass the GED Tests must remain consistent over time, the equating of new test forms must reflect the performance levels of the original (2001) norm group. To ensure this consistency, raw-to-standard score conversions were first established for the forms in the 2001 standardization and norming study using the percentile ranks on the anchor form from that year. However, in assigning percentile ranks for scores on the subsequently developed forms, the percentile ranks for the 2001 administration of the anchor form were used, not the percentile ranks from the current year's administration of the anchor form. Thus, if, in the current year, a raw score of 37 on the anchor form had a percentile rank of 49, but in 2001 it had a percentile rank of 48, the percentile rank of 48 was reported. If interpolation was required to do the equating, then it was also required in this process as well. In this way, the standard scores for all new forms are comparable to the standard score scale of the anchor form, and the percentile ranks for all new forms indicate how examinees compare to the norming sample that took the anchor form in 2001.

SCALING AND EQUATING OF THE LANGUAGE ARTS, WRITING TEST

Deriving and Scaling the Writing Composite Score

The standard score that is reported for the Language Arts, Writing Test is a weighted composite of a multiple-choice raw score and an essay raw score (the multiple-choice and essay standard scores are not reported separately to examinees). The multiple-choice portion of the Language Arts, Writing Test of the anchor form was scaled using the same procedure as the other four tests (by normalizing the smoothed cumulative raw score distribution and linearly converting to a standard score scale). For the essay portion, as described in Chapter 2, two readers scored each essay on a four-point scale, and the total essay raw score was the average of the two readings. In cases in which the discrepancy between the two readers was more than one point, the Chief Reader scored the essay and indicated which reader's score he or she agreed with; the total essay score was then equal to the average of those two essay scores.

The operational topics used in the 2001 standardization and norming study were selected from topics that had been selected according to the judgmental criteria described in Chapter 2. In the 2001 standardization and norming study, these topics were subjected to statistical criteria (i.e., equivalent raw score distributions and consistent essay/multiple-choice score correlations). Those that were finally approved for operational use have been considered "anchor" essay topics.

[13] The anchor form for Mathematics, Social Studies, and Language Arts, Reading was IA for each equating study. For Science and Language Arts, Writing the anchor form was IC. Thus, for example, in 2002, Forms ID, IE, IF, and IA were spiraled for the Mathematics; Social Studies; and Language Arts, Reading Tests and Forms ID, IE, IF, and IC were spiraled for the Science and Language Arts, Writing Tests.

Rather than equate essay topics to the anchor topic, a single raw-to-standard score conversion was used with all essay topics. The standard score scale for essay topics was developed in a manner similar to the score scale for the multiple-choice portion of the test. That is, the cumulative frequency distribution of raw scores was pre-smoothed using the log-linear method. The resulting smoothed distribution was then converted to a normal distribution of standard scores with a mean of 500 and standard deviation of 100.

In subsequent equating studies, one or more anchor topics were included in the data collection design. This was done for two reasons. First, variation in essay scores across equating samples was evaluated. Second, the inclusion of anchor topics ensures that a sufficient number of records with valid essays were available for computing the composite writing score (in the event that a high number of field-tested topics was rejected). New topics needed to meet two criteria: (a) Their raw score distributions must be statistically identical to the anchor topics, and (b) standard scores must correlate with multiple-choice standard scores to the same extent as the anchor topics.

Once raw-to-standard score conversions were developed for both the multiple-choice and essay portions of the Language Arts, Writing Test, a composite standard score for the test was formed by defining weights. The GED Testing Service Advisory Committee stated that the essay should be weighted at least .35 (out of a possible 1.0; Patience & Swartz, 1987, p. 5). These relative weights were calculated in such a way that the essay was weighted as much as possible without diminishing the composite reliability below .86, which was the estimated test-retest reliability for the 1988 series multiple-choice writing test. Under these criteria, nominal (relative) weights of .35 and .65 were chosen for the essay and multiple-choice portions, respectively.

The relative weights of .35 and .65 were chosen for the raw scores, and thus it was necessary to adjust the relative weights in order to maintain the desired standard deviation of 100 for the composite standard score. To do this, the formula for the variance of a weighted composite was adapted through simultaneously solving the following equations (Patience & Swartz, 1987):

$$\frac{W_{es}}{W_{es} + W_{mc}} = 0.35, \text{ and} \qquad \textbf{(3.1)}$$

$$1 = {W_{es}}^2 + {W_{mc}}^2 + 2(r_{mc,es})W_{es}W_{mc} \qquad \textbf{(3.2)}$$

where W_{es} is the essay weight, W_{mc} is the multiple-choice weight, and $r_{mc,es}$ is the correlation between essay and multiple-choice standard scores.

Solving these equations for W_{es} and W_{mc} yielded operational weights for each writing test form. In the administration of the Language Arts, Writing Test, several topics were spiraled with each test form. For the purpose of developing weights in each equating study, data from all acceptable topics for each test form were pooled. The pooled data were then used to compute operational weights for each writing test form.

Once the weights were determined, a conversion table was developed for each form of the Language Arts, Writing Test. In this two-way table, a standard score was created for every possible combination of multiple-choice and essay raw scores. In addition, for the anchor form in 2001, percentile ranks were computed for each possible standard score between 200 and 800. This table was based on the cumulative percent distribution of Language Arts, Writing Test composite standard scores in the norm group. In subsequent years, the 2001 composite standard score-to-percentile rank table appears with the raw-to-standard score conversion table for each Language Arts, Writing Test form. This permits, as in the case of the other four GED Tests, references to the 2001 norming sample's distribution of standard scores.

Equating Forms of the Language Arts, Writing Test

For the anchor form in 2001, percentile ranks were computed from the distribution of weighted composite Language Arts, Writing Test scores. In order to equate subsequent forms to the anchor form, the equating of new multiple-choice forms to the multiple-choice portion of the anchor form was carried out, using equipercentile equating in the manner described above for the other four GED Tests. For acceptable essay topics, the 2001 raw-to-standard score conversion for anchor topics was used. Using the equations above, operational weights were then computed. The resulting composite standard scores were then referenced to the 2001 percentile rank distribution.

Chapter 4: Reliability

Reliability refers to the consistency, or stability, of the test scores when the test is repeatedly administered to groups of examinees (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). If a given test yields widely discrepant scores for the same individual on separate test administrations, and the individual has not changed significantly on the attribute that is measured, then the scores on the test are not reliable. Conversely, if a test produces the same or similar scores for an individual on separate administrations, then the scores from the test are considered reliable. Reliability is inversely related to the amount of measurement error associated with test scores. That is, the more measurement error present in test scores, the less reliable the test scores.

Because reliability is a crucial index of test quality, test developers are required to evaluate and report the reliability of their test scores. Several procedures are used to evaluate reliability and each account for different sources of measurement error and thus produce different reliability coefficients. The reliability of scores from the multiple-choice portions of the GED Tests is evaluated by calculating estimates of the internal consistency reliability, the standard error of measurement, and alternate-form reliability. The reliability of the essay portion of the Language Arts, Writing Test is evaluated using additional criteria. More complete descriptions of reliability estimation can be found in Anastasi (1988), Feldt and Brennan (1989), and Lord and Novick (1968).

The results of the reliability analyses for the 2002 Series GED Tests are presented in this chapter. The data presented herein are from the 2001 standardization of Forms IA, IB, and IC, and subsequent equating studies introducing Forms ID through IK. All studies used a random sampling of graduating high school seniors from across the United States, as described in Chapter 3. Brief descriptions of reliability indices are also provided.

RELIABILITY OF GED TEST SCORES

Internal Consistency Reliability

Estimates of the internal consistency reliability of the GED test scores, with the exception of the Language Arts, Writing Test composite score, are based on the K-R 20 reliability coefficient (Kuder & Richardson, 1937), which is a special case of the more general coefficient alpha (Cronbach, 1951). The K-R 20 coefficient is used primarily with tests containing dichotomously scored multiple-choice items. It is an estimate of the extent to which all the items on a test correlate positively with one another. K-R 20 can also be considered an estimate of the expected correlation of a test with an alternate or parallel test form of the same length (Nunnally, 1978).

The formula for the coefficient alpha reliability coefficient (α) is:

$$\alpha = \frac{k}{k-1}\left[1-\left(\frac{\Sigma\sigma_i^2}{\sigma_t^2}\right)\right] \qquad \textbf{(4.1)}$$

where k = the number of items on the test, σ_i^2 = the variance of item i, and σ_t^2 = the variance of the total scores on the test. When the test items are dichotomously scored, the variance for an item becomes the proportion of examinees answering the item correctly (p) multiplied by 1 minus p (referred to as q). Substituting for σ_i^2 in Equation 4.1, the formula for the K-R 20 reliability coefficient for dichotomously scored multiple-choice tests becomes:

$$\text{K-R 20} = \frac{k}{k-1}\left[1-\left(\frac{\Sigma p_i q_i}{\sigma_t^2}\right)\right] \qquad \textbf{(4.2)}$$

where p_i = proportion of examinees answering item i correctly, and q_i = 1 - p_i.

The K-R 20 coefficient ranges from zero to one. As can be seen from Equation 4.1, three factors can affect the magnitude of the K-R 20 coefficient: the homogeneity of the test content (affects σ_i^2), the homogeneity of the examinee population tested (affects σ_t^2), and the number of items on the test (k). Tests comprising items that measure similar (i.e., homogenous) content areas will have higher K-R 20s than tests comprising items measuring diverse content areas because the covariance among the items is likely to be lower when the items measure widely different concepts or skills. Conversely, examinee populations that are highly homogenous can reduce the magnitude of the K-R 20 coefficient because the covariance among the items is limited by the amount of total variance in the examinee population. Assuming that all items correlate positively with one another, adding items to a test increases item covariance, and thus, the K-R 20 reliability coefficient is also increased by adding more items. Because the GED Tests measure highly interrelated content areas, and because of the heterogeneity of the norming populations and the GED examinee population, the K-R 20 reliability estimates for the GED Tests are not as likely to be attenuated by content heterogeneity or examinee homogeneity. However, the K-R 20 coefficients are influenced by differences in the number of items on the content area GED Tests.

Standard Error of Measurement

The standard error of measurement (SEM) is an estimate of the average amount of error that is associated with scores derived from a test. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) defines the SEM as "the standard deviation of a hypothetical distribution of measurement errors that arises when a given population is assessed via a particular test or measure" (p. 27). The SEM is often used to describe how far an examinee's observed test score may be, on average, from his or her "true" score (i.e., a score that is free from measurement error). Therefore, a smaller SEM is preferable to a larger one. The SEM can be used to form a confidence interval around an observed test score to suggest a score interval within which an examinee's true score may fall. Because the SEM is the standard deviation of a hypothetical, normal distribution of measurement errors, in most cases, it is expected that an examinee's observed score will be found within one SEM unit of his or her true score about 68 percent of the time.

The SEM is a function of the standard deviation of the test scores and of the reliability of the test scores. The equation for the SEM is:

$$SEM = \sigma_t \sqrt{1 - r_{tt}} \qquad \textbf{(4.3)}$$

where σ_t = the standard deviation of test scores, and r_{tt} = the reliability coefficient (for the SEM reported here, the reliability coefficient used is the K-R 20). From Equation 4.3, it can be seen that a test with a small standard deviation and large reliability yields a smaller SEM. Because the SEM is a function of the standard deviation of test scores, it is not an absolute measure of error; rather, it is expressed in raw score units. Therefore, unlike reliability coefficients, SEM cannot be compared across tests without considering the unit of measurement, range, and standard deviation of the tests' raw scores.

K-R 20 and SEM Results for the GED Tests

Table 4.1 presents the standard score means, standard deviations, and SEM for the test forms in the 2002 test series. It should be noted that the numbers in Table 4.1 for the Language Arts, Writing Test refer only to the multiple-choice portion of the test (the reliability of the essay scores and Language Arts, Writing Test composite score is discussed later in Chapter 4). The data presented in Table 4.1 facilitate comparison among the five subject tests by presenting the statistics reported in standard score units. Raw score data and K-R 20s are also presented in Table 4.1. The K-R 20s were computed for raw scores only. Because the transformation of raw scores to standard scores (described in Chapter 3) is nonlinear, it is not possible to compute these statistics directly for standard scores. However, the raw score to standard score transformation maintains the rank order of the examinees, and thus, the differences in K-R 20 would be negligible (American College Testing, 1988). The SEM, on the other hand, would be quite different because it is a function of the standard deviation of scores, as well as the reliability coefficient.

The information in Table 4.1 is based on the performance of the nationally representative sample of graduating high school seniors across the United States who took the GED Tests as part of the standardization project in 2001, as well as the equating projects in 2002, 2003, and 2005 (see Appendix H for corresponding reliability information obtained via adult GED examinee data). Data from Forms IA, IB, and IC originated from the 2001 standardization, data from Forms ID, IE, and IF originated from the 2002 equating study, data from IG, IH, and II originated from the 2003 equating study, and data from IJ and IK originated from the 2005 equating study. The results presented in Table 4.1 indicate that all U.S. forms have K-R 20s of at least .92; over 80 percent of the test forms have a K-R 20 of .94 or higher.

TABLE 4.1
SAMPLE SIZE (N), SCORE MEAN, STANDARD DEVIATION (SD), STANDARD ERROR OF MEASUREMENT (SEM), AND K-R 20 ESTIMATES FOR THE 2002 SERIES ENGLISH-LANGUAGE GED TESTS: U.S. GRADUATING HIGH SCHOOL SENIOR DATA

		STANDARD SCORES			RAW SCORES			
TEST/FORM	N	Mean	SD	SEM	Mean	SD	SEM	K-R 20
Language Arts, Writing								
Form IA	391	495.5	102.1	25.8	34.6	11.0	2.8	.94
Form IB	322	493.1	104.1	26.2	35.5	10.8	2.7	.94
Form IC	363	478.8	109.9	26.3	35.7	11.2	2.7	.94
Form ID	357	479.7	120.5	27.5	36.7	11.3	2.6	.95
Form IE	347	501.6	119.4	28.8	36.8	10.8	2.6	.94
Form IF	337	496.0	120.2	27.1	34.2	11.2	2.7	.95
Form IG	369	472.0	117.2	25.8	35.3	11.8	2.6	.95
Form IH	313	463.4	120.6	27.5	35.7	11.6	2.6	.95
Form II	327	499.4	119.4	30.6	36.9	10.2	2.6	.94
Form IJ	835	492.4	117.1	29.8	37.3	10.2	2.6	.94
Form IK	806	495.0	118.4	28.1	36.8	11.0	2.6	.94
Social Studies								
Form IA	462	496.9	103.4	27.9	34.3	10.3	2.8	.93
Form IB	453	499.5	105.4	27.6	33.8	10.8	2.8	.93
Form IC	456	497.9	98.6	25.9	33.5	10.8	2.8	.93
Form ID	385	464.0	106.4	25.0	31.8	11.3	2.8	.95
Form IE	331	471.8	109.6	26.4	30.7	11.4	2.9	.94
Form IF	413	479.6	116.0	26.6	31.2	12.1	2.8	.95
Form IG	579	470.8	114.1	26.0	30.8	12.4	2.8	.95
Form IH	535	468.6	118.3	26.2	30.4	12.8	2.8	.95
Form II	296	478.8	120.3	30.0	28.1	11.8	2.9	.94
Form IJ	826	477.7	119.7	30.9	31.6	11.2	2.9	.93
Form IK	893	481.0	120.5	29.3	33.4	11.4	2.8	.94
Science								
Form IA	105	501.0	101.4	24.7	35.0	11.2	2.7	.94
Form IB	187	494.3	101.1	25.6	34.6	10.8	2.7	.94
Form IC	236	505.8	108.1	28.2	35.7	10.5	2.7	.93
Form ID	169	482.5	115.9	29.1	31.6	11.6	2.9	.94
Form IE	192	487.9	115.0	26.4	32.7	12.0	2.8	.95
Form IF	349	485.3	116.6	28.6	32.2	11.4	2.9	.94
Form IG	571	456.4	115.4	26.5	31.1	12.4	2.8	.95
Form IH	531	453.4	117.1	28.2	29.6	12.0	2.9	.94
Form II	288	455.4	117.1	26.8	29.4	12.7	2.9	.95
Form IJ	818	479.2	122.9	26.2	32.3	13.0	2.8	.96
Form IK	871	480.6	129.1	27.6	33.8	12.7	2.7	.95
Language Arts, Reading								
Form IA	433	518.2	125.0	33.1	29.7	8.8	2.3	.93
Form IB	422	518.0	121.4	30.7	29.7	9.1	2.3	.94
Form IC	411	517.3	123.1	34.3	28.1	8.8	2.4	.92
Form ID	588	497.8	123.2	33.2	26.1	9.3	2.5	.93
Form IE	522	492.5	124.2	31.0	28.3	9.7	2.4	.94
Form IF	553	494.9	124.1	30.9	27.4	9.8	2.4	.94
Form IG	579	509.9	124.9	30.5	28.1	9.8	2.4	.94
Form IH	533	510.9	132.4	29.8	30.0	9.8	2.2	.95
Form II	302	507.0	134.3	31.9	26.1	10.4	2.5	.94
Form IJ	837	508.4	127.6	34.6	30.4	8.5	2.3	.93
Form IK	906	507.9	129.3	37.4	30.2	8.0	2.3	.92
Mathematics								
Form IA	258	512.3	124.3	28.0	34.0	12.1	2.7	.95
Form IB	208	526.7	126.9	28.0	34.6	12.1	2.7	.95
Form IC	278	501.3	101.0	24.3	33.8	11.6	2.8	.94
Form ID	530	485.2	110.6	26.4	33.5	11.7	2.8	.94
Form IE	514	477.9	112.4	26.3	30.4	12.3	2.9	.95
Form IF	505	492.7	114.7	29.0	33.5	11.1	2.8	.94
Form IG	683	471.5	122.7	28.8	29.9	12.6	2.9	.95
Form IH	541	488.4	116.4	27.4	33.0	12.0	2.8	.95
Form II	635	481.2	123.3	29.6	30.1	12.2	2.9	.94
Form IJ	878	495.7	115.2	27.1	33.2	11.9	2.8	.95
Form IK	848	494.6	116.9	27.5	33.6	11.9	2.8	.95

In the equating studies, an anchor test form was administered to a subsample of students within the participating sample of graduating high school seniors. The anchor form for the Language Arts, Writing and Science Tests is Form IC (administered in the 2001 standardization and norming study). For the Social Studies, Language Arts, Reading, and Mathematics Tests, the anchor form is Form IA (also administered in the 2001 standardization and norming study) of the corresponding content area test. That is, the Mathematics Test Form IA is the anchor form for subsequent Mathematics forms. K-R 20s and standard errors of measurement were obtained from the administrations of the anchor form to provide an estimate of the stability of the reliability of this form over successive administrations. Raw and standard score means, standard deviations, SEMs, as well as K-R 20s for the forms from the 2001 standardization and their subsequent use as anchor forms in the equating studies are presented in Table 4.2.

The statistics in Table 4.2 indicate that although there has been some variation in score performance on the anchor form across the study samples, K-R 20s and SEMs have remained consistent. The raw score SEMs within content areas differed by 0.2 raw score units or less across all years; standard score SEMs for within content areas differed by 0.5 to 2.4 standard score units across all years. Nunnally and Bernstein (1994) have noted that "the standard error of measurement is almost one-third as large as the overall standard deviation of test scores even when the reliability is .90" (p. 265). The data in Tables 4.1 and 4.2 indicate that the SEMs for all test forms are typically about 25 percent of the magnitude of the standard deviations with reliability coefficients greater than .90.

TABLE 4.2
SAMPLE SIZE (N), SCORE MEAN, STANDARD DEVIATION (SD), STANDARD ERROR OF MEASUREMENT (SEM), AND K-R 20 ESTIMATES FOR THE 2002 SERIES ENGLISH-LANGUAGE GED TESTS ANCHOR FORMS: U.S. GRADUATING HIGH SCHOOL SENIOR DATA

		STANDARD SCORES			RAW SCORES			
TEST/YEAR	N	Mean	SD	SEM	Mean	SD	SEM	K-R 20
Language Arts, Writing								
2001	363	478.8	109.9	26.3	35.7	11.2	2.7	.94
2002	92	503.2	105.7	25.4	36.2	11.1	2.7	.94
2003	331	488.7	108.0	25.8	34.8	11.5	2.7	.94
2005	854	492.4	100.0	24.9	35.7	10.9	2.7	.94
Social Studies								
2001	462	496.9	103.4	27.9	34.3	10.3	2.8	.93
2002	359	469.5	111.3	27.7	31.4	11.6	2.9	.94
2003	558	465.0	120.7	28.1	30.5	12.3	2.9	.95
2005	828	479.2	119.4	28.2	31.9	12.0	2.8	.94
Science								
2001	236	505.8	108.1	28.2	35.7	10.5	2.7	.93
2002	287	488.6	116.0	26.4	33.7	12.0	2.8	.95
2003	545	453.0	117.6	26.0	29.7	13.0	2.9	.95
2005	809	474.3	120.5	25.8	32.2	12.9	2.8	.95
Language Arts, Reading								
2001	433	518.2	125.0	33.1	29.7	8.8	2.3	.93
2002	363	493.3	124.1	31.1	27.6	9.7	2.4	.94
2003	543	501.1	130.6	31.7	28.2	9.8	2.4	.94
2005	824	513.4	131.4	33.0	29.1	9.4	2.4	.94
Mathematics								
2001	258	512.3	124.3	28.0	34.0	12.1	2.7	.95
2002	353	487.0	113.8	26.4	31.7	12.2	2.8	.95
2003	689	485.6	119.1	27.7	31.4	12.3	2.9	.95
2005	903	494.6	119.2	26.9	32.6	12.4	2.8	.95

Conditional Standard Errors of Measurement

As described above, the SEM provides an estimate of the average amount of error associated with an examinee's observed test score. However, the amount of error associated with test scores may differ at various points along the score scale. For this reason, the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) states:

> Conditional standard errors of measurement should be reported at several score levels if constancy cannot be assumed. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score. (p. 35)

As described in Chapter 1, the passing standard requirements for a GED credential are set at the jurisdiction level. However, for the individual content area GED Tests, the minimum score requirements are usually along the standard score interval of 410 to 450. Thus, it is important to estimate the amount of error of measurement along this score interval.

Conditional standard errors of measurement (CSEMs, i.e., SEMs at specific points or intervals along the score scale) were estimated using an approximation procedure described by Feldt and Qualls (1998). The information required for these calculations includes K-R 20 and K-R 21 for the raw scores, the mean and standard deviation of the raw scores, and a constant, C, which is determined a priori (as recommended by Feldt and Qualls, a constant value of 4 was used for these analyses). This process involves estimating the number of conditional SEMs within the range of $X_0 \pm C$, where X_0 refers to the raw score of interest. The assumption is that the same range of corresponding standard scores will have the same number of SEMs in scale score units. The CSEM for the raw score, $CSEM_{R(X)}$, was calculated as

$$CSEM_{R(X)} = \left[\left(\frac{1-KR_{20}}{1-KR_{21}}\right)\left(\frac{X_0(k-X_0)}{k-1}\right)\right]^{1/2}. \qquad \textbf{(4.4)}$$

In Equation 4.4, k is equal to the test length. This raw score standard error at point X_0 is used in the following equation to estimate the standard score conditional standard error of measurement.

$$CSEM_{SS(X)} = \left(\frac{SS_U - SS_L}{2C}\right) CSEM_{R(X)} \qquad \textbf{(4.5)}$$

In Equation 4.5, SS_U (standard score-upper value) and SS_L (standard score-lower value) are obtained using the raw-to-standard score conversion tables. Specifically, SS_U corresponds with the scaled score associated with the raw score at location $X_0 + C$. SS_L is obtained in a similar manner using $X_0 - C$.

The Language Arts, Writing Test score is derived by combining weighted multiple-choice and essay portions. As such, the raw-to-standard score conversions are not as direct as with other content area GED Tests. Therefore, the approximation method described above could not be applied to the Language Arts, Writing Test. The reliability of the essay scores and writing test composite score is evaluated more thoroughly in subsequent sections of this chapter.

The scale score CSEM for values between 400 and 460 are provided in Table 4.3. The scale score CSEMs for adult GED examinee data are provided in Appendix I. The values in Table 4.3 were derived using data collected during the standardization and equating studies, which utilized graduating high school senior data. Most of the CSEM estimates are lower than the SEM estimates provided in Table 4.1, above.

The CSEM can be used to estimate the margin of error associated with a given test score. The test score is used as an estimate of a person's true score, which is the theoretical average score a person would receive if he or she took parallel versions of a test an infinite number of times. Because the test score is not perfectly reliable, there is a certain level of measurement error associated with each test score. The CSEM can be used to provide a range of values within which the person's true score would fall. For example, if a test-taker receives a score of 450 on Science Form IA, his or her true score will fall within ± 1 standard error of measurement (20.6) of that score 68 percent of the time. In this case, the interval for this score would range from 429 to 471. In other words, if this person took the same test (or a parallel version of it) 100 times, his or her standard score would be expected to fall within the range of 429 to 471 about 68 times.

TABLE 4.3
STANDARD SCORE CONDITIONAL STANDARD ERRORS OF MEASUREMENT AT VARIOUS STANDARD SCORES FOR THE 2002 SERIES ENGLISH-LANGUAGE GED TESTS: U.S. GRADUATING HIGH SCHOOL SENIOR DATA

	STANDARD SCORE						
TEST/FORM	400	410	420	430	440	450	460
Social Studies							
Form IA	24.6	24.6	24.6	24.4	24.3	24.1	27.5
Form IB	25.5	30.0	25.8	25.8	25.7	25.5	25.3
Form IC	25.3	25.5	25.6	25.7	25.6	25.5	25.1
Form ID	18.5	22.1	21.9	21.7	25.0	20.8	20.4
Form IE	20.3	24.3	24.2	24.0	23.6	27.3	23.1
Form IF	24.0	24.2	24.2	24.2	24.0	23.9	23.7
Form IG	20.6	24.8	24.9	24.8	24.7	24.5	24.4
Form IH	25.6	30.0	25.8	25.8	25.6	25.3	25.1
Form II	32.1	24.3	24.5	24.7	24.8	24.8	24.8
Form IJ	30.3	26.0	25.9	25.8	25.7	25.5	25.3
Form IK	29.6	25.3	25.1	24.9	24.7	28.5	24.1
Science							
Form IA	21.4	21.4	17.1	17.0	21.0	20.6	20.3
Form IB	20.2	20.3	20.2	20.0	19.9	19.5	19.2
Form IC	21.7	17.3	21.6	21.4	21.3	20.8	20.6
Form ID	25.2	21.0	16.8	20.9	16.7	20.7	20.6
Form IE	24.4	20.4	16.3	20.4	20.2	20.0	15.9
Form IF	33.1	20.7	16.6	20.7	16.5	20.5	20.3
Form IG	24.9	20.5	20.4	16.2	20.0	19.7	23.3
Form IH	25.7	17.0	21.1	21.0	16.6	20.6	20.3
Form II	21.0	21.0	16.7	20.8	20.6	16.3	20.2
Form IJ	24.2	20.2	16.2	16.1	20.0	19.8	19.4
Form IK	27.0	17.9	22.1	17.5	21.7	25.3	24.8
Language Arts, Reading							
Form IA	23.1	23.0	22.8	18.6	22.0	21.6	24.1
Form IB	22.4	22.3	22.1	21.9	21.4	21.0	20.5
Form IC	23.2	23.3	19.4	23.2	22.8	18.8	22.2
Form ID	22.2	18.5	22.1	22.0	21.8	17.9	21.2
Form IE	19.1	18.7	22.1	21.7	21.2	24.1	26.7
Form IF	19.0	22.7	22.6	18.4	21.8	21.4	20.9
Form IG	11.6	19.4	23.1	22.9	22.2	21.8	21.4
Form IH	15.3	22.6	22.0	21.6	24.0	26.6	32.0
Form II	19.9	23.9	23.9	23.8	23.7	23.5	19.3
Form IJ	22.3	21.7	21.3	20.9	23.7	26.3	31.7
Form IK	25.2	21.3	20.9	20.4	23.2	25.7	31.0
Mathematics							
Form IA	25.1	25.4	25.4	25.4	25.4	25.3	25.1
Form IB	16.7	25.3	25.7	25.8	25.7	25.5	24.8
Form IC	21.6	30.4	26.2	26.2	26.0	25.9	25.4
Form ID	17.1	25.8	26.3	26.4	26.3	26.0	25.3
Form IE	24.1	24.4	24.4	24.5	24.4	24.3	24.1
Form IF	20.5	24.6	24.5	24.3	20.1	23.9	23.6
Form IG	24.9	25.1	25.1	25.0	24.9	20.5	24.4
Form IH	20.5	24.6	24.6	24.5	24.3	23.9	23.7
Form II	30.6	26.3	26.4	26.4	26.4	26.2	26.1
Form IJ	24.4	24.4	24.5	24.4	24.3	24.1	20.0
Form IK	20.5	24.7	24.6	24.5	24.3	20.1	23.9

Alternate-Form Reliability

Alternate-form reliability refers to the correlation between the scores derived from two different forms of a test that are administered to the same group of examinees. Because the two test forms are designed to measure the same proficiency (i.e., are developed from the same content specifications and are designed to have identical psychometric characteristics), examinees should receive similar scores on both test forms. The greater the similarity of examinee scores on the two test forms, the greater the alternate-form reliability.

An alternate-form reliability study was conducted in the spring of 2004 using forms administered in the 2003 equating study. Ninety-six schools (88 public and eight private) in the United States with a graduating senior class were invited to participate in the study. The schools were offered a cash honorarium and GED test score summary reports as incentives for participation. Participation in the study required each school to test a minimum of 35 graduating seniors (who would not require testing accommodations) over two testing sessions, each up to two hours in length and within any one-week period, but not on the same day, in one of the following content areas: reading, writing, mathematics, science, or social studies.

Eighty schools agreed to participate in the alternate-form reliability study, and 77 schools provided usable test data. Data from three schools (two public, one private) were not used because the test administrators at those schools reported that approximately only 25 percent of the students attempted to "do their best" on the tests. If a student completed both tests and met an inclusion rule (i.e., if he or she had answered at least one-third of the items and achieved a raw score greater than zero), his or her data were included in the analysis. Students who took the Language Arts, Writing Tests were required to have valid essay scores on both tests (in order to derive standard scores) for their data to be included in the analysis.

The final analysis file contained data from 2,557 graduating high school seniors from 77 schools (73 public, four private). Table 4.4 lists the gender and race/ethnicity of the students in the study. As shown in Table 4.5, 50 of the schools (65 percent; 62 percent of the students) administered the two test sessions with either one or two days between test sessions. Four schools (5 percent; 6 percent of the students) reported test session dates outside the designated time interval or did not report test dates; these schools were nevertheless included in the analyses.

TABLE 4.4
GENDER AND RACE/ETHNICITY OF GRADUATING HIGH SCHOOL SENIORS IN 2004 ALTERNATE-FORM RELIABILITY STUDY

	N	Sample %
Gender		
Male	1,233	48.2
Female	1,301	50.9
No response	23	0.9
Race/Ethnicity		
Alaskan Native	3	0.1
American Indian	8	0.3
Asian	75	2.9
African American	292	11.4
Pacific Islander	21	0.8
White	1,925	75.3
Other	174	6.8
No/Invalid response	59	2.3
Hispanic		
Yes	206	8.1
No	1,969	77.0
No response	382	14.9

TABLE 4.5
NUMBER OF STUDENTS/SCHOOLS TESTING AT VARIOUS TIME INTERVALS BETWEEN TEST SESSIONS

TIME BETWEEN	STUDENTS		SCHOOLS	
TEST SESSIONS	N	Sample %	N	Sample %
0 days	66	2.6	2	2.6
1 day	1,135	44.4	35	45.5
2 days	464	18.1	15	19.5
3 days	265	10.4	7	9.1
4 days	47	1.8	1	1.3
5 days	134	5.2	4	5.2
6 days	112	4.4	4	5.2
7 days	225	8.8	7	9.1
8 days	20	0.8	1	1.3
Missing data*	89	3.5	1	1.3

*No test dates were provided, or an invalid school code was written on the answer sheet.

Three equated parallel forms of each GED content area test were used in the alternate-form reliability study. The pairing and administration order of parallel test forms within a content area followed a counterbalanced design spiraled (see Table 4.6) within each school. In addition to student performance on the GED Tests, the study also collected information on students' gender, race/ethnicity, and high school coursework.

The number of students who met all eligibility requirements within each pairing of parallel test forms is shown in Table 4.6. Form 1 was most likely to be the first form administered and least likely to be the second form administered; form 3 was least likely to be the first form administered and most likely to be the second form administered. The Language Arts, Writing Test had the fewest number of students, perhaps because Language Arts, Writing Test cases were excluded if the student did not—on both test sessions—obtain a valid essay score (a score of 2 or greater on a 1–4 point scale), answer at least one-third of the items, and achieve a raw score greater than zero.

Table 4.7 reports the GED Tests standard score descriptive statistics and the alternate-form reliability correlation coefficient (Pearson correlation). For four of the GED Tests, mean performance on the first form administered was slightly higher (less than .25 standard deviation) than the mean performance on the second form administered. Alternate-form reliability correlation coefficients were highest for Social Studies, Science, and Mathematics (.82) and lowest for the Language Arts, Writing (.74) and Language Arts, Reading Tests (.74 and .72, respectively). Although not reported in Table 4.7, note also that the correlation between essay scores was estimated as .52. The reliability coefficients obtained are typical of those found with other multiple-choice tests of academic achievement.

TABLE 4.6
NUMBER OF STUDENTS TESTED WITHIN EACH TEST FORM PAIRING

GED Test/Form Pairing	N	%
Language Arts, Writing		
Form 1, Form 2	77	19.6
Form 1, Form 3	89	22.7
Form 2, Form 1	36	9.2
Form 2, Form 3	85	21.6
Form 3, Form 1	39	9.9
Form 3, Form 2	67	17.1
Total	393	
Social Studies		
Form 1, Form 2	125	24.5
Form 1, Form 3	70	13.7
Form 2, Form 1	80	15.7
Form 2, Form 3	111	21.8
Form 3, Form 1	73	14.3
Form 3, Form 2	51	10.0
Total	510	
Science		
Form 1, Form 2	124	26.2
Form 1, Form 3	78	16.5
Form 2, Form 1	69	14.6
Form 2, Form 3	101	21.3
Form 3, Form 1	62	13.1
Form 3, Form 2	40	8.4
Total	474	
Language Arts, Reading		
Form 1, Form 2	122	19.0
Form 1, Form 3	136	21.2
Form 2, Form 1	72	11.2
Form 2, Form 3	146	22.9
Form 3, Form 1	82	12.8
Form 3, Form 2	83	13.0
Total	641	
Mathematics		
Form 1, Form 2	105	19.5
Form 1, Form 3	103	19.1
Form 2, Form 1	82	15.2
Form 2, Form 3	115	21.3
Form 3, Form 1	68	12.6
Form 3, Form 2	66	12.2
Total	539	

TABLE 4.7
GED TESTS STANDARD SCORE DESCRIPTIVE STATISTICS AND ALTERNATE-FORM RELIABILITY CORRELATION COEFFICIENTS

GED Tests	Form Order	N	Mean	SD	Median	Min	Max	*r*
Language Arts, Writing	1st	393	560.7	106.9	560	310	800	
	2nd	393	538.7	105.5	530	240	800	.74
Social Studies	1st	510	474.9	102.4	470	200	800	
	2nd	510	455.1	107.1	440	200	800	.82
Science	1st	474	512.9	112.4	520	210	800	
	2nd	474	497.2	118.1	500	230	800	.82
Language Arts, Reading	1st	641	522.3	127.3	520	210	800	
	2nd	641	508.7	131.7	490	210	800	.72
Mathematics	1st	539	502.6	102.3	510	200	800	
	2nd	539	503.5	117.3	510	210	800	.82

The proportion of agreement and Cohen's Kappa coefficients calculated for pass status (a student passed the test if the standard score was 410 or higher) on the first and second test forms administered are presented in Table 4.8. Cohen's Kappa is a measure of agreement that corrects for chance and has a range from zero to one, with larger values indicating greater reliability. The high proportion of agreement for the Language Arts, Writing Test (93 percent of the students either passed both tests or failed both tests) reduces the interpretability of Cohen's Kappa for that test. Kappa coefficients ranging from .40 to .59 are considered "fair," and those ranging from .60 to .74 are considered "good," according to Fleiss (1981). However, this measure does not give weight to how far apart the pass/fail scores are on the parallel test forms. Agreement of pass status between the two test sessions occurred with 84 percent to 93 percent of the students, depending on the particular GED content area test. Most of the Kappa coefficients were in the middle to upper part of the "fair" range or higher.

TABLE 4.8
PROPORTION AGREEMENT AND KAPPA COEFFICIENTS FOR PASS STATUS

GED Tests	Proportion Agreement	Cohen's Kappa	95% Confidence Interval Lower	Upper
Language Arts, Writing	.93	.48	.28	.67
Social Studies	.84	.60	.52	.68
Science	.87	.55	.45	.65
Language Arts, Reading	.87	.57	.48	.66
Mathematics	.90	.63	.52	.73

In addition to the 2004 alternate-form study, a subsample of seniors participating in the 2003 equating study was administered two half-length practice tests. Correlations between the scores from the half-length practice tests were then obtained. Because of the differences in length between the half-length practice forms and the full-length forms, practice tests and full-length forms are not strictly "parallel." Raw scores on the two half-length tests were correlated and then adjusted using the Spearman-Brown prophecy formula (using a factor of two). The Spearman-Brown prophecy formula to obtain the corrected estimate of the reliability coefficient of the full-length test given a correlation between half-length tests (from Gulliksen, 1950, p. 63) is presented below:

$$r_c = \frac{2r_{12}}{1 + r_{12}} \qquad \textbf{(4.6)}$$

where r_c = the estimated alternate-form reliability coefficient and r_{12} = the correlation of scores from the two half-length test forms. The alternate-form reliability estimates are presented in Table 4.9, along with the unadjusted correlations.

The alternate-form reliability estimates reported in Table 4.9 are lower than the K-R 20 reliabilities reported for the same content area full-length tests. This finding is consistent with previous research (e.g., Whitney, Malizio, & Patience, 1986). However, it should be noted that because these reliabilities were estimated using adjustments to correlations obtained from half-length and full-length test forms, they are less precise than estimates that would have been obtained had "true" parallel forms been used. Thus, the adjusted correlations reported above are dependent upon the degree to which the half-length practice tests are representative of the content of the full-length operational tests.

TABLE 4.9
ALTERNATE-FORM RELIABILITY CORRELATION COEFFICIENTS: OFFICIAL GED PRACTICE TESTS

		r	
Test	N	Unadjusted	Corrected
Language Arts, Writing	255	.83	.91
Social Studies	271	.82	.90
Science	265	.80	.89
Language Arts, Reading	262	.83	.91
Mathematics	205	.84	.91

Note: The Language Arts, Writing; Social Studies; Science; and Mathematics Official GED Practice Tests had 25 items each. The Language Arts, Reading Official GED Practice Test had 20 items.

RELIABILITY OF ESSAY SCORES ON THE LANGUAGE ARTS, WRITING TEST

The reliability of the essay portion of the Language Arts, Writing Test was evaluated by analyzing reader agreement, or inter-rater reliability, and scoring stability. Essay scoring sessions must show evidence of reader agreement and scoring stability. Reader agreement refers to the degree of agreement of scores assigned among different readers scoring the same essays. Inter-rater reliability increases as the number of essays that require attention from the Chief Reader (due to differences between two readers' scores being greater than one point) decreases. Scoring stability refers to how well the scoring sites maintain the scoring standards established by the 2002 Series Writing Advisory Committee and presented in the 2002 Series GED Writing Test Official Essay Scoring Guide.

Maintaining scoring consistent with official GED essay scoring standards is essential in an essay scoring session. The standards for scoring GED essays must remain fixed, regardless of when the essay is administered, where it is scored, or what specific procedures are used in the scoring session itself. A high degree of inter-rater reliability does not ensure scoring stability. Just because readers agree with each other on the assignment of essay scores does not necessarily mean they are assigning the scores according to the standards defined in the scoring guide.

To achieve inter-rater reliability and scoring stability, the essay scoring standards are regularly reinforced. As the readers score the essays, a Chief Reader selects scored essays at random to verify that the readers' scoring is consistent with the definitions in the scoring guide. In cases of a disagreement in assigned essay scores, the Chief Reader discusses the essay with both readers. The monitoring process continues throughout the entire scoring session. Through this system of checks and rechecks, assurance is gained that readers are scoring according to the standards defined in the scoring guide.

Site Monitoring and Scoring Stability

To facilitate scoring stability, Chief Reader training, and site certification, site monitoring procedures were incorporated into the essay scoring process. Chief Reader training and site certification are described in Chapter 5, and site monitoring is described below.

The administration and scoring of the essays are decentralized, both in location and in frequency of scoring sessions. The reliability of the GED essay score is evaluated with respect to the congruence between the essay scores assigned by the scoring site readers and those assigned by the GED Testing Service Writing Advisory Committee (scoring stability).

In the past, GEDTS has conducted two types of monitoring: random monitoring and systematic monitoring. In random monitoring, a randomly selected set of 40 scored essays from a scoring site is rescored by the Writing Advisory Committee. In systematic monitoring, a common set of 40 essays, scored by the Writing Advisory Committee, is sent to each essay scoring site where the site's readers rescore the essays. In both types of monitoring, the site is evaluated by determining the congruence of its readers' essay scores to the Writing Advisory Committee's essay scores.

Scoring sites must demonstrate scoring stability, or adherence to the scoring standards established by the Writing Advisory Committee, in order to become a certified essay scoring site. Scoring sites are certified only after they demonstrate a required level of proficiency on several scoring stability criteria. In 2002, in order to be certified by GED Testing Service, each scoring site was required to have at least 90 percent of the scores on the selected essays assigned a score within one point of the Writing Advisory Committee scores *and* a correlation between reader and committee scores of .70 or higher. In addition, each reader

was required to have at least 35 percent of his or her essay scores equal to the score assigned by the Writing Advisory Committee *and* have no more than 7 percent of his or her essay scores differ by more than one point from the Writing Advisory Committee scores. Beginning in 2005, the criterion for a site's correlation between reader and committee scores was raised to .80 or higher. Additionally in 2005, each reader now was required to have at least 50 percent of his or her essay scores equal to the score assigned by the Writing Advisory Committee *and* have no more than 5 percent of his or her essay scores differ by more than one point from the Writing Advisory Committee scores.

Table 4.10 shows the results of the systematic site monitoring of GED essay scoring sites in 2008 (results for 2002 through 2007 are provided in Appendix J). The identities of specific scoring sites have not been revealed; instead, sites have been randomly assigned a number between 1 and *s*, where *s* equals the number of sites. Based on the scoring sites represented, the median percentages of agreement between essay scores from the scoring sites and the Writing Advisory Committee from 2002 through 2008 ranged as follows: 73–81 percent of scores were equal, 99.7–100 percent of scores differed by one point or less, and 0.0–0.3 percent of scores differed by more than one point (on the four-point holistic scale). Median correlations between the sites' readers' scores and the Writing Advisory Committee scores ranged from .89 to .95 across the seven years.

The agreement rates varied from site to site, as did the correlations between scores. In spite of this variation across sites, all but a single site met and exceeded GED Testing Service's criteria to become a certified essay scoring site.[14] Thus, the results support that it is possible to maintain score scale stability across multiple sites even when the distribution of scores across sites varies. This evidence also substantiates the ability of different scoring sites to apply the official scoring scale and still remain true to the 2002 Series GED Writing Test Official Essay Scoring Guide.

TABLE 4.10
2008 SYSTEMATIC SITE MONITORING RESULTS (FOUR-POINT HOLISTIC SCORING) FOR ENGLISH-LANGUAGE GED ESSAYS

		AGREEMENT OF SCORING SITE'S ESSAY SCORES WITH GEDTS WRITING ADVISORY COMMITTEE SCORES			
Site	Number of Readers	% Scores Equal	% Scores Within One Point	% Scores Differing by > One Point	Correlation*
1	3	66.7	98.3	1.7	.92
2	16	81.3	99.8	0.2	.95
3	8	85.6	100	0.0	.95
4	12	79.8	100	0.0	.93
5	7	78.2	99.6	0.4	.89
6	8	75.0	100	0.0	.93
7	6	83.3	100	0.0	.94
8	7	77.1	100	0.0	.91
9	7	83.9	100	0.0	.97
10	21	85.6	100	0.0	.95
11	4	79.4	100	0.0	.94
12	14	82.1	100	0.0	.96
13	7	81.4	100	0.0	.93
14	3	72.5	100	0.0	.90
15	5	74.5	99.5	0.5	.90
16	5	90.5	100	0.0	.95
17	3	83.3	100	0.0	.94
Mean	8	80.0	99.8	0.2	.93
Median	7	81.3	100	0.0	.94
Minimum	3	66.7	98.3	0.0	.89
Maximum	21	90.5	100	1.7	.97

* Pearson correlation between readers' scores and GEDTS Writing Advisory Committee scores

[14] A single site failed to meet the certification criteria in 2006. This site did not participate in essay scoring during any subsequent years.

Essay Score Inter-rater Reliability

As described above, inter-rater reliability refers to the consistency with which two or more essay readers assign scores to the same essay, and scoring stability refers to the consistency with which the readers conform to the scoring guidelines established by the Writing Advisory Committee. Inter-rater reliability for the essay portion of the Language Arts, Writing Test was estimated by first calculating the polychoric correlation between the readers' two scores for each essay in the standardization and equating studies. Second, since the essay raw score on the Language Arts, Writing Test is the average of the two readers' scores, the correlation was adjusted using the Spearman-Brown prophecy formula with a factor of two (Equation 4.6).

Sample sizes for the Language Arts, Writing Test forms used in the 2001 standardization, and 2002, 2003, and 2005 equating studies ranged from 313 to 835. Inter-rater reliability coefficients ranged from .95 to .99 (see Table 4.11).

TABLE 4.11
SPEARMAN-BROWN CORRECTED CORRELATION BETWEEN ASSIGNED GED ESSAY SCORES AND SAMPLE SIZE

Form	*r*	N
Form IA	.95	391
Form IB	.98	322
Form IC	.97	363
Form ID	.98	357
Form IE	.97	347
Form IF	.98	337
Form IG	.97	369
Form IH	.98	313
Form II	.99	327
Form IJ	.98	835
Form IK	.97	806

Neither reader agreement reliability nor scoring stability describes the reliability of a particular writing sample provided by an individual examinee. An estimate of an examinee's writing skill is made from the scoring of two essays; that is, the extent to which the obtained essay reflects the examinee's "true" writing skill. This estimate was ascertained in the alternate-form reliability study (see earlier in this chapter) through the administration of two different but parallel essays to each examinee. The Spearman-Brown corrected correlation between scores on the first and second essay was .68.

RELIABILITY OF THE LANGUAGE ARTS, WRITING TEST COMPOSITE SCORE

The reliability of the essay score is influenced by variability due to essay topics, differences among readers' score assignments, and adherence to the GED essay scoring guide. However, the reliability of the Language Arts, Writing Test composite score is influenced by all of the factors that affect the reliability of both the essay and the multiple-choice portions of the test. The reliability of the multiple-choice and essay scores, the agreement of essay readers, and how well the readers maintain the scoring standards established by the GED Testing Service Writing Advisory Committee were all described in previous sections of this chapter. The strict judgmental and statistical procedures used to ensure that the operational essay topics are qualitatively and statistically similar to one another were described in Chapter 2. This section describes the reliability of the composite writing score obtained by combining the essay and multiple-choice scores. The reliability of the writing test score is of primary importance because the composite score is the crucial score that is used in evaluating whether an examinee has met the minimum score requirement on the Language Arts, Writing Test.

Weighting the Essay and Multiple-Choice Scores

Several studies were conducted during the 1988 Series GED Tests to determine the proper weighting of the multiple-choice and essay portions of the Language Arts, Writing Test. Swartz, Patience, and Whitney (1985) and Swartz and Whitney (1985) evaluated correlations between scores on experimental essay topics and GED multiple-choice writing skills test scores. Moderate to high correlations between the two test scores were found (ranging from .55 to .69) and Swartz, Patience, and Whitney concluded that the two formats "are measuring related but clearly different sets of skills" (p. 11). Patience and Swartz (1987) reported correlations between five operational GED topics and multiple-choice writing skills test scores. They used the results of these correlational analyses to select the weights for the multiple-choice and essay portions of the Language Arts, Writing Test composite score.

In determining the proper weights for the essay and multiple-choice portions of the test, Patience and Swartz reported that weights were desired that would "weight the essay as much as possible without diminishing the estimated test-retest reliability [of the composite Writing Skills Test score] below a level professionally acceptable" (p. 5). Using Nunnally's (1978) formula for the reliability of weighted linear combinations (described in Equation 4.6), and based on calculations of essay/multiple-choice correlations, essay score reliability, and multiple-choice score reliability, Patience and Swartz recommended that the weights of .35 and .65 be used for the essay and multiple-choice scores, respectively. They reported that, all other things being equal, these weights should ensure a minimum reliability of .86 for the composite score.

Because Patience and Swartz based their calculations on data obtained from the 1988 Series GED Tests, it was important to determine whether these relative weights were still relevant to the current test series. A subsequent analysis was performed using data obtained via the alternate-form reliability study (described above), which utilized 2002 series GED content area test forms. Using the alternate-forms reliability estimates (Table 4.7), the relative weights of .35 and .65 for the essay and multiple-choice portions of the Language Arts, Writing Test, respectively, are still applicable to the 2002 Series GED Tests.

Nunnally's Formula

Nunnally (1978) provided formulae for estimating the reliability of a linear combination and the reliability of a weighted linear combination. Because the Language Arts, Writing Test composite score is a weighted sum of the essay and multiple-choice scores, Nunnally's formula for the reliability of a weighted sum was used to estimate the reliability of the composite score. This formula uses the variance of the composite score, the reliabilities of the component scores, the correlations among component scores, and the weights of each component. Nunnally's formula for the reliability of a weighted sum is:

$$r_{yy} = 1 - \frac{\Sigma b_i^2 - \Sigma b_i^2 r_{ii}}{\sigma_y^2} \qquad \textbf{(4.7)}$$

where r_{yy} = the reliability of the weighted linear composite, b_i = the weight for component i, r_{ii} = the reliability of component i, and σ_y^2 = the variance of the weighted linear composite. The variance of the weighted linear composite is equal to the sum of the squared weights (b_i^2) and the correlation between each pair of components multiplied by the products of the weights for the two components (p.250).

The weights for the essay and multiple-choice scores are .35 and .65, respectively. To estimate the reliability of the composite score, it is necessary to estimate the reliabilities of the essay and multiple-choice scores, the correlation between these two scores, and the variance of the composite scores. Alternate-form reliability estimates were used for this purpose. As described in Table 4.7, the alternate-form reliability of the Language Arts, Writing Test was estimated as .74. The alternate-form reliability coefficient for the essay portion was estimated as .52. Correlations between performance on the multiple-choice and essay portions of the Language Arts, Writing Test form was .85. The estimated reliability of the weighted composite score of the Language Arts, Writing Test form using Equation 4.7 is .82.

DECISION CONSISTENCY

Standard 2.15 in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) states:

> When a test or combination of measures is used to make categorical decisions, estimates should be provided of the percentage of examinees that would be classified in the same way on two applications of the procedure, using the same form or alternate forms of the instrument. (p. 35)

GED Testing Service uses a minimum score requirement for each content area test simultaneously with an average score requirement for the entire battery. Therefore, it is necessary to adhere to Standard 2.15 and provide appropriate measures of classification consistency (the extent to which examinees would be classified consistently across replications or alternate forms of the tests) at both levels.

Decision Consistency Based on Content Area Tests

The decision consistency for each of the five content area tests was examined using data obtained via the standardization and equating studies (i.e., using high school senior data). Because the Language Arts, Writing Test includes multiple-choice and essay portions, the Livingston and Lewis (1995) procedure was used. This procedure was implemented using the BB-Class software program developed by Brennan (2004).

The percentages of test-takers meeting and not meeting the minimum score requirements, the probability of correct classification (decision consistency), and false positive and negative classifications are presented in Table 4.12 for graduating high school seniors, for each test form in the 2002 series (see Appendix K for decision consistency associated with GED examinees). In terms of decision consistency, values range from zero to one, with values closer to one preferred.

The decision consistency rates for high school seniors varied markedly across test form and content area test. Overall, the consistency rates range from a low of .67 (Form IH of the Language Arts, Reading Test) to a high value of .97 (Forms ID and IG of the Social Studies and Science Tests, respectively).

The false positive rates listed in Tables 4.12 and Appendix K reflect the probability of an examinee incorrectly passing the test form, given their true score is below the minimum score. Conversely, the false negative rates indicate the probability that an examinee will not meet the minimum score requirement for the test form, given their true score is above the cut-score. In both cases, values closer to zero are preferable. For most of the forms administered to graduating high school seniors, the results indicate that there are many more seniors who incorrectly met or exceeded the minimum score requirement (false positives) than those who incorrectly failed to meet the minimum requirement (false negatives).

TABLE 4.12
PROBABILITY OF CORRECT CLASSIFICATION, FALSE POSITIVE, AND FALSE NEGATIVE RATES FOR THE 2002 SERIES ENGLISH-LANGUAGE GED TESTS: U.S. GRADUATING HIGH SCHOOL SENIOR DATA

Test/Form	N	Percent Not Meeting Minimum Score	Percent Meeting Minimum Score	Probability of Correct Classification	False Positive	False Negative
Language Arts, Writing						
Form IA	391	19	81	.81	†	.19
Form IB	322	21	79	.79	†	.21
Form IC	363	27	73	.80	.20	†
Form ID	357	26	74	.77	.23	†
Form IE	347	21	79	.74	.26	†
Form IF	337	23	77	.71	.29	†
Form IG	369	30	70	.88	.12	†
Form IH	313	32	68	.88	.12	†
Form II	327	24	76	.76	†	.24
Form IJ	835	24	76	.76	.24	*
Form IK	806	24	76	.71	.29	*
Social Studies						
Form IA	462	18	82	.76	.24	*
Form IB	453	18	82	.69	.31	†
Form IC	456	17	83	.73	.27	*
Form ID	385	32	68	.97	.03	†
Form IE	331	30	70	.86	.14	*
Form IF	413	28	72	.92	.08	*
Form IG	579	31	69	.97	.03	†
Form IH	535	32	68	.95	.05	†
Form II	296	27	73	.87	.13	†
Form IJ	826	29	71	.88	.12	*
Form IK	893	28	72	.92	.08	†
Science						
Form IA	105	17	83	.74	.26	†
Form IB	187	18	82	.81	.19	*
Form IC	236	18	82	.82	†	.18
Form ID	169	28	72	.89	.11	†
Form IE	192	26	74	.92	.08	†
Form IF	349	26	74	.94	.06	†
Form IG	571	38	62	.97	.03	†
Form IH	531	40	60	.95	.05	†
Form II	288	39	61	.97	.03	†
Form IJ	818	27	73	.95	.05	†
Form IK	871	30	70	.87	.13	†
Language Arts, Reading						
Form IA	433	17	83	.83	†	.17
Form IB	422	18	82	.82	†	.18
Form IC	411	17	83	.83	†	.17
Form ID	588	28	72	.68	.32	†
Form IE	522	28	72	.71	.29	†
Form IF	553	26	74	.73	.27	*
Form IG	579	20	80	.94	.03	.03
Form IH	533	20	80	.67	.33	*
Form II	302	27	73	.75	.25	*
Form IJ	837	21	79	.79	†	.21
Form IK	906	23	77	.77	†	.23

Continued on next page

Table 4.12 continued

Test/Form	N	Percent Not Meeting Minimum Score	Percent Meeting Minimum Score	Probability of Correct Classification	False Positive	False Negative
Mathematics						
Form IA	258	18	82	.77	.23	†
Form IB	208	14	86	.86	†	.14
Form IC	278	19	81	.81	†	.19
Form ID	530	24	76	.87	.12	*
Form IE	514	27	73	.91	.09	†
Form IF	505	25	75	.80	.20	†
Form IG	683	31	69	.92	.08	†
Form IH	541	26	74	.74	.26	†
Form II	635	31	69	.82	.18	*
Form IJ	878	23	77	.84	.16	*
Form IK	848	24	76	.88	.11	*

* Value is less than 0.01.
† Value is less than 0.001.

Decision Consistency Based on Entire GED Test Battery

The decision consistency estimates described above are appropriate for making classification decisions based on each individual GED content area test. However, the GED credential is awarded only to those individuals who score at least 410 on *each* subject test as well as a 450 average battery score (although these criteria may be more stringent in certain jurisdictions—see Appendix B). Thus, the classification decision is necessarily a "complex" one, as defined by Chester (2003). In that sense, it is also appropriate to examine the decision consistency of the overall or complex decision in addition to examining each individual GED content area test.

Established methods for examining consistency of complex decisions are not as common as those for decisions made based on a single assessment instrument. Douglas (2007), however, proposed a method for examining complex decision consistency and accuracy and applied the procedure to the GED Tests.[15] Candidate data obtained via Form IG were used for the study. Results of the procedure indicated that approximately "88 percent of examinees would receive the same overall decision if they took two parallel forms of the GED test battery" and "91 percent are accurately classified in regard to mastery." However, the analysis did not account for the essay portion of the Language Arts, Writing Test or the fact that GED examinees can attempt to pass the test(s) on more than one occasion. Both exclusions most likely indicate that the results of the decision-consistency procedure may be somewhat overestimated.

[15] The details of the procedure are beyond the scope of this document, but are described precisely in the original document.

Chapter 5: Validity

Investigating test validity requires the accumulation of evidence suggesting that a specific test score interpretation, or use, is a valid one. Validity is not a property of the test itself, but rather a description of the appropriateness of the interpretations made from test scores. Because validity describes the utility and appropriateness of test score interpretations, it is of paramount importance that test developers provide evidence of validity. As stated in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999):

> Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations. (p. 9)

According to the *Standards*, an ideal validation is one that includes several types of evidence which, when combined, best reflect the value of a test for an intended purpose: "Validity is a unitary concept. It is the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose" (p. 11). The *Standards* suggests that test developers report several types of validity evidence, when appropriate. Specifically, evidence may be provided based on:

- Test content.
- Response processes.
- Internal structure.
- Relations to other variables.
- Consequences of testing.

The sources of validity evidence included in this manual are those based on test content, internal structure, and relations to other variables.

As clearly noted in the *Standards,* evidence of validity reported by test developers should reflect the purpose(s) of the test and the types of inferences that are to be made from the test scores. Therefore, in evaluating the validity of the GED test scores, the purpose of the tests must be considered first.

PURPOSE OF THE GED TESTS

As reported in Chapter 1, the purpose of the GED Tests is to measure major academic skills and knowledge in core content areas that are learned during four years of high school. The validation of GED test scores must be made with respect to this purpose. Thus, the sources of validity evidence reported in this chapter help evaluate the ability of GED test scores to determine whether a GED examinee has attained the knowledge and skills that are typically acquired through completion of a normal high school academic program of study. The sources of validity evidence presented in this chapter report (1) the extent to which the content of the GED Tests represents standards that support high school curricula, (2) the degree to which the test items conform to the construct being measured, (3) the relationship of the test scores to other external variables, and (4) the extent to which the processes of the essay scorers are consistent with the intended scoring rubric.

EVIDENCE BASED ON TEST CONTENT

Passing the GED test battery is a prerequisite to earning or receiving a credential. Thus, a crucial component of validity evidence is in demonstrating the ability of a test to represent adequately the content domain it purports to measure. Evidence based on test content is usually provided through rational defense of the test specifications and test development plan (Nunnally, 1978). Such evidence is demonstrated by the extent to which the questions on the GED Tests reflect the major content of a high school program of study.

Evidence of validity based on test content often rests on subjective analyses of test content made by subject-matter experts (Osterlind, 1989; Thorndike, 1982). Thus, to ensure adequate content representation of the GED Tests, nationally representative groups of experts were used to develop the current test specifications and to evaluate each operational test form.

Development of GED Tests Specifications

The development of the 2002 Series GED Tests specifications began with an extensive review of the current national and state curriculum standards. Specifically, the GED Testing Service test specialists examined national curriculum standards established by such organizations as the National Council of Teachers in English and the International Reading Association; the National Council of Teachers in Mathematics; the National Research Council; and the National Council for the Social Studies, the Center for Civic Education, the National Center for History in the Schools, several geographic organizations (the National Council for Geographic Education, the National Geographic Society, the Association of American Geographers, and the American Geographical Society), and the National Council on Economic Education; the American Association for the Advancement of Science, the National Science Teachers Association, the New Standards Project, and the National Assessment of Educational Progress. The test specialists also studied state trends in curriculum standards to determine which were most commonly cited. The results of this analysis are described extensively in the *Alignment of National and State Standards* (ACE, 1999).

To make recommendations toward the new test series' specifications, a Tests Specifications Committee was created. Because of the responsibility entrusted to members of this committee, it was essential that nominations for membership originated with a wide range of affected groups. Thus, the GEDTS staff began by inviting nominations from an advisory committee, national educational organizations (e.g., the National Council of Mathematics Teachers, Association for Supervision and Curriculum Development, and Council of Basic Education), state adult education directors, and GED Administrators. The nomination form was included with a letter suggesting that consideration be given to persons who were practicing high school teachers or supervisors, professors of education, state curriculum specialists, educators with adult education expertise, assessment specialists, leaders in professional organizations, and persons who participated in a combination of these fields. Specifically, it was recommended that members of the 2002 Series GED Tests Specifications Committee had the following qualities:

- Comprehensive subject matter expertise in high school curriculum and instruction.
- Extensive knowledge of national and state standards initiatives and of current research in their academic disciplines.
- Practical knowledge of what graduating high school seniors would actually know and be able to do in the year 2000 and beyond.
- Broad skills in serving as an active participant in a consensus project.

Ultimately, 29 committee members were selected from over 200 nominations. The committee members were assigned to one of four content area panels (English language arts, mathematics, science, and social studies) based on their background and expertise.

Each panel, comprising seven to eight members, convened for three days in January 1997 and was asked to prepare a report containing recommendations for the next series of GED Tests. Each panel relied on a series of research reports that contained the results of the analysis on national and state curriculum standards. Each panel's recommendations were to include definitions of the broad content areas to be included on each test, descriptions of the specific skills and tasks to be tested within the broad content areas, and relative weights to be assigned to the broad content areas and specific skills and tasks. Each panel authored a final report containing its recommendations for the 2002 Series GED Tests.

Content Review of GED Tests and Items

Additional evidence based on test content comes from the involvement of many secondary school educators in writing and reviewing the items and assembled tests. As described in Chapter 2, content specialists were recruited from across the United States to create items that met the test specifications. After these items were reviewed and edited internally by GEDTS staff, they were sent out to a carefully selected group of content reviewers who were familiar with the test specifications. These reviewers determined whether the items were congruent with the test specifications and whether they were classified correctly according to the specific subject areas they were presumed to measure. Items that were considered inappropriate by one or more content reviewers were rewritten or discarded. Items that were deemed appropriate in terms of their content representation were routed for item tryout studies.

When new forms of the GED Tests were developed, a preliminary "final form" was distributed to a content review committee for each content area test. These Final Form Review Committees verified the content and cognitive classifications of each test item and determined whether the preliminary form adequately represented the test specifications. Items deemed unacceptable by a Final Form Review Committee were eliminated and replaced by new items that had received approval from testing service staff. The Final Form Review Committees contributed the following: verification of content and cognitive classifications of test items, evaluation of preliminary form representation of test specifications, determination of reading level appropriateness, and review of item context appropriateness.

The content review procedures described above were part of the multi-stage item review process described more thoroughly in Chapter 2. Each stage of this review process helped ensure the content representativeness of the GED Tests.

GED Tests Content and Workplace Skills

As stated in the purpose of the GED Tests described above, the GED test scores are intended as a measure of the knowledge and skills associated with a traditional high school program of study. Therefore, the GED test scores are not intended to represent a level of workplace readiness. However, given that a number of educators believe that high schools should provide the knowledge and skills necessary to compete in the workplace, it may be reasonable that there would be substantial overlap between the knowledge and skills measured by the GED Tests and those deemed by high schools as necessary for employment.

To evaluate the congruence of the knowledge and skills measured by the 2002 Series GED Tests and the knowledge and skills considered essential for workplace readiness, a comparison was conducted between the types of basic and thinking skills measured by the GED Tests with those required for the workplace. The U.S. Department of Labor (1991) Secretary's Commission on Achieving Necessary Skills report, "What Work Requires of Schools: A SCANS Report for America 2000," provided a list of foundational skills for the workplace. "Although the commission completed its work in 1992, its findings and recommendations continue to be a valuable source of information for individuals and organizations involved in education and workforce development" (U.S. Dept of Labor, http://wdr.doleta.gov/SCANS, retrieved March 14, 2005). Table 5.1 describes both the basic and thinking skills described in the commission's report, along with indicators demonstrating where the GED Tests have matching requirements. Table 5.1 demonstrates that most of the basic skills and all of the thinking skills are represented on the 2002 Series GED Tests.

TABLE 5.1
SCANS SKILLS AND GED TESTS REQUIREMENTS

Foundational Skills for the Workplace	Required by GED Tests
BASIC SKILLS	
Reading. Locates, understands, and interprets written information in prose and documents including manuals, graphs, and schedules to perform tasks; learns from text by determining the main idea or essential message; identifies relevant details, facts, and specifications; infers or locates the meaning of unknown or technical vocabulary; and judges the accuracy, appropriateness, style, and plausibility of reports, proposals, or theories of other writers.	YES
Writing. Communicates thoughts, ideas, information, and messages in writing; records information completely and accurately; composes and creates documents such as letters, directions, manuals, reports, proposals, graphs, flow charts; uses language, style, organization, and format appropriate to the subject matter, purpose, and audience. Includes supporting documentation and attends to level of detail; checks, edits, and revises for correct information, appropriate emphasis, form, grammar, spelling, and punctuation.	YES
Arithmetic/Mathematics. Performs basic computations; uses basic numerical concepts such as whole numbers and percentages in practical situations; makes reasonable estimates of arithmetic results without a calculator, and uses tables, graphs, diagrams, and charts to obtain or convey quantitative information. Approaches practical problems by choosing appropriately from a variety of mathematical techniques; uses quantitative data to construct logical explanations for real world situations; expresses mathematical ideas and concepts orally and in writing; and understands the role of chance in the occurrence and prediction of events.	YES
Listening. Receives, attends to, interprets, and responds to verbal messages and other cues such as body language in ways that are appropriate to the purpose, for example, to comprehend, learn from, critically evaluate, appreciate, or support the speaker.	INDIRECTLY
Speaking. Organizes ideas and communicates oral messages appropriate to listeners and situations; participates in conversation, discussion, and group presentations; selects an appropriate medium for conveying a message; uses verbal language and other cues such as body language appropriate in style, tone, and level of complexity to the audience and the occasion; speaks clearly and communicates a message; understands and responds to listener feedback; and asks questions when needed.	NO
THINKING SKILLS	
Creative Thinking. Uses imagination freely, combines ideas or information in new ways, makes connections between seemingly unrelated ideas, and reshapes goals in ways that reveal new possibilities.	YES
Decision Making. Specifies goals and constraints, generates alternatives, considers risks, and evaluates and chooses best alternatives.	YES
Problem Solving. Recognizes that a problem exits (i.e., there is a discrepancy between what is and what should or could be), identifies possible reasons for the discrepancy, and devises and implements a plan of action to resolve it. Evaluates and monitors progress, and revises plan as indicated by findings.	YES
Seeing Things in the Mind's Eye. Organizes and processes symbols, pictures, graphs, objects or other information; for example, sees a building from a blueprint, a system's operation from schematics, the flow of work activities from narrative descriptions, or the taste of food from reading a recipe.	YES
Knowing How to Learn. Recognizes and can use learning techniques to apply and adapt new knowledge and skills in both familiar and changing situations. Involves being aware of learning tools such as personal learning styles (visual, aural, etc.), formal learning strategies (note taking or clustering items that share some characteristics), and informal learning strategies (awareness of unidentified false assumptions that may lead to faulty conclusions).	YES
Reasoning. Discovers a rule or principle underlying the relationship between two or more objects and applies it in solving a problem. For example, uses logic to draw conclusions from available information, extracts rules or principles from a set of objects or written text; applies rules and principles to a new situation, or determines which conclusions are correct when given a set of facts and a set of conclusions.	YES

EVIDENCE BASED ON INTERNAL STRUCTURE

Factor Analyses

GEDTS reports a single standard score for each test. This score reporting structure assumes that each test score represents a single construct and all items on that test measure this same construct. When a single construct underlies the responses to the items on a test, we describe that test as being unidimensional. An important component in making a validity argument for test scores is assessing the internal structure, or dimensionality, of the test. One can assess whether each test form is unidimensional using a nonlinear factor model with dichotomous data.

To assess the dimensionality of each test form, nonlinear exploratory factor analysis (NFA) was performed. Using high school seniors' data obtained during an equating study, an NFA was performed only on forms IJ and IK for each content area test. Because the stability of the final estimates is a function of sample size, conducting an NFA with the remaining forms may provide misleading results due to the low sample sizes.

The TESTFACT software (Wood, Wilson, Gibbons, Schilling, Muraki, & Bock, 2003) program was used for these analyses. TESTFACT is based on multidimensional item response theory models, which are equivalent to nonlinear factor analysis models (Glöckner-Rist & Hoijtink, 2003; McDonald, 1999). TESTFACT calculates a matrix of tetrachoric correlations that is subsequently used as the input matrix for the NFA. Tetrachoric correlations assume that the dichotomous measure is a crude approximation of a normally distributed, continuous variable. A benefit to using TESTFACT is that it can adjust the tetrachoric correlations when guessing is present. Incorporating a guessing model into the estimation procedure has been examined by Stone and Yeh (2006).

Because the GED Tests are administered to high school seniors in a low-stakes setting, it was anticipated that a certain amount of guessing was present in the data. Each test item contains five response options. However, instead of using a guessing estimate of 1 divided by the number of response options (i.e., 0.20), a more conservative estimate of 0.15 was used. In each analysis, a Promax rotation was requested as the factors within each test were likely correlated.

Table 5.2 shows the results of the nonlinear factor analyses for forms IJ and IK associated with each test given to high school seniors. As can be seen, the results indicate that a single dominant factor underlies each of the test forms. Across all test forms, the proportion of common (i.e., shared) variance accounted for ranges from 0.41 to 0.60. Although not shown in the table, initial values for the first extracted eigenvalues (which indicate the amount of variance explained by the factor) ranged from 22.0 to 30.8; second initial eigenvalues ranged from 1.7 to 2.7. Moreover, each of the items loaded heavily and primarily onto the first factor. Factor loadings on subsequent factors were minimal. Finally, correlations among the extracted factors were all positive.

TABLE 5.2
NUMBER OF SALIENT FACTORS, PROPORTION OF VARIANCE ACCOUNTED FOR BY INITIAL FACTOR, AND SAMPLE SIZE FOR U.S. GRADUATING HIGH SCHOOL SENIOR SAMPLES

TEST/FORM	Number of Salient Factors	Proportion of Variance	N
Language Arts, Writing			
Form IJ	1	.48	835
Form IK	1	.53	806
Social Studies			
Form IJ	1	.48	826
Form IK	1	.53	893
Science (50 items)			
Form IJ	1	.60	818
Form IK	1	.57	871
Language Arts, Reading			
Form IJ	1	.45	837
Form IK	1	.41	906
Mathematics			
Form IJ	1	.51	878
Form IK	1	.49	848

Similar procedures were used with GED examinee data. Whereas the sample sizes associated with the majority of the test forms given to graduating high school seniors were too small to perform an NFA, the sample sizes associated with the examinees were considerably larger. Therefore, the dimensionality of each of the 11 test forms was assessed using the examinee data. In order to make the analyses more manageable, simple random samples of 5,000 examinees were drawn for each test form. Table 5.3 shows the results of the NFA for the examinee data.

Across all test forms, the proportion of common (i.e., shared) variance accounted for ranged from 0.22 to 0.68. Although not shown in the table, initial values for the first extracted eigenvalues ranged from 11.9 to 23.6 and second initial eigenvalues ranged from 1.3 to 2.6. As with the graduating high school senior data, each of the items loaded heavily and primarily onto the first factor and factor loadings on subsequent factors were minimal. Finally, correlations among the extracted factors were all positive.

TABLE 5.3
NUMBER OF SALIENT FACTORS AND PROPORTION OF VARIANCE ACCOUNTED FOR BY INITIAL FACTOR: U.S. GED EXAMINEES

TEST/FORM	Number of Salient Factors	Proportion of Variance
Language Arts, Writing		
Form IA	1	.31
Form IB	1	.34
Form IC	1	.33
Form ID	1	.32
Form IE	1	.33
Form IF	1	.31
Form IG	1	.30
Form IH	1	.22
Form II	1	.25
Social Studies		
Form IA	1	.35
Form IB	1	.31
Form IC	1	.37
Form ID	1	.38
Form IE	1	.30
Form IF	1	.33
Form IG	1	.40
Form IH	1	.38
Form II	1	.36
Science		
Form IA	1	.35
Form IB	1	.25
Form IC	1	.38
Form ID	1	.30
Form IE	1	.36
Form IF	1	.45
Form IG	1	.36
Form IH	1	.27
Form II	1	.35
Language Arts, Reading		
Form IA	1	.38
Form IB	1	.40
Form IC	1	.31
Form ID	1	.26
Form IE	1	.36
Form IF	1	.35
Form IG	1	.29
Form IH	1	.68
Form II	1	.30
Mathematics		
Form IA	1	.40
Form IB	1	.44
Form IC	1	.32
Form ID	1	.39
Form IE	1	.39
Form IF	1	.38
Form IG	1	.37
Form IH	1	.41
Form II	1	.42

Differential Item Functioning

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) indicates that test developers must assess the quality of test items in terms of fairness. In other words, it is incumbent upon the test developer to ensure, within reason, that the likelihood of producing a correct answer does not depend on the characteristics of an examinee that are external to the measure of interest. Items that result in differential likelihoods of success for different subgroups are described as having *differential item functioning* (DIF). However, final judgment as to whether an item is biased toward one group over another is relegated to a panel of expert reviewers.

GEDTS conducted DIF analyses after GED examinee data were collected and scored. Ideally, DIF analyses would occur during item tryout studies so that items exhibiting DIF (and deemed biased by a review panel) could be revised or removed from operational test forms. However, the sample sizes associated with the item tryout studies (as well as the norming and equating studies) have not been large enough to permit such analyses. Nevertheless, it is still important to assess DIF as it is key evidence for test score validity.

The process of assessing DIF is a statistical one and several methods are available. The Mantel-Haenszel (M-H) statistic (Holland & Thayer, 1988) is a commonly used measure that has been shown to be a sufficient method for detecting uniform DIF (Hambleton & Rogers, 1989; Narayanan & Swaminathan, 1994).

The M-H statistic is essentially an analysis of contingency table data. The procedure matches examinees from two differing groups on a criterion—typically the total test score—and compares the likelihood of success for each group within each level of the criterion. The two groups are usually classified as the focal group (e.g., females), which is of primary interest, and a reference group (e.g., males), to which the focal group is compared.

The null hypothesis for the M-H states that the odds of correctly answering the item at a given ability level is the same for both the focal and reference groups. The corresponding alternative hypothesis (*Ha*) is

$$Ha : \frac{P_{rj}}{Q_{rj}} = \alpha \frac{P_{fj}}{Q_{fj}},$$

where P_{rj} and P_{fj} are the probabilities of a correct answer for the reference and focal groups, respectively, at score level j ($j = 0,\ldots, k$) of the criterion, Q_{rj} and Q_{fj} are the probabilities of an incorrect response, and α is the common odds ratio estimated as

$$\hat{\alpha} = \frac{\sum_{j=0}^{k} R_{rj} W_{fj} / T_j}{\sum_{j=0}^{k} R_{fj} W_{rj} / T_j}. \qquad \textbf{(5.1)}$$

In Equation 5.1, R_{rj} and W_{rj} represent the number of examinees in the reference group at score level j who answered the item correctly or incorrectly, respectively. Further, R_{fj} and W_{fj} represent the number of examinees in the focal group at score level j who correctly or incorrectly responded to the item, respectively. Finally, T_j represents the total number of examinees with score j.

The common odds ratio given in Equation 5.1 represents the ratio of the reference group's odds of a correct response to the focal group's odds of a correct response, after conditioning on the total score. Therefore, when an item favors the reference group, the common odds ratio takes on values between one and positive infinity. Conversely, values between zero and one indicate the item favors the focal group.

A chi-square test, with one degree of freedom, is used to assess the null hypothesis. The chi-square statistic is calculated as

$$MH - \chi^2 = \frac{\left[\left|\sum_j R_{rj} - \sum_j E\left(R_{rj}\right)\right| - .5\right]^2}{\sum_j \mathrm{var}\left(Rr_j\right)}, \qquad \textbf{(5.2)}$$

where

$$E\left(R_{rj}\right) = n_{rj} m_{cj} / T_j$$

and

$$\mathrm{var}\left(R_{rj}\right) = \left(n_{rj} m_{cj} n_{fj} m_{Ij}\right) / \left[\left(T_j^2\right)\left(T_j - 1\right)\right] \qquad \textbf{(5.3)}$$

(Clauser & Mazor, 1998; Dorans & Holland, 1993). In Equation 5.3, n_{rj} is the total number of examinees in the reference group at score level *j*, m_{cj} is the total number of correct responses given at score level *j*, n_{fj} is the total number of examinees in the focal group at score level *j*, and m_{Ij} is the total number of incorrect responses given at score level *j*.

The estimated common odds ratio can be rescaled to make results more interpretable. The most common transformation is to take the log of $\hat{\alpha}$ and multiply it by the value -2.35. This transformation puts $\hat{\alpha}$ onto what is commonly referred to as the Educational Testing Service's (ETS) delta scale (Holland & Thayer, 1988) and is symbolized as Δ_{MH}. After rescaling to Δ_{MH}, items that favor the focal group will have values ranging from zero to positive infinity while items favoring the reference group will have values from zero to negative infinity.

It is well known that the chi-square test is sensitive to sample size. Identifying items as having DIF on the basis of the chi-square test alone would likely result in flagging items that were statistically significant, yet not practically significant. The Δ_{MH}, being a measure of effect size, can be used in conjunction with the chi-square test to flag items for DIF. ETS has developed a classification system to help judge whether an item should be flagged for review (Zieky, 1993). The classification system is based on three tiers or levels, A, B, or C and is as follows:

Level A: Δ_{MH} is not significantly different from zero or $|\Delta_{MH}| < 1.0$;

Level B: Δ_{MH} is significantly different from zero and $1.0 \le |\Delta_{MH}|$ and either $|\Delta_{MH}| < 1.5$ or $|\Delta_{MH}|$ is not significantly greater than 1.0;

Level C: Δ_{MH} is significantly greater than 1.0 and $1.5 \le |\Delta_{MH}|$.

Items classified at level C are of most concern and should be flagged for expert review (Clauser & Mazor, 1998).

The M-H procedure with a two-stage purification step (Clauser, Mazor, & Hambleton, 1993; Holland & Thayer, 1988) was performed using examinee data collected on test forms IA to IK. In total, seven variables were examined for DIF, including gender, primary language, and five race/ethnicity comparisons. These seven variables, along with the focal and reference categories, are presented in Table 5.4.

TABLE 5.4
VARIABLES EXAMINED FOR DIFFERENTIAL ITEM FUNCTIONING

Variable	Focal Group	Reference Group
Gender	Female	Male
Primary language	Language Other than English	English
Race/ethnicity (1)	Hispanic	White
Race/ethnicity (2)	American Indian or Alaska Native	White
Race/ethnicity (3)	Asian	White
Race/ethnicity (4)	Black	White
Race/ethnicity (5)	Native Hawaiian or Pacific Islander	White

Except for those cases with missing data on the variable of interest, all examinees who granted GEDTS research access to their data were included in the DIF analysis.

The first stage of the M-H procedure involved flagging items for DIF using the total raw score as the matching criterion. The common odds ratio was subsequently converted to the ETS delta scale. Because the entire dataset was used to examine each variable for DIF, only items with absolute delta values greater than 1.5 and significantly greater than 1 were flagged for DIF (i.e., level C in the ETS classification system).

Items flagged for level C DIF in the first stage were removed from the matching variable in the second stage. For example, if two items were flagged for DIF in stage one, these two items were subtracted from the total raw score to obtain a "purified" matching criterion.[16] The second stage proceeded by performing the M-H procedure again while using the purified criterion. Items were again flagged for DIF if the absolute delta values were greater than 1.5 and significantly greater than 1.

In total, there were 18,480 opportunities for potential DIF (11 forms, 240 items per form, seven examined variables). Of these, 435, or 2.4 percent, were ultimately flagged for DIF after the two-stage procedure. The numbers of DIF occurrences per form and by analysis variable are provided in Table 5.5. The Language Arts, Writing Test (all forms) accounted for approximately 41 percent of all DIF occurrences, followed by the Science Test (18 percent), Social Studies Test (17 percent) and Mathematics Test (14 percent). The Language Arts, Reading Test accounted for the smallest percentage of DIF occurrences, with approximately 11 percent.

With respect to the variables of interest, the white-Asian comparison accounted for over 42 percent of the DIF occurrences, more than any other analysis variable. Moreover, the white-Asian comparison consistently resulted in more DIF occurrences across all tests. In addition, primary language accounted for more than a quarter of the DIF occurrences. Only a single item was flagged for DIF when comparing the white and American Indian/Alaska Native subgroups.

[16] A number of previous studies have demonstrated that the item to be studied should be included in the matching criterion, despite whether it was flagged for DIF in the initial stage of detection (e.g., Donoghue, Holland, & Thayer, 1993). Failure to do so may result in Type I error rates, or incorrectly flagging items as exhibiting DIF (Lewis, 1993). However, in this case, such a procedure would require an excessively large number of iterations and it was therefore decided to exclude all DIF items in the second stage of detection. Empirical results have demonstrated the two-stage procedure used in this study to be adequate for DIF detection (Navas-Ara & Gómez-Benito, 2002).

TABLE 5.5
NUMBER OF ITEMS FLAGGED FOR DIF, BY TEST/FORM AND ANALYSIS VARIABLE

	DIF VARIABLE							
			White					
TEST/FORM	Gender	Primary Language	Hispanic	AIAN[a]	Asian	Black	NHPI[b]	Total
Language Arts, Writing								
Form IA	---	2	---	---	5	1	1	9
Form IB	---	5	2	---	6	3	3	19
Form IC	---	4	---	---	6	---	---	10
Form ID	---	5	2	---	9	2	3	21
Form IE	---	4	---	---	5	3	4	16
Form IF	---	4	1	---	8	1	1	15
Form IG	---	6	1	---	10	4	5	26
Form IH	---	5	2	---	10	1	2	20
Form II	---	7	---	---	8	1	3	19
Form IJ	---	3	---	---	---	7	---	10
Form IK	---	7	---	---	---	5	1	13
Total	0	52	8	0	67	28	23	178
Social Studies								
Form IA	---	2	---	---	2	1	1	6
Form IB	2	---	---	---	1	1	1	5
Form IC	1	2	1	---	2	1	---	7
Form ID	---	3	1	---	4	---	---	8
Form IE	---	1	---	---	2	---	---	3
Form IF	---	2	---	---	2	1	---	5
Form IG	---	2	---	---	3	1	---	6
Form IH	---	2	---	---	5	---	1	8
Form II	---	1	---	---	1	2	---	4
Form IJ	---	1	---	---	5	6	---	12
Form IK	---	2	---	---	5	1	---	8
Total	3	18	2	0	32	14	3	72
Science								
Form IA	3	1	1	---	2	2	---	9
Form IB	1	---	---	---	---	---	---	1
Form IC	---	1	---	---	2	---	---	3
Form ID	2	1	---	---	4	2	---	9
Form IE	---	1	---	---	---	1	1	3
Form IF	1	2	1	---	3	1	---	8
Form IG	3	1	---	---	2	1	---	7
Form IH	2	1	---	---	1	---	---	4
Form II	3	2	---	---	3	---	---	8
Form IJ	---	---	---	---	8	12	---	20
Form IK	---	---	---	---	6	1	---	7
Total	15	10	2	0	31	20	1	79

Continued on next page

Table 5.5 continued

	DIF VARIABLE							
			White					
TEST/FORM	Gender	Primary Language	Hispanic	AIAN[a]	Asian	Black	NHPI[b]	Total
Language Arts, Reading								
Form IA	---	1	---	---	1	1	---	3
Form IB	---	1	---	---	1	---	---	2
Form IC	---	---	---	---	---	---	---	0
Form ID	---	1	---	---	1	---	---	2
Form IE	---	---	---	---	---	---	---	0
Form IF	---	2	---	---	4	---	---	6
Form IG	---	1	---	---	1	1	---	3
Form IH	---	1	---	---	4	1	---	6
Form II	---	4	---	---	8	---	2	14
Form IJ	---	---	---	1	1	3	---	5
Form IK	---	1	---	---	---	4	---	5
Total	0	12	0	1	21	10	2	46
Mathematics								
Form IA	---	2	---	---	6	2	---	10
Form IB	---	2	---	---	4	---	---	6
Form IC	---	1	---	---	3	---	---	4
Form ID	---	---	---	---	2	1	---	3
Form IE	---	1	---	---	4	1	---	6
Form IF	1	5	---	---	3	---	---	9
Form IG	---	---	---	---	---	---	---	0
Form IH	---	---	---	---	---	---	---	0
Form II	---	---	---	---	---	---	---	0
Form IJ	---	---	3	---	6	5	---	14
Form IK	---	2	---	---	5	1	---	8
Total	1	13	3	0	33	10	0	60
Grand total	19	114	15	1	184	82	29	435

Note: Items were flagged for DIF if they favored either the focal or the reference group.
[a] American Indian, Alaska Native.
[b] Native Hawaiian or Pacific Islander.

As mentioned, items flagged for DIF are not necessarily biased. The final determination of whether an item was biased was left to a panel of expert reviewers. The items flagged for DIF in Table 5.5 above were examined by this panel. Thirteen items were deemed biased by a majority of reviewers. Additional details and results of the bias review were not available at the time this manuscript was published. Any items that were designated as biased will not count toward an examinee's total score for the remainder of the 2002 series.

EVIDENCE BASED ON RELATIONS WITH OTHER VARIABLES

Evidence based on relations with other variables refers to how well a test relates to other tests or criteria that are designed to measure the same or similar attributes. Therefore, several studies were performed that focused on the relationship between GED test scores and other related measures of academic proficiency. These studies have focused on the performance of graduating high school seniors and GED examinees on the GED Tests, and the performance of GED examinees on other measures of academic proficiency.

Correlations Among Content Area GED Tests

Because of time commitment, participation in GED standardization and equating studies do not require the students to take every GED content area test. Therefore, a complete correlation matrix of standard scores is not available. Table 5.6 presents the correlations among GED Tests based on GED examinees who completed the GED test battery in 2007. The magnitude of the correlations suggests that the five content area GED Tests are related, but distinct.

TABLE 5.6
STANDARD SCORE CORRELATIONS AMONG GED TESTS IN 2007 (U.S. GED EXAMINEE DATA)

	Social Studies	Science	Language Arts, Reading	Mathematics
Language Arts, Writing	.47	.46	.46	.45
Social Studies		.78	.71	.61
Science			.69	.67
Language Arts, Reading				.54

Note: Sample size for correlations is 521,002.

The Relationship Between GED Test Scores and High School Grades

Because the GED Tests are designed to measure academic knowledge and skills that are taught in a traditional high school program of study, it is important that they demonstrate a positive relationship with other measures of high school–level academic performance. To investigate this relationship, the self-reported grades of graduating high school seniors participating in the standardization and norming study and equating studies were collected and compared with the performance of these same seniors on the GED Tests. Students were asked to list the overall grades they received since ninth grade through the current term for five content areas: English literature, English composition, social studies, science, and mathematics.

The correlation between self-reported grades and GED test scores are reported in Table 5.7. The correlations reported in Table 5.7 vary across both year and content area. The variation in the correlations could be due to the fact that higher correlations may be expected for those tests that represent a greater proportion of content taught in the high school curriculum. Due to the fact that the letter grades were self-reported, the correlations may be somewhat lower than might be found if their official letter grades from the school had been used.

TABLE 5.7
CORRELATIONS OF U.S. GRADUATING HIGH SCHOOL SENIORS' ENGLISH-LANGUAGE GED TEST STANDARD SCORE WITH SELF-REPORTED LETTER GRADES IN THE SAME CONTENT AREA: 2001 STANDARDIZATION AND 2002, 2003, AND 2005 EQUATING STUDIES

	2001		2002		2003		2005	
	N	*r*	N	*r*	N	*r*	N	*r*
Language Arts, Writing	1,076	.49[a]	1,133	.45[a]	1,530	.44[a]	2,423	.47[c]
Language Arts, Writing	1,076	.45[b]	1,133	.46[b]	1,440	.44[b]	--	--
Social Studies	1,371	.37	1,488	.37	2,317	.38	2,452	.42
Science	524	.36	997	.38	2,285	.36	2,398	.36
Language Arts, Reading	1,264	.39[a]	2,020	.42[a]	2,310	.34[a]	2,484	.41[c]
Language Arts, Reading	1,264	.33[b]	2,020	.39[b]	2,106	.34[b]	--	--
Mathematics	--	--	1,734	.47	2,648	.46	2,541	.48

Note: All correlations were significant at $p < .001$. Letter grades are reported as Mostly A, Mostly B, Mostly C, Mostly D, and Mostly Below D. To compute the correlations, letter grades were recoded as Mostly A=4, Mostly B=3, Mostly C=2, Mostly D=1, Mostly Below D=0. Data for Mathematics Test were not available in 2001.

[a] Correlation with self-reported grades in English literature.

[b] Correlation with self-reported grades in English composition.

[c] Correlation with self-reported grades in English.

The grades reported by the graduating high school seniors were also compared with their performance at selected values along the GED standard score scale. The purpose of this analysis was to identify approximate GPA or letter-grade levels that correspond to levels of performance on the GED Tests. Table 5.8 presents, for each GED content area test, the percentages of soon-to-be graduating seniors meeting selected GED score standards for each letter grade (tables for additional equating studies are located in Appendix L). For example, the first row of Table 5.8 indicates that 99 percent of the seniors whose reported grades were "Mostly A" scored at or above a GED standard score of 350 on the Language Arts, Writing Test. The second row of the table shows that 39 percent of the "Mostly B" seniors achieved a score of at least 500.

TABLE 5.8
PERCENTAGE OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2001 ENGLISH-LANGUAGE STANDARDIZATION AND NORMING STUDY AT SELF-REPORTED GRADE LEVELS ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

SELF-REPORTED GRADES	N	GED Standard Score ≥ 350	410	450	500
		Language Arts, Writing Test			
English Literature					
Mostly A	309	99	94	91	74
Mostly B	493	93	79	62	39
Mostly C	238	83	59	39	21
Mostly D	33	94	55	36	15
Mostly Below D	3	†	†	†	†
		Language Arts, Writing Test			
English Composition					
Mostly A	293	98	93	87	72
Mostly B	508	93	78	62	40
Mostly C	238	84	63	44	24
Mostly D	36	89	56	36	11
Mostly Below D	1	†	†	†	†
		Social Studies Test			
Social Studies					
Mostly A	424	97	92	88	71
Mostly B	609	94	83	70	48
Mostly C	293	88	71	52	32
Mostly D	40	78	53	23	8
Mostly Below D	5	†	†	†	†
		Science Test			
Science					
Mostly A	132	97	90	88	79
Mostly B	225	93	85	68	46
Mostly C	142	89	75	62	41
Mostly D	22	82	68	55	27
Mostly Below D	3	†	†	†	†
		Language Arts, Reading Test			
English Literature					
Mostly A	336	96	92	87	77
Mostly B	543	94	85	74	50
Mostly C	345	91	74	54	34
Mostly D	36	83	58	39	31
Mostly Below D	4	†	†	†	†
		Language Arts, Reading Test			
English Composition					
Mostly A	328	95	91	85	74
Mostly B	607	94	85	73	50
Mostly C	301	90	72	53	34
Mostly D	27	85	56	41	33
Mostly Below D	1	†	†	†	†

† Indicates that the statistic was not calculated because of small sample size.
Note: Data for the Mathematics Test were not available.

Overall, Table 5.8, along with the additional tables in Appendix L, illustrates that the higher the high school grade, the higher the GED score, and therefore, the greater the likelihood of passing the particular GED content area test.

The results presented in Table 5.8 (and those in Appendix L) indicate that the passing standards established on the GED Tests do discriminate between higher and lower achieving high school students. Therefore, the results support both the validity of the GED test scores, and the validity of the GED standard setting procedure.

The Relationship Between GED Test Scores and Prior Instruction

To evaluate whether the content areas tested by the GED Tests are related to the content areas presumed to be taught in high schools across the nation, the graduating high school seniors in the 2001 standardization and norming study and subsequent equating studies were asked to report the number of years of instruction they received in various content areas. It was hypothesized that, if the GED Tests are accurate measures of content taught in a regular program of high school study, then a positive relationship would be observed between scores on each GED content area test and the amount of instruction received by students in the content area related to each test.

The graduating high school seniors participating in the standardization and norming study and equating studies were asked to indicate the number of years of English literature, English composition, social studies, science, and mathematics courses they had taken from ninth grade to the current term. The seniors were asked to indicate whether they had taken one year or less, two, three, or four years or more of coursework in each content area. In addition, they were asked to specify the types of courses they had taken in each content area. For example, for social studies, the seniors were asked to indicate whether they had taken behavioral sciences, civics, economics, geography, political science, national history, and/or world history.

Table 5.9 contains the percentage of graduating high school seniors (2001 standardization and norming study) at self-reported total years of study by various minimum standard scores. For example, 83 percent of graduating seniors who reported taking only one year or less of English literature scored at or above a standard score of 350 on the Language Arts, Writing Test. Additionally, 93 percent of those graduating seniors who took four years or more of English literature scored at least 350 on the Language Arts, Writing Test. As expected, the percentages decrease across each row as the standard score increases. Additionally, as the number of self-reported total years of study increases, the percentage of seniors generally increases within any given standard score category. For example, 81 percent of graduating seniors with one year or less of English composition scored at least 350 on the Language Arts, Writing Test. However, a larger percentage (94 percent) of those seniors with at least four years of English composition scored at least 350 on the Language Arts, Writing Test. Results from subsequent equating studies can be found in Appendix M.

TABLE 5.9
PERCENTAGE OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2001 ENGLISH-LANGUAGE STANDARDIZATION AND NORMING STUDY AT SELF-REPORTED TOTAL YEARS OF STUDY ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
SELF-REPORTED TOTAL YEARS OF STUDY	N	350	410	450	500
		Language Arts, Writing Test			
English Literature					
1 year or less	64	83	63	39	27
2 years	116	94	78	63	43
3 years	122	90	71	55	39
4 years or more	751	93	80	68	47
		Language Arts, Writing Test			
English Composition					
1 year or less	151	89	77	64	43
2 years	134	93	78	61	46
3 years	89	87	66	44	34
4 years or more	647	94	81	69	47
		Social Studies Test			
Social Studies					
1 year or less	36	81	61	44	28
2 years	159	94	77	64	36
3 years	550	94	83	70	51
4 years or more	605	94	85	75	57
		Science Test			
Science					
1 year or less	7	†	†	†	†
2 years	77	90	74	62	42
3 years	248	92	81	67	46
4 years or more	185	96	90	82	68
		Language Arts, Reading Test			
English Literature					
1 year or less	81	93	79	70	44
2 years	161	94	83	66	52
3 years	140	90	81	66	48
4 years or more	858	94	84	73	54
		Language Arts, Reading Test			
English Composition					
1 year or less	183	93	83	73	51
2 years	175	92	81	65	46
3 years	113	90	77	65	50
4 years or more	720	95	86	74	56

† Indicates that the statistic was not calculated because of small sample size.
Note: Data for Mathematics Test not available for 2001.

Table 5.10 reports average GED standard scores for the Language Arts, Writing; Social Studies; Science; and Language Arts, Reading Tests. These results were derived from the 2001 standardization and norming study (tables for the subsequent equating studies are provided in Appendix N). The average standard scores are broken down according to the four levels of amount of prior instruction (from one year or less to four years or more). For example, those seniors with one year or less of English literature instruction obtained an average GED standard score of 491 on the Language Arts, Reading Test, while those with four years or more achieved an average score of 522.

TABLE 5.10
AVERAGE GED STANDARD SCORES OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2001 ENGLISH-LANGUAGE STANDARDIZATION AND NORMING STUDY, BY YEARS OF INSTRUCTION IN CONTENT AREA

YEARS INSTRUCTION IN CONTENT AREA	Language Arts, Writing	Social Studies	Science	Language Arts, Reading
1 year or less	490	†	†	491
	(151)	(36)	(7)	(81)
2 years	484	473	473	522
	(134)	(159)	(77)	(161)
3 years	459	497	487	513
	(89)	(550)	(248)	(140)
4 years or more	498	512	541	522
	(647)	(605)	(185)	(858)

† Indicates that the statistic was not calculated because of small sample size.
Note: Numbers in parentheses refer to the number of seniors. Averages for the Language Arts, Writing Test are based on the numbers of years of instruction in English composition; averages for the Language Arts, Reading Test are based on the number of years of instruction in English literature. Data were not available for the Mathematics Test.

Tables 5.11 through 5.14 provide the average standard scores by specific courses of study for the 2001 standardization and norming study data (data were not available for the 2001 Mathematics Test). Tables for subsequent equating studies can be found in Appendix O. Here, the expectation is that those graduating high school seniors who had taken related courses should have scored higher than those who had not taken these courses. Although the majority of the average standard scores follow this pattern, several do not. The most dramatic difference, for example, occurs between those who had and had not taken General Mathematics (see tables in Appendix O). Seniors who had not taken this course scored approximately 50 standard score points higher than those who had taken this course.

Tables 5.11 through 5.14 also provide the percentages of high school seniors at various GED standard score levels by specific instructional courses.

TABLE 5.11
PERCENT OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2001 ENGLISH-LANGUAGE STANDARDIZATION AND NORMING STUDY SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, WRITING TEST, BY INSTRUCTION IN GRAMMAR AND LANGUAGE COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Grammar/Composition	Taken	501	859	94	82	69	49
	Not Taken	442	217	84	60	44	25
Spanish	Taken	496	645	95	82	68	46
	Not Taken	478	431	88	71	58	41
French	Taken	496	207	94	79	65	48
	Not Taken	488	869	92	77	64	43
German	Taken	504	77	91	79	70	52
	Not Taken	488	999	92	78	63	43
Latin	Taken	536	43	98	93	74	58
	Not Taken	487	1,033	92	77	64	43

TABLE 5.12
PERCENT OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2001 ENGLISH-LANGUAGE STANDARDIZATION AND NORMING STUDY SCORING AT OR ABOVE STANDARD SCORES ON SOCIAL STUDIES TEST, BY INSTRUCTION IN SOCIAL STUDIES COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Behavioral Science	Taken	526	372	97	89	80	63
	Not Taken	488	999	92	80	67	46
Civics	Taken	502	492	93	83	72	53
	Not Taken	496	879	93	82	70	50
Economics	Taken	504	684	94	85	73	53
	Not Taken	492	687	92	80	67	49
Geography	Taken	500	701	95	86	72	52
	Not Taken	496	670	91	79	68	50
Political Science	Taken	526	323	96	88	79	64
	Not Taken	489	1,048	92	81	68	47
History	Taken	508	1,190	94	85	74	55
	Not Taken	435	181	86	65	44	22
World History	Taken	501	1,151	94	83	71	52
	Not Taken	483	220	90	77	65	45

TABLE 5.13
PERCENT OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2001 ENGLISH-LANGUAGE STANDARDIZATION AND NORMING STUDY SCORING AT OR ABOVE STANDARD SCORES ON SCIENCE TEST, BY INSTRUCTION IN SCIENCE COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Biology	Taken	507	486	93	84	73	54
	Not Taken	430	42	81	62	38	19
Chemistry	Taken	518	338	94	86	77	59
	Not Taken	470	190	89	75	59	37
Earth Science	Taken	499	238	93	82	70	50
	Not Taken	502	290	92	82	71	53
General Science	Taken	500	141	92	88	74	48
	Not Taken	501	387	93	80	69	53
Genetics	Taken	542	14	93	93	86	79
	Not Taken	500	514	92	82	70	51
Physical Science	Taken	500	241	93	83	70	51
	Not Taken	501	287	92	82	70	52
Physics	Taken	546	142	98	90	82	68
	Not Taken	484	386	91	79	66	45
Zoology/Botany	Taken	518	27	96	96	81	59
	Not Taken	500	501	92	81	70	51

TABLE 5.14
PERCENT OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2001 ENGLISH-LANGUAGE STANDARDIZATION AND NORMING STUDY SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, READING TEST, BY INSTRUCTION IN ENGLISH AND LANGUAGE COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Literature	Taken	523	1,123	94	84	73	54
	Not Taken	480	143	89	72	55	35
European Literature	Taken	554	307	94	90	80	66
	Not Taken	506	959	93	81	67	48
World Literature	Taken	531	406	94	84	73	57
	Not Taken	511	860	93	82	69	49
Spanish	Taken	518	788	94	84	71	52
	Not Taken	518	478	92	81	69	52
French	Taken	553	232	97	90	80	66
	Not Taken	510	1,034	92	81	69	49
German	Taken	572	69	96	91	86	72
	Not Taken	515	1,197	93	82	70	51
Latin	Taken	560	48	98	88	83	75
	Not Taken	516	1,218	93	83	70	51

Predictive Utility of the GED Credential

The GED test battery has not been validated for, nor is it intended to predict success in the workplace or secondary education; rather, the GED Tests serve as a measure of major academic skills and knowledge in core content areas that are learned during four years of high school. However, the extensive acceptance of the GED credential in place of a standard high school diploma in the workplace and institutions of higher education point to the GED credential as an accepted and trusted measure of high school achievement. As mentioned in Chapter 1, approximately 96 percent of U.S. employers consider those who have earned GED credentials the same as traditional high school graduates with regard to hiring, salary, and opportunities for advancement (Society for Human Resource Management, 2002) and nearly all U.S. colleges and universities accept GED test score reports as being equivalent to high school transcripts (Annual Survey of Colleges, 2007).

Comparisons Between GED Credential Recipients and Graduating High School Seniors

Because a nationally representative sample of graduating high school seniors took the GED Tests as part of the standardization and norming process, the performance of these seniors on the GED Tests can be compared directly to the performance of GED credential recipients. Table 5.15 compares the performance of GED examinees in 2002 and graduating high school seniors who participated in the 2001 norming study on content area GED Tests (minimum score requirement of 410) and the GED test battery (minimum average test battery score of 450 or higher and a score of 410 or higher on each test). The results suggest that GED credential recipients scored slightly higher than graduating high school seniors on the Social Studies, Science, and Language Arts, Reading Tests. However, GED credential recipients scored lower than the graduating high school seniors on the Language Arts, Writing and Mathematics Tests. The percentage meeting the passing criterion for the entire test battery was higher for the GED credential recipients, as well.

TABLE 5.15
MEAN STANDARD SCORES AND PERCENTAGE MEETING PASSING CRITERION FOR U.S. GRADUATING HIGH SCHOOL SENIORS (2001) AND GED CREDENTIAL RECIPIENTS (2002)

	Graduating High School Seniors			GED Credential Recipients		
TEST	Mean	SD	Percentage Meeting Minimum Score	Mean	SD	Percentage Meeting Minimum Score
Language Arts, Writing	500	100	84	478	73	91
Social Studies	500	100	84	510	84	92
Science	500	100	84	512	84	93
Language Arts, Reading	500	100	84	529	104	93
Mathematics	500	100	84	472	82	83
GED test battery			60*	505	70	71

Note: Additional results are provided in the Annual Statistical Reports (GEDTS, 2005, 2006, 2007, 2008).
* Estimated; fewer examinees will meet the battery passing standard (minimum 450 average and 410 on each test) than the individual content area test minimum score (410).

A recent study by George-Ezzelle and Hsu (2007) also compared the performance between graduating high school seniors in the U.S. 2001 norm group, U.S. GED examinees who took one or more tests in 2002, 2003, or 2004, and U.S. GED examinees who passed the test battery in 2002, 2003, or 2004. One of the primary goals of this report was to provide evidence of the academic value of the GED credential, in turn allowing employers and admissions officers to evaluate GED credential recipients similarly to those with high school diplomas.

Comparisons were made using standard scores, percentile rank distributions, and item difficulty statistics for the tests and each of the content and cognitive levels measured by the GED Tests. The results demonstrated "that examinees who passed the GED Tests met and, in many test areas, exceeded that of the lower 40 percent of graduating high school seniors" (p. 33). In addition,

> GED Tests passers outperformed seniors at a statistically significant level in every content and cognitive level in every test except the Mathematics Test. In the Mathematics Test, GED Tests passers outperformed seniors only in the content areas of number operations and number sense and data analysis, statistics, and probability and on items measuring the conceptual cognitive level. Furthermore, examinees outperformed seniors in 9 of the 23 content areas (mainly in writing, social studies, and science) and 12 of the 18 cognitive levels (in every test except the Mathematics Test), and seniors outperformed examinees only in the content area of algebra, functions, and patterns. (p. 33-34)

Readers are referred to the full report (located at www.GEDtest.org) for additional details.

VALIDITY ISSUES SPECIFIC TO THE LANGUAGE ARTS, WRITING TEST

Although the essay scores from Part II of the Language Arts, Writing Test are not interpreted independently from the multiple-choice portion, the essay score is used to determine whether the examinee can write as well as those students who are expected to graduate in their senior year of high school. To this end, the process of assigning essay scores should be defensible in the sense that the essay readers are properly trained and monitored for adherence to the rubric. In addition, the rubric development process must also be defensible such that the rubric itself can provide the necessary information (i.e., score) used for making valid inferences.

The paragraphs that follow describe the processes involved in scoring the essays, the development of the scoring rubric, the selection and training of essay readers, as well as the site certification and monitoring processes.

Essay Scoring Sessions

Each GED Testing Service-certified Official Scoring Site employs a Chief Reader who oversees the essay scoring processes. Scoring sessions, overseen by the Chief Reader, occur at various times throughout the calendar year. At the beginning of each essay scoring session, the Chief Reader gives an essay topic, the 2002 Series GED Writing Test Official Essay Scoring Guide, four anchor essays (one for each score point), and sets of recalibration essays. After a discussion of the tasks required by the topic, a review of the qualities enumerated on the scoring guide, and an examination of the description of each score point as it applies to the four anchor essays, readers are asked to read one or two sets of recalibration essays and score them. The scoring and discussion of recalibration essays continues until the Chief Reader is satisfied that all readers are aligned with the scoring guide and can confidently score the essays. At this point, the recalibration session ends and the scoring of actual operational essays begins.

When scoring essays, readers are encouraged to read quickly because a slow, deliberate reading of an essay tends toward an analytical, rather than holistic, evaluation of writing. However, readers are reminded that accuracy is more important than speed.

How the Scoring Standards Are Defined

The standards for an essay scoring session were developed by the GED Testing Service Writing Advisory Committee, which comprises language arts educators. The standards are defined within the 2002 Series GED Writing Test Official Essay Scoring Guide (see Chapter 1) and by sample essays called *anchor essays,* which illustrate the different points on the scoring scale. Essays that serve as anchor essays for a topic include only those on which the GEDTS Writing Advisory Committee members' scores agreed.

Readers who score essays on Part II of the Language Arts, Writing Test rely on a four-point scoring guide that was developed by the Writing Advisory Committee. Although this scoring guide "describes" the writing of graduating high school seniors, it does not "prescribe" writing standards. The Writing Advisory Committee reviewed hundreds of essays written by the norming sample of graduating high school seniors, rank-ordered the essays in four distinct categories that represented the full range of ability (where a top score of 4 was assigned to the highest category of the essays, a score of 1 to the weakest essays), and then identified the key components found within each category.

The characteristics of essays at each score on the scale are described in general terms. While the sample essays illustrate standards only for the specific topic for which they are written, the scoring guide defines characteristics that all essays should exhibit regardless of the essay topic.

Selecting and Training Chief Readers and Essay Readers

When GED Testing Service decided to add direct assessment of writing to the previous series' Writing Skills Test, a decentralized program was maintained. GED Administrators were offered four possible scoring configurations for their jurisdiction: (a) a central scoring site within the jurisdiction, (b) multiple scoring sites within the jurisdiction, (c) a commercial scoring site, or (d) the GEDTS essay scoring site. Because all of the configurations were chosen, essays are currently scored at 17 decentralized essay-scoring sites, in addition to five Spanish-scoring and French-scoring sites each, with Chief Readers at each site trained and certified by GEDTS staff members. Each jurisdiction's GED Administrator is responsible for selecting a potential Chief Reader who meets the *GEDTS Qualifications for Chief and Site Readers* (see Appendix P).

Each Chief Reader must undergo a two-day, modified holistic training and qualifying session conducted by a GED Testing Service language arts test specialist, with assistance from Writing Advisory Committee members. Training includes an orientation to the background and purpose of the assessment, review of the training manual and guidelines for holistic scoring, discussion of reader objectivity issues, exposure to and participation in the Writing Advisory Committee's standard-setting procedures, practice in scoring four five-essay training sets, and an opportunity to serve as a Chief Reader.

The Writing Advisory Committee selects the essays used to train, certify, recalibrate, and monitor Chief Readers, essay readers, and scoring sites. The committee selects the essays from stratified, random national samples of direct writing from graduating high school seniors who participate in the standardization study. Standard essays identified by the committee comprise the four Chief Reader sets (five essays in each set) and one certification set (20 essays). Committee members read and discuss each essay at the same time; members then score each essay independently and compare their scores. If at least 80 percent of the

committee's members agree on the same score for an essay, it is assigned a committee score and may be included in the training set. For each essay, the committee also includes a commentary describing why it received a particular score.

The language arts test specialist uses the initial training set to guide potential Chief Readers through a simulated Writing Advisory Committee essay selection exercise, in which participants articulate score selection using 2002 Series GED Writing Test Official Essay Scoring Guide language. This exercise shows potential Chief Readers the correct procedure involved in assigning essay scores, thus emphasizing the importance of the scoring guide. The exercise also allows the trainer to hear the rationale offered by a Chief Reader when assigning a score to an essay. The test specialist and Writing Advisory Committee members then train Chief Reader candidates to apply the scoring criteria by guiding the candidates through four additional training sets.

To be certified as a Chief Reader, a candidate reads and scores 20 papers and must agree exactly with at least 10 of the scores set by the Writing Advisory Committee. The remaining scores must be within one score point. A candidate may have no discrepant scores (that is, scores that differ by two or more points from the Writing Advisory Committee's scores); if the candidate does have a discrepant score, he or she must take a second certification set and meet the standard. If a candidate fails to certify on the second attempt, he or she cannot be certified to score Language Arts, Writing Test essays and must wait at least six months before attending another training and certification session.

In order to train readers to discriminate correctly between scoring points, the training packets assembled by the GEDTS language arts test specialist include a disproportionate number of borderline essays at each scoring point. In addition, the training sets include examples of problematic essays, such as those that cause extensive deliberation by the committee, those that would probably result in discrepant scores, and those that would require third readings.

The Chief Reader training session serves as a model for Chief Readers to use when training potential essay readers for the decentralized scoring sites. After their training and certification, Chief Readers are responsible for recruiting site essay readers, using minimum GEDTS qualifications (see Appendix P). Training and certifying site essay readers follow the same methods and use the same materials and criteria for certifying Chief Readers. Following this procedure reinforces consistency between training and certifying Chief Readers and site essay readers.

Site Certification

In addition to the certification of Chief Readers and essay readers, GEDTS staff members must also certify each particular scoring site. Certified sites are qualified to read and score GED examinee essays. To qualify for certification, a site must demonstrate the quality of its essay scoring in two respects: (a) reader agreement—the consistency with which different readers, in a given reading, award the same scores to a given set of essays; and (b) scoring stability—the degree to which readers uniformly apply the GEDTS score scale (the scoring guide) in evaluating essays both within and across scoring sessions. Stability is determined by matching the total score awarded an essay by each reader at the potential site with the total score awarded an essay by the GED Testing Service Writing Advisory Committee.

To meet these standards on sample sets of essays that have been scored with at least an 80 percent agreement by the Writing Advisory Committee, a scoring site must achieve at least 90 percent agreement with the scores awarded by the Writing Advisory Committee. Agreement is defined as the percentage of a site's combined scores within one point (on the four-point scale) of the Writing Advisory Committee's scores. For more detailed information regarding the site certification process, see either the *Examiner's Manual for the Tests of General Educational Development* (GED Testing Service, 2005a) or Chapter 4 of this manual.

Site Monitoring

To ensure scoring stability over time and to eliminate scale drift, GEDTS has instituted four site-monitoring strategies, namely, recalibration monitoring, Chief Reader monitoring, systematic site monitoring, and random site monitoring. For recalibration monitoring, the Writing Advisory Committee develops seven different recalibration sets for each operational essay topic. Having seven sets allows the Chief Reader to rotate the sets over time, preventing essay readers from memorizing the correct scores. Recalibration sets

are used at the beginning of each scoring session and at the introduction of each new topic to assess the essay readers' adherence to the standards of the 2002 Series GED Writing Test Official Essay Scoring Guide. All readers are required to score the set of five to seven recalibration essays. Essay readers must compare their scores with those of the Writing Advisory Committee to see if they are consistent with the 2002 Series GED Writing Test Official Essay Scoring Guide standards and to determine if they are consistently scoring essays higher or lower than other readers. When a Chief Reader detects a reader who is not scoring according to the 2002 Series GED Writing Test Official Essay Scoring Guide, then the reader is retrained. If after retraining the reader continues to score inconsistently, the reader may no longer score essays for GEDTS.

The Chief Reader serves a critical site monitoring function on a weekly basis. After each reader completes assigned readings, the Chief Reader is required to complete a number of second readings to identify a reader's tendency to score higher or lower than the expected score as dictated by the 2002 Series GED Writing Test Official Essay Scoring Guide. In addition, the Chief Reader resolves all discrepant scores (scores differing by two or more points) and records the number of the reader with whom the Chief Reader agrees. If a reader shows a tendency to award a discrepant score more than once during a scoring session, the Chief Reader must help get the reader back on scale through personal guidance.

Systematic monitoring is identical to site certification procedures, using the same guidelines and criteria for passing a site. Every reader at each site must read and score a monitoring set of essays (two sets of 20 essays each). Site Chief Readers and essay readers do not know what scores the Writing Advisory Committee has given to the essays in the monitoring sets. Because all readers are scoring the same 40 essays, GEDTS can evaluate scoring stability on a site-by-site basis and across sites to ensure that an essay score awarded at any one site would receive a similar score at any other site.

Systematic site monitoring evaluates the scoring site and its readers' scoring stability. Scoring stability refers to the similarity of scores given to an essay by readers at a scoring site (site essay scores) and scores given to that essay by the Writing Advisory Committee (GEDTS essay scores). Site monitoring reports provide the results of the site's performance on four scoring stability criteria and the performance of individual readers at the site.

The four scoring stability criteria for the year 2005 and beyond are as follows (year 2002–2004 criteria are in parentheses):[17]

1. *Percent agreement with GEDTS essay scores.* This criterion indicates the percent of site essay scores equal to or within 1 point of the GEDTS essay scores. A site must have at least 90 percent agreement with GEDTS essay scores.
2. *Percent of site essay scores equal to GEDTS essay scores.* A site must have at least 50 percent of its essay scores equal to the GEDTS essay scores (raised from 35 percent).
3. *Percent of discrepant scores.* Discrepant scores are defined as essay scores that differ by more than 1 point from the GEDTS essay score. The percent of discrepant scores at a site must be 5 percent or less (dropped from 7 percent).
4. *Intraclass correlation between GEDTS and scoring site.* The intraclass correlation reflects the strength of agreement between site essay scores and the GEDTS essay scores. A site must have an intraclass correlation of 0.80 or higher (raised from 0.70).

In order to continue to be certified to score essays for GEDTS, a site must meet certification criteria 1 and 4. In addition, each reader at the site must meet certification criteria 2 and 3.

The empirical results of site monitoring were reported and discussed in Chapter 4.

[17] These criteria were revisited prior to 2005, when it was decided that an increase in standards was necessary.

Chapter 6: Accommodations for GED Examinees with Disabilities

The GED testing program has long provided accommodations to examinees with disabilities and is committed to complying with the requirements of the Americans With Disabilities Act of 1990 (ADA). In an effort to make the GED Tests accessible to all applicants, accommodations are made for examinees with diagnosed physical, learning, or psychological disabilities who can provide appropriate documentation from a qualified professional of their impairment and its effect on their ability to take the GED Tests under standard conditions (see Table 6.1).

TABLE 6.1
EXAMPLES OF ACCOMMODATED DISABILITIES AND DOCUMENTATION SOURCES

Disability	Examples of Disability	Licensed/Certified Professionals Providing Documentation
Physical disabilities	blindness, low vision, deafness, impaired hearing, mobility impairments	physician; specialists in a particular area, such as audiologists
Learning disabilities	dyslexia, dyscalculia, receptive aphasia, written language disorder	psychologist; school psychologists; educational specialist with advanced training
Attention-deficit/ hyperactivity disorder	attention-deficit/hyperactivity disorder	Psychologists with advanced training; psychiatrists; physicians
Psychological disabilities	bipolar syndrome, Tourette's syndrome	psychiatrists; psychologists; school psychologists

Under ADA, entities that administer standardized assessments must offer them in a place and manner that allows access to persons with disabilities. This may require reasonable modifications to the manner in which the test is administered, such as extended testing time, as well as appropriate auxiliary aids and services (i.e., testing accommodations). The goal is to ensure that, for individuals with documented disabilities, the "test results accurately reflect the individual's aptitude or achievement level or whatever other factor the examination purports to measure, rather than reflecting the individual's impaired sensory, manual or speaking skills (except where those skills are the factors that the test purports to measure)" (Americans With Disabilities Act of 1990, Section 12112, subsection A, § 7).

Consistent with ADA, GED Testing Service has long believed that every examinee should have a fair opportunity to demonstrate his or her knowledge and skills under appropriate test conditions. Some examinees with disabilities may not be able to fully demonstrate their knowledge and skills under standard testing conditions. Conditions such as a physical, psychological, or learning disability, or attention-deficit/hyperactivity disorder may create particular challenges for certain examinees.

AVAILABLE ACCOMMODATIONS FOR GED EXAMINEES

All potential GED examinees must be made aware of the availability of test accommodations and the process for requesting such accommodations. Even though it is the responsibility of GED Chief Examiners and GED Examiners to disseminate information about test accommodations, GED Testing Service print and web publications also provide this detailed information.

The following accommodations are available to examinees with documented disabilities:

- Audiocassette edition (with large-print reference copy)
- Braille edition
- Use of video equipment
- Use of a talking calculator or abacus
- Sign-language interpreter/use of a scribe
- Extended time/supervised extra breaks
- Use of a private room
- One-on-one testing at a health facility or at home
- Other reasonable accommodations as warranted, based on individual need

The Chief Examiner at each examination site may permit the use of certain adaptations and devices without prior approval from the GED Administrator, GEDTS, or GEDTS-trained and certified personnel. In general, the following accommodations require no documentation:

- Colored transparent overlays
- Clear transparent overlays and highlighter
- Temporary adhesive notes (e.g., Post-it® Notes)
- Earplugs
- Large-print test
- Magnifying glass
- One test per day
- Straightedge
- Other devices as deemed appropriate

EFFECT OF TEST ACCOMMODATIONS ON GED TEST PERFORMANCE

In an effort to contribute to the research on the effect of test accommodations on GED Tests performance, George-Ezzelle and Skaggs (2004) examined the comparability of test performance across examinees who did not receive any test accommodation and examinees who received combinations of accommodations including extended time, private room, and supervised breaks. The remaining sections of this chapter summarize George-Ezzelle and Skaggs's research findings.

Overview

Current testing standards call for test developers to provide evidence that testing procedures and test scores and the inferences made based on the test scores show evidence of validity and are comparable across subpopulations (AERA, APA, & NCME, 1999). Evidence of the comparability of test validity across subpopulations can be collected through examination of (a) the representation of the content domain being tested, (b) the relationship of test performance to other variables, (c) the internal structure of the test, (d) the response processes of examinees, and (e) the consequences of test score use. The following research focused on examining the internal structure of the test and the response processes of examinees with and without accommodations of test scheduling (e.g., extra time, breaks) and physical setting (private room).

Numerous studies on the comparability of tests across examinee subpopulations using the above approaches have been conducted. Specific subpopulations researched include those based on gender, ethnicity, primary language, cultural backgrounds, and the use of test accommodations. Summaries of the research on the effect of test accommodations on test performance have yielded inconsistent findings. "One thing that is clear from our review is that there are no unequivocal conclusions that can be drawn regarding the effects, in general, of accommodations on students' test performance. The literature is clear that accommodations and students are both heterogeneous" (Sireci, Li, & Scarpati, 2003, p. 16).

George-Ezzelle and Skaggs examined the comparability of GED Tests performance across examinees who did not receive any test accommodation and examinees who received an accommodation of either (a) extended time only; (b) extended time and private room only; or (c) extended time, private room, and

supervised breaks only. These test accommodations represent situations in which there is a change in the scheduling/timing or physical setting of the exam versus the presentation (e.g., audio) or response format (dictated response) of the exam. The equivalence of an academic achievement test's psychometric properties across accommodated and non-accommodated examinees was examined through the calculation of group descriptive statistics, reliability estimates, standard errors of measurement, and differential item functioning (DIF).

Method

The data analyzed were from the 2002 examination cycle of the GED Tests. At the time of analysis, the database contained test and examinee data from GED Tests administrations in 48 states and the District of Columbia (Ohio and Connecticut data were not included). Test and examinee data from the English-language version of the GED Tests administered in the United States during the 2002 examination cycle were the base source of the study's sample. This study referred to the three operational forms as Form 1, Form 2, and Form 3. In 2002, approximately 140,000 to 160,000 examinees took each test form. Note that the Mathematics Test was not included in this study.

Within this base source, a small sample of examinees requested and received some form of accommodation in the administration of the tests. Prior to testing, examinees who requested test accommodations were required to complete a form to include documentation of their disabilities. Approval of accommodation use was granted after review by state-level GED Administrators or GEDTS staff. The accommodation sample consisted of test and examinee data from examinees who took the GED Tests with the following test scheduling and/or setting accommodations: (a) extended time only; (b) extended time and private room only; or (c) extended time, private room, and supervised breaks only. The number of examinees who received these accommodations is presented in Table 6.2.

TABLE 6.2
NUMBER OF GED EXAMINEES RECEIVING SCHEDULING AND/OR SETTING ACCOMMODATIONS

TEST FORM	Extended Time Only	Extended Time & Private Room Only	Extended Time, Private Room, & Supervised Breaks Only	Total
Language Arts, Writing Form 1	60	42	22	124
Language Arts, Writing Form 2	51	34	7	92
Language Arts, Writing Form 3	52	26	9	87
Social Studies Form 1	60	49	18	127
Social Studies Form 2	56	26	7	89
Social Studies Form 3	53	39	7	99
Science Form 1	49	40	17	106
Science Form 2	41	28	7	76
Science Form 3	50	34	7	91
Language Arts, Reading Form 1	71	48	18	137
Language Arts, Reading Form 2	61	36	7	104
Language Arts, Reading Form 3	64	42	13	119

In addition to test scores from the GED Tests, the study also accessed demographic information provided by each examinee. Such demographic information included age, gender, race, geographic region of residence, and highest level of education completed. Descriptive statistics on demographic characteristics of the accommodated and non-accommodated samples are presented in Table 6.3. The non-response rate for the demographic questions was higher (41 percent to 56 percent) for the accommodated sample than for the non-accommodated sample (1 percent to 12 percent). Based on all responses, examinees in the accommodated sample were, on average, younger, less likely to be African American or of Hispanic descent, and more likely to have an educational level lower than examinees in the non-accommodated sample.

TABLE 6.3
DEMOGRAPHIC CHARACTERISTICS OF GED EXAMINEES RECEIVING SCHEDULING AND/OR SETTING ACCOMMODATIONS

	Sample Group (percent)	
DEMOGRAPHIC	Accommodated	Non-Accommodated
Age		
16-<20 years	34	17
20-<25 years	12	48
25-<30 years	3	13
30-<35 years	3	6
35-<40 years	2	7
40-<50 years	3	6
50-<60 years	2	2
60+ years	<1	1
Missing/Invalid	42	<1
Gender		
Male	42	57
Female	16	41
Missing	42	2
Ethnicity/Race		
Hispanic origin or descent	3	12
American Indian or Alaskan Native	<1	2
Asian	<1	1
Black/African American	5	21
Native Hawaiian or Pacific Islander	<1	<1
White	41	51
Missing	50	12
Highest Educational Level		
None	<1	0
K–6th grade	<1	1
7th grade	5	1
8th grade	8	7
9th grade	14	16
10th grade	10	25
11th grade	4	32
12th grade	1	6
Missing	56	10
Geographic Region		
Northeast	23	20
Midwest	16	16
South	18	41
West	2	21
Missing	41	2

DIF analyses were conducted to evaluate whether individual items or groups of items performed differentially for accommodated vs. non-accommodated examinees. The SIBTEST procedure (Shealy & Stout, 1993) was used for all DIF analyses. SIBTEST evaluates differences in item functioning between two groups: the reference group and the focal group. For this study, the focal group consisted of those examinees who received the specified scheduling and/or setting accommodations. The reference group was a sample of examinees who did not receive any accommodations. Because more than 140,000 examinees per test form received no accommodations, a random sample of 500 examinees was selected from each test form to make the group sizes more comparable.

SIBTEST conducts a DIF analysis on a suspect subtest containing one or more items. Most traditional DIF analyses focus on a suspect subtest of one item or, in other words, an individual item analysis. However, the suspect subtest can consist of groups of items. In this study, both types of analyses were carried out. Subtest groupings of items were based on item content; for example, poetry items in the Language Arts, Reading Test forms were analyzed as a group. SIBTEST uses a valid subtest to match examinee ability levels. The valid subtest is a group of items that are assumed to be free of DIF. In this study, the valid subtest for each SIBTEST run consisted of all the items that were not part of the suspect subtest. In other words, in individual item analyses, the valid subtest was the other 49 (or 39 for the Language Arts, Reading Test) items on the test. For content subtest analyses, the valid subtest was the remaining items on the test.

The end product of a SIBTEST analysis is the calculation of a statistic, β_{UNI}. β_{UNI} has the following form:

$$\beta_{UNI} = \sum_{i=1}^{k} p_i \left(\overline{Y}_{Ri}^{*} - \overline{Y}_{Fi}^{*} \right) \qquad \textbf{(6.1)}$$

where k is the number of items on the valid subtest, p_i is the proportion of focal group examinees obtaining raw score i, and $\overline{Y}_{Ri}^{*}$ and $\overline{Y}_{Fi}^{*}$ are the mean raw scores on the suspect subtest for the reference and focal groups, respectively, with raw score i on the valid subtest. The means are adjusted by a regression correction that effectively controls for an inflation to the Type I error rate that would occur due to measurement error. A useful feature of the β_{UNI} statistic is that its sign indicates the direction of DIF. A positive value favors the reference group, and a negative value favors the focal group. In addition, an asymptotic standard error is available for β_{UNI}. Dividing β_{UNI} by its standard error yields a z statistic that is normally distributed, thus providing a statistical test for the significance of the magnitude of β_{UNI}.

Results

Raw score descriptive statistics, K-R 20, and the standard error of measurement were calculated for the accommodated and non-accommodated samples. These results are shown in Table 6.4. In all 12 test forms, the examinees testing under standard administration had higher mean raw scores than the examinees receiving scheduling and/or setting accommodations. The differences in mean raw scores between the two groups ranged approximately from one-tenth to one-third of a standard deviation. The K-R 20s and SEMs were about the same between the two samples across test forms, providing evidence of equal reliability of the tests for both groups.

TABLE 6.4
RAW SCORE DESCRIPTIVE STATISTICS, K-R 20S, AND STANDARD ERRORS OF MEASUREMENT FOR ACCOMMODATED AND NON-ACCOMMODATED SAMPLES

	N	Mean	Median	SD	Min	Max	K-R 20	SEM
Language Arts, Writing Form 1								
Accommodated	124	31.07	32	8.99	8	48	.89	3.0
Non-accommodated	500	34.45	37	9.65	6	50	.91	2.9
Language Arts, Writing Form 2								
Accommodated	92	32.01	33	7.78	14	50	.85	3.0
Non-accommodated	500	35.83	37	8.20	6	50	.88	2.8
Language Arts, Writing Form 3								
Accommodated	87	31.32	32	9.29	10	50	.90	3.0
Non-accommodated	500	34.67	36	9.16	7	50	.90	2.9
Social Studies Form 1								
Accommodated	127	32.76	35	10.06	4	49	.92	2.9
Non-accommodated	500	35.47	37	8.60	6	50	.89	2.9
Social Studies Form 2								
Accommodated	89	32.64	34	9.14	6	49	.90	2.9
Non-accommodated	500	33.77	35	10.09	1	50	.92	2.9
Social Studies Form 3								
Accommodated	99	31.12	32	9.98	6	49	.91	3.0
Non-accommodated	500	35.29	37	9.31	4	50	.91	2.8
Science Form 1								
Accommodated	106	33.68	36	10.68	6	50	.93	2.9
Non-accommodated	500	37.07	39	9.33	5	50	.92	2.6
Science Form 2								
Accommodated	76	35.04	35	7.93	11	50	.87	2.9
Non-accommodated	500	35.54	37	8.34	10	50	.88	2.9
Science Form 3								
Accommodated	91	33.98	35	9.78	8	48	.91	2.9
Non-accommodated	500	36.54	39	9.19	6	50	.91	2.8
Language Arts, Reading Form 1								
Accommodated	137	28.18	29	7.46	6	40	.89	2.5
Non-accommodated	500	29.86	32	7.88	8	40	.91	2.4
Language Arts, Reading Form 2								
Accommodated	104	28.26	30	6.93	7	39	.86	2.6
Non-accommodated	500	31.04	33	7.25	6	40	.90	2.3
Language Arts, Reading Form 3								
Accommodated	119	27.82	29	7.14	3	39	.87	2.6
Non-accommodated	500	29.53	31	6.70	5	40	.86	2.5

Individual Items

In order to control for the Type I error rate within each form, the Bonferroni correction was used; an item was referred to content specialists for further review if the p-value for β_{UNI} was less than .05 divided by the number of items. Using this criterion, 11 items across the 12 test forms were identified as exhibiting substantial DIF. Table 6.5 lists the items exhibiting substantial DIF and contains item descriptions.

GEDTS test specialists examined the flagged items listed in Table 6.5 for any plausible reason to explain why the items favored the indicated group. For the three items from the Language Arts, Writing Tests, the specialists could identify no characteristic of the items that would have provided an advantage to one group over another. The single social studies item had a map with symbols that required analysis and evaluation. The specialists speculated that examinees with time/setting accommodations might have benefited by using extra time on this question to gain a better understanding of the map and symbols.

Two science items were flagged in favor of examinees with accommodations. One of the items had a food chain graphic, required analysis and inference, and was located toward the end of the test (number 40 out of 50). As speculated for the Social Studies Test item, examinees with time/setting accommodations may have benefited by using extra time to fully understand the graphic. The second science item had a topographic map with several legends; however, the cognitive requirements of the item appeared to be very low. No characteristic of the item that would have been advantageous to examinees with scheduling and/or setting test accommodations or disadvantageous to examinees under standard test administration was identified.

TABLE 6.5
INDIVIDUAL ITEMS FLAGGED FOR DIF BY SIBTEST

Test/Form	Item	Group Favored	Description of Item
Language Arts, Writing Form 1	27	Accommodated	Usage
Language Arts, Writing Form 2	6	Accommodated	Usage
Language Arts, Writing Form 3	18	Non-accommodated	Mechanics
Social Studies Form 2	16	Accommodated	Analyze map
Science Form 1	40	Accommodated	Make inference about graphic
Science Form 3	34	Accommodated	Analyze map
Language Arts, Reading Form 1	19	Accommodated	Reading comprehension
Language Arts, Reading Form 1	21	Non-accommodated	Analysis
Language Arts, Reading Form 1	40	Accommodated	Extended synthesis
Language Arts, Reading Form 2	6	Accommodated	Reading comprehension
Language Arts, Reading Form 3	17	Non-accommodated	Analysis

The Language Arts, Reading Test had the greatest number of flagged items; three items were flagged in favor of examinees with accommodations, and two items were flagged in favor of examinees testing under standard administration. All flagged items were attached to works of prose fiction (versus nonfiction or poetry). Two of the three questions that favored examinees with accommodations were reading comprehension items; the other was an expanded synthesis question that required examinees to use additional information in the item stem and synthesize it with the passage information in order to arrive at a correct answer. Test specialists speculated that examinees with extended time might have been more likely to go back through the items and check the accuracy of comprehension and analysis items. However, the two items in favor of non-accommodated examinees were also analysis items, suggesting that additional time may have resulted in examinees' mistrusting or second-guessing their first interpretation of the passage or their initial answer to the item.

Subtest

In addition to running the SIBTEST procedure on individual items, DIF analyses were extended to clusters of items (subtests) grouped by content areas. Because the same content areas were covered by each form within a test, it was possible to examine the consistency of subtest DIF results across test forms.

The one consistent finding was DIF in favor of non-accommodated examinees for the Mechanics subtest in all three forms of the Language Arts, Writing Test. Test specialists hypothesized that examinees testing under standard administration may be more accustomed to the requirements of unassisted editing for capitalization, spelling, and punctuation. There were several other statistically significant, but inconsistent, findings. In Social Studies Form 1, the Civics and Government subtest exhibited DIF in favor of non-accommodated examinees. In Science Form 2, the Life Science subtest exhibited DIF in favor of accommodated examinees, and in Science Form 3, the Physical Science subtest exhibited DIF in favor of non-accommodated examinees. In Language Arts, Reading Form 3, post-1960 fiction favored accommodated examinees, and drama favored non-accommodated examinees. The only pattern in the Language Arts, Reading Test DIF results was that the direction of DIF favored non-accommodated examinees on all three forms for poetry and drama, although the only statistically significant result was for drama on Form 3.

Discussion

Examination of raw score statistics indicated that, while the sample of GED examinees testing under standard administration procedures consistently (across all content area tests and test forms) achieved a higher average raw score than the sample testing under scheduling and/or setting accommodations, the differences in average scores were small, ranging in size from one-tenth to one-third of a standard deviation. Reliability estimates also indicated small differences in the values of K-R 20s and SEMs between the accommodated and non-accommodated samples. These small empirical differences in raw scores and reliability estimates provided evidence supporting the validity and comparability of test scores obtained under test scheduling and/or setting accommodations.

DIF analyses on the individual item responses of examinees testing under scheduling and/or setting accommodations and standard administration procedures flagged 11 items exhibiting substantial DIF. Five of the 11 flagged items were found across two of the three Language Arts, Reading Test forms, and three of the items were found across the three Language Arts, Writing Test forms. Less than 1 percent of any test form's items were flagged for exhibiting substantial DIF. In content areas where more than one item was flagged for DIF, the number of items favoring one group versus the other was nearly equal. Furthermore, while test content specialists were sometimes able to hypothesize about item characteristics that might have advantaged or disadvantaged one group over the other, the small number of flagged items and the variability of the characteristics of the items caution that these hypotheses require further research. Even discussion of the results of DIF analyses on content area subtests, where the Language Arts, Writing Tests' Mechanics subtest was consistently flagged for DIF in favor of examinees testing under standard administration, test content specialists were tentative about possible reasons for such differences.

Limitations of this study are connected to the definition of inclusion in the accommodated group and sample comparability. The accommodated group included examinees with either single or multiple scheduling and/or setting accommodations. Further, multiple-accommodation administrations sometimes involved the use of a setting accommodation (private room) in addition to scheduling accommodation(s) (extended time, breaks). Had the data been analyzed using only single-accommodation data or using only scheduling accommodations, results may have differed. Comparability of the accommodated and non-accommodated groups is questionable because nearly half of the examinees in the accommodated group did not respond to the demographic questions of age, gender, race/ethnicity, and highest educational level achieved. Analysis of non-missing responses showed group differences in age, gender, ethnicity, and highest educational level achieved, several of which are demographic characteristics that may influence performance on an educational achievement test and, therefore, affect the results of this study. As much as aggregate previous test accommodation research results are inconsistent, this study provided partial evidence of the comparability of test scores from the GED Tests administered under frequently used scheduling and setting accommodations.

In conclusion, the results of this study provided support that GED test scores in writing, reading, social studies, and science show evidence of validity under test accommodations of (a) extended time only; (b) extended time and private room only; and (c) extended time, private room, and supervised breaks only. Further research on the validity of GED test score interpretations under test accommodations is recommended and should attempt to address this study's limitations, particularly the relationship of sample demographics to performance.

Chapter 7: English-Language Canadian GED Tests

OVERVIEW

Chapter 1 provided a brief history of the GED testing program. Chapters 2 through 5 presented technical information pertaining to the development, norming, scaling, equating, reliability, and validity of the English-language U.S. GED Tests. This chapter describes the aspects of the GED Tests that are specific to the English-language Canadian GED Tests.

The purpose of the Canadian version of the GED Tests is to provide an opportunity to adults in Canada to certify their attainment of high school–level academic knowledge and skills and earn their jurisdiction's high school equivalency credential, diploma, or certificate. The development of the English-language Canadian GED Tests is similar to the development of the English-language U.S. GED Tests. However, because these tests serve different populations of GED examinees, they are normed on Canadian graduating high school seniors. A description of the history, development, reliability, and validity of these tests is provided in the next sections.

HISTORY OF THE ENGLISH-LANGUAGE CANADIAN GED TESTS

A thorough description of the early history of the GED testing program in Canada is provided by Quigley (1987). As Quigley noted, the first Canadian administration of the GED Tests occurred in the province of Nova Scotia in 1969. Similar to the circumstances that provided the impetus for the GED testing program in the United States, the needs of former military personnel prompted the demand for a high school equivalency program in Canada. The GED testing program in Canada expanded to Saskatchewan in 1970, Prince Edward Island in 1971, Manitoba in 1972, British Columbia in 1973, New Brunswick and Newfoundland in 1974, Northwest Territories in 1975, Yukon Territory in 1976, and Alberta in 1981.

TEST SPECIFICATIONS AND DEVELOPMENT

The content specifications for the English-language Canadian GED Tests are identical to those in four of the five English-language U.S. GED Tests: Language Arts, Writing; Science; Language Arts, Reading; and Mathematics. The content specifications for the Social Studies Test differ slightly from the U.S. counterpart: About 50 percent of the Social Studies Test measures content specific to Canadian history, political science, economics, and geography. The content specifications for the Canadian version of the Social Studies Test were presented in Chapter 2. As these specifications indicate, both items with specific Canadian focus and items that relate to the global community are included on this test.

Though the content specifications for the other four tests are identical for both the English-language U.S. and Canadian versions, some contextual differences exist between the U.S. and Canadian version Science and Mathematics Tests. On the Canadian versions of the Science and Mathematics Tests, International System of Units (SI units) are used throughout the test (e.g., metres, litres, grams), whereas on the U.S. version, Imperial units are primarily used (e.g., feet, gallons, pounds). Similarly, spaces, rather than commas, are used in the Canadian versions to denote triads of digits in long numbers and decimals. For example, in U.S. versions, the number *twenty-one thousand* would be displayed as 21,000, while in the Canadian versions it would be displayed as 21 000. It is assumed that these contextual differences do not result in different concepts being measured by the two versions of these tests.

Because the content of the English-language Canadian GED Tests is essentially identical to the U.S. version for four of the five content area tests, the test development process for the English-language Canadian versions of these tests was identical to that of their U.S. counterparts. As mentioned in Chapter 2, Canadian content specialists were included in the development of the 2002 Series GED Tests specifications. Item writers and content reviewers for the Canadian-specific Social Studies Test items were teachers, test specialists, and curriculum experts from Canada. In addition, as the English-language U.S. test forms were developed (i.e., after the items passed all content, measurement, and statistical reviews), they were reviewed and evaluated by Canadian content reviewers to ensure that the content of the tests was appropriate for Canadian GED examinees.

STANDARDIZATION AND NORMING

The English-language Canadian GED Tests were initially normed on a sample of graduating high school seniors across Canada who took the GED Tests during March, April, and May of 2001. Because of low participation in the 2001 standardization and norming study, the English-language Canadian GED Tests were again normed on a sample of graduating high school seniors across Canada during March, April, and May of 2002. The number of schools participating in this second study is presented in Table 7.1. The percentile ranks for standard scores obtained via test forms developed subsequent to 2002 have been based on the performance of this latter norm group.

TABLE 7.1
NUMBER OF SCHOOLS PARTICIPATING IN THE 2002 CANADIAN STANDARDIZATION AND NORMING STUDY AND THE 2003 AND 2004 EQUATING STUDIES FOR THE ENGLISH-LANGUAGE CANADIAN GED TESTS

	2002		2003		2004	
	N	%	N	%	N	%
Alberta	23	14.8	21	12.5	18	12.7
British Columbia	8	5.2	10	6.0	7	4.9
Manitoba	10	6.5	14	8.3	12	8.5
New Brunswick	22	14.2	22	13.1	16	11.3
Newfoundland and Labrador	22	14.2	22	13.1	20	14.1
Northwest Territories	0	0.0	0	0.0	0	0.0
Nova Scotia	18	11.6	25	14.9	20	14.1
Ontario	27	17.4	22	13.1	19	13.4
Prince Edward Island	5	3.2	6	3.6	7	4.9
Quebec	5	3.2	6	3.6	5	3.5
Saskatchewan	14	9.0	19	11.3	17	12.0
Yukon	1	0.6	1	0.6	1	0.7
Total Schools	155		168		142	
Total Students	3,256		3,743		3,042	

SCALING AND EQUATING

As with the English-language U.S. GED Tests, the raw scores from the English-language Canadian GED Tests were converted to a scale ranging from 200 to 800, with a mean of 500 and standard deviation of 100. The raw-to-standard score conversions for the Language Arts, Writing; Science; Language Arts, Reading; and Mathematics Tests are the same as the English-language U.S. raw-to-standard score conversions for those tests.

In 2001, data were collected via two Canadian forms, namely, IA and IC. No equating process was implemented with these forms because of insufficient sample sizes. To obtain the standard scores for these two forms, the conversion tables from the English-language U.S. GED Tests were used. Percentile ranks were calculated based on the combined data from Forms IA and IC.

As mentioned, the sample size obtained in the 2001 Canadian standardization study was insufficient, and thus a second standardization study was conducted in 2002 (see Table 7.1). During this study, Social Studies Test Form IA and all content area tests for Form ID were administered to Canadian graduating high school seniors. The Social Studies Test data were scaled, normed, and equated using the same procedures outlined in Chapter 3. The norms for the remaining four tests were also obtained, yet the conversion tables from the English-language U.S. GED Tests were again used to obtain standard scores.

In a 2003 equating study, all content area tests for Forms IE and IF, and Social Studies Form IA were administered to graduating high school seniors in Canada. Forms IE and IF were not equated because the English-language U.S. tests conversion tables were used again. Social Studies Forms IE and IF were equated, however, to Form IA. In a 2004 equating study, Forms IG and IH were administered alongside Social Studies Form IA. Again, no equating occurred as the English-language U.S. tests conversion tables were used. However, Social Studies Forms IG and IH were equated to Form IA.

RELIABILITY

The reliability of the English-language Canadian GED test scores was analyzed using the same methods that were applied to the English-language U.S. GED Tests. These methods are described thoroughly in Chapter 4. The reliability of the scores from the multiple-choice portions of the English-language Canadian GED Tests are evaluated by calculating the K-R 20 reliability coefficient (Kuder & Richardson, 1937), the standard error of measurement (SEM), and decision consistency. The reliability of the essay portion of the Language Arts, Writing Test was evaluated using additional criteria described in Chapter 4.

The results of the reliability analyses for the 2002 Series English-Language Canadian GED Tests are presented in this chapter. The Canadian data presented herein are from Forms ID through IH, which correspond with the Canadian standardization and norming study performed in 2002 and subsequent equating studies in 2003 and 2004. All studies used a random sampling of graduating high school seniors from across Canada, as described above.

K-R 20 and SEM Results

Table 7.2 presents the score means, standard deviations, SEM, and K-R 20 estimates for the test forms in the 2002 Series English-Language Canadian GED Tests. It should be noted that the numbers in Table 7.2 for the Language Arts, Writing Test refer only to the multiple-choice portion of the test (the reliability of the essay scores and Language Arts, Writing Test composite score is the same as that reported in Chapter 4). The results presented in Table 7.2 are reported in both standard and raw score units. Because the transformation of raw scores to standard scores (described in Chapter 3) is nonlinear, it is not possible to compute K-R 20 directly for standard scores. Thus, K-R 20 estimates are for raw scores only.

The information in Table 7.2 is based on the performance of the sample of graduating high school seniors across Canada who took the GED Tests as part of the standardization and equating studies in years 2002 through 2004. Data from Form ID originated from the 2002 standardization. Data from Forms IE and IF originated from the 2003 equating study, and data from IG and IH originated from the 2004 equating study. The results presented in Table 7.2 indicate that all but one English-language Canadian test form have K-R 20s of at least .90.

TABLE 7.2
SAMPLE SIZES (N), SCORE MEANS, STANDARD DEVIATIONS (SD), STANDARD ERRORS OF MEASUREMENT (SEM), AND K-R 20 ESTIMATES FOR THE 2002 SERIES ENGLISH-LANGUAGE GED TESTS: CANADIAN GRADUATING HIGH SCHOOL SENIOR DATA

		STANDARD SCORES			RAW SCORES			
TEST/FORM	N	Mean	SD	SEM	Mean	SD	SEM	K-R 20
Language Arts, Writing								
Form ID	122	535.5	90.1	22.1	41.2	6.4	2.4	.86
Form IE	750	532.2	99.3	28.1	39.6	7.7	2.5	.90
Form IF	888	537.2	91.1	24.1	40.9	7.8	2.4	.90
Form IG	833	536.9	112.1	29.7	39.7	8.4	2.5	.91
Form IH	458	526.6	113.4	30.0	39.1	8.3	2.5	.91
Social Studies								
Form ID	1,251	493.5	99.5	26.3	34.8	10.5	2.8	.93
Form IE	932	487.4	105.7	28.0	34.6	10.3	2.8	.93
Form IF	908	486.6	106.5	28.2	34.9	10.2	2.8	.93
Form IG	544	489.8	100.5	28.4	33.7	10.1	2.9	.92
Form IH	638	487.3	102.1	28.9	31.4	10.7	2.9	.92
Science								
Form ID	596	538.0	99.7	29.9	37.0	9.0	2.7	.91
Form IE	683	551.7	109.6	29.0	38.9	9.6	2.5	.93
Form IF	675	533.6	106.8	28.3	37.5	10.1	2.6	.93
Form IG	565	552.0	115.6	28.3	40.0	9.6	2.4	.94
Form IH	550	543.6	113.0	29.9	38.2	9.4	2.6	.93
Language Arts, Reading								
Form ID	637	547.0	120.8	38.2	29.8	7.5	2.4	.90
Form IE	678	542.5	115.6	36.6	32.4	7.0	2.2	.90
Form IF	671	553.8	122.9	34.7	31.8	7.9	2.2	.92
Form IG	645	568.1	118.2	37.4	32.5	6.7	2.3	.90
Form IH	557	550.4	117.9	33.4	33.6	7.0	2.0	.92
Mathematics								
Form ID	638	529.4	95.0	28.5	38.4	8.8	2.6	.91
Form IE	661	505.2	108.1	26.5	33.6	11.6	2.8	.94
Form IF	665	517.4	112.8	29.8	36.0	10.4	2.7	.93
Form IG	544	517.5	125.2	30.7	34.5	11.5	2.8	.94
Form IH	532	531.9	116.6	28.6	37.4	10.4	2.6	.94

Note: Data unavailable for Forms IA and IC.

Conditional Standard Errors of Measurement

The conditional standard errors of measurement (CSEM) were calculated for various standard scores using the same methods applied to the English-language U.S. GED Tests (see Chapter 4). The passing standard requirement for the English-language Canadian GED Tests is 450 for each of the five tests and an average of 450 for the battery for all provinces and territories. The estimated standard score CSEM for the English-language Canadian data are presented in Table 7.3.

TABLE 7.3
STANDARD SCORE CONDITIONAL STANDARD ERRORS OF MEASUREMENT AT VARIOUS STANDARD SCORES FOR THE 2002 SERIES ENGLISH-LANGUAGE GED TESTS: CANADIAN GRADUATING HIGH SCHOOL SENIOR DATA

	STANDARD SCORE						
TEST/FORM	400	410	420	430	440	450	460
Social Studies							
Form ID	21.1	25.5	25.4	25.3	25.1	24.7	28.5
Form IE	33.0	24.7	24.6	24.4	24.2	27.7	23.4
Form IF	28.7	24.4	24.3	24.1	23.9	27.6	23.0
Form IG	25.3	25.3	25.2	25.1	25.0	24.8	28.4
Form IH	25.8	30.5	26.2	26.2	26.1	26.0	25.7
Science							
Form ID	25.4	21.2	17.0	21.1	16.8	20.9	20.8
Form IE	24.7	20.6	16.5	20.5	20.3	20.2	16.0
Form IF	32.8	20.5	16.4	20.4	16.3	20.2	20.1
Form IG	25.2	20.8	20.7	16.4	20.3	20.0	23.6
Form IH	26.2	17.3	21.5	21.4	16.9	20.9	20.7
Language Arts, Reading							
Form ID	22.6	18.8	22.5	22.3	22.1	18.2	21.6
Form IE	19.5	19.1	22.6	22.1	21.7	24.6	27.3
Form IF	19.8	23.7	23.5	19.2	22.7	22.3	21.8
Form IG	11.3	18.8	22.3	22.1	21.5	21.1	20.7
Form IH	15.3	22.6	22.0	21.6	24.0	26.6	32.1
Mathematics							
Form ID	16.7	25.3	25.7	25.8	25.7	25.5	24.8
Form IE	25.8	26.0	26.1	26.1	26.0	25.9	25.8
Form IF	21.6	25.9	25.8	25.6	21.2	25.2	24.9
Form IG	26.1	26.3	26.3	26.2	26.1	21.5	25.6
Form IH	20.9	25.1	25.0	24.9	24.7	24.3	24.1

Note: Data for Forms IA and IC not available.

Reliability of Essay Scores on the Language Arts, Writing Test

The reliability of the Language Arts, Writing Test essay scores is the same as that reported in Chapter 4 for the English-language U.S. GED Tests. Because the test specifications are the same for both the English-language U.S. and English-language Canadian versions of the Language Arts, Writing Test, and because essay scoring is performed by language (i.e., English, French, and Spanish), the reliability of essay scores is reported only by language (see Chapters 8 and 9 for reliability information regarding the French- and Spanish-language versions).

Decision Consistency

The decision consistency for each of the five content area tests was examined using data obtained via the Canadian norming and equating studies (i.e., using high school senior data). The same procedure used with the English-language U.S. GED Tests was also applied to the English-language Canadian data (i.e., Livingston and Lewis procedure via the BB-Class software program, see Chapter 4).

The decision consistency (probability of correct classification) estimates for Forms ID through IH are provided in Table 7.4. As did the estimates from the U.S. data, the decision consistency rates for Canadian graduating high school seniors varied markedly across test form and content area. Across all test forms,

values ranged from a low of .72 (Language Arts, Reading Test Form IF) to a high of 1.0 (Social Studies Forms IE and IF).

The false positive rates given in Table 7.4 reflect the probability of an examinee incorrectly meeting the minimum score on the test form, given that his or her true score is below the criterion. Conversely, the false negative rates indicate the probability that an examinee will not meet the minimum score on the test form, given that his or her true score is above the criterion. In both cases, values closer to zero are preferable.

TABLE 7.4
PROBABILITY OF CORRECT CLASSIFICATION, FALSE POSITIVE, AND FALSE NEGATIVE RATES FOR THE 2002 SERIES ENGLISH-LANGUAGE GED TESTS: CANADIAN GRADUATING HIGH SCHOOL SENIOR DATA

TEST/FORM	N	Percent Not Meeting Minimum Score	Percent Meeting Minimum Score	Probability of Correct Classification	False Positive	False Negative
Language Arts, Writing						
Form ID	122	19	81	.81	0	.19
Form IE	750	20	80	.80	0	.20
Form IF	888	16	84	.84	0	.16
Form IG	833	26	74	.74	0	.26
Form IH	458	30	70	.73	.27	†
Social Studies						
Form ID	1,251	33	67	.96	.04	†
Form IE	932	32	68	1.00	*	†
Form IF	908	33	67	1.00	*	0
Form IG	544	31	69	.98	.02	†
Form IH	638	34	66	.99	.01	†
Science						
Form ID	596	18	82	.80	.20	†
Form IE	683	15	85	.80	.20	†
Form IF	675	17	83	.97	.03	†
Form IG	565	17	83	.84	.16	†
Form IH	550	21	79	.85	.15	†
Language Arts, Reading						
Form ID	637	22	78	.78	0	.22
Form IE	678	19	81	.81	0	.19
Form IF	671	18	82	.72	.28	†
Form IG	645	16	84	.84	0	.16
Form IH	557	18	82	.82	0	.18
Mathematics						
Form ID	638	22	78	.73	.27	†
Form IE	661	30	70	.98	.02	†
Form IF	665	27	73	.94	.06	†
Form IG	544	26	74	.94	.06	0
Form IH	532	23	77	.75	.25	†

* Value is less than 0.01.
† Value is less than 0.001.
Note: Data unavailable for Forms IA and IC.

VALIDITY

Readers should refer to Chapter 5 for background on the definition of validity. The English-language Canadian GED Tests were subjected to many of the same validity analyses as the English-language U.S. GED Tests. Again, the validation of the English-language Canadian GED test scores must be made with respect to its purpose: to measure major academic skills and knowledge in core content areas that are learned during four years of high school. Therefore, analyses must be undertaken to demonstrate that the GED test scores can be used to evaluate whether an examinee has attained the knowledge and skills associated with the completion of a normal high school academic program of study.

The sources of validity evidence presented in this chapter report the following: the extent to which the content of the GED Tests represents standard high school curricula, the degree to which the test items conform to the construct being measured, and the relationship of the test scores to other external variables.

It should be noted that no additional validity evidence is provided regarding Part II of the Language Arts, Writing Test. The essay prompts used for the English-language Canadian version are the same as those used for the English-language U.S. GED Tests. The scoring of the English-language Canadian version occurs at the same locations as the English-language U.S. tests (details on the Spanish- and French-language GED Tests are provided in Chapters 8 and 9). Thus, the validity evidence provided in Chapter 5 regarding the writing essays is also applicable to the English-language Canadian GED Tests.

Evidence Based on Test Content

The content of the English-language Canadian GED Tests is based on the same set of specifications as those for the English-language U.S. GED Tests. The only exception is found within the Social Studies Test, in which differences occur within the government and civics and the history sections. Thus, the evidence based on test content and specifications given in Chapter 5 is applicable to the English-language Canadian GED Tests. The development of the Canadian government and civics and the history sections was similar to that of the other sections, except that Canadian teachers, test specialists, and curriculum experts provided the relevant test development guidance.

Evidence Based on Internal Structure

To assess the dimensionality of the English-language Canadian GED Tests, nonlinear exploratory factor analysis (NFA) was performed in a manner similar to that for the English-language U.S. GED Tests. NFA was performed on Forms ID through IH for each test using Canadian graduating high school seniors.

Table 7.5 shows the results of the nonlinear factor analyses for Forms ID through IH given to Canadian graduating high school seniors. The results indicate that a single dominant factor underlies each of the test forms. Across all test forms, the proportion of common (i.e., shared) variance accounted for ranged from 0.36 to 0.50. Although not shown in the table, initial values for the first extracted eigenvalues (which indicate the amount of variance explained by the factor) ranged from 17.0 to 25.7, and second initial eigenvalues ranged from 1.4 to 2.3. Moreover, each of the items loaded heavily and mainly onto the first factor. Factor loadings on subsequent factors were minimal. Finally, correlations among the extracted factors were all positive, with the exception of Social Studies Form IH and Mathematics Form IG.

TABLE 7.5
NUMBER OF SALIENT FACTORS, PROPORTION OF VARIANCE ACCOUNTED FOR BY INITIAL FACTOR, AND SAMPLE SIZE: CANADIAN GED EXAMINEES

TEST/FORM	Number of Salient Factors	Proportion of Variance	N
Language Arts, Writing			
Form ID	1	---	122
Form IE	1	.37	750
Form IF	1	.37	888
Form IG	1	.39	833
Form IH	1	.37	458
Social Studies			
Form ID	1	.49	1,251
Form IE	1	.43	932
Form IF	1	.44	908
Form IG	1	.41	544
Form IH	2	.50	638
Science			
Form ID	1	.41	596
Form IE	1	.43	683
Form IF	1	.40	675
Form IG	1	.49	565
Form IH	1	.43	550
Language Arts, Reading			
Form ID	1	.41	637
Form IE	1	.40	678
Form IF	1	.46	671
Form IG	1	.38	645
Form IH	1	.49	557
Mathematics			
Form ID	1	.36	638
Form IE	1	.49	661
Form IF	1	.45	665
Form IG	3	.57	544
Form IH	1	.43	532

Note: The sample size for Language Arts, Writing Form ID was insufficient for this analysis. Proportion of variance explained within Social Studies Form IH and Mathematics Form IG refers to the number of factors in the best-fitting model. Data unavailable for Forms IA and IC.

Additional analyses were performed on Social Studies Form IH and Mathematics Form IG to further examine the dimensionality. In the case of Social Studies Form IH, it was determined that a two-factor model fit the data best (compared to a three-factor and a single-factor model). About half the items loaded onto each factor. The correlation between the two factors was .75, suggesting the two factors are measuring very similar constructs. For Form IG of the Mathematics Test, a three-factor model fit best (compared to a four-factor and two-factor model). However, the correlations among the three factors in the three-factor model were all high (.65, .72, and .74) and positive, suggesting somewhat substantial overlap between the factors. These findings suggest that the test data for these forms may not be essentially unidimensional. These forms will continue to be investigated in the future.

Evidence Based on Relations with Other Variables

The validity of English-language Canadian GED test scores was also assessed by comparing the performance of Canadian graduating high school seniors on the English-language Canadian GED Tests with other measures of academic proficiency.

The Relationship Between GED Test Scores and High School Grades

Because the GED Tests are designed to measure academic knowledge and skills that are taught in a regular high school program of study, it is important that they demonstrate a positive relationship to other measures of high school–level academic performance. To investigate this relationship, the self-reported grades of Canadian graduating high school seniors participating in the standardization and norming study and equating studies were collected and compared with the performance of these same seniors on the English-language Canadian GED Tests. Students were asked to list the overall grades they received since ninth grade through the current term for five content areas: English literature, English composition, social studies, science, and mathematics.

The correlations between self-reported grades and English-language Canadian GED test scores are reported in Table 7.6. The correlations reported in Table 7.6 vary across both year and subject area, which may be because higher correlations would be expected for those tests that represent a greater proportion of content taught in the high school curriculum. Because the letter grades were self-reported, the correlations may be somewhat lower than might be found if official letter grades from the school had been used.

TABLE 7.6
CORRELATIONS OF CANADIAN GRADUATING HIGH SCHOOL SENIORS' ENGLISH-LANGUAGE GED TEST STANDARD SCORE WITH SELF-REPORTED LETTER GRADES IN THE SAME CONTENT AREA: 2002 STANDARDIZATION AND 2003 AND 2004 EQUATING STUDIES

	2002		2003		2004	
	N	*r*	N	*r*	N	*r*
Language Arts, Writing	122	.51[a]	1,589	.48[a]	1,220	.47[a]
Language Arts, Writing	122	.41[b]	1,511	.48[b]	1,187	.47[b]
Social Studies	1,251	.38	1,777	.41	1,123	.45
Science	595	.38	1,303	.39	1,064	.42
Language Arts, Reading	637	.39[a]	1,283	.40[a]	1,107	.41[a]
Language Arts, Reading	637	.36[b]	1,242	.38[b]	1,121	.41[b]

Note: All correlations were significant at $p < .001$. Letter grades are reported as Mostly A, Mostly B, Mostly C, Mostly D, and Mostly Below D. To compute the correlations, letter grades were recoded as Mostly A=4, Mostly B=3, Mostly C=2, Mostly D=1, Mostly Below D=0.
[a] Correlation with self-reported grades in English literature.
[b] Correlation with self-reported grades in English composition.

The grades reported by the graduating high school seniors were also compared with their performance at selected values along the GED standard score scale. This analysis helps identify the approximate GPA or letter-grade levels that correspond to levels of performance on the GED Tests. Table 7.7 presents, for each English-language Canadian GED Test, the percentages of soon-to-be graduating seniors meeting selected GED score standards for each letter grade. For example, the first row of Table 7.7 indicates that 100 percent of the seniors whose reported grades were "Mostly A" scored at or above a GED standard score of 350 on the Language Arts, Writing Test. The second row of the table shows that 82 percent of the "Mostly B" seniors achieved a score of at least 500.

Overall, the results reported in Table 7.7 illustrate that the higher the high school grade, the higher the GED score, and therefore, the greater the likelihood of earning the minimum score required for the particular GED Test. The results presented in Table 7.7 indicate that the passing standards established on the English-language Canadian GED Tests do discriminate between higher and lower achieving high school students. Therefore, the results support both the validity of the English-language Canadian GED test scores, and the validity of the GED standard setting procedure.

TABLE 7.7
PERCENTAGE OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE STANDARDIZATION STUDY AT SELF-REPORTED GRADE LEVELS ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
	N	350	410	450	500
		Language Arts, Writing Test			
Self-reported Grades in English Literature					
Mostly A	24	100	100	100	91
Mostly B	49	100	100	94	82
Mostly C	35	100	94	66	37
Mostly D	12	†	†	†	†
Mostly Below D	2	†	†	†	†
		Language Arts, Writing Test			
Self-reported Grades in English Composition					
Mostly A	24	100	100	100	92
Mostly B	50	100	98	90	74
Mostly C	34	100	94	65	44
Mostly D	12	†	†	†	†
Mostly Below D	2	†	†	†	†
		Social Studies Test			
Self-reported Grades in Social Studies					
Mostly A	238	97	95	89	79
Mostly B	534	94	86	72	52
Mostly C	335	90	73	56	37
Mostly D	129	86	61	40	24
Mostly Below D	15	80	67	47	20
		Science Test			
Self-reported Grades in Science					
Mostly A	87	100	97	92	84
Mostly B	212	98	92	87	77
Mostly C	206	99	87	78	63
Mostly D	75	96	81	71	43
Mostly Below D	15	93	73	60	27
		Language Arts, Reading Test			
Self-reported Grades in English Literature					
Mostly A	92	100	99	98	95
Mostly B	268	98	92	85	71
Mostly C	191	97	82	70	50
Mostly D	73	97	71	51	32
Mostly Below D	13	†	†	†	†
		Language Arts, Reading Test			
Self-reported Grades in English Composition					
Mostly A	97	100	98	96	92
Mostly B	283	98	91	83	69
Mostly C	180	97	82	73	51
Mostly D	66	97	71	47	32
Mostly Below D	11	†	†	†	†
		Mathematics Test			
Self-reported Grades in Mathematics					
Mostly A	126	99	98	97	94
Mostly B	232	99	94	83	69
Mostly C	176	98	87	72	53
Mostly D	89	97	76	56	31
Mostly Below D	15	100	67	60	33

† Indicates that the statistic was not calculated because of small sample size.

The Relationship Between GED Test Scores and Prior Instruction

If the English-language Canadian GED Tests are accurate measures of subjects taught in a regular program of Canadian high school study, then a positive relationship should be observed between scores on content area GED Tests and the amount of instruction received by students in the content area related to each test.

The Canadian graduating high school seniors participating in the 2002, 2003, and 2004 studies were asked to indicate the number of years of English literature, English composition, social studies, science, and mathematics courses they had taken from ninth grade to the current term. The students were asked to indicate whether they had taken one year or less, two, three, or four years or more of coursework in each content area. In addition, they were also asked to specify the types of courses they had taken in each content area. For example, for social studies, the students were asked to indicate whether they had taken behavioral sciences, civics, economics, geography, political science, national history, or world history.

Table 7.8 contains the percentage of Canadian graduating high school seniors in the 2002 study at self-reported total years of study by various minimum standard scores. As expected, the percentages generally decrease across each row as the standard score increases. Additionally, as the number of self-reported total years of study increases, the percentage of seniors generally increases within any given standard score category. For example, 85 percent of graduating seniors with one year or less of social studies scored at least 350 on the Social Studies Test. However, a larger percentage (92 percent) of those seniors with at least four years of social studies scored at least 350 on the Social Studies Test.

TABLE 7.8
PERCENTAGE OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE STANDARDIZATION STUDY AT SELF-REPORTED TOTAL YEARS OF STUDY ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
	N	350	410	450	500
		Language Arts, Writing Test			
Self-reported Total Years of Study in English Literature					
1 year or less	9	†	†	†	†
2 years	27	7	7	7	7
3 years	117	19	19	17	11
4 years or more	478	21	20	16	13
		Language Arts, Writing Test			
Self-reported Total Years of Study in English Composition					
1 year or less	15	33	33	33	33
2 years	35	9	9	9	6
3 years	107	20	20	16	11
4 years or more	448	19	19	15	12
		Social Studies Test			
Self-reported Total Years of Study in Social Studies					
1 year or less	20	85	60	45	25
2 years	110	86	73	55	34
3 years	509	96	85	70	50
4 years or more	601	92	80	68	54
		Science Test			
Self-reported Total Years of Study in Science					
1 year or less	4	†	†	†	†
2 years	31	94	84	71	48
3 years	152	99	84	72	52
4 years or more	393	98	93	88	76
		Language Arts, Reading Test			
Self-reported Total Years of Study in English Literature					
1 year or less	10	†	†	†	†
2 years	17	100	94	88	59
3 years	123	96	85	70	58
4 years or more	474	98	88	80	65
		Language Arts, Reading Test			
Self-reported Total Years of Study in English Composition					
1 year or less	16	100	88	75	63
2 years	27	100	93	78	59
3 years	119	97	86	71	56
4 years or more	442	98	88	81	66

Continued on next page

Table 7.8 continued

	N	GED Standard Score ≥ 350	410	450	500
		Mathematics Test			
Self-reported Total Years of Study in Mathematics					
1 year or less	1	†	†	†	†
2 years	14	†	†	†	†
3 years	118	100	87	75	59
4 years or more	493	98	92	81	67

† Indicates that the statistic was not calculated because of small sample size.

In Table 7.9, average GED standard scores are reported for the English-language Canadian GED Tests. These results were derived from the 2002 Canadian standardization and norming study. The average standard scores are broken down according to the four levels of amount of prior instruction (from one year or less to four years or more). For example, those seniors with three years of English literature instruction obtained an average GED standard score of 536 on the Language Arts, Reading Test, while those with four years or more achieved an average score of 551.

TABLE 7.9
AVERAGE GED STANDARD SCORES OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE STANDARDIZATION STUDY, BY YEARS OF INSTRUCTION IN CONTENT AREA

YEARS INSTRUCTION IN SUBJECT AREA	Language Arts, Writing	Social Studies	Science	Language Arts, Reading	Mathematics
1 year or less	†	†	†	†	†
	(5)	(20)	(4)	(10)	(1)
2 years	†	465	†	†	†
	(3)	(110)	(31)	(17)	(14)
3 years	†	499	512	536	529
	(21)	(509)	(152)	(123)	(118)
4 years or more	531	496	556	551	534
	(87)	(601)	(393)	(474)	(493)

† Indicates that the statistic was not calculated because of small sample size.
Note: Numbers in parentheses refer to the number of seniors. Averages for the Language Arts, Writing Test are based on the numbers of years of instruction in English composition; averages for the Language Arts, Reading Test are based on the number of years of instruction in English literature.

Tables 7.10 through 7.14 provide the average standard scores by specific courses of study for the 2002 Canadian standardization and norming study. Here, the expectation is that those graduating high school seniors who have taken related courses should score higher than those who have not taken these courses. Although the majority of the average standard scores follow this pattern, several do not. The most dramatic difference, for example, occurs between those who have and have not taken earth science. High school seniors who have not taken this course scored nearly 30 standard score points higher than those who have taken this course.

Tables 7.10 through 7.14 also provide the percentages of Canadian graduating high school seniors at various GED standard score levels by specific instructional courses.

TABLE 7.10
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE STANDARDIZATION STUDY SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, WRITING TEST, BY INSTRUCTION IN GRAMMAR AND LANGUAGE COURSES

COURSE		Mean	N	GED Standard Score ≥ 350	410	450	500
Grammar/Composition	Taken	544	105	100	98	84	68
	Not Taken	484	17	100	88	65	41
French	Taken	541	91	100	96	82	69
	Not Taken	519	31	100	100	77	48

Note: Sample sizes for Spanish, German, and Latin were too small for reporting.

TABLE 7.11
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE STANDARDIZATION STUDY SCORING AT OR ABOVE STANDARD SCORES ON SOCIAL STUDIES TEST, BY INSTRUCTION IN SELECTED SOCIAL STUDIES COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Behavioral Science	Taken	493	1,072	96	83	73	54
	Not Taken	498	179	92	88	67	49
Civics	Taken	505	152	93	86	73	55
	Not Taken	492	1,099	93	81	67	49
Economics	Taken	504	462	94	84	72	55
	Not Taken	488	789	92	80	65	47
Geography	Taken	493	782	92	81	68	50
	Not Taken	495	469	94	81	67	49
Political Science	Taken	516	265	94	86	77	62
	Not Taken	488	986	92	80	65	47
History	Taken	496	964	93	81	69	51
	Not Taken	486	287	92	82	63	45
World History	Taken	502	775	94	82	71	55
	Not Taken	480	476	91	80	62	42

TABLE 7.12
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE STANDARDIZATION STUDY SCORING AT OR ABOVE STANDARD SCORES ON SCIENCE TEST, BY INSTRUCTION IN SELECTED SCIENCE COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Biology	Taken	547	472	99	93	85	72
	Not Taken	504	124	96	76	69	52
Chemistry	Taken	562	387	99	95	88	78
	Not Taken	493	209	96	79	71	49
Earth Science	Taken	519	183	98	87	79	59
	Not Taken	547	413	98	90	83	72
General Science	Taken	540	368	98	89	83	70
	Not Taken	534	228	99	90	80	65
Genetics	Taken	554	113	99	96	91	76
	Not Taken	534	483	98	88	80	66
Physical Science	Taken	533	122	96	88	83	71
	Not Taken	539	474	99	90	82	67
Physics	Taken	570	322	99	95	90	81
	Not Taken	501	274	97	82	73	53
Zoology/Botany	Taken	543	47	96	91	91	72
	Not Taken	538	549	98	89	81	67

TABLE 7.13

MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE STANDARDIZATION STUDY SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, READING TEST, BY INSTRUCTION IN SELECTED ENGLISH AND LANGUAGE COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Literature	Taken	551	584	98	87	78	64
	Not Taken	504	53	98	87	74	47
European Literature	Taken	567	150	98	91	85	72
	Not Taken	541	487	98	86	76	60
World Literature	Taken	565	143	99	90	84	72
	Not Taken	542	494	97	86	76	60
Spanish	Taken	593	18	100	100	100	78
	Not Taken	546	619	98	87	77	62
French	Taken	554	457	98	89	81	67
	Not Taken	528	180	97	82	71	53

Note: Sample sizes for German and Latin were too small for reporting.

TABLE 7.14

MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE STANDARDIZATION STUDY SCORING AT OR ABOVE STANDARD SCORES ON MATHEMATICS TEST, BY INSTRUCTION IN SELECTED MATHEMATICS COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Algebra I	Taken	538	502	99	91	82	68
	Not Taken	496	136	96	85	64	47
Algebra II	Taken	549	426	10	94	8	73
	Not Taken	489	212	96	81	61	45
Business Math	Taken	532	150	97	91	78	63
	Not Taken	529	488	99	89	78	64
Calculus	Taken	583	209	100	99	97	88
	Not Taken	503	429	98	85	69	52
General Math	Taken	529	427	98	89	77	63
	Not Taken	539	211	99	91	82	65
Geometry	Taken	540	449	99	93	82	68
	Not Taken	504	189	97	81	69	52
Trigonometry	Taken	545	458	100	95	85	71
	Not Taken	489	180	96	76	62	45

Chapter 8: French-Language GED Tests

OVERVIEW

Chapter 1 provided a brief history of the GED testing program. Chapters 2 through 5 presented technical information pertaining to the development, norming, scaling, equating, reliability, and validity of the English-language U.S. GED Tests. This chapter describes the aspects of the GED Tests that are specific to the French-language GED Tests.

The purpose of the French-language GED Tests is to provide an opportunity to adults who have French as their primary language to certify their attainment of high school–level academic knowledge and skills and earn their jurisdiction's high school equivalency credential, diploma, or certificate. The development of the French-language GED Tests is similar to the development of the English-language Canadian GED Tests. However, because these tests are intended to serve different populations of GED examinees, they are normed on French-speaking Canadian graduating high school seniors. A description of the history, development, reliability, and validity of these tests is provided in the next sections.

HISTORY OF THE FRENCH-LANGUAGE GED TESTS

The province of New Brunswick was the first to request a French-language version of the GED Tests to serve its population of French-speaking adults. This request led to the development of the French-language GED Tests that were introduced in New Brunswick in 1974. Currently, the French-language GED Tests are administered throughout the provinces and territories that participate in the GED testing program. They are also administered in some parts of the United States when requested by jurisdictions. The vast majority of French-language GED testing in Canada occurs in New Brunswick. Yet, the majority of all Canadian GED examinees take the English-language Canadian version.

The development of the French-language GED Tests followed that of the English-language Canadian GED Tests. A brief history of the English-language Canadian GED Tests is provided in Chapter 7. Forms of the 2002 Series French-Language GED Tests were first administered to adult examinees in 2004. In 2007, there were approximately 800 GED examinees who were administered the majority of their tests via the French-language version (GEDTS, 2008; examinees may be allowed to take the five content area tests in different languages, depending on the jurisdiction's policy).

TEST SPECIFICATIONS AND DEVELOPMENT

As stated in Chapter 2, the French-language GED Tests have specifications similar to the English-language Canadian GED Tests. The French-language version of the Social Studies; Science; Language Arts, Reading; and Mathematics Tests are direct translations of the respective English-language Canadian versions. The content and cognitive specifications for these tests are identical to the English-language Canadian versions. The Language Arts, Writing Test was developed independently by the Quebec Ministry of Education and has different content and cognitive test specifications. The Language Arts, Writing Test comprises Spelling and Grammar (50 percent), Syntax and Punctuation (35 percent), and Organization of Text and Ideas (15 percent). The testing times for each of the French-language GED Tests are listed in Chapter 1.

Following the specifications for the Canadian GED Tests, the French-language versions use International System of Units (SI units) throughout the test (e.g., mètres, litres, grammes), whereas in the U.S. version, Imperial units are primarily used (e.g., feet, gallons, pounds). Similarly, spaces, rather than commas, are used in the French-language versions to denote triads of digits in long numbers and decimals. For example,

in U.S. versions, the number *twenty-one thousand* would be displayed as 21,000, while in the French-language versions, it would be displayed as 21 000.

Because the content of the French-language GED Tests is essentially identical to that of most of the English-language Canadian GED Tests, the test development process for the French-language GED Tests was identical to that of their Canadian counterparts. Thus, the test development procedures described in detail in Chapter 2 also apply to the development of the French-language GED Tests. Once the English-language Canadian Social Studies, Science, and Mathematics Test forms were developed (i.e., after the items passed all content, measurement, and statistical reviews), they were directly translated into Canadian French, reviewed, and evaluated by independent consultants. As mentioned above, the Language Arts, Writing Test was developed independently by the Quebec Ministry of Education.

STANDARDIZATION AND NORMING

The French-language GED Tests were normed on samples of French-speaking graduating high school seniors across Canada who took the GED Tests during March, April, May, September, and October of 2003. A school was eligible for the French standardization and norming study if it offered courses for senior or 12th grade students, enrolled students who were fluent in French, graduated a senior class, and awarded high school diplomas. A student in his or her last year of schooling was eligible if he or she expected to receive a high school diploma by the end of his or her schooling, spoke French fluently, and did not require testing accommodations (such as large print, extended time, etc.). The number of schools participating in this study and the equating study in 2005 is presented in Table 8.1.

TABLE 8.1
NUMBER OF SCHOOLS PARTICIPATING IN THE 2003 CANADIAN STANDARDIZATION AND NORMING AND 2005 EQUATING STUDIES FOR THE FRENCH-LANGUAGE GED TESTS

	2003		2005	
	N	%	N	%
Alberta	6	8.2	1	2.9
British Columbia	5	6.8	3	8.8
Manitoba	5	6.8	3	8.8
New Brunswick	24	32.9	10	29.4
Newfoundland and Labrador	5	6.8	4	11.8
Northwest Territories	0	0.0	0	0.0
Nova Scotia	3	4.1	2	5.9
Ontario	11	15.1	2	5.9
Prince Edward Island	7	9.6	4	11.8
Quebec	6	8.2	3	8.8
Saskatchewan	1	1.4	2	5.9
Yukon	0	0.0	0	0.0
Total Schools	73		34	
Total Students*	525		439	

*Because of concerns regarding fluency in French, only records of students who indicated they were equally fluent in French and English or more fluent in French were analyzed.

SCALING AND EQUATING

As with the English-language U.S. and Canadian GED Tests, the raw scores from the French-language GED Tests were converted to a scale ranging from 200 to 800, with a mean of 500 and standard deviation of 100. The raw-to-standard score conversions for the Social Studies, Science, and Mathematics Tests are the same as the English-language Canadian raw-to-standard score conversions for those tests.

In 2003, data were collected via two French-language forms, namely IA and IC. During this study, all content area tests were administered to French-speaking Canadian graduating high school seniors. The intention was to scale, norm, and equate the Language Arts, Writing and Language Arts, Reading Tests data using the same procedures outlined in Chapter 3; however, due to insufficient sample sizes for those tests, the English-language Canadian raw-to-standard score conversions for those tests had to be used until a 2005 equating study was performed. The norms for the remaining three tests were also obtained, and the conversion tables from the English-language Canadian GED Tests were used to obtain standard scores.

In the 2005 equating study, Forms ID, IE, Language Arts, Writing Form IA and Language Arts, Reading Form IC were administered to French-speaking graduating high school seniors in Canada. The Language Arts, Writing and Language Arts, Reading Tests data were scaled, normed, and equated using the same procedures outlined in Chapter 3. The norms for the remaining three tests were also obtained, and the conversion tables from the English-language Canadian GED Tests were used to obtain standard scores.

RELIABILITY

The reliability of the French-language GED test scores was analyzed using the same methods that were applied to the English-language U.S. and Canadian GED Tests. These methods are described thoroughly in Chapter 4. The reliability of the scores from the multiple-choice portions of the French-language Canadian GED Tests was evaluated by calculating the K-R 20 reliability coefficient (Kuder & Richardson, 1937), the standard error of measurement (SEM), and decision consistency. The reliability of the essay portion of the Language Arts, Writing Test was evaluated using additional criteria discussed below.

The results of the reliability analyses for the 2002 Series French-Language GED Tests are presented in this chapter. The French-language Canadian data presented herein are from Forms IA, IC, ID, and IE, which correspond with the French Canadian standardization and norming study performed in 2003 and the subsequent equating study in 2005. All studies used a random sample of French-speaking graduating high school seniors from across Canada, as described above.

K-R 20 and SEM Results

Table 8.2 presents the score means, standard deviations, SEM, and K-R 20 estimates for the test forms in the 2002 Series French-Language GED Tests. It should be noted that the numbers in Table 8.2 for the Language Arts, Writing Test refer only to the multiple-choice portion of the test. The results presented in Table 8.2 are reported in both standard and raw score units. Because the transformation of raw scores to standard scores (described in Chapter 3) is nonlinear, it is not possible to compute K-R 20 directly for standard scores. Thus, K-R 20 estimates are for raw scores only.

The information in Table 8.2 is based on the performance of the sample of French-speaking graduating high school seniors across Canada who took the GED Tests as part of the standardization and equating studies in years 2003 and 2005, respectively. Data from Form IA and IC originate from the 2003 standardization and data from Forms ID and IE originate from the 2005 equating study. The results presented in Table 8.2 indicate that all French-language test forms have a K-R 20 of .83 or higher, and all but five French-language test forms have K-R 20s of at least .90.

TABLE 8.2
SAMPLE SIZES (N), SCORE MEANS, STANDARD DEVIATIONS (SD), STANDARD ERRORS OF MEASUREMENT (SEM), AND K-R 20 ESTIMATES FOR THE 2002 SERIES FRENCH-LANGUAGE GED TESTS: CANADIAN GRADUATING HIGH SCHOOL SENIOR DATA

		Standard Scores			Raw Scores			
TEST/FORM	N	Mean	SD	SEM	Mean	SD	SEM	K-R 20
Language Arts, Writing								
Form IA	96	557.3	111.9	27.4	29.6	12.3	2.9	.94
Form IC	71	549.2	94.4	26.7	29.5	10.3	2.9	.92
Form ID	166	498.5	101.7	39.4	23.9	7.9	3.1	.85
Form IE	155	489.2	99.9	36.0	22.2	8.8	3.1	.87
Social Studies								
Form IA	38	466.1	97.9	25.9	31.6	11.0	2.9	.93
Form IC	27	449.3	90.3	23.9	28.2	11.3	3.0	.93
Form ID	60	404.0	91.2	24.1	24.1	11.3	3.1	.93
Form IE	53	420.8	96.6	29.0	27.9	10.1	3.0	.91
Science								
Form IA	34	490.3	104.8	23.4	33.9	12.2	2.8	.95
Form IC	26	495.4	85.3	25.6	35.7	9.2	2.8	.91
Form ID	60	496.7	114.1	30.2	32.9	10.8	2.8	.93
Form IE	56	486.8	102.1	25.0	33.3	11.1	2.8	.94

Continued on next page

Table 8.2 continued

TEST/FORM	N	Standard Scores			Raw Scores			
		Mean	SD	SEM	Mean	SD	SEM	K-R 20
Language Arts, Reading								
Form IA	36	493.9	135.8	56.0	17.6	6.8	2.8	.83
Form IC	121	515.1	125.8	39.8	20.9	8.4	2.7	.90
Form ID	98	467.1	116.1	46.4	16.5	7.0	2.8	.84
Form IE	99	476.5	110.5	42.8	16.9	7.2	2.8	.85
Mathematics								
Form IA	76	524.5	82.0	24.6	36.9	9.0	2.6	.91
Form IC	59	519.8	82.3	24.7	37.1	9.1	2.7	.91
Form ID	59	565.1	95.3	30.1	41.3	7.6	2.4	.90
Form IE	56	549.1	104.5	25.6	37.9	10.3	2.6	.94

Conditional Standard Errors of Measurement

The conditional standard errors of measurement (CSEM) were calculated for various standard scores using the same methods applied to the English-language U.S. GED Tests (see Chapter 4). The passing standard requirement for the French-language GED Tests is dependent upon the jurisdiction (see Appendix B). The estimated standard score CSEM for the French-language GED Tests are presented in Table 8.3.

TABLE 8.3
STANDARD SCORE CONDITIONAL STANDARD ERRORS OF MEASUREMENT AT VARIOUS STANDARD SCORES FOR THE 2002 SERIES FRENCH-LANGUAGE GED TESTS: CANADIAN GRADUATING HIGH SCHOOL SENIOR DATA

TEST/FORM	Standard Score						
	400	410	420	430	440	450	460
Social Studies							
Form IA	25.4	25.4	25.4	25.2	25.1	24.9	28.5
Form IC	25.1	25.2	25.4	25.4	25.3	25.2	24.9
Form ID	20.9	25.2	25.1	25.0	24.8	24.4	28.2
Form IE	33.4	25.0	24.9	24.7	24.5	28.1	23.7
Science							
Form IA	21.3	21.4	17.1	17.0	21.0	20.5	20.3
Form IC	20.9	16.7	20.8	20.6	20.5	20.1	19.8
Form ID	25.4	21.1	16.9	21.1	16.8	20.9	20.7
Form IE	24.4	20.4	16.3	20.4	20.2	20.0	15.9
Language Arts, Reading							
Form IA	75.8	75.8	75.8	70.4	70.4	64.5	64.5
Form IC	67.1	73.9	68.6	68.6	62.9	56.8	56.8
Form ID	67.9	81.5	76.7	76.7	71.3	71.3	65.3
Form IE	64.3	81.2	76.4	76.4	71.0	71.0	65.1
Mathematics							
Form IA	24.9	25.2	25.2	25.2	25.2	25.1	24.9
Form IC	21.2	29.8	25.7	25.7	25.5	25.4	25.0
Form ID	17.3	25.9	25.8	25.6	25.2	20.5	20.2
Form IE	24.9	25.1	25.2	25.2	25.1	25.0	24.9

Reliability of Essay Scores on the Language Arts, Writing Test

The reliability of the essay portion of the Language Arts, Writing Test was evaluated by analyzing reader agreement, or inter-rater reliability, and scoring stability. Essay scoring sessions must show evidence of reader agreement and scoring stability. Reader agreement refers to the degree of agreement of scores assigned among different readers scoring the same essays. Inter-rater reliability increases as the number of essays that require attention from the Chief Reader (due to differences in two readers' scores being greater than one point) decreases. Scoring stability refers to how well the scoring sites maintain the scoring standards established by the GEDTS Writing Advisory Committee and presented in the 2002 Series GED Writing Test Official Essay Scoring Guide.

Maintaining scoring consistent with official GED essay scoring standards is essential in an essay scoring session. The standards for scoring GED essays must remain fixed, regardless of when the essay is administered, where it is scored, or what specific procedures are used in the scoring session itself. A high degree of inter-rater reliability does not ensure scoring stability. Just because readers agree with each other on the assignment of essay scores does not necessarily mean they are assigning the scores according to the standards defined in the scoring guide.

To achieve inter-rater reliability and scoring stability, the essay scoring standards are regularly reinforced. As readers score the essays, a Chief Reader selects scored essays at random to verify that the readers' scoring is consistent with the definitions in the scoring guide. In cases of a disagreement in assigned essay scores, the Chief Reader discusses the essay with both readers. The monitoring process continues throughout the entire scoring session. This system of checks and rechecks ensures that readers are scoring according to the standards defined in the scoring guide.

Site Monitoring and Score Scale Stability

To facilitate score scale stability, Chief Reader training, and site certification, site monitoring procedures were incorporated into the essay scoring process. Chief Reader training and site certification are described in Chapter 5, and site monitoring for the French-language GED Tests is described below.

In the past, GED Testing Service has conducted two types of monitoring: random monitoring and systematic monitoring. In random monitoring, a randomly selected set of 40 scored essays from a scoring site is rescored by the GED Testing Service Writing Advisory Committee. In systematic monitoring, a common set of 40 essays, scored by the Writing Advisory Committee, is sent to each essay scoring site where the site's readers rescore the essays. In both types of monitoring, the site is evaluated by determining the congruence of its readers' essay scores to the Writing Advisory Committee's essay scores.

As described in Chapter 4, scoring sites must demonstrate scale score stability, or adherence to the scoring standards established by the Writing Advisory Committee, in order to become a certified essay scoring site. Scoring sites are certified only after they demonstrate a required level of proficiency on several score stability criteria.

Table 8.4 shows the results of the systematic site monitoring of essay scoring sites in 2008 for the French-language essays. The identities of specific sites have not been revealed; instead, sites have been randomly assigned a number between 1 and *s*, where *s* equals the number of sites.

TABLE 8.4
2008 SYSTEMATIC SITE MONITORING RESULTS (FOUR-POINT HOLISTIC SCORING) FOR FRENCH-LANGUAGE GED ESSAYS

		AGREEMENT OF SCORING SITE'S ESSAY SCORES WITH GEDTS WRITING ADVISORY COMMITTEE SCORES			
SITE	Number of Readers	% Scores Equal	% Scores Within One Point	% Scores Differing by > One Point	Correlation*
1	3	74.2	100	0	.86
2	3	85.0	100	0	.94
3	3	78.3	100	0	.93
4	3	58.3	100	0	.81
5	5	67.5	100	0	.92
Mean	3.4	72.7	100	0	.89
Median	3	74.2	100	0	.92
Minimum	3	58.3	100	0	.81
Maximum	5	85.0	100	0	.94

* Pearson correlation between readers' scores and GEDTS Writing Advisory Committee scores.

Essay Score Inter-rater Reliability

The inter-rater reliability was calculated by using the polychoric correlation between readers' scores on the French-language essays. Because the essay raw score on the Language Arts, Writing Test is the average of the two readers' scores, the correlation was adjusted using the Spearman-Brown prophecy formula with a factor of two. The inter-rater reliability associated with the 2005 equating study was estimated as .98 (data were not available for 2003).

Decision Consistency

The decision consistency for each of the five content area tests was examined using data obtained via the French Canadian norming and equating studies (i.e., using high school senior data). The same procedure used with the English-language U.S. GED Tests was also applied to the French-language data (i.e., Livingston and Lewis procedure via the BB-Class software program; see Chapter 4).

The decision consistency (probability of correct classification) estimates for Forms IA, IC, ID, and IE are provided in Table 8.5. The false positive rates given in Table 8.5 reflect the probability of an examinee incorrectly earning the minimum score required on the test form, given that their true score is below the criterion. Conversely, the false negative rates indicate the probability that an examinee will not earn the minimum score required on the test form, given that their true score is above the criterion. In both cases, values closer to zero are preferable.

TABLE 8.5
PROBABILITY OF CORRECT CLASSIFICATION, FALSE POSITIVE, AND FALSE NEGATIVE RATES FOR THE 2002 SERIES FRENCH-LANGUAGE GED TESTS: CANADIAN GRADUATING HIGH SCHOOL SENIOR DATA

TEST/FORM	N	Percent Not Meeting Minimum Score	Percent Meeting Minimum Score	Probability of Correct Classification	False Positive	False Negative
Language Arts, Writing						
Form IA	55	13	87	.88	.12	†
Form IC	53	6	94	.96	.02	.01
Form ID	125	20	80	.80	.00	.20
Form IE	113	22	78	.78	.00	.22
Social Studies						
Form IA	38	21	79	1.00	.00	†
Form IC	27	30	70	.99	.01	.00
Form ID	60	47	53	1.00	†	*
Form IE	53	40	60	.99	.01	†
Science						
Form IA	34	24	76	.87	.13	†
Form IC	26	15	85	.79	.21	†
Form ID	60	20	80	.85	.15	†
Form IE	56	21	79	.90	.10	†
Language Arts, Reading						
Form IA	36	28	72	.99	.01	*
Form IC	121	21	79	.92	.08	†
Form ID	98	28	72	.99	*	*
Form IE	99	29	71	.93	.07	†
Mathematics						
Form IA	76	9	91	.91	.00	.09
Form IC	59	7	93	.93	.00	.07
Form ID	59	5	95	.95	.00	.05
Form IE	56	9	91	.91	.00	.09

* Value is less than 0.01.
† Value is less than 0.001.

VALIDITY

Readers should refer to Chapter 5 for background on the definition of validity. The French-language GED Tests were subjected to many of the same validity analyses as the English-language U.S. GED Tests. Again, the validation of the French-language GED test scores must be made with respect to its purpose: to measure major academic skills and knowledge in core content areas that are learned during four years of high school. Therefore, analyses must be undertaken to demonstrate that the GED test scores can be used to evaluate whether an examinee has attained the knowledge and skills associated with the completion of a normal high school academic program of study.

The sources of validity evidence presented in this chapter report: the extent to which the content of the GED Tests represents standard high school curricula and the relationship of the test scores to other external variables.

Evidence Based on Test Content

The content of the French-language Social Studies; Science; Language Arts, Reading; and Mathematic Tests is based on the same set of specifications as those for the English-language Canadian GED Tests. Thus, the evidence based on test content and specifications given in Chapters 5 and 7 for these three tests is applicable to the French-language versions as well. As mentioned above and in Chapter 2, the Language Arts, Writing Test was developed independently by the Quebec Ministry of Education and has different content and cognitive test specifications.

Evidence Based on Relations with Other Variables

The validity of the French-language GED test scores was also assessed by comparing the performance of French-speaking Canadian high school seniors on the GED Tests with other measures of academic proficiency.

The Relationship Between GED Test Scores and High School Grades

Because the GED Tests are designed to measure academic knowledge and skills that are taught in a regular high school program of study, it is important that they demonstrate a positive relationship with other measures of high school–level academic performance. To investigate this relationship, the self-reported grades of French-speaking Canadian graduating high school seniors participating in the standardization and norming study and equating study were collected and compared with the performance of these same seniors on the French-language GED Tests. Students were asked to list the overall grades they received since ninth grade through the current term for five content areas: English literature, English composition, social studies, science, and mathematics. The correlations between self-reported grades and French-language GED test scores are reported in Table 8.6.

TABLE 8.6
CORRELATIONS OF CANADIAN GRADUATING HIGH SCHOOL SENIORS' FRENCH-LANGUAGE GED TEST STANDARD SCORES WITH SELF-REPORTED LETTER GRADES IN THE SAME CONTENT AREA: 2003 AND 2005 STUDIES COMBINED

TEST	N	*r*
Language Arts, Writing	320	.17 [a]
Language Arts, Writing	326	.22 [b]
Social Studies	166	.22
Science	164	.34
Language Arts, Reading	324	.25 [a]
Language Arts, Reading	326	.23 [b]
Mathematics	225	.41

Note: Data are from Forms IA and IC (collected in 2003) and ID and IE (collected in 2005). All correlations were significant at $p < .01$. Letter grades are reported as Mostly A, Mostly B, Mostly C, Mostly D, and Mostly Below D. To compute the correlations, letter grades were recoded as Mostly A=4, Mostly B=3, Mostly C=2, Mostly D=1, Mostly Below D=0.
[a] Correlation with self-reported grades in French literature.
[b] Correlation with self-reported grades in French composition.

The grades reported by the graduating high school seniors were also compared with performance at selected values along the GED standard score scale. This analysis helps identify the approximate GPA or letter-grade levels that correspond to levels of performance on the GED Tests. Table 8.7 presents, for each French-language GED Test, the percentages of soon-to-be graduating French-speaking Canadian seniors meeting selected GED score standards for each letter grade. For example, the first row of Table 8.7 indicates that 100 percent of the seniors whose reported grades were "Mostly A" scored at or above a GED standard score of 350 on the Language Arts, Writing Test. The second row of the table shows that 57 percent of the "Mostly B" seniors achieved a score of at least 500.

Overall, Table 8.7 illustrates that the higher the high school grade, the higher the GED score, and therefore, the greater the likelihood of meeting the minimum score requirements for the particular GED Test. These results indicate that the passing standards established on the GED Tests do discriminate between higher and lower achieving French-speaking Canadian high school students. Therefore, the results support both the validity of the GED test score interpretations, and the validity of the GED standard setting procedure.

TABLE 8.7
PERCENTAGE OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 FRENCH-LANGUAGE GED TESTS STUDIES AT SELF-REPORTED GRADE LEVELS ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

SELF-REPORTED GRADES	N	GED Standard Score ≥ 350	410	450	500
		Language Arts, Writing Test			
French Literature					
Mostly A	64	100	92	80	70
Mostly B	149	97	83	70	57
Mostly C	95	100	80	62	42
Mostly D	11	†	†	†	†
Mostly Below D	1	†	†	†	†
		Language Arts, Writing Test			
French Composition					
Mostly A	62	100	94	84	74
Mostly B	152	97	84	70	56
Mostly C	99	100	79	61	41
Mostly D	12	†	†	†	†
Mostly Below D	1	†	†	†	†
		Social Studies Test			
Social Studies					
Mostly A	39	90	74	56	38
Mostly B	76	78	66	54	22
Mostly C	39	62	46	33	15
Mostly D	12	†	†	†	†
Mostly Below D	0	†	†	†	†
		Science Test			
Science					
Mostly A	36	94	86	83	75
Mostly B	77	91	84	78	57
Mostly C	44	91	68	48	32
Mostly D	5	†	†	†	†
Mostly Below D	2	†	†	†	†
		Language Arts, Reading Test			
French Literature					
Mostly A	65	91	86	78	66
Mostly B	143	90	77	70	55
Mostly C	96	83	70	63	50
Mostly D	17	82	53	29	12
Mostly Below D	3	†	†	†	†
		Language Arts, Reading Test			
French Composition					
Mostly A	63	89	83	78	63
Mostly B	138	92	80	72	57
Mostly C	100	85	75	65	54
Mostly D	21	71	38	24	14
Mostly Below D	4	†	†	†	†
		Mathematics Test			
Mathematics					
Mostly A	54	100	98	94	91
Mostly B	85	100	93	88	79
Mostly C	62	100	95	71	58
Mostly D	21	100	76	57	48
Mostly Below D	3	†	†	†	†

† Indicates that the statistic was not calculated because of small sample size.

The Relationship Between GED Test Scores and Prior Instruction

If the French-language GED Tests are accurate measures of subjects taught in a regular program of French Canadian high school study, then a positive relationship should be observed between scores on content area GED Tests and the amount of instruction received by students in the content area related to each test.

The French-speaking Canadian graduating high school seniors participating in the 2003 and 2005 studies were asked to indicate the number of years of French literature, French composition, social studies, science, and mathematics courses they had taken from ninth grade to the current term. The students were asked to indicate whether they had taken one year or less, two, three, or four years or more of coursework in each content area. In addition, they were asked to specify the types of courses they had taken in each content area. For example, for social studies, the students were asked to indicate whether they had taken behavioral sciences, civics, economics, geography, political science, national history, or world history.

Table 8.8 contains the percentage of French-speaking Canadian graduating high school seniors in the 2003 and 2005 studies at self-reported total years of study by various minimum standard scores. As expected, the percentages generally decrease across each row as the standard score increases. Additionally, as the number of self-reported total years of study increases, the percentage of seniors generally increases within any given standard score category. For example, 89 percent of graduating seniors with two years of science scored at least 350 on the Science Test. However, a larger percentage (96 percent) of those seniors with at least four years of science scored at least 350 on the Science Test.

TABLE 8.8
PERCENTAGE OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 FRENCH-LANGUAGE GED TESTS STUDIES AT SELF-REPORTED TOTAL YEARS OF STUDY ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
SELF-REPORTED TOTAL YEARS OF STUDY	N	350	410	450	500
		Language Arts, Writing Test			
French Literature					
1 year or less	3	†	†	†	†
2 years	7	†	†	†	†
3 years	118	97	87	72	58
4 years or more	202	98	80	65	50
		Language Arts, Writing Test			
French Composition					
1 year or less	1	†	†	†	†
2 years	14	†	†	†	†
3 years	107	98	87	73	61
4 years or more	220	97	81	66	51
		Social Studies Test			
Social Studies					
1 year or less	8	†	†	†	†
2 years	30	87	67	47	23
3 years	58	79	71	55	28
4 years or more	79	72	58	47	24
		Science Test			
Science					
1 year or less	8	†	†	†	†
2 years	37	89	78	68	49
3 years	32	84	75	59	47
4 years or more	94	96	81	70	51
		Language Arts, Reading Test			
French Literature					
1 year or less	2	†	†	†	†
2 years	8	†	†	†	†
3 years	131	92	85	74	62
4 years or more	200	83	66	59	46
		Language Arts, Reading Test			
French Composition					
1 year or less	1	†	†	†	†
2 years	14	†	†	†	†
3 years	122	93	86	76	63
4 years or more	211	83	67	61	47

Continued on next page

Table 8.8 continued

SELF-REPORTED TOTAL YEARS OF STUDY	N	GED Standard Score ≥ 350	410	450	500
		Mathematics Test			
Mathematics					
1 year or less	23	100	100	91	83
2 years	11	†	†	†	†
3 years	55	100	89	78	69
4 years or more	157	100	94	82	71

† Indicates that the statistic was not calculated because of small sample size.

In Table 8.9, average GED standard scores are reported for the French-language GED Tests. The average standard scores are broken down according to the four levels of amount of prior instruction (from one year or less to four years or more). For example, those seniors with three years of science instruction obtained an average GED standard score of 468 on the Science Test, while those with four years or more achieved an average score of 500.

TABLE 8.9
AVERAGE GED STANDARD SCORES OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 FRENCH-LANGUAGE GED TESTS STUDIES, BY YEARS OF INSTRUCTION IN CONTENT AREA

YEARS INSTRUCTION IN SUBJECT AREA	Language Arts, Writing	Social Studies	Science	Language Arts, Reading	Mathematics
1 year or less	† (1)	† (8)	† (8)	† (2)	† (23)
2 years	† (14)	† (30)	490 (37)	† (8)	† (11)
3 years	527 (107)	437 (58)	468 (32)	521 (131)	533 (55)
4 years or more	507 (220)	423 (79)	500 (94)	465 (200)	539 (157)

† Indicates that the statistic was not calculated because of small sample size.
Note: Numbers in parentheses refer to the number of seniors. Averages for the Language Arts, Writing Test are based on the number of years instruction in composition; averages for the Language Arts, Reading Test are based on the number of years instruction in literature.

Tables 8.10 through 8.14 provide the average standard scores by specific courses of study for the 2003 and 2005 French Canadian studies. Here, the expectation is that those graduating high school seniors who have taken related courses should score higher than those who have not taken these courses. Although the majority of the average standard scores follow this pattern, several do not. The most dramatic difference, for example, occurs between those who have and have not taken business math. High school seniors who have not taken this course tended to score more than 30 standard score points higher than those who have taken this course.

TABLE 8.10
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 FRENCH-LANGUAGE GED TESTS STUDIES SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, WRITING TEST, BY INSTRUCTION IN GRAMMAR AND LANGUAGE COURSES

COURSE		Mean	N	GED Standard Score ≥ 350	410	450	500
Grammar/Composition	Taken	518	252	98	84	70	56
	Not Taken	497	94	97	80	66	49
Spanish	Taken	551	78	97	88	79	74
	Not Taken	501	268	97	81	66	49

Note: Sample sizes for German and Latin were too small for reporting purposes.

TABLE 8.11
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 FRENCH-LANGUAGE GED TESTS STUDIES SCORING AT OR ABOVE STANDARD SCORES ON SOCIAL STUDIES TEST, BY INSTRUCTION IN SELECTED SOCIAL STUDIES COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Behavioral Science	Taken	434	48	83	65	50	23
	Not Taken	427	130	75	63	48	25
Civics	Taken	461	35	83	77	69	40
	Not Taken	421	143	76	60	44	21
Economics	Taken	443	51	78	71	57	29
	Not Taken	423	127	77	61	46	23
Geography	Taken	446	119	82	71	60	31
	Not Taken	395	59	68	47	27	12
Political Science	Taken	456	31	90	77	58	26
	Not Taken	423	147	75	61	47	24
History	Taken	439	143	80	68	55	28
	Not Taken	387	35	66	46	26	11
World History	Taken	432	90	76	62	50	26
	Not Taken	426	88	80	65	48	24

TABLE 8.12
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 FRENCH-LANGUAGE GED TESTS STUDIES SCORING AT OR ABOVE STANDARD SCORES ON SCIENCE TEST, BY INSTRUCTION IN SELECTED SCIENCE COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Biology	Taken	491	148	92	80	68	51
	Not Taken	500	28	89	79	75	57
Chemistry	Taken	504	143	94	83	74	57
	Not Taken	440	33	82	67	45	33
Earth Science	Taken	481	38	87	82	63	42
	Not Taken	495	138	93	79	70	55
General Science	Taken	493	101	90	77	67	55
	Not Taken	491	75	93	83	71	48
Genetics	Taken	509	32	94	81	69	56
	Not Taken	488	144	91	79	69	51
Physical Science	Taken	486	46	83	72	65	50
	Not Taken	494	130	95	82	70	53
Physics	Taken	506	107	91	81	76	62
	Not Taken	471	69	93	77	58	38
Zoology/Botany	Taken	482	10	90	80	70	30
	Not Taken	493	166	92	80	69	54

TABLE 8.13
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 FRENCH-LANGUAGE GED TESTS STUDIES SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, READING TEST, BY INSTRUCTION IN SELECTED ENGLISH AND LANGUAGE COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Literature	Taken	496	290	88	76	67	54
	Not Taken	458	64	80	69	61	45
European Literature	Taken	505	125	91	74	68	53
	Not Taken	480	229	84	74	65	52
World Literature	Taken	490	125	88	72	64	50
	Not Taken	488	229	86	76	67	54
Spanish	Taken	544	80	93	84	79	68
	Not Taken	473	274	85	72	62	48
German	Taken	489	7	100	71	71	43
	Not Taken	489	347	86	74	66	53
Latin	Taken	496	11	100	73	73	45
	Not Taken	489	343	86	74	66	53

TABLE 8.14
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF CANADIAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 FRENCH-LANGUAGE GED TESTS STUDIES SCORING AT OR ABOVE STANDARD SCORES ON MATHEMATICS TEST, BY INSTRUCTION IN SELECTED MATHEMATICS COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Algebra I	Taken	550	167	100	96	86	77
	Not Taken	516	83	99	86	70	61
Algebra II	Taken	552	130	100	96	87	77
	Not Taken	524	120	99	88	74	66
Business Math	Taken	514	53	98	85	70	60
	Not Taken	545	197	100	94	84	75
Calculus	Taken	571	108	100	99	93	87
	Not Taken	514	142	99	87	72	60
General Math	Taken	540	208	100	94	81	71
	Not Taken	532	42	98	83	79	74
Geometry	Taken	548	165	99	95	85	76
	Not Taken	519	85	100	88	73	62
Trigonometry	Taken	549	165	100	95	85	76
	Not Taken	518	85	99	88	72	62

Chapter 9: Spanish-Language GED Tests

OVERVIEW

Chapter 1 provided a brief history of the GED testing program. Chapters 2 through 5 presented technical information pertaining to the development, norming, scaling, equating, reliability, and validity of the English-language U.S. GED Tests. This chapter describes the aspects of the GED Tests that are specific to the Spanish-language GED Tests.

The purpose of the Spanish-language GED Tests is to provide an opportunity to adults who have Spanish as their primary language to certify their attainment of high school–level academic knowledge and skills and earn their jurisdictions' high school equivalency credential, diploma, or certificate. The development of the Spanish-language GED Tests is similar to the development of the English-language U.S. GED Tests. However, these tests are intended to serve different populations of GED examinees and may be normed on different groups of graduating high school seniors, depending on the test. A description of the history, development, reliability, and validity of these tests is provided in the next sections.

HISTORY OF THE SPANISH-LANGUAGE GED TESTS

In 1969, the Commission on Educational Credit and Credentials authorized the development of the Spanish-language versions of the GED Tests. The tests were developed primarily for Puerto Rico, but the commission authorized usage for other Spanish-speaking residents of the United States and its territories. The first generation of the Spanish-language GED Tests was developed from 1969 to 1970 and was first administered in 1971. These tests were revised along with the second and third generations of English-language U.S. GED Tests in 1979 and 1988. Similar to the French-language GED Tests, the 2002 Series Spanish-Language GED Tests were first administered to adult examinees in 2004. In 2007, there were over 28,000 GED examinees who were administered the majority of their tests via the Spanish-language GED Tests (e.g., took at least three of the five tests in the Spanish-language; GEDTS, 2008).

TEST SPECIFICATIONS AND DEVELOPMENT

As stated in Chapter 2, the Spanish-language GED Tests follow most of the same specifications as those for the English-language U.S. GED Tests. More specifically, the Spanish-language versions of the Social Studies; Science; Language Arts, Reading; and Mathematics Tests are a direct translation of the English-language U.S. versions.[18] Almost 90 percent of the Language Arts, Writing Test items were also direct translations. A select few (less than 10 percent) of the English-language U.S. version of the Language Arts, Writing Test items were replaced altogether to avoid translation issues. The testing times for Spanish-language GED Tests are listed in Chapter 1.

Because the content of the Spanish-language GED Tests is essentially identical to that of most of the English-language U.S. GED Tests, the test development process for the Spanish-language GED Tests was identical to that of their U.S. counterparts. Thus, the test development procedures described in detail in Chapter 2 also apply to the development of the Spanish-language GED Tests. Once the English-language U.S. Social Studies; Science; Language Arts, Reading; and Mathematics Test forms were developed (i.e., after the items had passed all content, measurement, and statistical reviews), they were directly translated into

[18] Spanish-language Social Studies Form IA and Language Arts, Reading Form IA are direct translations of English-language U.S. Social Studies Form IB and Language Arts, Reading Form IB.

the Spanish language (as spoken in Mexico and Central America), reviewed, and evaluated by independent consultants.

STANDARDIZATION AND NORMING

The Spanish-language GED Tests have two sets of norms: one for tests administered in the United States and one for tests administered in Puerto Rico.

The standardization and norming data for tests administered in Puerto Rico are based on samples of Spanish-speaking graduating high school seniors across Puerto Rico who took the Spanish-language GED Tests during March, April, and May of 2003. A school was eligible to participate if it offered courses for senior or 12th grade students, graduated a senior class, and awarded high school diplomas. A student was eligible if he or she would have received a high school diploma before September 1, 2003; spoke Spanish fluently; and would not require testing accommodations. Forty-six schools (2,142 students) participated in the Puerto Rico Spanish-language GED Tests standardization and norming study.

The standardization for tests administered in the United States was conducted with samples of Spanish-speaking graduating high school seniors across the United States who took the Spanish-language GED Tests during March, April, and May of 2003. A school was eligible to participate if it offered courses for senior or 12th grade students who were bilingual in Spanish and English, graduated a senior class, and awarded high school diplomas. A student was eligible if he or she would have received a high school diploma before September 1, 2003; spoke Spanish and English fluently; and would not require testing accommodations (such as large print, extended time, etc.). The number of schools participating in this study is presented in Table 9.1. Thirty-seven schools (976 students) participated in the U.S. Spanish-language GED Tests standardization study.

TABLE 9.1
NUMBER OF SCHOOLS PARTICIPATING IN THE 2003 U.S. STANDARDIZATION AND 2005 EQUATING STUDIES FOR THE SPANISH-LANGUAGE GED TESTS, BY STATE

	2003		2005	
	N	%	N	%
Arizona	24	2.5	0	0.0
California	249	25.5	275	38.5
Colorado	27	2.8	12	1.7
Florida	36	3.7	0	0.0
Georgia	23	2.4	56	7.8
Illinois	0	0.0	30	4.2
Indiana	41	4.2	19	2.7
Iowa	12	1.2	0	0.0
Kansas	13	1.3	0	0.0
Louisiana	0	0.0	23	3.2
Maryland	36	3.7	35	4.9
Massachusetts	24	2.5	23	3.2
Michigan	17	1.7	15	2.1
New York	24	2.5	33	4.6
North Carolina	19	1.9	37	5.2
Ohio	22	2.3	0	0.0
Oklahoma	39	4.0	0	0.0
Oregon	30	3.1	19	2.7
Pennsylvania	17	1.7	20	2.8
South Carolina	16	1.6	0	0.0
Tennessee	0	0.0	8	1.1
Texas	263	26.9	72	10.1
Washington	7	0.7	0	0.0
Wisconsin	37	3.8	37	5.2
Total Schools	37		38	
Total Students	976		714	

SCALING AND EQUATING

As with the English-language U.S. GED Tests, the raw scores from the Spanish-language GED Tests were converted to a scale ranging from 200 to 800, with a mean of 500 and standard deviation of 100. The raw-to-standard score conversions for all tests except the Language Arts, Writing Test, regardless of whether they were administered in Puerto Rico or the United States, are the same as the English-language U.S. raw-to-standard score conversions for those tests.

In 2003, in both the United States and Puerto Rico, data were collected via two Spanish-language forms, namely IA and IC. During these studies, all content area tests were administered to the participating Spanish-speaking Puerto Rican and U.S. graduating high school seniors. The intention was to scale, norm, and equate the Language Arts, Writing Test data separately for the United States and Puerto Rico using the same procedures outlined in Chapter 3; however, due to insufficient sample sizes, the United States and Puerto Rico data were combined in order to establish the Spanish-language raw-to-standard score conversions for the Language Arts, Writing Test forms.

In 2005, in both the United States and Puerto Rico, Forms ID and IE of all five tests were administered to Spanish-speaking graduating high school seniors; in the United States, Form IA of each content area test was also administered. Again, the intention was to scale, norm, and equate the Language Arts, Writing Test data separately for the United States and Puerto Rico; however, due to insufficient sample sizes, the United States and Puerto Rican data were combined in order to establish the Spanish-language raw-to-standard score conversions for the Language Arts, Writing Test forms.

The norms for Spanish-language GED Tests administered in the United States are based on the 2001 norming of the English-language U.S. GED Tests. The norms for Spanish-language GED Tests administered in Puerto Rico are based on the 2003 standardization and norming of the Spanish-language GED Tests in Puerto Rico.

RELIABILITY

The reliability of the Spanish-language GED test scores was analyzed using the same methods that were applied to the English-language U.S. GED Tests. These methods are described thoroughly in Chapter 4. The reliability of the scores from the multiple-choice portions of the Spanish-language GED Tests was evaluated by calculating the K-R 20 reliability coefficient (Kuder & Richardson, 1937), the standard error of measurement (SEM), and decision consistency. The reliability of the essay portion of the Language Arts, Writing Test was evaluated using additional criteria discussed below.

The Spanish-language data presented herein are from Forms IA, IC, ID, and IE, which correspond with the Spanish studies performed in both the United States and Puerto Rico in 2003 and 2005. The analyses that follow utilized data combined from the U.S and Puerto Rican studies and, because of concerns regarding fluency in Spanish, only records of students who indicated they were equally fluent in Spanish and English or more fluent in Spanish were analyzed.

K-R 20 and SEM Results

Table 9.2 presents the score means, standard deviations, SEM, and K-R 20 estimates for the test forms in the 2002 Series Spanish-Language GED Tests. It should be noted that the numbers in Table 9.2 for the Language Arts, Writing Test refer only to the multiple-choice portion of the test. The results presented in Table 9.2 are reported in both standard and raw score units. Because the transformation of raw scores to standard scores (described in Chapter 3) is nonlinear, it is not possible to compute K-R 20 directly for standard scores. Thus, K-R 20 estimates are for raw scores only.

The information in Table 9.2 is based on the performance of the sample of Spanish-speaking graduating high school seniors across the United States and Puerto Rico who took the Spanish-language GED Tests as part of the studies in years 2003 and 2005. Data from Form IA and IC originate from the 2003 studies and data from Forms ID and IE originate from the 2005 equating studies. The results presented in Table 9.2 indicate that 85 percent of the test forms have K-R 20 scores of at least .90 and all are greater than .87.

TABLE 9.2
SAMPLE SIZES (N), SCORE MEANS, STANDARD DEVIATIONS (SD), STANDARD ERRORS OF MEASUREMENT (SEM), AND K-R 20 ESTIMATES FOR THE 2002 SERIES SPANISH-LANGUAGE GED TESTS: U.S. AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIOR DATA

		Standard Scores			Raw Scores			
TEST/FORM	N	Mean	SD	SEM	Mean	SD	SEM	K-R 20
Language Arts, Writing								
Form IA	139	495.3	98.1	32.5	29.1	9.2	3.1	.89
Form IC	98	492.0	94.5	29.9	29.5	9.4	3.0	.90
Form ID	217	507.3	100.4	31.7	31.0	9.3	3.0	.90
Form IE	247	488.5	87.2	30.2	31.9	8.7	3.0	.88
Social Studies								
Form IA	191	392.0	91.5	27.5	21.8	10.2	3.0	.91
Form IC	193	377.3	85.6	25.7	19.5	10.0	3.1	.91
Form ID	144	357.6	80.8	24.2	20.3	10.1	3.1	.91
Form IE	177	364.7	61.4	22.1	18.9	8.5	3.1	.87
Science								
Form IA	214	412.3	85.3	22.6	24.2	11.7	3.0	.93
Form IC	165	401.0	99.2	24.3	24.1	12.2	3.0	.94
Form ID	162	400.9	94.2	28.3	22.8	10.3	3.0	.91
Form IE	170	401.4	99.7	24.4	23.5	12.1	3.0	.94
Language Arts, Reading								
Form IA	342	430.3	99.2	26.2	22.7	9.9	2.6	.93
Form IC	347	422.0	99.4	28.1	20.7	9.6	2.7	.92
Form ID	90	408.4	99.4	28.1	18.1	9.6	2.7	.92
Form IE	268	398.1	84.0	23.8	20.6	9.6	2.7	.92
Mathematics								
Form IA	273	424.4	100.4	26.6	25.0	11.4	3.0	.93
Form IC	197	430.4	75.5	21.4	24.7	10.7	3.0	.92
Form ID	188	414.6	88.7	25.1	25.6	10.8	3.1	.92
Form IE	183	410.5	95.8	25.3	23.2	11.3	3.0	.93

Note: Because of concerns regarding fluency in Spanish, only records of students who indicated they were equally fluent in Spanish and English or more fluent in Spanish were analyzed.

Conditional Standard Errors of Measurement

As described in Chapter 4, the SEM provides an estimate of the average amount of error associated with an examinee's observed test score. However, the amount of error associated with test scores may differ at various points along the score scale.

As with the English-language U.S. GED Tests, the passing standard requirements for the individual content area Spanish-language GED Tests are usually along the standard score interval of 410 to 450. Thus, it is important to estimate the amount of error of measurement along this score interval. Conditional standard errors of measurement (CSEMs, i.e., SEMs at specific points or intervals along the score scale) for the Spanish-language GED Tests were estimated (see Chapter 4) and are presented in Table 9.3.

TABLE 9.3
STANDARD SCORE CONDITIONAL STANDARD ERRORS OF MEASUREMENT AT VARIOUS STANDARD SCORES FOR THE 2002 SERIES SPANISH-LANGUAGE GED TESTS: U.S. AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIOR DATA

	Standard Score						
TEST/FORM	400	410	420	430	440	450	460
Social Studies							
Form IA	25.2	25.3	21.2	25.4	21.0	25.1	20.5
Form IC	25.0	25.1	21.1	21.1	21.0	20.9	20.6
Form ID	21.1	21.0	20.7	20.5	24.4	19.7	23.3
Form IE	21.1	21.0	20.9	25.0	20.5	24.3	20.0
Science							
Form IA	21.8	21.9	17.5	17.4	21.5	21.0	20.8
Form IC	21.4	17.1	21.3	21.1	21.0	20.6	20.3
Form ID	25.5	21.3	17.0	21.2	16.9	21.0	20.9
Form IE	25.4	21.2	17.0	21.2	20.9	20.8	16.5
Language Arts, Reading							
Form IA	22.9	19.0	22.6	22.4	25.5	25.0	24.5
Form IC	23.1	23.2	23.2	19.3	22.8	22.5	25.9
Form ID	23.4	23.3	19.4	23.1	22.9	22.6	26.0
Form IE	19.1	22.5	25.9	25.4	24.8	24.2	30.2
Mathematics							
Form IA	25.1	25.3	25.4	25.4	25.3	25.2	25.1
Form IC	20.7	29.1	25.1	25.1	24.9	24.8	24.4
Form ID	16.9	25.4	25.2	25.1	24.7	20.1	19.8
Form IE	24.9	25.1	25.2	25.2	25.1	25.0	24.9

Note: Because of concerns regarding fluency in Spanish, only records of students who indicated they were equally fluent in Spanish and English or more fluent in Spanish were analyzed.

Reliability of Essay Scores on the Language Arts, Writing Test

The reliability of the essay portion of the Spanish-language Language Arts, Writing Test was evaluated by analyzing reader agreement, or inter-rater reliability, and scoring stability. Essay scoring sessions must show evidence of reader agreement and scoring stability. Reader agreement refers to the degree of agreement of scores assigned among different readers scoring the same essays. Inter-rater reliability increases as the number of essays that require attention from the Chief Reader (due to differences in two readers' scores being greater than one point) decreases. Scoring stability refers to how well the scoring sites maintain the scoring standards established by the GED Testing Service Writing Advisory Committee and presented in the 2002 Series GED Writing Test Official Essay Scoring Guide.

Maintaining scoring consistent with official GED essay scoring standards is essential in an essay scoring session. The standards for scoring GED essays must remain fixed, regardless of when the essay is administered, where it is scored, or what specific procedures are used in the scoring session itself. A high degree of inter-rater reliability does not ensure scoring stability. Just because readers agree with each other on the assignment of essay scores does not necessarily mean they are assigning the scores according to the standards defined in the scoring guide.

To achieve inter-rater reliability and scoring stability, the essay scoring standards are regularly reinforced. As the readers score essays, a Chief Reader selects scored essays at random to verify that the readers' scoring is consistent with the definitions in the scoring guide. In cases of a disagreement in assigned essay scores, the Chief Reader discusses the essay with both readers. The monitoring process continues throughout the entire scoring session. Through this system of checks and rechecks, assurance is gained that readers are scoring according to the standards defined in the scoring guide.

Site Monitoring and Score Scale Stability

To facilitate score scale stability, Chief Reader training, and site certification, site monitoring procedures were incorporated into the essay scoring process. Chief Reader training and site certification are described in Chapter 5, and site monitoring for the Spanish-language GED Tests is described below.

In the past, GEDTS has conducted two types of monitoring: random monitoring and systematic monitoring. In random monitoring, a randomly selected set of 40 scored essays from a scoring site is rescored by the Writing Advisory Committee. In systematic monitoring, a common set of 40 essays, scored by the Writing Advisory Committee, is sent to each essay scoring site, where the site's readers rescore the essays. In both types of monitoring, the site is evaluated by determining the congruence of its readers' essay scores to the Writing Advisory Committee's essay scores.

As described in Chapter 4, scoring sites must demonstrate scale score stability, or adherence to the scoring standards established by the Writing Advisory Committee, in order to become a certified essay scoring site. Scoring sites are certified only after they demonstrate a required level of proficiency on several score stability criteria.

Table 9.4 shows the results of the systematic site monitoring of essay scoring sites in 2008 for the Spanish-language essays. The identities of specific sites have not been revealed; instead, sites have been randomly assigned a number between 1 and *s*, where *s* equals the number of sites.

TABLE 9.4
2008 SYSTEMATIC SITE MONITORING RESULTS (FOUR-POINT HOLISTIC SCORING) FOR SPANISH-LANGUAGE GED ESSAYS

		AGREEMENT OF SCORING SITE'S ESSAY SCORES WITH GEDTS WRITING ADVISORY COMMITTEE SCORES			
SITE	Number of Readers	% Scores Equal	% Scores Within One Point	% Scores Differing by > One Point	Correlation*
1	3	85.0	100	0	.96
2	3	78.3	100	0	.89
3	4	88.8	100	0	.97
4	3	88.3	100	0	.97
5	4	81.9	100	0	.95
Mean	3.4	84.5	100	0	.95
Median	3	85.0	100	0	.96
Minimum	3	78.3	100	0	.89
Maximum	4	88.8	100	0	.97

* Pearson correlation between readers' scores and GEDTS Writing Advisory Committee scores.

Essay Score Inter-rater Reliability

To obtain inter-rater reliability coefficients, the polychoric correlation was first calculated between the readers' scores on the Spanish-language essays. Next, the Spearman-Brown prophecy formula was applied to obtain the reliability estimate. For the 2003 and 2005 studies, the inter-rater reliability estimates were calculated as .97 and .98, respectively.

Decision Consistency

The decision consistency for each of the five content area Spanish-language GED Tests was examined using data obtained via the Spanish-language GED Tests 2003 and 2005 studies in the United States and Puerto Rico (i.e., using graduating high school senior data). The same procedure used with the English-language U.S. GED Tests was also applied to the Spanish-language data (i.e., Livingston and Lewis procedure via the BB-Class software program; see Chapter 4).

The decision consistency (probability of correct classification) estimates for Forms IA, IC, ID, and IE are provided in Table 9.5. The false positive rates given in Table 9.5 reflect the probability of an examinee incorrectly earning the minimum score required on a test form, given that their true score is below the criterion. Conversely, the false negative rates indicate the probability that an examinee will not earn the minimum score required on the test form, given that their true score is above the criterion. In both cases, values closer to zero are preferable.

TABLE 9.5
PROBABILITY OF CORRECT CLASSIFICATION, FALSE POSITIVE, AND FALSE NEGATIVE RATES FOR THE 2002 SERIES SPANISH-LANGUAGE GED TESTS: U.S AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIOR DATA

TEST/FORM	N	Percent Not Meeting Minimum Score	Percent Meeting Minimum Score	Probability of Correct Classification	False Positive	False Negative
Language Arts, Writing						
Form IA	110	15	85	.85	.00	.15
Form IC	98	20	80	.80	.00	.20
Form ID	217	18	82	.82	.00	.18
Form IE	247	16	84	.84	.00	.16
Social Studies						
Form IA	173	57	43	1	†	†
Form IC	193	64	36	1	*	*
Form ID	144	72	28	.91	†	.09
Form IE	177	79	21	.98	*	.02
Science						
Form IA	193	52	48	.99	.01	†
Form IC	165	58	42	1	†	†
Form ID	162	62	38	1	*	*
Form IE	170	55	45	1	†	†
Language Arts, Reading						
Form IA	304	41	59	.99	.01	†
Form IC	347	43	57	.99	*	†
Form ID	90	59	41	.94	.06	†
Form IE	268	58	42	1	*	†
Mathematics						
Form IA	253	42	58	.99	.01	†
Form IC	197	41	59	.83	.16	.01
Form ID	188	46	54	.99	.01	†
Form IE	183	48	52	1	†	†

* Value is less than 0.01.
† Value is less than 0.001.
Note: Because of concerns regarding fluency in Spanish, only records of students who indicated they were equally fluent in Spanish and English or more fluent in Spanish were analyzed.

VALIDITY

Readers should refer to Chapter 5 for background on the definition of validity. The Spanish-language GED Tests were subjected to many of the same validity analyses as the English-language U.S. GED Tests. Again, the validation of the Spanish-language GED test scores must be made with respect to its purpose: to measure major academic skills and knowledge in core content areas that are learned during four years of high school. Therefore, analyses must be undertaken to demonstrate that the scores from the GED Tests can be used to evaluate whether an examinee has attained the knowledge and skills associated with the completion of a normal high school academic program of study.

The sources of validity evidence presented in this chapter report: the extent to which the content of the Spanish-language GED Tests represents standard high school curricula and the relationship of the test scores to other external variables.

Evidence Based on Test Content

The content of all but one of the Spanish-language GED Tests is based on the same set of specifications as those for the English-language U.S. GED Tests. Thus, the evidence based on test content and specifications given in Chapters 5 for these three tests is applicable to the Spanish-language versions as well. As mentioned in Chapter 2, almost 90 percent of the Language Arts, Writing Test items were also direct translations. A select few (fewer than 10 percent) of the English-language U.S. version of the Language Arts, Writing Test items were replaced altogether to avoid translation issues.

Evidence Based on Relations with Other Variables

The validity of the Spanish-language GED test scores was also assessed by comparing the performance of high school seniors on the GED Tests with other measures of academic proficiency.

The Relationship Between GED Test Scores and High School Grades

Because the GED Tests are designed to measure academic knowledge and skills that are taught in a regular high school program of study, it is important that they demonstrate a positive relationship to other measures of high school–level academic performance. To investigate this relationship, the self-reported grades of Spanish-speaking U.S. and Puerto Rican graduating high school seniors participating in the studies in 2003 and 2005 were collected and compared with the performance of these same seniors on the Spanish-language GED Tests. Students were asked to list the overall grades they received since ninth grade through the current term for five content areas: literature, composition, social studies, science, and mathematics. The correlations between self-reported grades and Spanish-language GED test scores are reported in Table 9.6.

TABLE 9.6
CORRELATIONS OF U.S. AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIORS' SPANISH-LANGUAGE GED TEST STANDARD SCORES WITH SELF-REPORTED LETTER GRADES IN THE SAME CONTENT AREA: 2003 AND 2005 STUDIES COMBINED

TEST	N	*r*
Language Arts, Writing	652	.41[a]
Language Arts, Writing	637	.41[b]
Social Studies	662	.35
Science	665	.35
Language Arts, Reading	964	.31[a]
Language Arts, Reading	918	.32[b]
Mathematics	792	.40

Note: Data are from Forms IA and IC (collected in 2003) and Forms IA, ID, and IE (collected in 2005). All correlations were significant at $p < .01$. Letter grades are reported as Mostly A, Mostly B, Mostly C, Mostly D, and Mostly Below D. To compute the correlations, letter grades were recoded as Mostly A=4, Mostly B=3, Mostly C=2, Mostly D=1, Mostly Below D=0.
[a] Correlation with self-reported grades in literature.
[b] Correlation with self-reported grades in composition.

The grades reported for the graduating high school seniors were also compared with their performance at selected values along the GED standard score scale. This analysis helps identify the approximate GPA or letter-grade levels that correspond to performance levels on the GED Tests. Table 9.7 presents, for each Spanish-language GED Test, the percentages of soon-to-be graduating, Spanish-speaking U.S. and Puerto Rican seniors meeting selected GED score standards for each letter grade. For example, the first row of Table 9.7 indicates that 100 percent of the seniors whose reported grades were "Mostly A" scored at or above a GED standard score of 350 on the Language Arts, Writing Test. The second row of the table shows that 33 percent of the "Mostly B" seniors achieved a score of at least 500.

TABLE 9.7
PERCENTAGE OF U.S. AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 SPANISH-LANGUAGE GED TESTS STUDIES AT SELF-REPORTED GRADE LEVELS ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
SELF-REPORTED GRADES	N	350	410	450	500
		Language Arts, Writing Test			
Literature					
Mostly A	313	100	93	85	61
Mostly B	243	97	79	58	33
Mostly C	87	94	67	44	23
Mostly D	8	†	†	†	†
Mostly Below D	1	†	†	†	†
		Language Arts, Writing Test			
Composition					
Mostly A	294	100	93	85	62
Mostly B	241	95	80	59	35
Mostly C	97	95	65	41	20
Mostly D	3	†	†	†	†
Mostly Below D	2	†	†	†	†
		Social Studies Test			
Social Studies					
Mostly A	285	81	49	28	12
Mostly B	254	55	28	16	8
Mostly C	113	39	9	2	1
Mostly D	10	†	†	†	†
Mostly Below D	0	†	†	†	†
		Science Test			
Science					
Mostly A	290	84	63	50	33
Mostly B	232	69	37	27	16
Mostly C	128	54	21	16	8
Mostly D	15	27	13	13	0
Mostly Below D	0	†	†	†	†
		Language Arts, Reading Test			
Literature					
Mostly A	395	88	69	54	33
Mostly B	397	75	48	33	17
Mostly C	156	68	35	19	8
Mostly D	14	†	†	†	†
Mostly Below D	2	†	†	†	†
		Language Arts, Reading Test			
Composition					
Mostly A	365	87	68	55	33
Mostly B	383	77	49	34	17
Mostly C	156	69	34	17	7
Mostly D	12	†	†	†	†
Mostly Below D	2	†	†	†	†

Continued on next page

Table 9.7 continued

SELF-REPORTED GRADES	N	GED Standard Score ≥ 350	410	450	500
		Mathematics Test			
Mathematics					
Mostly A	345	91	74	53	30
Mostly B	276	85	56	34	19
Mostly C	138	62	31	13	4
Mostly D	31	45	19	10	6
Mostly Below D	2	†	†	†	†

† Indicates that the statistic was not calculated because of small sample size.

Overall, Table 9.7 illustrates that the higher the high school grade, the higher the GED score, and therefore, the greater the likelihood of earning the minimum score required for the particular GED Test. These results indicate that the passing standards established on the GED Tests do discriminate between higher and lower achieving Spanish-speaking graduating high school students in the United States and Puerto Rico. Therefore, the results support both the validity of the GED test scores and the validity of the GED standard setting procedure.

The Relationship Between GED Test Scores and Prior Instruction

If the Spanish-language GED Tests are accurate measures of subjects taught in a regular program of high school study, then a positive relationship should be observed between scores on content area GED Tests and the amount of instruction received by students in the content area related to each test. The Spanish-speaking U.S. and Puerto Rican graduating high school seniors participating in the 2003 and 2005 studies were asked to indicate the number of years of English literature, English composition, social studies, science, and mathematics courses they had taken from ninth grade to the current term. The students were asked to indicate whether they had taken one year or less, two, three, or four years or more of coursework in each content area. In addition, they were also asked to specify the types of courses they had taken in each content area. For example, for social studies, the students were asked to indicate whether they had taken behavioral sciences, civics, economics, geography, political science, national history, or world history.

Table 9.8 contains the percentage of U.S. and Puerto Rican graduating high school seniors in the 2003 and 2005 studies at self-reported total years of study by various minimum standard scores. As expected, the percentages generally decrease across each row as the standard score increases. Additionally, as the number of self-reported total years of study increases, the percentage of seniors generally increases within any given standard score category. For example, 39 percent of graduating seniors with two years of mathematics scored at least 410 on the Mathematics Test. However, a larger percentage (62 percent) of those seniors with at least four years of mathematics scored at least 410 on the Mathematics Test.

TABLE 9.8
PERCENTAGE OF U.S. AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 SPANISH-LANGUAGE GED TESTS STUDIES AT SELF-REPORTED TOTAL YEARS OF STUDY ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
SELF-REPORTED TOTAL YEARS OF STUDY	N	350	410	450	500
		Language Arts, Writing Test			
Literature					
1 year or less	54	96	67	46	33
2 years	55	96	82	71	51
3 years	50	98	92	70	52
4 years or more	518	98	84	69	44
		Language Arts, Writing Test			
Composition					
1 year or less	45	93	76	62	44
2 years	70	97	79	61	43
3 years	77	95	78	61	44
4 years or more	473	98	85	70	45
		Social Studies Test			
Social Studies					
1 year or less	30	43	23	13	3
2 years	90	67	29	12	6
3 years	123	59	28	15	4
4 years or more	449	64	36	21	10
		Science Test			
Science					
1 year or less	41	61	29	12	10
2 years	74	70	27	18	8
3 years	149	63	38	26	16
4 years or more	435	74	50	40	26
		Language Arts, Reading Test			
Literature					
1 year or less	76	70	38	25	12
2 years	51	80	65	39	12
3 years	102	71	47	35	18
4 years or more	767	79	55	39	23
		Language Arts, Reading Test			
Composition					
1 year or less	126	71	50	39	25
2 years	91	84	64	41	16
3 years	94	77	46	35	17
4 years or more	643	78	53	38	22
		Mathematics Test			
Mathematics					
1 year or less	17	35	24	12	6
2 years	33	82	39	15	3
3 years	185	75	46	24	12
4 years or more	592	83	62	42	24

In Table 9.9, average GED standard scores are reported for the Spanish-language GED Tests. The average standard scores are broken down according to the four levels of amount of prior instruction (from one year or less to four years or more). For example, those seniors with three years of English literature instruction obtained an average GED standard score of 409 on the Language Arts, Reading Test, while those with four years or more achieved an average score of 424.

TABLE 9.9
AVERAGE GED STANDARD SCORES OF U.S. AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 SPANISH-LANGUAGE GED TESTS STUDIES, BY YEARS OF INSTRUCTION IN CONTENT AREA

YEARS INSTRUCTION IN SUBJECT AREA	Language Arts, Writing	Social Studies	Science	Language Arts, Reading	Mathematics
1 year or less	488	†	372	391	†
	(45)	(30)	(41)	(76)	(17)
2 years	483	371	379	419	392
	(70)	(90)	(74)	(51)	(33)
3 years	486	363	394	409	397
	(77)	(123)	(149)	(102)	(185)
4 years or more	503	380	417	424	432
	(473)	(449)	(435)	(767)	(592)

† Indicates that the statistic was not calculated because of small sample size.
Note: Numbers in parentheses refer to the number of seniors. Averages for the Language Arts, Writing Test are based on the number of years instruction in composition; averages for the Language Arts, Reading Test are based on the number of years instruction in literature.

Tables 9.10 through 9.14 provide the average standard scores by specific courses of study for the 2003 and 2005 U.S. and Puerto Rico Spanish-language GED Tests studies. Here, the expectation is that those graduating high school seniors who have taken related courses should score higher than those who have not taken these courses. Although the majority of the average standard scores follow this pattern, several do not. The most dramatic difference, for example, occurs between those who have and have not taken general math. High school seniors who have not taken this course scored nearly 30 standard score points higher than those who have taken this course.

TABLE 9.10
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF U.S. AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 SPANISH-LANGUAGE GED TESTS STUDIES SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, WRITING TEST, BY INSTRUCTION IN GRAMMAR AND LANGUAGE COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Grammar/Composition	Taken	505	557	98	85	70	48
	Not Taken	462	144	94	75	56	31
Spanish	Taken	498	196	97	81	64	45
	Not Taken	495	505	97	83	68	44
German	Taken	506	219	98	85	69	48
	Not Taken	492	482	97	82	66	42
Latin	Taken	504	178	97	83	67	47
	Not Taken	494	523	97	82	67	43

TABLE 9.11
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF U.S. AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 SPANISH-LANGUAGE GED TESTS STUDIES SCORING AT OR ABOVE STANDARD SCORES ON SOCIAL STUDIES TEST, BY INSTRUCTION IN SELECTED SOCIAL STUDIES COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Behavioral Science	Taken	399	139	73	44	28	14
	Not Taken	368	566	60	30	16	7
Civics	Taken	405	75	75	52	32	13
	Not Taken	370	630	61	30	17	7
Economics	Taken	377	141	65	33	18	9
	Not Taken	373	564	62	32	18	8
Geography	Taken	376	429	64	33	18	8
	Not Taken	371	276	59	33	18	9
Political Science	Taken	383	199	66	36	21	9
	Not Taken	371	506	61	31	17	8
History	Taken	378	613	63	35	20	9
	Not Taken	350	92	54	20	5	3
World History	Taken	392	360	69	42	25	13
	Not Taken	355	345	54	23	11	3

TABLE 9.12
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF U.S. AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 SPANISH-LANGUAGE GED TESTS STUDIES SCORING AT OR ABOVE STANDARD SCORES ON SCIENCE TEST, BY INSTRUCTION IN SELECTED SCIENCE COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Biology	Taken	405	674	71	44	33	21
	Not Taken	388	37	57	38	27	19
Chemistry	Taken	415	555	74	48	37	23
	Not Taken	368	156	54	27	18	11
Earth Science	Taken	419	335	76	51	39	25
	Not Taken	391	376	65	36	28	16
General Science	Taken	405	248	72	40	32	21
	Not Taken	404	463	69	45	33	21
Genetics	Taken	411	25	64	40	28	24
	Not Taken	404	686	70	43	33	20
Physical Science	Taken	414	46	70	50	35	17
	Not Taken	404	665	70	43	33	21
Physics	Taken	404	146	66	42	31	18
	Not Taken	404	565	71	44	33	21
Zoology/Botany	Taken	418	449	72	49	41	27
	Not Taken	382	262	66	34	20	10

TABLE 9.13
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF U.S. AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 SPANISH-LANGUAGE GED TESTS STUDIES SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, READING TEST, BY INSTRUCTION IN SELECTED ENGLISH AND LANGUAGE COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Literature	Taken	422	865	78	54	39	22
	Not Taken	398	182	71	43	28	12
European Literature	Taken	439	399	85	58	45	28
	Not Taken	404	648	72	49	32	16
World Literature	Taken	436	467	83	58	45	28
	Not Taken	402	580	72	47	31	15
French	Taken	390	171	76	39	23	10
	Not Taken	423	876	77	55	40	23
German	Taken	431	208	84	53	39	25
	Not Taken	414	839	75	52	36	19
Latin	Taken	400	128	80	41	28	13
	Not Taken	420	919	76	54	38	22

TABLE 9.14
MEAN STANDARD SCORE, SAMPLE SIZE, AND PERCENT OF U.S. AND PUERTO RICAN GRADUATING HIGH SCHOOL SENIORS IN 2003 AND 2005 SPANISH-LANGUAGE GED TESTS STUDIES SCORING AT OR ABOVE STANDARD SCORES ON MATHEMATICS TEST, BY INSTRUCTION IN SELECTED MATHEMATICS COURSES

				GED Standard Score ≥			
COURSE		Mean	N	350	410	450	500
Algebra I	Taken	420	760	81	58	36	19
	Not Taken	426	81	72	46	36	27
Algebra II	Taken	431	646	83	62	41	24
	Not Taken	385	195	70	37	22	6
Business Math	Taken	433	55	91	65	40	20
	Not Taken	420	786	79	56	36	20
Calculus	Taken	500	107	93	85	72	51
	Not Taken	409	734	78	52	31	16
General Math	Taken	406	414	78	51	30	15
	Not Taken	434	427	82	61	42	25
Geometry	Taken	426	738	82	59	38	22
	Not Taken	382	103	64	37	22	10
Trigonometry	Taken	460	269	91	76	54	33
	Not Taken	402	572	75	47	28	14

References

American College Testing. (1988). *ACT Assessment technical manual.* Iowa City, IA: American College Testing Program.

American Council on Education. (1999). *Alignment of national and state standards: A report by the GED Testing Service.* Washington, DC: Author.

American Council on Education. (1964). *Examiners manual for the Tests of General Educational Development, high school level.* Washington, DC: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Americans with Disabilities Act of 1990, 42 U.S.C.A. § 12101 *et seq.* (West 1993).

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Annual Survey of Colleges: 2007 [Data file]. New York, NY: The College Board.

Benners, G. A., & George-Ezzelle, C. E. (2006). *An exploration of GED® standard score stability: 2001 through 2005.* Washington, DC: American Council on Education.

Bloom, B. (Ed.). (1956). *Taxonomy of educational objectives: Book 1, cognitive domain. Handbook I: cognitive domain.* New York: McKay.

Brennan, R. L. (2004). Manual for BB-Class: A computer program that uses the beta-binomial model for classification consistency and accuracy (CASMA Rep. No. 9) [Computer software manual]. Retrieved from www.education.uiowa.edu/casma/computer_programs.htm#classification.

Chester, M. D. (2003). Multiple measures and high-stakes decisions: A framework for combining measures. *Educational Measurement: Issues and Practice, 22*(2), 32–41.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice, 17*(1), 31–44.

Clauser, B. E., Mazor, K. M., & Hambleton, R. K. (1993). The effects of purification of the matching criterion on identification of DIF using the Mantel-Haenszel procedure. *Applied Measurement in Education, 6*, 269–279.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Cui, Z., & Chien, Y-M. (2004). RAGE-RGEQUATE [Computer software]. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment.

Donoghue, J. R., Holland, P. W., & Thayer, D. T. (1993). A Monte Carlo study of factors that affect the Mantel-Haenszel and standardization measures of differential item functioning. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 137–166). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum Associates.

Douglas, K. M. (2007). *A general method for estimating the classification reliability of complex decisions based on configural combinations of multiple assessment scores.* Unpublished doctoral dissertation, University of Maryland, College Park.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105–146). Phoenix: American Council on Education/Oryx Press Series on Higher Education.

Feldt, L. S., & Qualls, A. L. (1998). Approximating scale score standard error of measurement from the raw score standard error. *Applied Measurement in Education, 11*(2), 159–177.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions.* New York: Wiley and Sons.

GED Testing Service. (2008a). *2007 GED testing program statistical report.* Washington, DC: American Council on Education.

GED Testing Service. (2008b). *GED Testing Service policies and procedures manual.* Washington, DC: American Council on Education.

GED Testing Service. (2007). *2006 GED testing program statistical report.* Washington, DC: American Council on Education.

GED Testing Service. (2006a). *Who passed the GED Tests? 2005 statistical report.* Washington, DC: American Council on Education.

GED Testing Service. (2006b). *Who passed the GED Tests? 2004 statistical report.* Washington, DC: American Council on Education.

GED Testing Service. (2005a). *Examiner's manual for the Tests of General Educational Development.* Washington, DC: American Council on Education.

GED Testing Service. (2005b). *Who passed the GED Tests? 2003 statistical report.* Washington, DC: American Council on Education.

GED Testing Service. (2004). *Who passed the GED Tests? 2002 statistical report.* Washington, DC: American Council on Education.

GED Testing Service. (2002a). *GED examiner's manual for the Tests of General Educational Development.* Washington, DC: American Council on Education.

GED Testing Service. (2002b). *Official GED practice tests administrator's manual.* Austin, TX: Steck-Vaughn.

GED Testing Service. (1993). *The Tests of General Educational Development: Technical manual* (1st ed.). Washington, DC: American Council on Education.

George-Ezzelle, C. E., & Hsu, Y.-C. (2007). *Examinee and high school senior performance on the GED® Tests.* Washington, DC: American Council on Education.

George-Ezzelle, C. E., & Skaggs, G. (2004). *Examining the validity of GED® Tests scores with scheduling and setting accommodations.* Washington, DC: American Council on Education.

Glöckner-Rist, A., & Hoijtink, H. (2003). The best of both worlds: Factor analysis of dichotomous data using item response theory and structural equation modeling. *Structural Equation Modeling, 10*(4), 544–565.

Gulliksen, H. (1950). *Theory of mental tests.* New York: John Wiley and Sons.

Hambleton, R. K., & Rogers, H. J. (1989). Detecting potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods. *Applied Measurement in Education, 2*(4), 313–334.

Hanson, B. A. (1992). RGEQUATE [Computer software]. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer-Verlag.

Kolen, M. J., & Whitney, D. R. (1982). Comparison of four procedures for equating the Tests of General Educational Development. *Journal of Educational Measurement, 19*, 279–293.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2,* 151–160.

Lewis, C. (1993). A note on the value of including the studied item in the test score when analyzing test items for DIF. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 317-320). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lindquist, E. F. (1944, October). The use of tests in the accreditation of military experience and in the educational placement of war veterans. *Educational Record*, 357–376.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32*, 179–197.

Lord, F. M., & Novick, M. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mullane, L. (2001). *Bridges of opportunity: A history of the Center for Adult Learning and Educational Credentials.* Washington, DC: American Council on Education.

Narayanan, P., & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item bias. *Applied Psychological Measurement, 18*(4), 315–328.

National Research Council, (1996). *National science education standards.* Washington, DC: National Academy Press.

Navas-Ara, M. J., & Gómez-Benito, J. (2002). Effects of ability scale purification on the identification of DIF. *European Journal of Psychological Assessment, 18*(1), 9–15.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Osterlind, S. J. (1989). *Constructing test items.* Norwell, MA: Academic Press.

Patience, W. P., & Swartz, R. (1987). *Essay score reliability: Issues in and methods of reporting GED writing skills scores.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Washington, DC.

Patience, W. M., & Whitney, D. R. (1982). *What do the GED Tests measure?* Washington, DC: American Council on Education.

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221–262). Phoenix: American Council on Education/Oryx Press Series on Higher Education.

Quigley, A. (1987). *The Canadianization of the GED: The history and development of the General Educational Development program in Canada.* Washington, DC: American Council on Education.

Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika, 58*, 159–194.

Sireci, S., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature* (Rep. No. 485). University of Massachusetts, Amherst: Center for Educational Assessment Research.

Society for Human Resource Management. (2002). *Does your company accept applicants with a GED credential for jobs requiring a high school degree?* [Online poll]. Retrieved January 25, 2008, from www.shrm.org/poll/results.asp?question=67.

Stone, C. A., & Yeh, C.-C. (2006). Assessing the dimensionality and factor structure of multiple-choice exams: An empirical comparison of methods using the Multistate Bar Exam. *Educational and Psychological Measurement, 66*(2), 193–214.

Swartz, R., Patience, W. M., & Whitney, D. R. (1985, July). *Adding an essay to the GED Writing Skills Test: Reliability and validity issues.* (GEDTS Research Studies, No. 7). Washington, DC: American Council on Education. (ERIC Document Reproduction Service No. ED266288)

Swartz, R., & Whitney, D. R. (1985, June). *The relationship between scores on the GED Writing Skills Test and on direct measures of writing.* (GEDTS Research Studies, No. 6). Washington, DC: American Council on Education. (ERIC Document Reproduction Service No. ED266289)

Thorndike, R. L. (1982). *Applied psychometrics.* Boston: Houghton Mifflin.

U.S. Department of Labor. (1991). *What work requires of schools: A SCANS report for America 2000.* Washington, DC: Author.

Whitney, D. R., Malizio, A. G., & Patience, W. M. (1986). Reliability of the GED Tests. *Educational and Psychological Measurement, 46,* 689–698.

Whitney, D. R., & Patience, W. M. (1981). *Work rates on the GED Tests: Relationships with examinee age and test time limits.* (GEDTS Research Studies, No. 6). Washington, DC: American Council on Education.

Wood, R., Wilson, D., Gibbons, R., Schilling, S., Muraki, E., & Bock, D. (2003). TESTFACT 4: Test scoring, item statistics, and item factor analysis [Computer software]. Mooresville, IN: Scientific Software.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–348). Hillsdale, NJ: Lawrence Erlbaum Associates.

Appendices

APPENDIX A

TABLE A.1
CANADIAN ENGLISH-LANGUAGE GED TESTS STANDARD SCORE (SS) TO PERCENTILE RANK (PR) CONVERSION TABLES: FORMS ID THROUGH IH

Language Arts, Writing				Social Studies				Science				Language Arts, Reading				Mathematics			
SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR
800	99	490	38	800	99	490	47	800	99	490	30	800	99	490	36	800	99	490	36
790	99	480	35	790	99	480	44	790	99	480	27	790	99	480	32	790	99	480	33
780	99	470	32	780	99	470	36	780	99	470	24	780	98	470	28	780	99	470	31
770	99	460	28	770	99	460	34	770	99	460	21	770	97	460	25	770	99	460	28
760	99	450	25	760	99	450	32	760	99	450	19	760	95	450	22	760	99	450	26
750	99	440	22	750	99	440	27	750	98	440	17	750	93	440	20	750	99	440	22
740	99	430	19	740	99	430	24	740	98	430	15	740	92	430	18	740	99	430	18
730	98	420	16	730	99	420	23	730	97	420	13	730	90	420	16	730	99	420	17
720	97	410	13	720	99	410	19	720	96	410	12	720	89	410	14	720	98	410	13
710	96	400	11	710	98	400	16	710	96	400	10	710	88	400	12	710	97	400	10
700	95	390	9	700	98	390	13	700	94	390	9	700	87	390	9	700	96	390	8
690	94	380	7	690	97	380	11	690	94	380	6	690	86	380	8	690	95	380	5
680	93	370	5	680	96	370	10	680	93	370	4	680	85	370	6	680	94	370	3
670	91	360	4	670	96	360	8	670	90	360	3	670	83	360	4	670	93	360	3
660	89	350	3	660	95	350	6	660	88	350	2	660	81	350	3	660	91	350	2
650	87	340	2	650	94	340	6	650	86	340	2	650	80	340	2	650	88	340	2
640	85	330	1	640	93	330	5	640	84	330	2	640	78	330	2	640	87	330	1
630	83	320	1	630	91	320	4	630	82	320	1	630	75	320	1	630	84	320	1
620	81	310	1	620	89	310	3	620	79	310	1	620	72	310	1	620	81	310	1
610	78	300	1	610	86	300	3	610	78	300	1	610	68	300	1	610	78	300	1
600	75	290	1	600	84	290	2	600	75	290	1	600	67	290	1	600	76	290	1
590	72	280	1	590	81	280	1	590	72	280	1	590	65	280	1	590	74	280	1
580	69	270	1	580	77	270	1	580	67	270	1	580	63	270	1	580	70	270	1
570	66	260	1	570	76	260	1	570	63	260	1	570	60	260	1	570	67	260	1
560	63	250	1	560	72	250	1	560	59	250	1	560	58	250	1	560	64	250	1
550	60	240	1	550	67	240	1	550	55	240	1	550	55	240	1	550	61	240	1
540	56	230	1	540	65	230	1	540	51	230	1	540	51	230	1	540	56	230	1
530	53	220	1	530	61	220	1	530	48	220	1	530	48	220	1	530	51	220	1
520	49	210	1	520	58	210	1	520	40	210	1	520	45	210	1	520	47	210	1
510	46	200	1	510	54	200	1	510	37	200	1	510	42	200	1	510	43	200	1
500	42			500	51			500	33			500	40			500	40		

TABLE A.2
FRENCH-LANGUAGE GED TESTS STANDARD SCORE (SS) TO PERCENTILE RANK (PR) CONVERSION TABLES: ALL FORMS

Language Arts, Writing				Social Studies				Science				Language Arts, Reading				Mathematics			
SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR
800	99	490	46	800	99	490	65	800	99	490	60	800	99	490	46	800	99	490	35
790	99	480	42	790	99	480	61	790	99	480	55	790	99	480	42	790	99	480	31
780	99	470	38	780	99	470	56	780	99	470	50	780	99	470	38	780	99	470	27
770	99	460	34	770	99	460	51	770	99	460	46	770	99	460	34	770	99	460	22
760	99	450	31	760	99	450	47	760	99	450	41	760	99	450	31	760	99	450	19
750	99	440	27	750	99	440	42	750	99	440	36	750	99	440	27	750	99	440	15
740	99	430	24	740	99	430	37	740	99	430	31	740	99	430	24	740	99	430	12
730	99	420	21	730	99	420	32	730	99	420	27	730	99	420	21	730	99	420	10
720	99	410	18	720	99	410	28	720	99	410	23	720	99	410	18	720	98	410	8
710	98	400	16	710	99	400	24	710	99	400	19	710	98	400	16	710	98	400	6
700	98	390	14	700	99	390	20	700	99	390	16	700	98	390	14	700	97	390	4
690	97	380	12	690	99	380	17	690	99	380	13	690	97	380	12	690	97	380	3
680	96	370	10	680	99	370	14	680	98	370	10	680	96	370	10	680	96	370	2
670	96	360	8	670	99	360	11	670	98	360	8	670	96	360	8	670	95	360	2
660	95	350	7	660	99	350	9	660	98	350	6	660	95	350	7	660	94	350	1
650	93	340	5	650	98	340	7	650	97	340	5	650	93	340	5	650	92	340	1
640	92	330	4	640	98	330	5	640	97	330	3	640	92	330	4	640	91	330	1
630	90	320	4	630	97	320	4	630	96	320	2	630	90	320	4	630	89	320	1
620	88	310	3	620	97	310	3	620	95	310	2	620	88	310	3	620	87	310	1
610	86	300	2	610	96	300	2	610	94	300	1	610	86	300	2	610	85	300	1
600	84	290	1	600	95	290	2	600	92	290	1	600	84	290	1	600	82	290	1
590	82	280	1	590	94	280	1	590	91	280	1	590	82	280	1	590	79	280	1
580	79	270	1	580	92	270	1	580	89	270	1	580	79	270	1	580	76	270	1
570	76	260	1	570	90	260	1	570	87	260	1	570	76	260	1	570	72	260	1
560	73	250	1	560	88	250	1	560	85	250	1	560	73	250	1	560	68	250	1
550	69	240	1	550	86	240	1	550	82	240	1	550	69	240	1	550	64	240	1
540	66	230	1	540	83	230	1	540	79	230	1	540	66	230	1	540	59	230	1
530	62	220	1	530	80	220	1	530	76	220	1	530	62	220	1	530	55	220	1
520	58	210	1	520	77	210	1	520	72	210	1	520	58	210	1	520	50	210	1
510	54	200	1	510	73	200	1	510	68	200	1	510	54	200	1	510	45	200	1
500	50			500	69			500	64			500	50			500	40		

TABLE A.3
PUERTO RICO (SPANISH-LANGUAGE GED TESTS) STANDARD SCORE (SS) TO PERCENTILE RANK (PR) CONVERSION TABLES: ALL FORMS

Language Arts, Writing				Social Studies				Science				Language Arts, Reading				Mathematics			
SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR	SS	PR
800	99	490	54	800	99	490	88	800	99	490	81	800	99	490	74	800	99	490	77
790	99	480	51	790	99	480	86	790	99	480	78	790	99	480	71	790	99	480	74
780	99	470	48	780	99	470	83	780	99	470	75	780	99	470	68	780	99	470	71
770	99	460	44	770	99	460	81	770	99	460	71	770	99	460	64	770	99	460	67
760	99	450	41	760	99	450	78	760	99	450	68	760	99	450	60	760	99	450	63
750	99	440	37	750	99	440	74	750	99	440	64	750	99	440	56	750	99	440	59
740	99	430	34	740	99	430	71	740	99	430	60	740	99	430	52	740	99	430	54
730	99	420	31	730	99	420	67	730	99	420	55	730	99	420	48	730	99	420	49
720	98	410	28	720	99	410	62	720	99	410	51	720	99	410	44	720	99	410	45
710	98	400	25	710	99	400	58	710	99	400	46	710	99	400	39	710	99	400	40
700	97	390	22	700	99	390	53	700	99	390	42	700	99	390	35	700	99	390	35
690	96	380	20	690	99	380	48	690	99	380	37	690	99	380	31	690	99	380	31
680	96	370	17	680	99	370	43	680	99	370	33	680	99	370	28	680	99	370	26
670	95	360	15	670	99	360	39	670	99	360	28	670	99	360	24	670	99	360	22
660	94	350	13	660	99	350	34	660	99	350	24	660	98	350	20	660	98	350	19
650	93	340	11	650	99	340	29	650	99	340	20	650	98	340	17	650	98	340	15
640	92	330	9	640	99	330	25	640	98	330	17	640	98	330	14	640	98	330	12
630	90	320	8	630	99	320	21	630	98	320	14	630	97	320	12	630	97	320	10
620	89	310	7	620	99	310	17	620	98	310	11	620	96	310	10	620	97	310	7
610	87	300	5	610	98	300	14	610	97	300	9	610	96	300	8	610	96	300	6
600	85	290	4	600	98	290	11	600	97	290	7	600	95	290	6	600	96	290	4
590	83	280	4	590	98	280	8	590	96	280	5	590	94	280	5	590	95	280	3
580	81	270	3	580	97	270	6	580	95	270	4	580	93	270	3	580	94	270	2
570	79	260	2	570	97	260	4	570	94	260	3	570	92	260	3	570	93	260	1
560	76	250	2	560	96	250	3	560	93	250	2	560	90	250	2	560	92	250	1
550	73	240	1	550	96	240	2	550	92	240	1	550	89	240	1	550	91	240	1
540	70	230	1	540	95	230	1	540	91	230	1	540	87	230	1	540	89	230	1
530	68	220	1	530	94	220	1	530	89	220	1	530	85	220	1	530	87	220	1
520	64	210	1	520	93	210	1	520	88	210	1	520	83	210	1	520	85	210	1
510	61	200	1	510	91	200	1	510	86	200	1	510	80	200	1	510	83	200	1
500	58			500	90			500	83			500	77			500	80		

APPENDIX B

LIST OF PARTICIPATING JURISDICTIONS, NUMBER OF OFFICIAL GED TESTING CENTERS, AND MINIMUM SCORE REQUIREMENTS

Jurisdiction	Active Official GED Testing Centers	Min. Scores Requirements
United States		
Alabama	50	*
Alaska	21	*
Arizona	33	*
Arkansas	61	*
California	190	*
Colorado	44	*
Connecticut	22	*
Delaware	6	*
District of Columbia	1	*
Florida	88	*
Georgia	48	*
Hawaii	12	*
Idaho	8	*
Illinois†	69	*
Indiana	70	*
Iowa	98	*
Kansas	26	*****
Kentucky	43	*
Louisiana	40	*
Maine	80	*
Maryland	20	*
Massachusetts	31	*
Michigan	121	*
Minnesota	60	*
Mississippi	37	*
Missouri	26	*
Montana	22	*
Nebraska	33	*
Nevada	22	*
New Hampshire	19	*
New Jersey	34	*
New Mexico	29	*
New York	317	*
North Carolina	74	*
North Dakota	19	*
Ohio	109	*
Oklahoma	43	*
Oregon	41	*
Pennsylvania	116	*
Rhode Island	11	*
South Carolina	1	*
South Dakota	17	*
Tennessee	38	*
Texas	157	*
Utah	21	*
Vermont	11	*
Virginia	80	*
Washington	57	*
West Virginia	68	*
Wisconsin	79	*
Wyoming	28	*
Insular Areas		
American Samoa	1	*
Federated States of Micronesia	NA	*
Guam†	1	*
Marshall Islands	NA	NA
Northern Mariana Islands	NA	NA
Palau	1	*
Puerto Rico†	11	*
Virgin Islands†	1	*

Continued on next page

Appendix B continued

Jurisdiction	Active Official GED Testing Centers	Min. Scores Requirements
Canada		
Alberta	17	**
British Columbia	1	**
Manitoba	1	**
New Brunswick	2	**
Newfoundland and Labrador	1	**
Northwest Territories†	1	**
Nova Scotia	1	**
Nunavut	1	**
Ontario	1	**
Prince Edward Island	1	**
Quebec	1	**
Saskatchewan	1	**
Yukon Territory†	1	**
Federal and Other Contracts		
DANTES	NA	*
Federal Bureau of Prisons†	115	***
International	100+	****
Michigan Prisons	43	*
VA Hospitals	NA	NA

Source: 2007 GED® Testing Service data.
* Minimum total score of 2,250 (450 average) on the battery of tests and a minimum of 410 on each content area test.
** 450 minimum on each content area test.
*** Minimum scores and other requirements depend on the jurisdiction of the Official GED Testing Center.
**** Passing standards vary by location.
***** Minimum total score of 2,250 (450 average) on the battery of tests and a minimum of 420 on each content area test.
† Information is from 2006.
NA = Not available.

APPENDIX C

OVERVIEW OF BLOOM'S TAXONOMY CATEGORIES APPLICABLE TO THE GED TESTS

KNOWLEDGE

Knowledge questions require the candidate to observe and recall information, including major ideas or concepts and a basic mastery of subject matter. Although the GED Tests do not assess basic recall of information, candidates should have knowledge of ideas and concepts that can be used in answering other questions.

COMPREHENSION

Comprehension questions require the candidate to understand the meaning and intent of written and visual text. Comprehension questions measure the ability to:

- Understand and restate information.
- Summarize ideas.
- Translate knowledge into new contexts.
- Make inferences.
- Draw conclusions.

APPLICATION

Application questions require the candidate to use information and ideas in a concrete situation. Other higher-order questions, such as those involving analysis or synthesis, require application as a part of the thinking process. Application questions measure the ability to:

- Use information in a new context.
- Solve problems that require skills or knowledge.

ANALYSIS

Analysis questions require the candidate to break down information and to explore the relationship between ideas. These questions measure the ability to:

- Identify patterns.
- Distinguish fact from opinion.
- Recognize hidden or unstated meaning.
- Identify cause and effect relationships.
- Make a series of related inferences.

SYNTHESIS

Synthesis questions require the candidate to produce information in the form of hypotheses, theories, stories, or compositions. Synthesis questions require the candidate to bring together pieces of information to create new ideas or thoughts. Synthesis questions measure the ability to:

- Use old ideas to create new ones.
- Make generalizations based on given facts.
- Relate knowledge from a variety of areas.
- Make predictions based on information provided.

EVALUATION

Evaluation questions require the candidate to make judgments about the validity and reliability of information based on criteria provided or assumed. These questions measure the candidate's ability to:

- Compare and discriminate among ideas.
- Assess the value of theories, evidence, and presentations.
- Make choices based on reasoned argument.
- Recognize the role that values play in beliefs and decision making.
- Indicate logical fallacies in arguments.

APPENDIX D

FINAL FORMS REVIEW SHEET

GED TEST: _____

FORMS _____

REVIEWER: _____ TEST FORM: _____
DATE: _____

Please review the enclosed GED Tests and respond to the following questions. These responses will be the basis for the forthcoming committee discussion. You are encouraged to critique this information fully and to bring any concerns to the attention of the committee and the test specialist.

OVERALL IMPRESSION:

CONTENT:

1. Does the test material represent the content and skills considered to be among the lasting outcomes of a high school education?

   Yes_____ No_____

   Comments:

2. Are the reading materials at an appropriate level of difficulty for high school seniors?

   Yes_____ No_____

3. Do any of the stimulus materials or test items portray any group unfavorably or stereotype individuals according to their age, gender, race, religion, or nationality?

   Yes_____ No_____

   Comments:

CONTEXT:

4. Are the item contexts appropriate for and likely to be familiar to adults?

   Yes_____ No_____

   Comments:

COHERENCE:

5. Does the impression left by the test as a whole suggest unity and clarity of purpose?

   Yes_____ No_____

   Comments:

In hopes of assisting you in evaluating the forms, I suggest that in thinking about Question #5 of the Final Forms Review Report you ask yourself these sub-questions:

(1) Is there anything about the test that would indicate that irrelevant sources of difficulty would seriously influence examinees' efforts on the test? Cite examples.

(2) Does the test "flow" from one area of questioning to the next, minimizing abrupt transitions? Cite examples.

(3) Does the test progress from easier to more difficult questions? Cite examples.

APPENDIX E

OFFICIAL GED PRACTICE TEST ESSAY TOPICS

Suppose you had the opportunity to teach something you know to someone else.

In your essay, identify what you would teach and explain how you would teach this. Use your personal observations, experience, and knowledge to support your essay.

Our opinions may change over a period of time.

Identify an opinion you once held but that you have given up or changed. Write an essay explaining how and why the change occurred. Use your personal observations, experience, and knowledge to support your essay.

What is one important goal you would like to achieve in the next few years?

In your essay, identify that one goal and explain how you plan to achieve it. Use your personal observations, experience, and knowledge to support your essay.

Everyone has at least one "rule" to live by.

In your essay, identify one rule that you believe is important to follow. Explain your reasons for following that rule. Use your personal observations, experience, and knowledge to support your essay.

We all have different views of what it means to be a successful person.

In your essay, identify someone whom you consider successful. Explain what qualities make that person a "success" in your view. Use your personal observations, experience, and knowledge to support your essay.

APPENDIX F

GED ESSAY TOPICS: SOME SPECIFICATIONS

Essay topics, like all other items to be used on the GED Tests, undergo a rigorous scrutiny before they are included on a final form. The process begins with the writing of "raw" topics that address the test specifications. In the case of the essay topics to be used on the Language Arts, Writing Test, the following general specifications must be met:

- The topic must be based upon information or a situation that is general enough to be familiar to most examinees. The topic must be accessible to as many examinees as possible—ideally to all.
- The topic must offer an idea that examinees view as worth writing about. In this sense, the situation defined by the topic's context should be realistic. It would be a simple matter to construct contexts consisting of hypothetical issues, but these could not be expected to fully engage the writers. The purpose of the topic is to elicit a sample of writing that displays the examinee's strengths; the kind of topic most likely to do this is one that presents a situation about which the examinee already has an opinion or some feelings, or can form them readily.
- The topic should not elicit an overly emotional response from the writer. While the situation provided as a prompt should trigger the writer's interest in some way, it should not go too far. Of course, the larger the population writing on a topic, the greater the chance that the topic will elicit an emotional response from someone. Emotional writing is rarely controlled writing and thus does not present an example of an examinee's best writing skill. For this reason, emotionally charged issues are not used as writing prompts.
- The topic should be clearly stated and contain only the amount of information necessary to provide the prompt for writing. When examinees finish reading the topic they should know exactly what they are being asked to write about. The objective is to state the topic clearly enough that students immediately know what they are going to write and spend their time working out how they are going to write it.

Only topics that meet these specifications become eligible for field-testing with high school seniors. After topics are field-tested, the papers written on those topics are holistically scored by experienced readers who use the 2002 Series GED Writing Test Official Essay Scoring Guide. This type of "scoring session" differs significantly from one concerned solely with producing scores for essays. In this "topic selection reading," readers are asked not only to score the papers using the Scoring Guide, but also to evaluate how well the topic is working. Readers evaluate topics according to the following specifications:

- The topic should elicit papers with characteristics comparable to those described in the essay scoring guide. If, for example, the best papers written on a field-tested topic seem significantly less accomplished than those described in the four categories in the 2002 Series GED Writing Test Official Scoring Guide, the topic may be more difficult for writers than desirable and therefore unusable. In the interests of making all of the topics as equal in difficulty as possible, all papers produced by the topics must conform to the 2002 Series GED Writing Test Official Scoring Guide. If the papers exhibit characteristics that are clearly different from those described in the 2002 Series GED Writing Test Official Scoring Guide, the topic will be rejected; the 2002 Series GED Writing Test Official Scoring Guide will not be revised to conform to a new topic.
- The topic should elicit papers that illustrate the full range of student writing ability. A good topic allows strong writers to display their skill in writing, yet still allows weak writers access to the question. Because the field tests involve significant numbers of high school seniors, the total pool of papers produced for a topic should include essays at all points on the four-point scale.
- The topic should elicit papers that clearly address the question provided. If papers written on a topic consistently show that writers are unclear about the question asked, or if an unusual number of writers are writing on a topic other than the one asked, the topic is ineffective. A topic that fails to meet this criterion was probably not clearly stated.

- The topic should elicit a variety of responses. To say that all papers must address the topic is not to say that all papers must appear the same. A good topic will yield papers with ideas that mirror the diversity of the population tested. Papers with a wide range of ideas will be less tedious for readers to score, and thus readers are likely to score more accurately. While scoring large numbers of papers inevitably becomes tiresome, a good topic produces papers that engage readers' interest to the extent possible.
- The topic should elicit fully developed responses. A topic that yields a large proportion of incomplete papers may be too demanding for the time allowed. If papers consistently exhibit a shallowness of thought and inadequate development, examinees might not have enough information immediately at hand to write well about the topic.
- The topic should not elicit an emotional or biased response from readers. In the same way that care is taken not to trigger emotional reactions from writers, the resulting papers must not fuel preconceptions or biases among readers. Essay readers are urged to evaluate the writing, not the writer or the writer's values, and the scoring process contains numerous checks of this standard. However, readers cannot be expected to remain immune from emotional reactions so an effective topic produces few papers that are likely to set off an emotional response in readers.
- The topic elicits papers for which readers can easily agree on scores. Where these distinctions are blurred, the topic itself may be at fault. Among the many statistical checks on topic performance is the rate at which readers agree on scores for particular papers; a significant rate of disagreement among readers often indicates that the topic is yielding papers that cannot reliably be scored using the standards provided.

APPENDIX G

2002 SERIES GED WRITING TEST OFFICIAL ESSAY SCORING GUIDE

	1	2	3	4
	INADEQUATE	**MARGINAL**	**ADEQUATE**	**EFFECTIVE**
	Reader has difficulty identifying or following the writer's ideas.	**Reader occasionally has difficulty understanding or following the writer's ideas.**	**Reader understands writer's ideas.**	**Reader understands and easily follows the writer's expression of ideas.**
Response to the Prompt	Attempts to address prompt but with little or no success in establishing a focus.	Addresses the prompt, though the focus may shift.	Uses the writing prompt to establish a main idea.	Presents a clearly focused main idea that addresses the prompt.
Organization	Fails to organize ideas.	Shows some evidence of an organizational plan.	Uses an identifiable organizational plan.	Establishes a clear and logical organization.
Development and Details	Demonstrates little or no development; usually lacks details or examples or presents irrelevant information.	Has some development but lacks specific details; may be limited to a listing, repetitions, or generalizations.	Has focused but occasionally uneven development; incorporates some specific detail.	Achieves coherent development with specific and relevant details and examples.
Conventions of EAE	May exhibit minimal or no control of sentence structure and the conventions of EAE	May demonstrate inconsistent control of sentence structure and the conventions of EAE.	Generally controls sentence structure and the conventions of EAE.	Consistently controls sentence structure and the conventions of Edited American English (EAE).
Word Choice	Exhibits weak and/or inappropriate words.	Exhibits a narrow range of word choice, often including inappropriate selections.	Exhibits appropriate word choice.	Exhibits varied and precise word choice.

APPENDIX H

SAMPLE SIZE (N), STANDARD ERROR OF MEASUREMENT (SEM), AND K-R 20 ESTIMATES FOR THE 2002 SERIES ENGLISH-LANGUAGE GED TESTS: ADULT GED EXAMINEE DATA

TEST/FORM	N	SEM	K-R 20
Language Arts, Writing			
Form IA	151,885	2.88	.88
Form IB	143,724	2.81	.88
Form IC	133,385	2.86	.89
Form ID	160,997	2.63	.87
Form IE	154,916	2.78	.88
Form IF	151,092	2.72	.88
Form IG	167,048	2.72	.88
Form IH	160,599	2.74	.87
Form II	159,340	2.62	.86
Form IJ	124,479	2.79	.95
Form IK	124,430	2.75	.96
Social Studies			
Form IA	148,879	2.83	.88
Form IB	141,159	2.85	.91
Form IC	132,457	2.88	.90
Form ID	162,807	2.78	.90
Form IE	158,162	2.88	.89
Form IF	154,944	2.80	.89
Form IG	164,828	2.73	.90
Form IH	158,580	2.82	.90
Form II	154,937	2.79	.89
Form IJ	124,910	2.95	.93
Form IK	124,911	2.87	.94
Science			
Form IA	147,046	2.73	.89
Form IB	139,244	2.80	.88
Form IC	130,372	2.78	.89
Form ID	160,407	2.80	.89
Form IE	156,091	2.71	.88
Form IF	149,171	2.83	.89
Form IG	163,675	2.70	.88
Form IH	158,084	2.77	.87
Form II	153,241	2.70	.88
Form IJ	121,895	2.78	.95
Form IK	121,482	2.66	.96
Language Arts, Reading			
Form IA	149,998	2.38	.90
Form IB	143,501	2.36	.89
Form IC	134,939	2.47	.86
Form ID	163,747	2.43	.86
Form IE	157,275	2.31	.86
Form IF	152,420	2.33	.86
Form IG	164,810	2.27	.85
Form IH	158,801	2.08	.88
Form II	153,720	2.34	.86
Form IJ	124,525	2.36	.95
Form IK	124,963	2.42	.94
Mathematics			
Form IA	137,593	2.97	.91
Form IB	133,085	2.89	.92
Form IC	122,271	2.97	.91
Form ID	153,141	3.05	.90
Form IE	152,685	3.04	.90
Form IF	152,022	2.99	.90
Form IG	165,178	3.09	.90
Form IH	159,698	2.97	.91
Form II	154,970	3.01	.90
Form IJ	109,230	2.93	.94
Form IK	109,380	2.93	.94

Note: Data were obtained via the U.S. English print edition only, during 2002, 2003, 2004, and 2006. Only those candidates who indicated that GEDTS may use their data for research purposes are included here.

APPENDIX I

STANDARD SCORE CONDITIONAL STANDARD ERRORS OF MEASUREMENT AT VARIOUS STANDARD SCORES FOR THE 2002 SERIES ENGLISH-LANGUAGE GED TESTS: ADULT GED EXAMINEE DATA

TEST/FORM	400	410	420	430	440	450	460
Social Studies							
Form IA	25.1	25.2	25.1	25.0	24.8	24.7	28.2
Form IB	24.8	29.2	25.1	25.1	25.0	24.8	24.6
Form IC	25.1	25.2	25.4	25.4	25.4	25.2	24.9
Form ID	21.2	25.3	25.0	24.7	28.5	23.8	23.3
Form IE	21.2	25.4	25.3	25.2	24.8	28.6	24.2
Form IF	24.9	25.1	25.1	25.1	24.9	24.8	24.6
Form IG	20.9	25.1	25.2	25.1	25.0	24.8	24.6
Form IH	24.8	29.1	25.0	25.0	24.8	24.5	24.3
Form II	32.1	24.3	24.4	24.6	24.7	24.8	24.7
Form IJ	29.8	25.6	25.5	25.4	25.3	25.1	24.8
Form IK	29.8	25.4	25.2	25.0	24.8	28.6	24.2
Science							
Form IA	21.1	21.2	17.0	16.8	20.8	20.4	20.1
Form IB	20.8	20.9	20.8	20.6	20.4	20.0	19.8
Form IC	21.1	16.9	21.1	20.9	20.7	20.3	20.0
Form ID	25.2	21.0	16.8	20.9	16.7	20.7	20.6
Form IE	25.3	21.1	16.9	21.1	20.9	20.7	16.4
Form IF	33.9	21.2	16.9	21.1	16.8	20.9	20.8
Form IG	25.0	20.7	20.5	16.3	20.1	19.8	23.4
Form IH	24.8	16.4	20.4	20.3	16.1	19.8	19.6
Form II	20.8	20.7	16.5	20.5	20.4	16.2	20.0
Form IJ	25.8	21.5	17.2	17.1	21.3	21.1	20.7
Form IK	25.5	17.0	20.9	16.6	20.5	23.9	23.5
Language Arts, Reading							
Form IA	22.9	22.8	22.7	18.5	21.9	21.5	23.9
Form IB	23.2	23.1	23.0	22.8	22.2	21.8	21.3
Form IC	22.8	22.9	19.0	22.8	22.4	18.5	21.8
Form ID	22.5	18.7	22.4	22.3	22.1	18.2	21.5
Form IE	19.1	18.7	22.2	21.8	21.3	24.2	26.8
Form IF	19.2	22.9	22.8	18.6	22.0	21.6	21.1
Form IG	11.3	18.9	22.4	22.2	21.6	21.2	20.8
Form IH	15.3	22.6	22.0	21.6	24.0	26.6	32.1
Form II	19.1	23.0	23.0	22.9	22.8	22.6	18.6
Form IJ	22.7	22.1	21.7	21.2	24.1	26.7	32.2
Form IK	26.2	22.1	21.7	21.2	24.1	26.8	32.3
Mathematics							
Form IA	24.4	24.7	24.7	24.7	24.7	24.6	24.4
Form IB	15.9	24.0	24.4	24.5	24.4	24.2	23.5
Form IC	20.3	28.6	24.7	24.7	24.5	24.4	24.0
Form ID	16.3	24.6	25.1	25.2	25.1	24.8	24.1
Form IE	24.7	25.0	25.0	25.0	25.0	24.9	24.7
Form IF	20.9	25.0	25.0	24.7	20.4	24.3	24.0
Form IG	25.3	25.4	25.4	25.4	25.3	20.8	24.7
Form IH	21.0	25.2	25.2	25.1	24.9	24.5	24.2
Form II	28.7	24.7	24.7	24.7	24.7	24.6	24.4
Form IJ	25.1	25.1	25.1	25.1	25.0	24.8	20.5
Form IK	21.1	25.3	25.3	25.1	25.0	20.7	24.6

Note: Data were obtained via the U.S. English print edition only, during 2002, 2003, 2004, and 2006. Only those candidates who indicated that GEDTS may use their data for research purposes are included here.

APPENDIX J

TABLE J.1
2002 SYSTEMATIC SITE MONITORING RESULTS (FOUR-POINT HOLISTIC SCORING) FOR ENGLISH-LANGUAGE GED ESSAYS

		AGREEMENT OF SCORING SITE'S ESSAY SCORES WITH GEDTS WRITING ADVISORY COMMITTEE SCORES			
SITE	Number of Readers	% Scores Equal	% Scores Within One Point	% Scores Differing by > One Point	Correlation*
1	8	78.1	99.4	0.7	.94
2	4	65.6	100	0.0	.91
3	14	75.2	99.5	0.5	.94
4	11	80.0	100	0.0	.91
5	5	77.5	100	0.0	.92
6	18	72.2	99.4	0.6	.94
7	7	71.4	98.9	1.1	.94
8	5	80.5	100	0.0	.92
9	10	76.5	100	0.0	.94
10	6	80.0	100	0.0	.91
11	3	68.3	100	0.0	.89
12	12	73.8	99.8	0.2	.95
13	5	79.0	100	0.0	.94
14	9	70.0	98.9	1.1	.92
15	34	79.6	100	0.0	.96
16	7	64.6	99.6	0.4	.92
17	16	70.8	98.9	1.1	.93
18	7	84.6	100	0.0	.92
19	2	68.8	100	0.0	.89
20	8	71.3	100	0.0	.93
21	3	65.8	97.5	2.5	.84
22	7	85.0	100	0.0	.91
23	4	69.4	99.4	0.6	.93
24	6	81.7	100	0.0	.92
Mean	8.8	74.6	99.6	0.4	.92
Median	7	74.5	100	0.0	.92
Minimum	2	64.6	97.5	0.0	.84
Maximum	34	85.0	100	2.5	.96

* Pearson correlation between readers' scores and GEDTS Writing Advisory Committee scores.

TABLE J.2
2003 SYSTEMATIC SITE MONITORING RESULTS (FOUR-POINT HOLISTIC SCORING) FOR ENGLISH-LANGUAGE GED ESSAYS

		AGREEMENT OF SCORING SITE'S ESSAY SCORES WITH GEDTS WRITING ADVISORY COMMITTEE SCORES			
SITE	Number of Readers	% Scores Equal	% Scores Within One Point	% Scores Differing by > One Point	Correlation*
1	3	80.0	100	0.0	.91
2	17	72.1	99.6	0.4	.92
3	11	76.8	100	0.0	.91
4	4	72.5	99.4	0.6	.85
5	14	72.7	100	0.0	.91
6	7	68.6	98.9	1.1	.90
7	5	73.0	100	0.0	.84
8	10	67.8	100	0.0	.90
9	6	75.4	100	0.0	.89
10	3	57.5	99.2	0.8	.86
11	8	74.7	99.7	0.3	.92
12	5	77.0	99.5	0.5	.89
13	9	71.7	99.7	0.3	.90
14	32	77.6	100	0.0	.93
15	5	74.5	99.0	1.0	.92
16	15	70.0	99.4	0.6	.91
17	7	78.9	100	0.0	.90
18	3	75.8	99.2	0.8	.90
19	8	70.6	99.7	0.3	.90
20	3	67.5	100	0.0	.90
21	5	79.5	97.5	2.5	.84
22	4	63.8	98.1	1.9	.83
23	4	79.4	100	0.0	.89
24	3	72.5	100	0.0	.90
25	8	75.6	99.4	0.6	.92
Mean	8.0	73.0	99.5	0.5	.89
Median	6	73.0	99.7	0.3	.90
Minimum	3	57.5	97.5	0.0	.83
Maximum	32	80.0	100	2.5	.93

* Pearson correlation between readers' scores and GEDTS Writing Advisory Committee scores.

TABLE J.3
2004 SYSTEMATIC SITE MONITORING RESULTS (FOUR-POINT HOLISTIC SCORING) FOR ENGLISH-LANGUAGE GED ESSAYS

		AGREEMENT OF SCORING SITE'S ESSAY SCORES WITH GEDTS WRITING ADVISORY COMMITTEE SCORES			
SITE	Number of Readers	% Scores Equal	% Scores Within One Point	% Scores Differing by > One Point	Correlation*
1	15	75.5	99.7	0.3	.95
2	12	89.4	100	0.0	.96
3	3	87.5	100	0.0	.93
4	12	84.2	99.8	0.2	.95
5	6	77.9	100	0.0	.95
6	6	82.1	100	0.0	.90
7	10	79.3	100	0.0	.94
8	6	81.7	100	0.0	.93
9	3	80.8	99.2	0.8	.90
10	10	80.8	100	0.0	.94
11	11	73.6	99.3	0.7	.94
12	9	79.7	100	0.0	.94
13	32	82.8	100	0.0	.95
14	4	81.9	100	0.0	.93
15	19	80.4	99.9	0.1	.96
16	8	76.9	100	0.0	.90
17	4	67.5	100	0.0	.90
18	8	79.7	100	0.0	.95
19	5	71.0	99.5	0.5	.92
20	5	93.5	100	0.0	.96
21	4	73.8	100	0.0	.92
22	4	75.6	100	0.0	.89
23	3	68.3	100	0.0	.88
24	8	81.3	100	0.0	.96
Mean	8.6	79.4	99.9	0.1	.93
Median	7	80.1	100	0.0	.94
Minimum	3	67.5	99.2	0.0	.88
Maximum	32	93.5	100	0.8	.96

* Pearson correlation between readers' scores and GEDTS Writing Advisory Committee scores.

TABLE J.4
2005 SYSTEMATIC SITE MONITORING RESULTS (FOUR-POINT HOLISTIC SCORING) FOR ENGLISH-LANGUAGE GED ESSAYS

		AGREEMENT OF SCORING SITE'S ESSAY SCORES WITH GEDTS WRITING ADVISORY COMMITTEE SCORES			
SITE	Number of Readers	% Scores Equal	% Scores Within One Point	% Scores Differing by > One Point	Correlation*
1	5	82.5	100	0.0	.91
2	15	79.2	99.8	0.2	.98
3	11	83.4	100	0.0	.94
4	5	79.5	99.0	1.0	.95
5	15	78.7	99.7	0.3	.95
6	6	76.3	99.6	0.4	.89
7	9	79.7	100	0.0	.95
8	6	80.8	98.3	1.7	.95
9	3	75.8	100	0.0	.93
10	8	77.5	99.7	0.3	.97
11	9	79.2	99.2	0.8	.95
12	9	81.1	100	0.0	.97
13	27	88.2	100	0.0	.98
14	4	76.9	100	0.0	.94
15	15	80.8	99.8	0.2	.97
16	6	87.1	100	0.0	.94
17	3	75.8	100	0.0	.94
18	7	80.0	100	0.0	.97
19	5	72.5	99.0	1.0	.91
20	5	86.5	100	0.0	.94
21	4	73.8	100	0.0	.93
22	4	86.9	100	0.0	.95
23	3	75.8	100	0.0	.92
24	8	76.6	99.4	0.6	.95
25	6	78.3	99.6	0.4	.97
Mean	7.9	79.7	99.7	0.3	.95
Median	6	79.2	100	0.0	.95
Minimum	3	72.5	98.3	0.0	.89
Maximum	27	88.2	100	1.7	.98

* Pearson correlation between readers' scores and GEDTS Writing Advisory Committee scores.

TABLE J.5
2006 SYSTEMATIC SITE MONITORING RESULTS (FOUR-POINT HOLISTIC SCORING) FOR ENGLISH-LANGUAGE GED ESSAYS

		AGREEMENT OF SCORING SITE'S ESSAY SCORES WITH GEDTS WRITING ADVISORY COMMITTEE SCORES			
SITE	Number of Readers	% Scores Equal	% Scores Within One Point	% Scores Differing by > One Point	Correlation*
1	7	82.1	100	0.0	.86
2	15	79.7	99.7	0.3	.90
3	11	84.1	100	0.0	.91
4	13	82.5	100	0.0	.92
5	6	85.8	100	0.0	.89
6	10	76.0	100	0.0	.90
7	6	85.0	100	0.0	.91
8	3	92.5	51.7	7.5	.62
9	5	79.0	100	0.0	.89
10	9	79.4	100	0.0	.90
11	25	89.1	100	0.0	.92
12	4	81.9	100	0.0	.89
13	14	82.7	100	0.0	.91
14	8	77.8	100	0.0	.87
15	1	90.0	100	0.0	.89
16	5	77.0	100	0.0	.89
17	5	92.0	100	0.0	.92
18	4	80.6	100	0.0	.86
19	3	77.5	100	0.0	.87
20	6	61.3	97.9	2.1	.88
Mean	8	81.8	97.5	0.5	.88
Median	6	82.0	100	0.0	.89
Minimum	1	61.3	51.7	0.0	.62
Maximum	25	92.5	100	7.5	.92

* Pearson correlation between readers' scores and GEDTS Writing Advisory Committee scores.

TABLE J.6
2007 SYSTEMATIC SITE MONITORING RESULTS (FOUR-POINT HOLISTIC SCORING) FOR ENGLISH-LANGUAGE GED ESSAYS

		AGREEMENT OF SCORING SITE'S ESSAY SCORES WITH GEDTS WRITING ADVISORY COMMITTEE SCORES			
SITE	Number of Readers	% Scores Equal	% Scores Within One Point	% Scores Differing by > One Point	Correlation*
1	3	70.0	96.7	3.3	.88
2	15	73.7	99.3	0.7	.93
3	11	70.7	100	0.0	.89
4	1	77.5	100	0.0	.88
5	14	77.9	99.3	0.7	.95
6	6	74.6	95.0	5.0	.83
7	9	66.1	99.7	0.3	.88
8	6	69.2	100	0.0	.86
9	8	74.7	99.7	0.3	.95
10	9	74.2	99.4	0.6	.92
11	26	82.2	100	0.0	.96
12	5	69.5	99.5	0.5	.90
13	15	76.0	99.2	0.8	.94
14	8	73.8	100	0.0	.90
15	4	66.9	100	0.0	.89
16	5	65.0	98.5	1.5	.88
17	5	78.0	100	0.0	.89
Mean	8.8	72.9	99.2	0.8	.90
Median	8	73.8	99.7	0.3	.89
Minimum	1	65.0	95.0	0.0	.83
Maximum	26	82.2	100	5.0	.96

* Pearson correlation between readers' scores and GEDTS Writing Advisory Committee scores.

APPENDIX K

PROBABILITY OF CORRECT CLASSIFICATION, FALSE POSITIVE, AND FALSE NEGATIVE RATES FOR THE 2002 SERIES ENGLISH-LANGUAGE GED TESTS: U.S. GED EXAMINEE DATA

TEST/FORM	N	Percent Not Meeting Minimum Score	Percent Meeting Minimum Score	Probability of Correct Classification	False Positive	False Negative
Language Arts, Writing						
Form IA	151,885	13	87	.91	.07	.02
Form IB	143,724	18	82	.90	.07	.02
Form IC	133,385	21	79	.90	.07	.03
Form ID	160,997	14	86	.91	.07	.02
Form IE	154,916	18	82	.91	.07	.03
Form IF	151,092	9	91	.92	.07	.01
Form IG	167,048	23	77	.90	.06	.03
Form IH	160,599	26	74	.90	.07	.03
Form II	159,340	11	89	.91	.07	.02
Form IJ	124,479	37	63	.99	.01	*
Form IK	124,430	38	62	1.00	*	†
Social Studies						
Form IA	148,879	10	90	.90	†	.10
Form IB	141,159	12	88	.88	†	.12
Form IC	132,457	9	91	.91	†	.09
Form ID	162,807	16	84	.84	†	.16
Form IE	158,162	15	85	.85	†	.15
Form IF	154,944	7	93	.93	†	.07
Form IG	164,828	8	92	.92	†	.08
Form IH	158,580	10	90	.90	†	.10
Form II	154,937	4	96	.96	†	.04
Form IJ	124,910	26	74	.98	.02	†
Form IK	124,911	29	71	.99	.01	†
Science						
Form IA	147,046	7	93	.93	†	.07
Form IB	139,244	9	91	.91	†	.09
Form IC	130,372	12	88	.88	†	.12
Form ID	160,407	10	90	.90	†	.10
Form IE	156,091	6	94	.94	†	.06
Form IF	149,171	11	89	.89	†	.11
Form IG	163,675	11	89	.89	†	.11
Form IH	158,084	12	88	.88	†	.12
Form II	153,241	8	92	.92	†	.08
Form IJ	121,895	21	79	.94	.06	†
Form IK	121,482	27	73	1.00	*	†

Continued on next page

Appendix K continued

TEST/FORM	N	Percent Not Meeting Minimum Score	Percent Meeting Minimum Score	Probability of Correct Classification	False Positive	False Negative
Language Arts, Reading						
Form IA	149,998	12	88	.88	†	.12
Form IB	143,501	10	90	.90	†	.10
Form IC	134,939	7	93	.93	†	.07
Form ID	163,747	10	90	.90	†	.10
Form IE	157,275	10	90	.90	†	.10
Form IF	152,420	8	92	.92	†	.08
Form IG	164,810	4	96	.96	†	.04
Form IH	158,801	5	95	.95	†	.05
Form II	153,720	5	95	.95	†	.05
Form IJ	124,525	26	74	.92	.08	†
Form IK	124,963	30	70	.89	.11	†
Mathematics						
Form IA	137,593	25	75	.79	.21	†
Form IB	133,085	15	85	.85	†	.15
Form IC	122,271	27	73	.73	†	.27
Form ID	153,141	35	65	.85	.15	†
Form IE	152,685	26	74	.83	.17	†
Form IF	152,022	28	72	.70	.30	*
Form IG	165,178	28	72	.65	.35	*
Form IH	159,698	27	73	.73	†	.27
Form II	154,970	27	73	.63	.37	†
Form IJ	109,230	40	60	1.00	†	†
Form IK	109,380	44	56	1.00	†	†

Note: Data were obtained via the U.S. English-language print edition only, during 2002, 2003, 2004, and 2006. Only those candidates who indicated that GEDTS may use their data for research purposes are included here.

* Value is less than 0.01.

† Value is less than 0.001.

APPENDIX L

TABLE L.1
PERCENTAGE OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE EQUATING STUDY AT SELF-REPORTED GRADE LEVELS ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
	N	350	410	450	500
		Language Arts, Writing Test			
Self-reported Grades in English Literature					
Mostly A	305	97	90	84	74
Mostly B	517	92	80	63	44
Mostly C	279	81	58	37	18
Mostly D	31	84	52	35	23
Mostly below D	0	-	-	-	-
		Language Arts, Writing Test			
Self-reported Grades in English Composition					
Mostly A	297	98	92	86	77
Mostly B	548	91	79	62	42
Mostly C	255	81	59	38	18
Mostly D	28	86	50	21	14
Mostly below D	5	†	†	†	†
		Social Studies Test			
Self-reported Grades in Social Studies					
Mostly A	539	95	86	76	65
Mostly B	652	83	65	50	35
Mostly C	264	79	55	40	23
Mostly D	31	74	42	29	10
Mostly below D	2	†	†	†	†
		Science Test			
Self-reported Grades in Science					
Mostly A	243	95	91	86	77
Mostly B	446	88	74	64	46
Mostly C	279	81	60	51	30
Mostly D	28	71	61	43	25
Mostly below D	1	†	†	†	†
		Language Arts, Reading Test			
Self-reported Grades in English Literature					
Mostly A	573	95	88	84	73
Mostly B	917	92	73	62	45
Mostly C	460	84	58	38	21
Mostly D	65	75	43	25	11
Mostly below D	5	†	†	†	†
		Language Arts, Reading Test			
Self-reported Grades in English Composition					
Mostly A	567	95	87	84	73
Mostly B	1,006	91	72	60	43
Mostly C	391	85	57	39	21
Mostly D	50	70	42	26	12
Mostly below D	6	†	†	†	†
		Mathematics Test			
Self-reported Grades in Mathematics					
Mostly A	489	98	91	85	76
Mostly B	684	91	77	65	45
Mostly C	445	84	63	45	24
Mostly D	112	72	43	26	16
Mostly below D	4	†	†	†	†

† Indicates that the statistic was not calculated because of small sample size.

TABLE L.2
PERCENTAGE OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2003 ENGLISH-LANGUAGE EQUATING STUDY AT SELF-REPORTED GRADE LEVELS ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
	N	350	410	450	500
		Language Arts, Writing Test			
Self-reported Grades in English Literature					
Mostly A	494	97	90	81	67
Mostly B	647	91	73	55	35
Mostly C	337	80	53	33	20
Mostly D	50	60	34	20	10
Mostly below D	2	†	†	†	†
		Language Arts, Writing Test			
Self-reported Grades in English Composition					
Mostly A	470	97	89	80	66
Mostly B	619	91	75	57	38
Mostly C	294	80	51	32	19
Mostly D	51	61	29	14	6
Mostly below D	6	†	†	†	†
		Social Studies Test			
Self-reported Grades in Social Studies					
Mostly A	813	92	84	77	64
Mostly B	966	85	68	58	37
Mostly C	478	77	52	39	21
Mostly D	56	61	21	13	7
Mostly below D	4	†	†	†	†
		Science Test			
Self-reported Grades in Science					
Mostly A	613	90	80	75	64
Mostly B	1,008	81	63	51	38
Mostly C	570	70	47	35	22
Mostly D	91	54	38	27	18
Mostly below D	3	†	†	†	†
		Language Arts, Reading Test			
Self-reported Grades in English Literature					
Mostly A	670	96	90	81	65
Mostly B	980	91	78	65	44
Mostly C	564	84	60	44	25
Mostly D	87	76	52	38	15
Mostly below D	9	†	†	†	†
		Language Arts, Reading Test			
Self-reported Grades in English Composition					
Mostly A	649	96	90	81	64
Mostly B	865	92	78	66	46
Mostly C	499	86	61	45	26
Mostly D	78	78	53	37	13
Mostly below D	15	†	†	†	†
		Mathematics Test			
Self-reported Grades in Mathematics					
Mostly A	673	95	89	84	74
Mostly B	1,016	91	78	65	47
Mostly C	767	81	57	41	22
Mostly D	182	71	42	21	11
Mostly below D	10	†	†	†	†

† Indicates that the statistic was not calculated because of small sample size.

TABLE L.3
PERCENTAGE OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2005 ENGLISH-LANGUAGE EQUATING STUDY AT SELF-REPORTED GRADE LEVELS ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
	N	350	410	450	500
		Language Arts, Writing Test			
Self-reported Grades in English [a]					
Mostly A	891	97	93	85	70
Mostly B	1,000	93	76	60	40
Mostly C	478	86	56	36	19
Mostly D	50	78	44	18	14
Mostly below D	4	†	†	†	†
		Social Studies Test			
Self-reported Grades in Social Studies					
Mostly A	928	93	87	81	71
Mostly B	1,044	85	68	54	35
Mostly C	433	73	53	37	22
Mostly D	47	72	45	32	17
Mostly below D	0	-	-	-	-
		Science Test			
Self-reported Grades in Science					
Mostly A	637	91	86	82	71
Mostly B	1,118	84	74	62	44
Mostly C	568	73	56	46	28
Mostly D	73	64	44	25	15
Mostly below D	2	†	†	†	†
		Language Arts, Reading Test			
Self-reported Grades in English [a]					
Mostly A	799	98	93	88	73
Mostly B	1,066	93	79	67	44
Mostly C	548	84	64	45	25
Mostly D	69	71	46	35	20
Mostly below D	2	†	†	†	†
		Language Arts, Reading Test			
Self-reported Grades in Mathematics					
Mostly A	739	96	92	88	82
Mostly B	1,012	92	80	70	55
Mostly C	648	84	62	45	26
Mostly D	138	71	43	27	14
Mostly below D	4	†	†	†	†

† Indicates that the statistic was not calculated because of small sample size.
[a] English Literature and English Composition were combined into a single English category in 2005.

APPENDIX M

TABLE M.1
PERCENTAGE OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE EQUATING STUDY AT SELF-REPORTED TOTAL YEARS OF STUDY ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
SELF-REPORTED TOTAL YEARS OF STUDY	N	350	410	450	500
		Language Arts, Writing Test			
English Literature					
1 year or less	72	83	74	56	32
2 years	133	92	72	59	40
3 years	130	89	72	52	39
4 years or more	791	91	79	65	48
		Language Arts, Writing Test			
English Composition					
1 year or less	164	87	73	55	38
2 years	171	89	74	58	37
3 years	68	94	76	47	35
4 years or more	678	92	80	67	51
		Social Studies Test			
Social Studies					
1 year or less	35	80	43	34	29
2 years	152	81	61	46	32
3 years	582	85	68	53	37
4 years or more	695	90	76	65	51
		Science Test			
Science					
1 year or less	11	†	†	†	†
2 years	107	84	62	48	24
3 years	450	86	70	60	42
4 years or more	416	91	82	78	63
		Language Arts, Reading Test			
English Literature					
1 year or less	175	89	58	47	29
2 years	253	90	78	66	51
3 years	240	92	72	59	45
4 years or more	1,318	91	75	64	48
		Language Arts, Reading Test			
English Composition					
1 year or less	300	91	70	59	40
2 years	289	88	75	64	52
3 years	146	92	71	58	41
4 years or more	1,128	92	76	65	49
		Mathematics Test			
Mathematics					
1 year or less	4	†	†	†	†
2 years	101	87	64	51	27
3 years	625	86	68	54	35
4 years or more	957	93	82	71	56

† Indicates that the statistic was not calculated because of small sample size.

TABLE M.2
PERCENTAGE OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2003 ENGLISH-LANGUAGE EQUATING STUDY AT SELF-REPORTED TOTAL YEARS OF STUDY ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
SELF-REPORTED TOTAL YEARS OF STUDY	N	350	410	450	500
		Language Arts, Writing Test			
English Literature					
1 year or less	86	90	70	56	35
2 years	180	90	72	56	39
3 years	157	89	69	57	46
4 years or more	1,150	89	73	57	41
		Language Arts, Writing Test			
English Composition					
1 year or less	192	88	72	60	47
2 years	205	89	71	54	37
3 years	121	90	67	51	34
4 years or more	976	90	74	59	42
		Social Studies Test			
Social Studies					
1 year or less	45	73	42	40	20
2 years	248	77	58	48	35
3 years	1,053	84	67	57	39
4 years or more	1,056	87	73	64	47
		Science Test			
Science					
1 year or less	25	†	†	†	†
2 years	325	64	43	35	26
3 years	1,018	78	60	48	35
4 years or more	997	84	70	60	48
		Language Arts, Reading Test			
English Literature					
1 year or less	138	89	68	54	38
2 years	261	91	76	64	44
3 years	281	86	72	61	43
4 years or more	1,709	90	76	63	44
		Language Arts, Reading Test			
English Composition					
1 year or less	382	89	73	58	40
2 years	314	91	80	65	46
3 years	170	90	69	58	44
4 years or more	1,363	91	77	65	46
		Mathematics Test			
Mathematics					
1 year or less	9	†	†	†	†
2 years	169	81	54	39	24
3 years	976	83	64	49	30
4 years or more	1,591	90	78	68	54

† Indicates that the statistic was not calculated because of small sample size.

TABLE M.3
PERCENTAGE OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2005 ENGLISH-LANGUAGE EQUATING STUDY AT SELF-REPORTED TOTAL YEARS OF STUDY ACHIEVING SELECTED GED STANDARD SCORES OR HIGHER

		GED Standard Score ≥			
SELF-REPORTED TOTAL YEARS OF STUDY	N	350	410	450	500
		Language Arts, Writing Test			
English					
1 year or less	3	†	†	†	†
2 years	19	†	†	†	†
3 years	148	88	67	47	28
4 years or more	2,312	93	78	64	48
		Social Studies Test			
Social Studies					
1 year or less	22	†	†	†	†
2 years	169	82	63	48	31
3 years	1,042	82	68	56	41
4 years or more	1,294	88	76	65	52
		Science Test			
Science					
1 year or less	11	†	†	†	†
2 years	274	76	61	50	30
3 years	1,086	80	67	57	41
4 years or more	1,101	87	78	70	55
		Language Arts, Reading Test			
English					
1 year or less	9	†	†	†	†
2 years	11	†	†	†	†
3 years	149	87	60	45	24
4 years or more	2,381	92	80	69	50
		Mathematics Test			
Mathematics					
1 year or less	9	†	†	†	†
2 years	126	79	54	41	27
3 years	825	86	71	56	38
4 years or more	1,651	93	81	73	62

† Indicates that the statistic was not calculated because of small sample size.

APPENDIX N

TABLE N.1
AVERAGE GED STANDARD SCORES OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE EQUATING STUDY, BY YEARS OF INSTRUCTION IN CONTENT AREA

Years Instruction in Subject Area	English Composition	Social Studies	Science	English Literature	Mathematics
1 year or less	473	†	†	453	†
	(164)	(35)	(11)	(175)	(4)
2 years	486	438	440	507	440
	(171)	(152)	(107)	(253)	(101)
3 years	469	460	469	489	455
	(68)	(582)	(450)	(240)	(625)
4 years or more	508	492	524	502	514
	(678)	(695)	(416)	(1,318)	(967)

† Indicates that the statistic was not calculated because of small sample size.
Note: Numbers in parentheses refer to the number of seniors. Averages for the Writing Test are based on the numbers of years of instruction in English composition; averages for the Reading Test are based on the number of years of instruction in English literature.

TABLE N.2
AVERAGE GED STANDARD SCORES OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2003 ENGLISH-LANGUAGE EQUATING STUDY, BY YEARS OF INSTRUCTION IN CONTENT AREA

Years Instruction in Subject Area	English Composition	Social Studies	Science	English Literature	Mathematics
1 year or less	484	†	†	478	†
	(192)	(45)	(25)	(138)	(9)
2 years	469	446	407	503	428
	(205)	(248)	(325)	(261)	(169)
3 years	458	463	449	488	446
	(121)	(1,053)	(1,018)	(281)	(976)
4 years or more	488	488	482	502	509
	(976)	(1,056)	(997)	(1,709)	(1,591)

† Indicates that the statistic was not calculated because of small sample size.
Note: Numbers in parentheses refer to the number of seniors. Averages for the Writing Test are based on the numbers of years of instruction in English composition; averages for the Reading Test are based on the number of years of instruction in English literature.

TABLE N.3
AVERAGE GED STANDARD SCORES OF U.S. GRADUATING HIGH SCHOOL SENIORS IN 2005 ENGLISH-LANGUAGE EQUATING STUDY, BY YEARS OF INSTRUCTION IN CONTENT AREA

Years Instruction in Subject Area	English	Social Studies	Science	English	Mathematics
1 year or less	†	†	†	†	†
	(3)	(22)	(11)	(9)	(9)
2 years	†	447	441	†	430
	(19)	(169)	(274)	(11)	(126)
3 years	447	467	462	448	461
	(148)	(1,042)	(1,086)	(149)	(825)
4 years or more	497	496	507	515	519
	(2,312)	(1,294)	(1,101)	(2,381)	(1,651)

† Indicates that the statistic was not calculated because of small sample size.
Note: Numbers in parentheses refer to the number of seniors. Averages for the Writing Test and Reading Test are based on the number of years instruction in English.

APPENDIX O

TABLE O.1
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, WRITING TEST, BY INSTRUCTION IN GRAMMAR AND LANGUAGE COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Grammar/Composition	Taken	504	945	92	80	65	48
	Not Taken	441	188	81	61	45	29
Spanish	Taken	499	705	93	80	64	46
	Not Taken	484	428	86	72	58	43
French	Taken	523	177	93	81	70	57
	Not Taken	488	956	90	76	60	43
German	Taken	507	70	96	86	69	49
	Not Taken	492	1,063	90	76	61	45
Latin	Taken	571	33	97	91	85	76
	Not Taken	491	1,100	90	76	61	44

TABLE O.2
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2003 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, WRITING TEST, BY INSTRUCTION IN GRAMMAR AND LANGUAGE COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Grammar/Composition	Taken	489	1,262	90	75	60	43
	Not Taken	446	333	84	62	44	29
Spanish	Taken	489	1,019	92	77	60	43
	Not Taken	464	576	84	64	50	35
French	Taken	491	272	92	75	61	44
	Not Taken	477	1,323	88	72	56	40
German	Taken	497	70	91	74	60	46
	Not Taken	479	1,525	89	72	56	40
Latin	Taken	566	34	91	85	79	71
	Not Taken	478	1,561	89	72	56	40

TABLE O.3
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2005 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, WRITING TEST, BY INSTRUCTION IN GRAMMAR AND LANGUAGE COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Creative Writing	Taken	502	742	93	80	65	49
	Not Taken	490	1,753	92	77	62	45
Journalism	Taken	511	301	95	81	68	53
	Not Taken	491	2,194	92	77	62	45
Language Study	Taken	497	1,924	93	79	64	48
	Not Taken	479	571	91	73	59	41
Speech/Debate	Taken	510	802	94	82	68	51
	Not Taken	486	1,693	92	75	61	44
Technical/Business Writing	Taken	488	189	90	75	62	46
	Not Taken	494	2,306	93	78	63	46

TABLE 0.4
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON SOCIAL STUDIES TEST, BY INSTRUCTION IN SELECTED SOCIAL STUDIES COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Behavioral Science	Taken	507	388	93	82	73	60
	Not Taken	459	1,100	84	66	52	37
Civics	Taken	465	478	84	67	53	40
	Not Taken	475	1,010	88	72	59	45
Economics	Taken	468	797	85	69	57	43
	Not Taken	476	691	89	71	58	43
Geography	Taken	469	828	86	70	58	43
	Not Taken	474	660	87	70	57	43
Political Science	Taken	517	211	92	83	73	58
	Not Taken	464	1,277	86	68	55	40
History	Taken	480	1,271	88	73	61	47
	Not Taken	422	217	81	52	37	21
World History	Taken	473	1,206	87	70	58	44
	Not Taken	464	282	84	69	55	40

TABLE 0.5
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2003 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON SOCIAL STUDIES TEST, BY INSTRUCTION IN SELECTED SOCIAL STUDIES COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Behavioral Science	Taken	511	629	92	82	73	56
	Not Taken	456	1,811	82	64	53	37
Civics	Taken	472	788	85	68	58	42
	Not Taken	470	1,652	84	68	58	42
Economics	Taken	470	1,233	84	68	58	42
	Not Taken	471	1,207	86	69	58	42
Geography	Taken	471	1,386	85	69	58	41
	Not Taken	470	1,054	84	68	59	43
Political Science	Taken	497	459	87	77	67	51
	Not Taken	464	1,981	84	66	56	40
History	Taken	478	2,128	86	71	61	45
	Not Taken	419	312	76	53	39	22
World History	Taken	472	1,975	85	69	59	43
	Not Taken	464	465	84	65	53	38

TABLE 0.6
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2005 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON SOCIAL STUDIES TEST, BY INSTRUCTION IN SELECTED SOCIAL STUDIES COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Behavioral Science	Taken	525	568	93	85	75	64
	Not Taken	466	1,979	83	67	55	40
Government/Civics	Taken	480	2,029	85	72	60	46
	Not Taken	476	518	83	68	58	45
Economics	Taken	479	1,327	85	71	60	45
	Not Taken	480	1,220	85	71	59	46
Geography	Taken	472	1,304	85	70	58	43
	Not Taken	487	1,243	86	72	62	48
Political Science	Taken	498	163	89	80	69	55
	Not Taken	478	2,384	85	70	59	45
History	Taken	491	2,133	87	74	64	50
	Not Taken	420	414	73	53	38	24
World History	Taken	483	2,109	85	72	61	46
	Not Taken	462	438	82	65	55	41

TABLE 0.7
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON SCIENCE TEST, BY INSTRUCTION IN SELECTED SCIENCE COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Biology	Taken	491	909	88	75	67	50
	Not Taken	442	88	78	61	48	32
Chemistry	Taken	515	597	91	80	75	60
	Not Taken	444	400	82	63	52	32
Earth Science	Taken	471	352	85	70	63	42
	Not Taken	495	645	88	75	66	52
General Science	Taken	490	272	87	76	65	50
	Not Taken	485	725	87	72	65	48
Genetics	Taken	504	34	85	79	71	62
	Not Taken	486	963	87	73	65	48
Physical Science	Taken	478	542	88	72	63	43
	Not Taken	496	455	86	75	68	55
Physics	Taken	558	259	97	91	87	75
	Not Taken	461	738	84	67	58	39
Zoology/Botany	Taken	489	42	86	79	74	50
	Not Taken	486	955	87	73	65	48

TABLE 0.8
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2003 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON SCIENCE TEST, BY INSTRUCTION IN SELECTED SCIENCE COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Biology	Taken	459	2,242	80	63	52	40
	Not Taken	408	157	60	45	36	21
Chemistry	Taken	479	1,560	84	70	60	47
	Not Taken	413	839	68	46	35	23
Earth Science	Taken	450	957	77	60	49	36
	Not Taken	460	1,442	80	63	53	41
General Science	Taken	447	589	78	60	48	36
	Not Taken	459	1,810	79	62	52	40
Genetics	Taken	519	71	92	76	72	62
	Not Taken	454	2,328	78	61	50	38
Physical Science	Taken	455	1,334	80	61	51	38
	Not Taken	458	1,065	77	62	52	41
Physics	Taken	492	690	84	70	63	53
	Not Taken	441	1,709	76	58	46	33
Zoology/Botany	Taken	456	132	83	64	50	36
	Not Taken	456	2,267	78	61	51	39

TABLE 0.9
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2005 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON SCIENCE TEST, BY INSTRUCTION IN SELECTED SCIENCE COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Biology	Taken	480	2,329	83	72	62	46
	Not Taken	447	169	73	59	50	34
Chemistry	Taken	501	1,684	86	78	69	53
	Not Taken	431	814	73	57	46	29
Earth Science	Taken	453	803	77	65	53	37
	Not Taken	490	1,695	84	74	65	50
Environmental Science	Taken	447	445	76	62	51	36
	Not Taken	485	2,053	83	73	64	47
General Science	Taken	463	1,066	79	67	58	41
	Not Taken	489	1,432	84	74	64	49
Introductory Physics and Chemistry	Taken	489	1,012	84	74	64	49
	Not Taken	471	1,486	80	69	59	43
Physics	Taken	528	754	91	84	75	61
	Not Taken	456	1,744	78	65	55	39

TABLE 0.10
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, READING TEST, BY INSTRUCTION IN SELECTED ENGLISH AND LANGUAGE COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Literature	Taken	500	1,713	91	75	64	48
	Not Taken	464	313	85	63	50	35
European Literature	Taken	423	435	92	81	70	57
	Not Taken	487	1,591	90	70	59	43
World Literature	Taken	502	671	92	76	63	48
	Not Taken	491	1,355	90	71	61	45
Spanish	Taken	506	1,167	92	77	67	51
	Not Taken	480	859	88	67	54	39
French	Taken	517	345	92	79	67	52
	Not Taken	490	1,681	90	72	60	45
German	Taken	516	102	91	78	74	50
	Not Taken	494	1,924	90	72	61	46
Latin	Taken	547	63	95	84	81	57
	Not Taken	493	1,963	90	72	61	46

TABLE 0.11
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2003 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, READING TEST, BY INSTRUCTION IN SELECTED ENGLISH AND LANGUAGE COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Literature	Taken	505	2,158	91	77	64	46
	Not Taken	447	268	82	60	47	26
European Literature	Taken	527	400	94	80	69	52
	Not Taken	493	2,026	89	74	61	42
World Literature	Taken	506	653	92	79	65	45
	Not Taken	495	1,773	89	73	61	43
Spanish	Taken	510	1,468	92	79	67	48
	Not Taken	479	958	86	68	55	38
French	Taken	511	401	88	76	66	50
	Not Taken	496	2,025	90	74	62	43
German	Taken	435	130	93	85	76	60
	Not Taken	496	2,296	89	74	62	43
Latin	Taken	484	40	98	85	80	75
	Not Taken	497	2,386	89	75	62	43

TABLE 0.12
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2005 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON LANGUAGE ARTS, READING TEST, BY INSTRUCTION IN SELECTED ENGLISH AND LANGUAGE COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
American Literature	Taken	533	1,426	94	84	74	55
	Not Taken	481	1,141	88	71	59	38
Language Study	Taken	517	1,955	93	80	70	49
	Not Taken	488	612	87	71	59	43
Technical/Business Writing	Taken	502	237	91	76	64	44
	Not Taken	511	2,330	91	78	68	48
World Literature	Taken	522	1,446	93	81	71	52
	Not Taken	494	1,121	90	75	62	42

TABLE 0.13
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2002 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON MATHEMATICS TEST, BY INSTRUCTION IN SELECTED MATHEMATICS COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Algebra I	Taken	484	1,483	90	75	62	45
	Not Taken	492	419	87	72	63	50
Algebra II	Taken	507	1,312	93	82	71	54
	Not Taken	439	590	82	58	44	29
Business Math	Taken	463	168	85	68	57	36
	Not Taken	488	1,734	90	75	63	47
Calculus	Taken	495	278	99	97	93	86
	Not Taken	467	1,624	88	71	57	39
General Math	Taken	450	413	86	65	50	32
	Not Taken	495	1,489	91	77	66	50
Geometry	Taken	501	1,367	93	80	68	52
	Not Taken	446	535	81	60	47	33
Trigonometry	Taken	559	601	99	95	88	75
	Not Taken	452	1,301	85	65	51	33

TABLE 0.14
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2003 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON MATHEMATICS TEST, BY INSTRUCTION IN SELECTED MATHEMATICS COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Algebra I	Taken	477	2,414	87	71	59	42
	Not Taken	501	359	84	69	60	50
Algebra II	Taken	502	2,154	91	78	67	51
	Not Taken	405	619	74	47	31	15
Business Math	Taken	462	284	86	66	51	37
	Not Taken	483	2,489	87	72	60	44
Calculus	Taken	582	519	97	92	87	79
	Not Taken	457	2,254	85	66	53	35
General Math	Taken	437	680	81	57	43	27
	Not Taken	495	2,093	89	75	64	48
Geometry	Taken	496	2,296	90	76	65	49
	Not Taken	406	477	73	47	31	16
Trigonometry	Taken	549	917	96	89	81	68
	Not Taken	446	1,856	82	62	48	31

TABLE 0.15
PERCENT OF U.S. HIGH SCHOOL SENIORS IN 2005 ENGLISH-LANGUAGE EQUATING STUDY SCORING AT OR ABOVE STANDARD SCORES ON MATHEMATICS TEST, BY INSTRUCTION IN SELECTED MATHEMATICS COURSES

				GED Standard Score ≥			
Course		Mean	N	350	410	450	500
Algebra I	Taken	485	2,177	89	75	63	49
	Not Taken	543	452	92	83	77	70
Algebra II	Taken	514	2,087	93	82	73	60
	Not Taken	423	542	76	54	38	23
Statistics	Taken	543	310	92	85	81	74
	Not Taken	489	2,319	89	75	63	49
Calculus	Taken	590	571	99	95	91	87
	Not Taken	469	2,058	87	71	58	43
General Math	Taken	450	710	83	64	49	33
	Not Taken	512	1,919	93	81	72	59
Geometry	Taken	507	2,233	92	81	70	57
	Not Taken	425	396	77	52	38	26
Trigonometry	Taken	559	990	97	91	85	77
	Not Taken	456	1,639	85	67	53	37

APPENDIX P

Qualifications for GED Tests Chief Readers and Essay Readers

CHIEF READER QUALIFICATION:

To apply for certification as a GED Tests Chief Reader, a candidate must

- Meet all essay reader qualifications.
- Have demonstrated leadership ability.
- Have strong communication skills.
- Have knowledge of holistic scoring procedures (participation in or leadership of scoring sessions preferred).

CHIEF READER CERTIFICATION:

To be certified as a GED Tests Chief Reader, a candidate must be

- Approved by the state or province administrator.
- Trained in holistic scoring procedures in accordance with GED Testing Service's Chief Reader guidelines by attending a GED Testing Service Chief Reader training session.
- Willing to supervise GED holistic scoring sessions, and be certified as a GED essay reader.

Candidates who qualify for certification will be issued a GED Tests Chief Reader Certificate by the GED Testing Service.

ESSAY READER QUALIFICATION:

To apply for certification as a GED Tests Essay Reader, a candidate must possess the following:

- A baccalaureate degree, preferably in English.
- At least two years' total experience teaching English language arts at the secondary or postsecondary levels.
- The ability to write effectively.
- A willingness to accept established essay scoring standards.
- An openness to the concepts and principles of holistic scoring.
- A demonstrated ability to work well in group situations.

ESSAY READER CERTIFICATION:

To be certified as a GED Tests Essay Reader, a qualified candidate must

- Attend a GED Testing Service–designed holistic scoring training session.
- Achieve acceptable scores on a set of reader certification papers provided by GED Testing Service.

Candidates who qualify for certification will be issued a GED Tests Essay Reader Certificate by the state or province administrator. (Persons currently serving as GED teachers may **not** participate in the reading of GED examinee papers during an actual scoring session.)

American Council on Education

One Dupont Circle NW
Washington, DC 20036-1193
(202) 939-9300
www.acenet.edu