Texas National Comparative Data Study

DRAFT

Stephen B. Dunbar
Michelle A. Mengeling
Iowa Testing Programs
and
Diane J. Signatur
The Riverside Publishing Company

July 8, 2008

This report was prepared for the Texas Education Agency. Requests for reprints should be directed to Riverside Publishing, 3800 Golf Road, Suite 100, Rolling Meadows, IL 60008.

Table of Contents

List of Tables

List of Figures

Executive Summary

The National Comparative Data Study (NCDS) provides a comparison of the achievement of Texas students to a national reference group as required by Texas Education Code Section 39.028. The Texas Education Agency (TEA) requires the periodic statewide administration of a norm-referenced test to a representative sample of students in the subject areas and grade levels measured by the Texas Assessment of Knowledge and Skills (TAKS). The Iowa Tests® were administered to approximately 12,500 students in each of grades 3 through 11 in the fall of school year 2004–2005, and were administered again in the fall of 2007 to allow for a comparison of student performance over time, as well as a comparison to the nation. This report focuses on the fall 2007 administration.

Average achievement of students in the NCDS sample is reported in terms of standard scores (SSs) and national percentile ranks (NPRs) based on 2005 national norms. Results of the 2007 NCDS were also compared to the results from the 1994, 1999, and 2004 studies in terms of Normal Curve Equivalents (NCEs). Analyses showed average performance of Texas students in Reading and Math to be similar to that in 1994 and 1999. Reading Comprehension in 2007 was above the national average in grades 3 to 5 and below the national average in grades 6 and above. Grade-to-grade growth in reading achievement was consistent with national data for grades 3 through 5. Less than average expected growth was seen between grades 5 and 6 and grades 8 and 9. Given the level of achievement at those grades, grade-to-grade growth was similar to that in the national data from grades 6 through 8 and grades 9 through 11. Math achievement was typically around the national average across grade levels. Students generally performed slightly better on Math Concepts and Problem Solving than they did on Math Computation. Grade-to-grade growth in achievement for Math was consistent with national data.

The NCDS sampling design incorporated key demographic variables for group analyses of statewide performance. Average achievement data and growth trajectories are reported for White, African American, and Hispanic students as well as for students in special education programs, students with limited English proficiency (LEP), and students who are economically disadvantaged.

Part 1 Introduction

The National Comparative Data Study (NCDS) for the Texas Assessment Program provides a comparison of the achievement of a stratified, representative sample of Texas students to a national reference group. Riverside Publishing conducted this study in connection with the Texas Education Code Section 39.028. The Texas statute requires the periodic statewide administration of a norm-referenced test to a representative sample of students in the subject areas and grade levels included in the Texas Assessment of Knowledge and Skills (TAKS). The NCDS component of the Texas Assessment Program was designed to inform Texas leaders and educators about how their students' achievement compares with the nation. These data are intended to supplement the data gathered from the TAKS, which is designed to measure the state content standards and is the primary instrument for school accountability. The Iowa Tests® were administered to approximately 12,500 students in each of grades 3 through 11 in the fall of school year 2004–2005, and were administered again in the fall of 2007 to allow for a comparison of student performance over time, as well as a comparison to the nation. This report focuses on the fall 2007 administration.

1.1 Content Coverage of the National Comparative Data Study

A new edition of The Iowa Tests (Form C) was administered to a representative sample of students in the same grade levels and content areas tested by the TAKS. The Iowa Tests are comprised of two batteries: the *Iowa Tests of Basic Skills® (ITBS®)*, designed for elementary and middle-school grades and the *Iowa Tests of Educational Development® (ITED®)*, for high-school grades. Form C is the new form of the current edition of The Iowa Tests developed in 2007 for large-scale assessment programs. The Complete Battery of the *ITBS* was administered in grades 3–8, and the Complete Battery of the *ITED* was administered in grades 9–11. Table 1.1 summarizes the content areas measured.

Table 1.1
Content Measured in the National Comparative Data Study

Subject	Grade(s)
Reading	3–11
English Language Arts	4, 7, 10, 11
Math	3–11
Social Studies	8, 10, 11
Science	5, 8, 10, 11

The content of The Iowa Tests reflects extensive empirical research by Iowa Testing Programs and current instructional materials, including textbooks as well as the established and emerging content standards listed in Table 1.2.

Riverside's alignments for The Iowa Tests indicate strong and thorough measurement of the Texas Essential Knowledge and Skills (TEKS) targets. This is due to the fact that both assessments utilize standards, guidelines, and recommendations of national organizations of educators such as the National Council of Teachers of English (NCTE), the International Reading Association (IRA), and the Association for Supervision and Curriculum Development

(ASCD). In addition, the blueprints for The Iowa Tests are designed to measure skills included in curriculums taught across the United States.

Moreover, The Iowa Tests are designed to measure the broadest range of skills possible in a norm-referenced test. A recent comparison of the number of items and working times on The Iowa Tests with comparable nationally normed achievement batteries has shown that The Iowa Tests provide greater coverage of skills through more test items than other batteries, within a similar amount of testing time.

The Iowa Tests are well aligned with the TEKS targets and thus are a highly effective and efficient measure of student progress relative to the Texas core curriculum.

Table 1.2
National Content Standards Aligned with The Iowa Tests

Content Area	Organization	Standards
Reading/ Language Arts	National Council of Teachers of English (NCTE) and International Reading Association (IRA)	Standards for the English Language Arts
Mathematics	National Council of Teachers of Mathematics (NCTM)	Principles and Standards for School Mathematics and Curriculum and Evaluation Standards for Mathematics
Social Studies	National Council for Social Studies (NCSS)	Expectations of Excellence: Curriculum Standards for Social Studies
	National Geographic Research & Exploration	Geography for Life: National Geography Standards
	Center for Civic Education	National Standards for Civics and Government
Science	National Research Council (NRC)	National Science Education Standards
	American Association for the Advancement of Science (AAAS)	Atlas of Science Literacy: Project 2061 and Benchmarks for Science Literacy

1.2 Overview of Study

The NCDS has four major components: selection of the sample, administration of the tests, processing and scoring, and development of the report.

The Student Assessment Division of TEA provided Riverside with summaries of enrollment by grade for the public schools in Texas. Riverside used this information to create a stratified sample of schools that met the sample size targets specified by TEA. Additionally, the Student Assessment Division provided information on critical variables that were used for the sampling process.

The demographically representative sample of Texas public schools included White, African American, and Hispanic, special education, limited English proficient (LEP), and economically disadvantaged students. The Riverside Research and Measurement Services staff worked with TEA Student Assessment Division staff to determine the number of campuses required to assure a representative sample.

After the sampling plan was approved by TEA, the TEA notified districts that had been selected for the study. Riverside in turn provided each of the selected districts with the *Directions for Administration*, training materials, and instructions for return of test materials two weeks prior to the test administration window.

After testing materials were received at Riverside Scoring Service®, a file of scored data was prepared and delivered to TEA. This research report was prepared by Iowa Testing Programs and Riverside Publishing.

Part 2 Test Administration

2.1 Overview

The Iowa Tests are designed to be administered to groups of students by classroom teachers. The tests were standardized on groups of normal class size (20–35 students) but can be administered to larger groups if more than one administrator is available. The tests may also be administered to individuals and small groups. Administrators need no special training in educational or psychological testing. Standard conditions for test administration are described in the *Directions for Administration* used in the National Comparative Data Study (NCDS).

2.2 Administrator Training

Prior to the administration of the test, Riverside staff provided instructions and prepared training materials, which were distributed to the District Test Coordinators. Detailed instructions for administering the test and suggestions for training of Test Administrators were provided in the NCDS *Test Coordinator's Manual.*

The *Test Coordinator's Manual* indicated that training should focus on the following topics:

- **Checking the Shipment of Testing Materials.** Each Test Administrator should check the quantities of materials provided to be sure they are adequate for the number of students to be tested. *These materials should always be stored in a secure place.*

- **Setting a Proper Testing Atmosphere.** Testing should be conducted in an environment like that of the classroom. Test Administrators should be informed that their attitude toward the tests and their administration of the tests affect student scores. Students should be informed that the tests are important; however, a tense atmosphere should be avoided.

- **Administering the Tests According to Established Procedures.** Test Administrators should be thoroughly familiar with the *Directions for Administration*. It is imperative that all tests be administered in a uniform manner. For this purpose, specific directions are provided for test administration. Each Test Administrator should follow the directions exactly to ensure a standardized administration.

- **Checking School and Student Identifying Information.** After testing is completed, each Test Administrator should make sure that the student and school identifying information on each scorable test booklet or answer document has been properly written and/or coded. Information about ethnicity is requested only for the purposes of conducting statistical analysis. No student is ever identified by name or by ethnic group for this study.

- **Returning the Testing Materials.** After testing is completed, each Test Administrator is responsible for returning all testing materials to the District Test Coordinator. It is not necessary that the test booklets or answer documents be alphabetized by student name. Booklets should be checked for completeness of erasures.

2.3 Handling Materials

Upon receipt of testing materials, District Test Coordinators were instructed to check the materials against the packing document to be sure the quantities received were correct. District Test Coordinators were instructed to follow the order form to ensure that the proper tests were available for each grade level. The original shipping containers should have been kept to return the testing materials after test administration was completed.

District Test Coordinators were to check the quantities delivered against the number of students to be tested to ensure that there were sufficient materials for *each* student and Test Administrator. Some overage was included in each shipment. Riverside stressed to District Test Coordinators that the shipments be verified immediately upon receipt so that additional materials could be provided in advance of the actual testing.

After the tests were administered, Test Administrators returned all materials to the District Test Coordinators. The *Test Coordinator's Manual* included information on how to pack materials, and how materials were to be picked up. District Test Coordinators should have packed and returned materials to Riverside in the original shipping boxes, following the printed guidelines for returning materials. Test Coordinators should have affixed the pre-identified shipping labels to the boxes to be returned to the Riverside Scoring Service®.

The Riverside Scoring Service received and opened all materials, checked contents, and sought resolution of missing inventory or other issues. Bar code labels on the boxes were scanned to log in each school's shipment and the number of boxes received. Boxes containing materials for scoring were opened immediately, logged in, and released for processing. Any discrepancies or problems encountered during receipt that could not be solved by the receiving staff were logged on an alert form and forwarded to the Riverside Project Director for resolution.

2.4 Test Security

In order to ensure security, Riverside distributes the state-selected test form to only those District Test Coordinators designated by the state. Likewise, Riverside sells test forms for grades *not* included in the statewide assessment program only to the person in each school district officially designated by the TEA. Form C is parallel to Form A—which is used by many districts in the state—but comprises all unique items.

2.5 Student Population

Students participated in the National Comparative Data Study (NCDS) for the Texas assessment program in subject areas and grade levels covered by the Texas Assessment of Knowledge and Skills (TAKS) in English. For example, a grade 5 student receiving special education services who takes the TAKS in reading but not mathematics and science tests participated only in the grade 5 reading component of the NCDS. Similarly, a grade 6 student who takes all of the TAKS tests in Spanish or is exempt from TAKS was not required to participate in the NCDS.

2.6 Accommodations and Modifications

Riverside provided instructions in the *Test Coordinator's Manual* about the types of accommodations allowed for testing and how to provide these accommodations. The same accommodations that were used in the 2000 national standardization of The Iowa Tests were allowed for this study. Riverside staff examined the criteria for exemption and for providing accommodations for the TAKS and determined that they were essentially the same as those used for the national standardization of The Iowa Tests.

During the national standardization, schools were given instructions on testing students in special education and English language learners (ELL). Among students with disabilities, nearly all were eligible for special education services and had an Individualized Education Program (IEP), an Individualized Accommodation Plan (IAP), or a Section 504 Plan.

An accommodation refers to a change in the procedures for administering the assessment and is intended to neutralize, as much as possible, the effect of the student's disability on the assessment process. For this study, Test Administrators were asked to use the guidelines contained in the *Test Coordinator's Manual*, examine the IEP or other plan for these students, decide whether the student should receive accommodations, and determine the nature of those accommodations. Test Administrators were told that they should use the same criteria that they used for TAKS in determining if a student should participate in this study.

Special Education Students

The following accommodations were permitted in this study for students with disabilities.

Method of Response. If a student has a temporary or permanent disabling condition that interferes with his or her ability to record machine-scorable responses, the examinee may respond orally to the test items, mark responses in the test booklet, or type responses. The Test Administrator must record these responses verbatim on a scorable answer document or scorable test booklet.

Extended Time. Some students may need to have time limits on some tests extended to reduce the effect of slow work rate on their test performance. Students who use magnifiers, have attention disorders, and/or need help with word identification or reading may need extended time.

Transferred Answers. Responses recorded in test booklets or those recorded by technologically assistive devices can be transferred by the Test Administrator onto the student's answer document or scorable test booklet in preparation for scoring.

Small-Group/Individual Administration. Students may be tested in small groups or individually; there is no minimum group size requirement for test administration. Students who

need extra breaks and those who might be disruptive in a classroom-testing situation may be tested in this manner.

Repeated Directions. For some students, directions may be read aloud as many times as is necessary to inform students of the proper procedures to follow in responding to test items.

Tests Read Aloud. Students with reading disabilities may need to have portions of tests or some complete tests read aloud to them so that their reading skills do not interfere with measuring their achievement. Under no circumstances should the Vocabulary or Reading Comprehension tests be read to a student. To do so would drastically change what the test measures and what the student's score means. Reading aloud should be used when the student's IEP specifies this as an accommodation and only rarely for those who do not have an IEP.

Other. Accommodations noted in a student's IEP but not included in the above list may be used. However, they should be used only if doing so will not alter the nature of the achievement the test is intended to measure.

Students with Limited English Proficiency (LEP)

For a student whose native language is not English and who has been in an English-only classroom for a limited time, two questions should have been considered prior to testing. First, is the student's English developed sufficiently to warrant testing, and second, should testing involve the use of any particular accommodations? In all instances, the guidelines in place in the state were to be implemented answering these questions.

The purpose of using testing accommodations for students with LEP is to measure skills and knowledge without significant interference due to language difficulties. Thus, those just beginning instruction in English are not likely to be able to answer many questions no matter what accommodations are used. Such students probably should not be tested until their English language skills become more fully developed. For those in their second or third year of instruction in an English as a Second Language (ESL) program, accommodations might be warranted to reduce the effect of language on test performance.

The types of accommodations permitted for students with LEP in the NCDS follow.

Extended Time. Students may need to have time limits extended to reduce the effect of slow work rate on test performance. The slower work rate may be due to limited vocabulary, the need to seek assistance, or the use of a glossary.

Small-Group/Individual Administration. Students may be tested in small groups or individually; there is no minimum group size requirement for test administration. Students who need extra time or those who might be intimidated by the speed with which their peers complete testing might benefit from a small-group administration.

Repeated Directions. For some students, directions may be read aloud as many times as is necessary to inform students of the proper procedures to follow in answering questions.

Test Administered by ESL Teacher or Individual Providing Language Services. This accommodation allows the student to be tested in the environment that is most comfortable for him or her.

Part 3 Processing and Scoring

The Riverside Scoring Service® process is designed to score test documents accurately, while tracking the progress and status of each order and minimizing turnaround time.

Significant emphasis is placed on making sure answer documents are processed correctly. There are several quality control (QC) steps strategically integrated throughout the process to ensure accuracy. In addition to these steps, the processes themselves have their own quality checks. For instance, a tracking system validates order movement, which prevents orders from being processed at any workstation without the previous stage being complete. Below is a summary of the steps that the Riverside Scoring Service took to scan and score the test documents from the National Comparative Data Study (NCDS).

3.1 Receiving and Check-In

The receiving staff of Riverside Scoring Service handled the unloading, sorting, and logging deliveries from Texas districts participating in the study. When test materials arrived, a clerk counted the boxes and verified the count with the carrier. If there was a discrepancy, it was resolved. The boxes were moved to the check-in area, where district information, building names, and class counts were entered into the computerized tracking system.

3.2 Document Preparation and Quality Control

Staff prepared answer documents for scanning by verifying and sorting them. Document totals were verified against what was received, and discrepancies sent to the Riverside Project Manager for resolution. After discrepancies were resolved, a member of the QC team checked each batch of documents to ensure that all documents were in the correct order and ready for scanning. The QC check ensured that no documents had problems that might interfere with the scanning process.

3.3 Slitting and Scanning

The answer documents were then slit to remove the saddle-stitched bindings in preparation for scanning. The slitting machine used by Riverside Scoring Service utilizes a digital microprocessor that accurately cuts the bindings from booklets to 1/100 of an inch. This type of slitter yields a consistent and accurate cut every time, a requirement that is vital to the success of the scanning process.

The documents were then scanned using Scan Optics 9000M scanners. Four cameras picked up both the bitonal and grayscale images of the top and bottom pages as the sheets were fed through the scanner. While documents were being scanned, an operator monitored the scanner screen for jams or scan errors. The scanner automatically verified that documents were facing the correct direction and that each document contained the correct number of pages.
Any document that could not be scanned (i.e., a document that was wrinkled or torn) was manually key-entered. When key-entry was necessary, two people entered the student responses

independently. The computer system then automatically verified the two entries and notified the operator if any discrepancies existed.

3.4 Editing and Quality Control

As each document proceeded through the scanning stage, editing rules were applied to alert operators to problems with scanning caused either by the student responses (e.g., double grids, excessive omits, or light marks) or by the scanner. Any flagged documents were placed into a queue for an editor to resolve. In addition, the scan system randomly flagged a small percentage of non-problematic documents to go to editing for review. This QC process ensured that the scanners and the editing rules were working properly.

An experienced editor resolved issues with documents in the editing queue. In addition, a QC editor (working independently with no knowledge of the first editor's actions) also resolved issues on the same documents. As the QC editor resolved each document issue, the system notified him or her of any discrepancies with the original editor's responses.

3.5 Scoring

Once the answer documents completed the scanning and editing phases of the process, an electronic file containing all student information and responses was sent to the scoring system. The scoring system first performed a check on the data to ensure that they would be accepted into the system. For example, the system checked to be sure that each student had provided enough responses to receive a score on each test (completion criteria). After this final check, the file was ready to be scored.

The answer strings were then scored against the answer key to produce raw scores. These raw scores were used to look up the derived scores for each individual student.

3.6 Data File

A scored data file was produced and sent to Iowa Testing Programs for analysis. This file contained the demographic variables as well as the scores for each individual student. These data were analyzed in several different ways to assemble the findings in this technical report.

Part 4 Sampling Methodology

4.1 Sampling Plan

The test administration window started on October 10, 2007, and ended November 2, 2007. Riverside shipped enough testing materials to each selected campus to test all students in the selected grades. Schools were allowed to pick the most convenient testing week during the test administration window. Instructions for this testing program were communicated in a custom *Test Coordinator's Manual*.

To aid in administration procedures and reduce overall scoring errors, each student took all of the content areas assigned to his or her grade. For example, in grade 4, all students took the reading, language, and mathematics tests of the *Iowa Tests of Basic Skills (ITBS*), Complete Battery, Form C. Table 4.1 provides the administration times by content area and grade level.

Table 4.1

Administration Times (in minutes) on The Iowa Tests

	ITBS **Complete Battery Form C**						***ITED*** **Complete Battery Form C**		
Grade	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Reading	70	70	70	70	70	70	55	55	55
Language		66			66			50	50
Mathematics	75	75	75	75	75	75	55	55	55
Social Studies						30		40	40
Science			30			30		40	40

Note. Shaded areas represent *ITBS*/*ITED* content areas and grade levels not assessed. Each student took all of the content areas assigned to his or her grade, unless the student was exempt from one or more TAKS content areas. In these cases, the student took only the content area(s) that he or she would take in TAKS.

4.2 Procedures for Selecting the Fall 2007 Representative Sample

The Student Assessment Division of TEA provided Riverside with summaries of enrollment by grade for the public schools in Texas. Riverside used this information to create a stratified sample of schools from Texas public schools in grades 3 through 11 that met the sample size targets specified by TEA. During the stratification and sampling procedure, campuses were grouped into one of six sampling regions in Texas: (1) Northeast Region, (2) Southeast Region, (3) Central-Valley Region, (4) North Central Region, (5) Northwest Region, and (6) Big 20—campuses from the 20 largest school districts.

Prior to sampling, TEA asked that some campuses not be selected for this study, as they were participating in another study. These campuses were removed from the list of campuses eligible for selection. Approximately 500 districts were sampled for a total of 12,500 students per grade.

After all testing materials were received by the Riverside Scoring Service®, the exact numbers and percentages of students in the total sample and in each stratification category were determined. Weights for these samples were determined by comparing the proportion of students statewide in each cohort to the corresponding sample proportion. These percentages were then adjusted by weights to more closely reflect the composition of the Texas school population by region and race/ethnicity.

The optimal weights for the Texas sample were determined by comparing the proportion of students in Texas by region and ethnicity to corresponding state proportions. Once the optimal weights were obtained, the stratification variables were simultaneously considered to assign final weights. These weights (integer values 1 through 9, with 3 denoting perfect proportional representation) were assigned to adjust the frequencies. As a result, the weighted distribution in the Texas sample closely approximates the distribution of the state student population with regard to region and ethnicity.

The total number and percent of students by geographic region who participated in the 2007 fall National Comparative Data Study (NCDS) for the Texas assessment program are given in Table 4.2 in the unweighted sample column. The table also provides the weighted sample percents and the Texas population percents for each region.

Table 4.2
Sample Sizes and Percents by Region,
Fall 2007 Texas Representative Sample

Region	Unweighted Sample	Unweighted Sample %	Weighted Sample %	Texas Population %
Northeast	19,509	19.2	17.3	17.4
Southeast	14,608	14.4	16.3	16.3
Central Valley	17,414	17.1	18.6	18.6
North Central	10,929	10.7	10.0	10.0
Northwest	7,233	7.1	6.3	6.5
Big 20	32,097	31.5	31.5	31.3
Total	101,790	100	100	100

The TEA Student Assessment Division provided two additional variables for reporting student performance at the state level: ethnicity and special populations. Within ethnicity, TEA identified the following categories: White, African American, Hispanic, and other. Table 4.3 summarizes the unweighted and weighted sample characteristics for the Texas fall 2007 NCDS based on the principal stratification variable of ethnicity.

Table 4.3
Sample Sizes and Percents by Ethnicity,
Fall 2007 Texas Representative Sample

	Unweighted Sample	Unweighted Sample %	Weighted Sample %	Texas Population %
White	36,956	36.3	37.6	37.9
African American	12,473	12.3	14.5	14.5
Hispanic	40,935	40.2	43.6	43.7
Other	11,426	11.2	4.2	3.9
Total	101,790	100	100	100

For the special populations variable, TEA defined the following three categories: special education, limited English proficient (LEP), and economically disadvantaged. Of these groups, students in special education and students with limited English proficiency (LEP) may have received accommodations in testing.

The special populations categories are not mutually exclusive; individual students may be represented in more than one category. The participation in testing of students receiving special education services or students identified as LEP and/or economically disadvantaged followed guidelines set by TEA. Table 4.4 summarizes the unweighted and weighted sample characteristics for the Texas fall 2007 NCDS for special populations.

Table 4.4
Sample Sizes and Percents by Special Populations,
Fall 2007 Texas Representative Sample

	Unweighted Sample	Unweighted Sample %	Weighted Sample %	Texas Population %
Special Education	5,774	5.7	5.8	9.3
Limited English Proficient (LEP)	7,593	7.5	7.7	9.8
Economically Disadvantaged	31,923	31.4	33.3	51.0

Note. The special populations categories are not mutually exclusive. Individual students may be included in more than one group.

4.3 Representation of Special Populations by Grade

In order to select the representative number of students per region for each special population, Riverside worked with the TEA to determine the number of students who could participate in each region at each grade level. This same process was followed in determining the target percents needed to represent the variables of students classified by ethnicity and special populations.

In grade 3, teachers were asked to indicate the ethnicity group to which students belong. In the remaining grades, students provided this information.

Schools were given detailed instructions on testing students in special education and students with LEP. Schools identified all students classified as such and determined if testing should occur. Accommodations in testing were determined on an individual student basis. Accommodations refer to changes in the procedures for administering the assessment to neutralize, as much as possible, the effect of the student's disability on the assessment process. When accommodations were allowed, test administrators recorded their use on each student's answer document.

Table 4.5 summarizes special population representation in the grade 3 through grade 11 weighted samples.

Table 4.5

Special Population Students, Fall 2007 Texas Representative Sample

	Texas Population	Special Education		Limited English Proficient		Economically Disadvantaged	
Grade	N	N	Weighted Sample %	N	Weighted Sample %	N	Weighted Sample %
3	47,986	2,134	4.5	6,715	14.1	17,325	36.4
4	54,418	2,656	5.5	6,053	12.5	18,097	37.3
5	51,519	2,801	5.4	5,260	10.1	19,006	36.6
6	31,363	2,447	7.3	1,597	4.8	12,568	37.5
7	52,695	2,781	5.6	2,849	5.8	18,824	38.2
8	49,060	3,255	6.6	2,035	4.1	15,218	30.8
9	33,991	1,978	5.9	2,463	7.3	11,199	33.2
10	42,254	2,601	6.2	1,483	3.5	9,147	21.9
11	53,470	1,282	5.1	735	2.9	5,949	23.5

Note. All *N*-counts are from the weighted sample. The unweighted number of students in the Texas sample is approximately 11,000 per grade.

Part 5 Reliability and Validity

5.1 Reliability

Test reliability can be quantified by a variety of statistical data. Such data, however, reduce to two basic indices. The first index is the reliability coefficient. In numerical value, the reliability coefficient is always between .00 and .99, and generally between .60 and .95. The closer the coefficient is to the upper limit, the greater is the evidence of a well-constructed test.

The second statistical index to describe test reliability is the standard error of measurement (SEM). This index measures the net effect of all factors leading to inconsistency in student test scores and to inconsistency in score interpretation. The standard error of measurement can be explained by a hypothetical example. Suppose students with the same reading ability were to take the same reading test. Despite their equal ability, they would not all get the same score. Instead, their scores would range over an interval. A few would get scores much higher than expected, a few much lower; the majority would get scores fairly close to their actual ability. Such variation in scores would be attributable to differences in motivation, attentiveness, and other factors. The standard error of measurement is an index of the variability of the scores of students having the same actual ability. It tells the degree of precision in placing a student at a point on the achievement continuum.

The reliability data presented on the following pages are based on Kuder–Richardson formula 20 (K–R 20). The means, standard deviations (SD), and SEMs are shown for number correct raw scores and for standard scores in Table 5.1. Also reported in Table 5.1 for each grade level is the national percentile rank (NPR) of the average student in the Texas sample.

In terms of general measurement precision in the NCDS, the *Iowa Tests of Basic Skills* (*ITBS)/Iowa Tests of Educational Development* (*ITED)* performed as would be expected from research results in national samples. The SEMs of all tests in the NCDS sample were compared to their values in the 2005 national standardization of the *ITBS/ITED* and found to be very similar in magnitude.

Table 5.1
Raw Score and Standard Score Summary Statistics
ITBS Complete Battery, Form C, 2005 Norms
Fall 2007 Texas Representative Sample

Grade 3		Reading			Mathematics			
		V	RC	RT	M1	M2	M3	MT+
Number of items		29	37		31	22	25	
RSs	Mean	15.5	19.4		15.4	11.5	13.6	
	SD	6.0	7.5		5.4	4.8	4.9	
	SEM	2.3	2.7		2.3	2.0	2.3	
SSs	Mean	174.3	180.0	177.3	175.0	174.6	173.1	174.5
	SD	19.9	22.5	19.8	16.7	21.5	14.6	15.2
	SEM	7.5	8.2	5.4	7.1	9.2	6.9	4.3
Reliability	K-R 20	.859	.867	.925	.818	.818	.776	.921
NPR*		*47*	*58*	*53*	*52*	*48*	*52*	*53*

Grade 4		Reading			Language					Mathematics			
		V	RC	RT	L1	L2	L3	L4	LT	M1	M2	M3	MT+
Number of items		34	41		32	26	26	33		36	24	27	
RSs	Mean	16.2	24.6		17.1	12.7	12.8	16.3		19.8	12.8	14.5	
	SD	6.5	8.9		6.6	5.1	5.4	6.3		6.3	4.9	5.2	
	SEM	2.6	2.9		2.6	2.3	2.3	2.6		2.8	2.2	2.0	
SSs	Mean	185.6	193.3	189.7	191.2	184.8	190.6	193.8	190.5	191.5	192.0	188.0	190.9
	SD	23.3	28.0	23.9	24.4	31.4	30.2	32.8	25.4	22.3	28.1	15.8	19.1
	SEM	9.4	9.2	6.5	9.7	14.3	12.7	13.8	6.3	10.0	12.7	6.0	5.4
Reliability	K-R 20	.838	.893	.926	.843	.792	.824	.824	.939	.797	.797	.858	.920
NPR*		*40*	*52*	*47*	*49*	*39*	*48*	*54*	*46*	*53*	*50*	*48*	*53*

Grade 5		Reading			Mathematics				Science
		V	RC	RT	M1	M2	M3	MT+	SC
Number of items		37	43		40	26	29		37
RSs	Mean	16.8	25.2		21.4	13.2	16.5		16.1
	SD	7.4	8.3		7.0	5.5	6.2		6.5
	SEM	3.0	2.7		3.0	2.5	2.4		3.0
SSs	Mean	197.6	207.4	202.6	205.2	207.5	201.2	205.2	210.7
	SD	28.2	28.8	26.4	25.8	33.5	22.5	23.7	32.6
	SEM	11.5	9.4	7.2	10.9	15.2	8.8	6.7	14.8
Reliability	K-R 20	.833	.894	.926	.822	.794	.846	.921	.793
NPR*		*38*	*53*	*46*	*51*	*52*	*45*	*50*	*57*

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

* NPR is the National Percentile Rank of the average Texas Student

Table 5.1 (continued)
Raw Score and Standard Score Summary Statistics
***ITBS* Complete Battery, Form C, 2005 Norms**
Fall 2007 Texas Representative Sample

Grade 6		Reading			Mathematics			
		V	RC	RT	M1	M2	M3	MT+
Number of items		39	45		43	28	30	
RSs	Mean	19.3	25.4		23.3	14.3	17.0	
	SD	8.0	8.8		8.6	4.6	5.3	
	SEM	3.0	2.9		3.0	2.2	2.3	
SSs	Mean	209.7	214.9	212.8	217.5	216.9	214.5	216.8
	SD	28.3	33.7	29.1	28.7	33.0	22.6	24.4
	SEM	10.5	11.2	7.6	10.1	15.7	9.4	6.7
Reliability	K-R 20	.862	.890	.931	.876	.775	.828	.924
NPR*		*36*	*45*	*41*	*49*	*47*	*46*	*48*

Grade 7		Reading			Language					Mathematics			
		V	RC	RT	L1	L2	L3	L4	LT	M1	M2	M3	MT+
Number of items		41	48		40	32	32	40		46	30	31	
RSs	Mean	20.4	26.3		21.5	15.7	15.6	20.3		26.3	13.3	15.0	
	SD	8.4	10.0		8.7	6.3	6.9	7.8		8.4	6.1	5.8	
	SEM	2.9	2.9		2.6	2.3	2.2	2.7		2.8	2.4	2.4	
SSs	Mean	223.4	229.0	226.4	231.3	218.5	228.9	231.1	227.9	236.7	233.0	225.5	232.1
	SD	30.0	35.7	31.0	31.3	41.2	42.6	45.1	35.2	28.0	37.6	29.8	28.1
	SEM	10.5	10.4	7.4	9.2	15.2	13.6	15.6	6.8	9.2	15.0	12.2	7.1
Reliability	K-R 20	.878	.915	.943	.913	.864	.898	.881	.963	.891	.840	.832	.936
NPR*		*39*	*48*	*43*	*50*	*40*	*47*	*50*	*46*	*58*	*51*	*44*	*51*

Grade 8		Reading			Mathematics				Social Studies	Science
		V	RC	RT	M1	M2	M3	MT+	SS	SC
Number of items		42	52		49	32	32		43	43
RSs	Mean	19.9	27.8		24.6	15.2	14.5		18.2	18.1
	SD	8.3	10.9		8.2	6.4	6.4		7.6	7.3
	SEM	2.8	3.1		3.1	2.5	2.6		2.7	2.7
SSs	Mean	233.9	241.0	237.7	246.9	244.4	237.7	243.5	232.0	241.5
	SD	30.0	39.4	32.6	31.6	41.0	36.5	31.7	37.2	37.7
	SEM	10.3	11.4	7.6	11.8	16.3	15.0	8.2	13.3	14.1
Reliability	K-R 20	.883	.917	.946	.860	.842	.831	.933	.873	.861
NPR*		*39*	*48*	*44*	*55*	*50*	*45*	*51*	*41*	*49*

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

* NPR is the National Percentile Rank of the average Texas Student

Table 5.1 (continued)
Raw Score and Standard Score Summary Statistics
***ITED* Complete Battery, Form C, 2005 Norms**
Fall 2007 Texas Representative Sample

Grade 9		Reading			Mathematics		
		V	RC	RT	M	M3	MT+
Number of items		40	44		40	30	
RSs	Mean	18.6	22.4		17.4	12.2	
	SD	8.7	10.0		8.3	5.3	
	SEM	2.6	2.9		2.7	2.2	
SSs	Mean	240.8	244.1	242.5	255.7	248.0	253.3
	SD	31.9	43.1	35.2	36.8	35.4	33.5
	SEM	9.6	12.3	7.8	12.1	14.5	7.6
Reliability	K-R 20	.910	.918	.951	.891	.833	.949
NPR*		*36*	*42*	*40*	*52*	*45*	*50*

Grade 10		Reading			Language		Mathematics			Social Studies	Science
		V	RC	RT	L1	RW	M	M3	MT+	SS	SC
Number of items		40	44		30	56	40	30		50	48
RSs	Mean	20.4	23.5		13.7	30.7	17.7	12.0		20.5	18.4
	SD	8.7	10.2		6.6	11.7	8.4	5.3		9.7	8.6
	SEM	2.8	2.9		2.3	3.3	2.9	2.0		3.2	3.3
SSs	Mean	250.2	253.0	251.8	258.1	256.7	257.6	248.2	254.8	246.0	248.9
	SD	35.2	46.5	38.1	36.5	43.5	40.5	35.1	35.5	46.3	48.2
	SEM	11.4	13.2	8.8	12.7	12.1	13.9	13.2	7.9	15.2	18.5
Reliability K-R 20		.895	.920	.947	.879	.922	.883	.859	.950	.892	.853
NPR*		*36*	*42*	*40*	*45*	*46*	*46*	*37*	*43*	*37*	*38*

Grade 11		Reading			Language		Mathematics			Social Studies	Science
		V	RC	RT	L1	RW	M	M3	MT+	SS	SC
Number of items		40	44		30	56	40	30		50	48
RSs	Mean	21.2	23.1		14.6	30.4	16.7	12.9		21.9	19.0
	SD	8.8	10.3		6.5	11.6	8.2	5.7		10.2	9.6
	SEM	2.8	2.9		2.2	3.1	2.8	1.9		3.2	3.1
SSs	Mean	261.7	264.3	263.3	270.2	271.0	271.1	264.0	269.4	258.5	264.0
	SD	36.2	47.5	38.8	36.7	44.1	40.6	33.7	35.2	47.7	47.1
	SEM	11.7	13.2	8.8	12.4	11.9	14.0	11.1	7.5	15.0	15.0
Reliability K-R 20		.896	.923	.948	.885	.927	.881	.892	.955	.901	.898
NPR*		*40*	*44*	*44*	*50*	*50*	*51*	*45*	*49*	*41*	*44*

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

* NPR is the National Percentile Rank of the average Texas Student

5.2 Validity

Validity is an attribute of information from tests that, according to the *Standards for Educational and Psychological Testing*, "refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999, p. 9).

The validity of an assessment depends on its purpose and the relevance of test results to that purpose. The purpose of assessment in the Texas National Comparative Data Study (NCDS) is to provide a national framework for understanding the overall achievement of students in Texas schools. The assessment tool must satisfy professional standards of technical quality related to test content and psychometric characteristics to have validity given the purpose of NCDS. Pertinent validity evidence is provided in research guides for The Iowa Tests (Hoover, Dunbar, & Frisbie, 2003; Forsyth, Ansley, Feldt, & Alnot, 2003).

Because the inferences from NCDS pertain to the relative performance of students in Texas and students in the nation at large, evidence from NCDS about the internal structure of The Iowa Tests and about cross-grade patterns of performance is of interest. The extent to which such characteristics of the assessment instrument are similar in the NCDS sample and the national standardization sample influences the validity of the instrument for the purpose at hand.

To examine the internal structure of The Iowa Tests in the NCDS and the national standardization, the methods of structural equation modeling were used. Specifically, a hierarchical latent-variable model was established to reflect the covariance structure of the subtests of the *ITBS* and the *ITED*. This model was then fit to data from grades 4, 7, and 11 in the NCDS and national samples. Results indicated good model fit at all grades in both samples, suggesting that relations among subtests in The Iowa Tests followed the same pattern in the Texas NCDS sample as they do in the nation. The details of results from the latent-variable modeling are beyond the scope of this report. They are available from Iowa Testing Programs.

Also pertinent to the validity of The Iowa Tests in the context of the NCDS are cross-sectional growth patterns over the years of schooling. If grade-to-grade changes followed a different pattern in the NCDS sample than in the nation, questions could be raised about alignment to the Texas Essential Knowledge and Skills (TEKS) or about the scope-and-sequence of test content vis-à-vis the TEKS. Such results may also raise questions about state and national differences in instruction and learning of interest to policymakers in the state.

Figures 5.1 and 5.2 show the general pattern of change for scores in Reading and Math, respectively, in grades 3 through 11 for the national standardization sample and the NCDS 2004 and 2007 samples. As can be seen in the figures, grade-to-grade growth in achievement is quite similar among samples. The results show a somewhat flatter growth trajectory in the upper elementary grades (5 through 7) than is observed nationally in Reading. The growth trajectories for Math are virtually identical for the NCDS 2007 representative sample and the national standardization sample from grade 3 through high school except in grade 10. Results from grade 10 show less growth and lower overall achievement than expected from national norms.

Figure 5.1
Reading Total Standard Score Means by Grade
National Standardization and Fall 2007 Texas Representative Samples

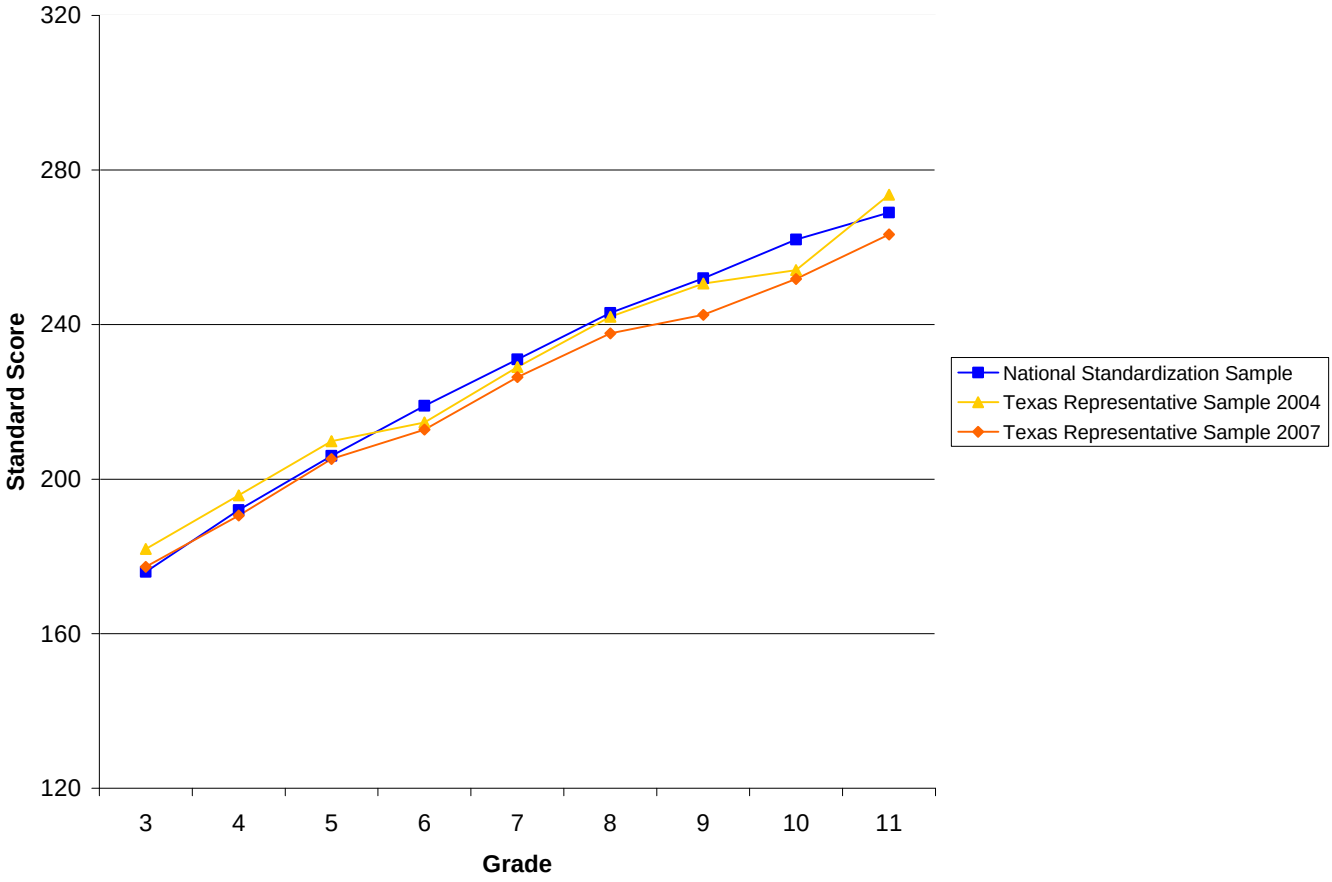
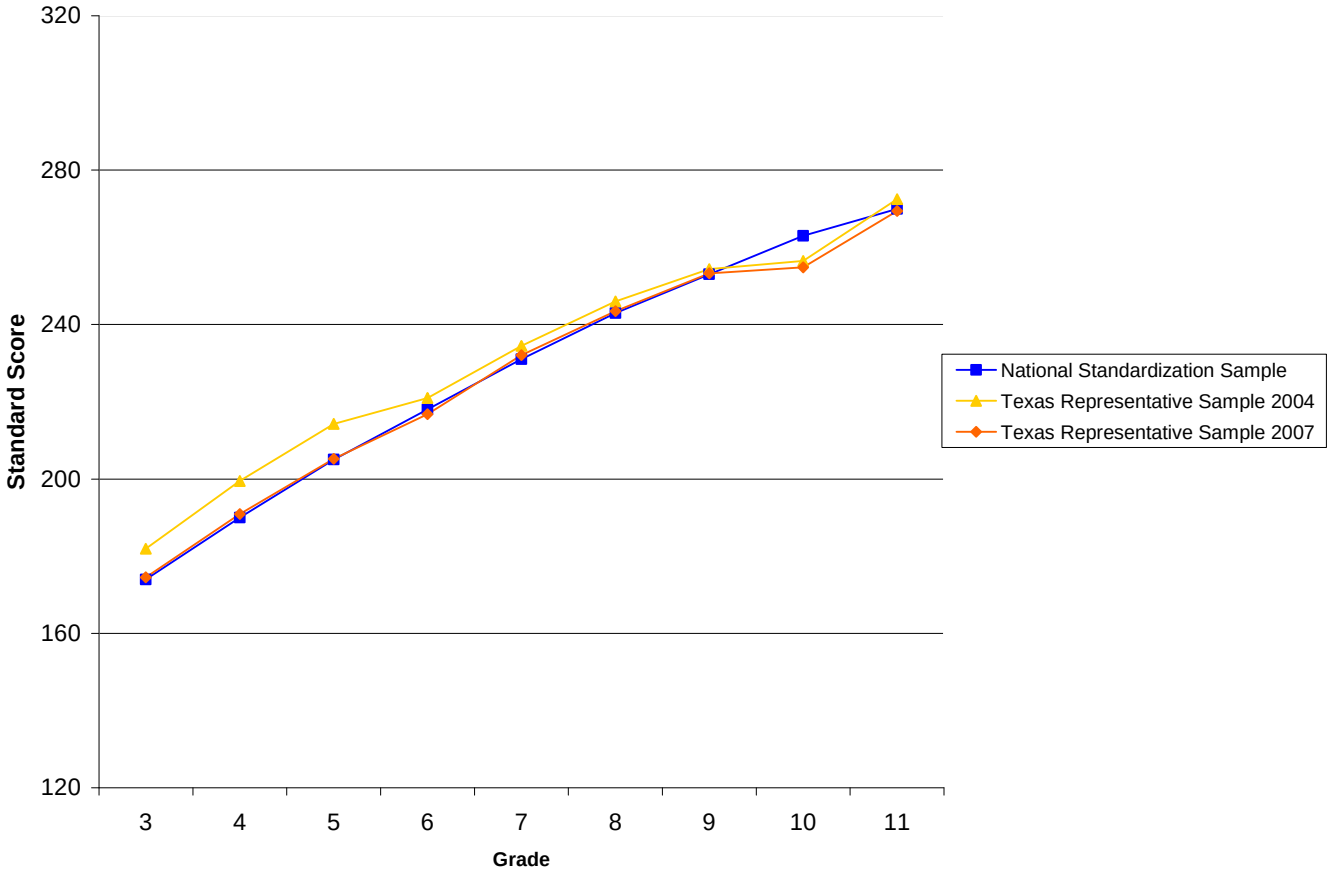


Figure 5.2
Math Total Standard Score Means by Grade
National Standardization and Fall 2007 Texas Representative Samples



A final source of evidence on validity exists in the NCDS results viewed over time. Table 5.2 reports average levels of achievement observed in NCDS samples from 1994, 1999, 2004, and 2007. These years were used because they represented approximately equal time intervals between studies and because three of the years (1994, 2004 and 2007) involved data from roughly parallel forms of The Iowa Tests. All achievement test results are expressed in Normal Curve Equivalents (NCEs), a transformation of percentile ranks.

Table 5.2
Mean Normal Curve Equivalents
Texas NCDS 1994, 1999, 2004, 2007

	Reading				Math			
Grade	**1994**	**1999**	**2004**	**2007**	**1994**	**1999**	**2004**	**2007**
3	51	50	57	52	56	55	61	52
4	52	52	54	48	55	59	61	52
5	50	51	53	48	53	57	59	50
6	50	49	47	45	51	55	51	49
7	50	50	48	46	49	55	49	51
8	50	48	50	47	50	53	52	51
9	—	—	49	45	—	—	51	50
10	48	52	46	45	—	55	47	46
11	—	—	53	47	—	—	51	49

The data in Table 5.2 reveal a similar pattern in average test scores over the period since the NCDS was established. There was a marked increase in average test scores in the elementary grades in 2004, but the 2007 results are similar to those shown for 1994 and 1999.

5.3 Difficulty of the Tests

To maximize the reliability of a ranking within a group, an achievement test must use nearly the entire range of possible scores; the raw scores on the test should range from near zero to the highest possible score. A few items included in the final test should be so easy that at least 80 percent of students answer them correctly. These should identify the least able students. Similarly, a few very difficult items should be included to challenge the most able students. Most items, however, should be of medium difficulty and should discriminate well at all levels of ability. In other words, the typical student will succeed on only a little more than half of the test items, while the least able students may succeed on only a few. A test constructed in this manner results in the widest possible range of scores and yields the highest reliability per unit of testing time. It is of interest to check that these characteristics of the *ITBS* and *ITED* national sample hold for the NCDS sample in Texas.

A summary of the difficulty indices for all tests and grades is presented in Table 5.3. These data are based on all students in the Texas representative sample. The difficulty indices reported are proportions (*p*-values) rather than percents correct. The mean item proportions correct are shown (in italics).

The difficulty indices reported in the table describe the distribution of item difficulty within each test. The mean proportions correct are shown in italics. The range of item difficulty within a test can be determined from the difference between p-values of extremely easy and extremely hard items. The rows labeled P_{90} and P_{10} indicate the proportions correct for such items. (P_{90} and P_{10} represent the 90^{th} and 10^{th} percentiles respectively, of the distributions of difficulty.) For example, the grade 3 Reading Comprehension (RC) test had an average p-value in Texas of .54. An easy item on this test was one with a p-value of .69 (i.e., P_{90}), and a hard item on this test was one with a p-value of .38 (i.e., P_{10}). These p-values reflect a range of item difficulty needed to reliably measure reading comprehension at all achievement levels and to produce score distributions without ceiling and floor effects. Generally speaking, the values reported in Table 5.3 are very similar to those obtained in the national standardization of the *ITBS* and *ITED*. This similarity is an important source of validity evidence in the context of NCDS.

Table 5.3
Summary of Difficulty (Proportion Correct), *ITBS* Complete Battery, Form C
Fall 2007 Texas Representative Sample

Grade 3	Reading						Mathematics				
	V	RC					M1	M2	M3		
Number of items	29	37					31	22	25		
Mean	*.56*	*.54*					*.53*	*.54*	*.61*		
P_{90}	.75	.69					.77	.72	.80		
Median	.57	.54					.50	.51	.66		
P_{10}	.30	.38					.27	.26	.28		

Grade 4	Reading		Language				Mathematics				
	V	RC	L1	L2	L3	L4	M1	M2	M3		
Number of items	34	41	32	26	26	33	36	24	27		
Mean	*.53*	*.63*	*.56*	*.53*	*.52*	*.52*	*.59*	*.56*	*.64*		
P_{90}	.80	.79	.80	.75	.77	.72	.86	.82	.84		
Median	.47	.64	.53	.51	.55	.51	.57	.51	.64		
P_{10}	.31	.48	.33	.29	.24	.26	.29	.30	.38		

Grade 5	Reading						Mathematics				Science
	V	RC					M1	M2	M3		SC
Number of items	37	43					40	26	29		37
Mean	*.51*	*.63*					*.60*	*.57*	*.71*		*.49*
P_{90}	.72	.87					.87	.78	.85		.64
Median	.50	.64					.62	.60	.72		.48
P_{10}	.31	.38					.30	.29	.51		.33

Grade 6	Reading						Mathematics				
	V	RC					M1	M2	M3		
Number of items	39	45					43	28	30		
Mean	*.55*	*.60*					*.62*	*.54*	*.64*		
P_{90}	.77	.79					.83	.86	.86		
Median	.54	.57					.63	.50	.69		
P_{10}	.32	.42					.39	.27	.31		

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

Table 5.3 (continued)
Summary of Difficulty (Proportion Correct),
***ITBS* Complete Battery, Form C**
Fall 2007Texas Representative Sample

Grade 7	Reading		Language				Mathematics				
	V	RC	L1	L2	L3	L4	M1	M2	M3		
Number of items	41	48	40	32	32	40	46	30	31		
Mean	*.54*	*.59*	*.56*	*.53*	*.53*	*.53*	*.61*	*.48*	*.57*		
P_{90}	.76	.78	.80	.75	.71	.72	.87	.67	.79		
Median	.55	.59	.55	.53	.53	.54	.62	.47	.57		
P_{10}	.31	.37	.36	.31	.27	.34	.34	.27	.32		

Grade 8	Reading						Mathematics			Social Studies	Science
	V	RC					M1	M2	M3	SS	SC
Number of items	42	52					49	32	32	43	43
Mean	*.51*	*.58*					*.55*	*.52*	*.56*	*.46*	*.46*
P_{90}	.70	.73					.80	.72	.80	.64	.65
Median	.52	.57					.53	.48	.56	.46	.45
P_{10}	.27	.36					.34	.33	.31	.30	.25

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

Table 5.3 (continued)
Summary of Difficulty (Proportion Correct),
***ITBS* Complete Battery, Form C**
Fall 2007Texas Representative Sample

Grade 9	Reading				Mathematics			
	V	RC			M	M3		
Number of items	40	44			40	30		
Mean	*.50*	*.58*			*.47*	*.49*		
P_{90}	.68	.71			.63	.72		
Median	.49	.57			.48	.47		
P_{10}	.30	.43			.27	.24		

Grade 10	Reading		Language		Mathematics		Social Studies	Science
	V	RC	L1	RW	M	M3	SS	SC
Number of items	40	44	30	56	40	30	50	48
Mean	*.55*	*.59*	*.58*	*.47*	*.48*	*.47*	*.44*	*.42*
P_{90}	.67	.73	.73	.61	.64	.67	.56	.53
Median	.53	.58	.63	.47	.50	.41	.44	.41
P_{10}	.37	.45	.36	.29	.27	.24	.35	.29

Grade 11	Reading		Language		Mathematics		Social Studies	Science
	V	RC	L1	SS	SC	M3	SS	SC
Number of items	40	44	30	56	40	30	50	48
Mean	*.56*	*.59*	*.58*	*.50*	*.45*	*.50*	*.47*	*.43*
P_{90}	.70	.72	.75	.66	.62	.69	.59	.57
Median	.55	.58	.61	.51	.44	.51	.47	.41
P_{10}	.40	.45	.35	.29	.23	.25	.34	.32

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME). (1999). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

Hoover, H. D., Dunbar, S. B., Frisbie, D. A. (2003). Guide to Research and Development. *Iowa Tests of Basic Skills.* Itasca, IL: Riverside Publishing.

Forsyth, R. A., Ansley, T. N., Feldt, L. S., Alnot, S. D. (2003). Guide to Research and Development. *Iowa Tests of Educational Development.* Itasca, IL: Riverside Publishing.

Part 6 Sampling Precision

The statistics in this report provide estimates of the achievement of Texas students sampled by school building. The precision of a sample estimate is given by the standard error. In the National Comparative Data Study (NCDS), the statistic of interest is the average achievement of Texas students as measured by standard scores. The accuracy of the standard score building average depends on the sampling design and the number of students tested. Because the NCDS used stratified sampling of schools rather than simple random sampling of students, special methods were required to determine sampling precision.

6.1 The Bootstrap Analysis

The purpose of the bootstrap analysis was to evaluate the stability of building averages given the state's need to disaggregate results by important demographic variables. Bootstrap standard errors of mean standard scores treat the representative sample of Texas schools in the NCDS as if it were an intact population (Efron & Tibshirani, 1986). To estimate the sampling error of the state mean, the relevant data set was simply a list of means for each building in the sample by grade. For estimating the sampling errors of subgroup means, a list of subgroup means was used. The bootstrap method consists of repeated random sampling, with replacement, from the appropriate list. Each such random sample yields an estimate of the overall mean for the state, and each subgroup. One hundred bootstrap samples were drawn (Efron, 1987), and the standard deviation of the building means across samples was used as the bootstrap estimate of the standard error. All standard error estimates for these analyses are expressed on the *Iowa Tests of Basic Skills* (*ITBS*) and *Iowa Tests of Educational Development* (*ITED*) standard score (SS) scales.

6.2 Bootstrap Results

Bootstrap standard errors of building means for all grades and subgroups are provided in Table 6.1. Several things should be noted in evaluating these results. First, the bootstrap estimates are uniformly larger (in many cases twice as large) as the usual estimates based on simple random sampling of students. This fact reflects the loss in precision that occurs when cluster sampling is used instead of simple random sampling. Second, standard errors tend to be larger for test scores than for total scores. Third, standard error estimates for subgroups are in some cases markedly larger than the corresponding estimates for the entire state. This reflects the fact that the building-level sample sizes for certain subgroups were small, so that the means across building showed greater variability.

6.3 Use of Bootstrap Standard Errors

The standard errors reported in Table 6.1 provide a direct description of the variability of building averages within the state of Texas. From the results it can be seen, for example, that school averages of Whites and Hispanics are more similar to each other than either is to the school averages of African Americans. The bootstrap method also indicated that at the early grades within-school averages of students in Special Education (perhaps because of small sample sizes) tended to be larger than standard errors of other subgroup building averages. At the higher grades, the Limited English Proficient subgroup tended to have the largest standard errors associated with building averages. The standard errors for males, which are usually larger than those for female subgroup – a reflection of the common finding in large representative samples that males are more variable than females on many measures of school achievement – tended to show more similarly sized standard errors relative to the female standard errors in this sample.

Quantifying sampling errors of building averages is important in the interpretation of mean differences between groups. The building mean give or take a standard error gives an approximate 68 percent confidence band for the true average. When the confidence bands for two group means overlap, there is considerable doubt about whether the difference is real or simply due to chance sampling fluctuations.

Table 6.1
Bootstrap Standard Errors of Building Averages
Fall 2007 Texas Representative Sample
2005 Norms

Grade 3									
Test	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
RV	0.62	0.90	1.01	0.88	1.38	1.45	0.88	0.69	0.79
RC	0.73	1.07	1.04	0.97	1.50	1.28	0.78	0.78	0.80
RT	0.68	0.92	0.98	0.83	1.45	1.27	0.80	0.69	0.85
M1	0.47	0.73	0.71	0.59	1.20	0.80	0.54	0.58	0.54
M2	0.70	1.04	0.92	0.84	1.28	1.07	0.70	0.82	0.74
MT-	0.61	0.88	0.82	0.68	1.13	0.94	0.60	0.71	0.60
M3	0.42	0.65	0.89	0.44	1.19	0.93	0.57	0.47	0.43
MT+	0.53	0.76	0.81	0.57	1.14	0.91	0.45	0.65	0.49

Grade 4									
Test	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
RV	0.72	0.92	1.07	0.81	1.89	1.53	1.00	0.81	0.79
RC	0.83	1.26	1.20	0.87	1.86	1.58	1.08	0.95	0.85
RT	0.73	1.03	1.18	0.89	1.75	1.33	0.90	0.91	0.73
L1	0.76	1.03	1.13	0.81	1.63	1.45	0.85	0.73	0.84
L2	0.92	1.49	1.26	0.97	1.85	2.00	1.48	0.92	1.04
L3	1.03	1.34	1.54	1.00	1.48	1.43	1.15	1.04	1.17
L4	1.04	1.34	1.40	0.94	1.68	1.47	1.05	0.99	1.08
LT	0.85	1.04	1.29	0.89	1.45	1.26	1.07	0.75	0.90
M1	0.83	0.99	1.17	0.76	1.48	1.53	0.96	0.87	0.76
M2	0.92	1.08	1.08	0.89	1.61	1.53	1.18	1.06	0.95
MT	0.76	1.02	1.25	0.79	1.43	1.26	0.91	0.88	0.79
M3	0.49	0.72	0.73	0.54	1.08	1.02	0.64	0.58	0.51
MT+	0.73	0.79	0.95	0.71	1.31	1.17	0.86	0.71	0.73

Note: A = Texas Representative sample; B = White subgroup; C = African American subgroup; D = Hispanic subgroup; E = Special Education subgroup; F = Limited English Proficient subgroup; G = Disadvantaged subgroup; H= Male subgroup; I = Female subgroup

Table 6.1 (continued)
Bootstrap Standard Errors of Building Averages
Fall 2007 Texas Representative Sample
2005 Norms

Grade 5									
Test	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
RV	1.17	1.22	1.75	1.26	1.55	1.86	1.28	1.14	1.15
RC	1.00	1.23	1.70	0.87	1.93	1.85	1.07	1.10	1.12
RT	1.10	1.18	1.70	1.05	1.72	1.77	1.17	1.24	1.12
M1	0.97	1.12	1.68	1.00	1.28	1.95	0.90	0.94	1.09
M2	1.14	1.75	2.04	1.03	1.88	2.52	0.90	1.06	1.39
MT-	1.03	1.30	1.74	1.10	1.52	2.14	0.97	1.02	1.13
M3	0.77	1.07	1.64	0.87	1.32	1.66	1.05	0.92	0.87
MT+	0.85	1.12	1.38	0.75	1.32	1.95	0.99	0.80	1.01
SC	1.04	1.52	1.90	1.07	1.72	1.97	1.16	1.13	1.28

Grade 6									
Test	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
RV	1.37	1.20	2.02	1.44	2.46	2.04	1.14	1.49	1.24
RC	1.59	1.60	2.13	1.59	2.62	2.37	1.54	1.65	1.45
RT	1.31	1.46	2.00	1.53	2.42	2.23	1.46	1.49	1.32
M1	1.22	1.23	1.93	1.23	2.14	2.37	1.18	1.27	1.27
M2	1.21	1.45	1.98	1.37	2.24	2.04	1.42	1.32	1.27
MT-	1.02	1.18	1.45	1.13	2.22	2.27	1.28	1.12	1.27
M3	0.77	1.28	1.39	1.07	1.82	1.73	1.08	0.96	0.94
MT+	1.06	1.10	1.37	0.94	1.79	1.83	1.00	1.12	0.95

Note: A = Texas Representative sample; B = White subgroup; C = African American subgroup; D = Hispanic subgroup; E = Special Education subgroup; F = Limited English Proficient subgroup; G = Disadvantaged subgroup; H= Male subgroup; I = Female subgroup

Table 6.1 (continued)
Bootstrap Standard Errors of Building Averages
Fall 2007 Texas Representative Sample
2005 Norms

Grade 7									
Test	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
RV	1.52	1.45	2.14	1.46	2.49	2.47	1.41	1.76	1.53
RC	1.66	2.12	3.00	1.98	2.44	2.90	1.70	1.78	1.73
RT	1.34	1.60	2.40	1.64	1.95	2.45	1.48	1.58	1.44
L1	1.00	1.51	2.40	1.05	1.61	2.43	1.14	1.02	1.17
L2	1.63	2.02	2.51	1.62	1.94	3.79	1.92	1.63	1.74
L3	1.55	1.92	2.46	1.67	1.87	3.22	1.71	1.63	1.71
L4	1.68	2.49	2.65	1.95	2.32	3.74	2.18	1.78	1.84
LT	1.39	1.97	2.08	1.37	1.53	2.73	1.66	1.44	1.51
M1	1.36	1.59	1.57	1.47	2.11	2.35	1.49	1.24	1.19
M2	1.44	2.24	2.03	1.18	2.17	2.95	1.45	1.58	1.48
MT-	1.23	1.96	1.67	1.12	2.28	2.56	1.32	1.24	1.22
M3	1.09	1.56	1.71	1.33	1.86	2.53	1.41	1.28	1.19
MT+	1.32	1.73	1.44	1.07	1.80	2.36	1.13	1.22	1.29

Grade 8									
Test	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
RV	1.35	1.72	2.26	1.76	2.07	3.43	2.38	1.51	1.44
RC	1.68	2.17	3.03	2.05	2.74	3.80	2.23	1.88	1.64
RT	1.32	1.72	2.28	1.67	2.14	3.43	2.20	1.31	1.49
M1	1.48	1.72	2.06	1.65	2.21	3.76	1.93	1.80	1.47
M2	1.58	1.88	2.95	1.82	2.01	3.83	1.77	1.72	1.72
MT-	1.68	1.83	2.40	1.94	1.59	3.51	1.94	1.91	1.71
M3	1.74	2.19	2.74	2.06	2.07	4.35	2.65	1.86	1.89
MT+	1.48	1.76	2.12	1.82	1.79	3.57	2.06	1.58	1.60
SS	1.47	1.72	2.87	1.87	1.69	3.25	1.87	1.58	1.72

Note: A = Texas Representative sample; B = White subgroup; C = African American subgroup; D = Hispanic subgroup; E = Special Education subgroup; F = Limited English Proficient subgroup; G = Disadvantaged subgroup; H= Male subgroup; I = Female subgroup

Table 6.1 (continued)
Bootstrap Standard Errors of Building Averages
Fall 2007 Texas Representative Sample
2005 Norms

Grade 9									
Test	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
RV	2.40	2.90	2.15	2.15	2.83	3.65	2.41	2.73	2.23
RC	2.71	2.89	2.81	3.04	3.05	4.79	2.02	2.81	2.50
RT	2.58	2.41	2.86	2.48	2.67	4.13	2.32	2.64	2.30
MT-	2.27	2.67	2.82	2.39	2.99	3.99	1.77	2.70	2.00
M3	2.60	2.69	2.69	2.76	3.26	2.65	3.01	2.55	2.39
MT+	2.61	2.66	2.57	2.41	2.31	3.20	2.04	2.83	2.55

Grade 10									
Test	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
RV	1.81	1.48	3.88	1.97	2.82	3.67	2.19	2.20	1.55
RC	2.23	2.24	3.65	2.66	2.86	3.88	3.24	2.35	2.36
RT	1.81	1.56	3.17	2.19	2.44	3.27	2.27	2.18	1.90
L1	1.59	1.87	3.10	2.35	2.40	3.44	1.90	1.85	1.70
L	2.03	1.96	3.46	2.90	2.87	3.45	2.56	2.34	2.14
M	1.46	1.84	3.79	2.24	2.44	3.55	2.41	2.07	1.69
M3	1.40	1.50	3.01	2.19	2.58	4.25	2.16	1.57	1.83
MT+	1.40	1.40	3.37	1.99	2.37	3.05	2.00	1.75	1.59
CT-	1.67	1.67	3.42	2.32	2.19	3.11	2.15	2.03	1.62
CT+	1.60	1.44	3.21	2.84	2.16	3.25	2.43	2.02	1.81
SS	2.04	2.09	4.43	3.09	2.83	3.26	3.25	2.28	2.27

Note: A = Texas Representative sample; B = White subgroup; C = African American subgroup; D = Hispanic subgroup; E = Special Education subgroup; F = Limited English Proficient subgroup; G = Disadvantaged subgroup; H= Male subgroup; I = Female subgroup

Table 6.1 (continued)
Bootstrap Estimates of Standard Errors
Fall 2007 Texas Representative Sample
2005 Norms

Grade 11									
Test	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
RV	1.73	1.67	2.94	2.66	2.71	4.59	2.81	1.85	1.78
RC	1.88	2.40	3.87	2.63	2.71	4.36	3.23	2.28	2.73
RT	1.95	2.04	3.41	2.54	2.08	4.39	2.73	2.22	2.12
L1	1.57	2.14	2.93	2.09	2.80	3.27	3.34	1.75	1.64
L	1.88	1.90	2.56	2.60	2.44	4.01	3.61	1.87	2.31
M	2.10	1.89	2.62	2.62	2.38	4.22	4.25	2.32	2.01
M3	2.18	2.28	2.22	2.03	3.40	4.41	3.23	2.27	2.30
MT+	2.10	1.96	2.38	2.14	2.24	4.27	4.29	2.34	1.96
CT-	1.81	1.67	2.40	2.24	1.95	3.40	3.09	1.84	2.12
CT+	1.76	1.77	2.39	1.84	2.09	3.08	3.07	1.82	2.22
SS	2.66	2.40	3.65	2.78	2.47	4.14	4.35	3.03	2.67
SC	2.98	2.93	3.46	2.70	3.06	4.44	3.96	3.14	2.77

Note: A = Texas Representative sample; B = White subgroup; C = African American subgroup; D = Hispanic subgroup; E = Special Education subgroup; F = Limited English Proficient subgroup; G = Disadvantaged subgroup; H= Male subgroup; I = Female subgroup

---

Efron, B. & Tibshirai, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science, 1* (1), 54-77.

Part 7 Statewide Results for Groups

The performance of special populations in Texas in terms of national norms is of interest to educators and policymakers in the state. The National Comparative Data Study (NCDS) used a detailed sampling design to enhance the interpretability of results for subpopulations. This part of the report provides comprehensive information on the distribution of achievement for Whites, African Americans, Hispanics, students in special education, students with limited English proficiency (LEP), and students who are economically disadvantaged.

Tables 7.1 contains summary statistics for groups on all tests administered to each grade in the NCDS. Standard score means, standard deviations (*SD*), and standard errors of measurement (*SEM*) are included.

Table 7.1
Standard Score Summary Statistics
Fall 2007 Texas Representative Sample by Groups

Grade 3 By Groups		Reading V	Reading RC	Reading RT	Mathematics M1	Mathematics M2	Mathematics M3	Mathematics MT+
Texas Sample	Mean	174.3	180.0	177.3	175.0	174.6	173.1	174.5
	SD	19.9	22.5	19.8	16.7	21.5	14.6	15.2
	SEM	7.5	8.2	5.4	7.1	9.2	6.9	4.3
White	Mean	183.5	188.6	186.2	180.6	183.0	174.7	179.6
	SD	19.5	24.0	20.3	17.2	22.5	14.6	15.9
	SEM	7.3	8.8	5.6	7.3	9.6	6.9	4.5
African American	Mean	170.9	174.7	173.0	170.4	168.5	171.4	170.4
	SD	17.7	19.7	17.3	15.0	18.7	14.6	13.8
	SEM	6.6	7.2	4.7	6.4	8.0	6.9	3.9
Hispanic	Mean	167.7	174.7	171.3	171.7	169.5	172.0	171.3
	SD	17.6	19.4	17.1	15.1	19.0	14.0	13.5
	SEM	6.6	7.1	4.7	6.4	8.1	6.6	3.8
Special Education	Mean	161.0	165.7	163.8	165.9	164.3	165.4	165.6
	SD	19.2	19.7	18.2	16.3	20.0	13.9	14.3
	SEM	7.2	7.2	5.0	7.0	8.5	6.6	4.0
LEP	Mean	162.3	171.1	166.8	170.5	167.1	172.1	170.0
	SD	17.1	18.6	16.3	15.0	18.9	14.3	13.4
	SEM	6.4	6.8	4.5	6.4	8.1	6.8	3.8
Economically Disadvantaged	Mean	168.1	174.0	171.2	171.5	169.0	171.3	170.9
	SD	17.7	19.3	17.0	15.1	18.8	14.0	13.4
	SEM	6.6	7.0	4.7	6.4	8.0	6.6	3.8

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

Table 7.1 (continued)
Standard Score Summary Statistics
Fall 2007 Texas Representative Sample by Groups

Grade 4 By Groups		Reading			Language					Mathematics			
		V	RC	RT	L1	L2	L3	L4	LT	M1	M2	M3	MT+
Texas Sample	Mean	185.6	193.3	189.7	191.2	184.8	190.6	193.8	190.5	191.5	192.0	188.0	190.9
	SD	23.3	28.0	23.9	24.4	31.4	30.2	32.8	25.4	22.3	28.1	15.8	19.1
	SEM	9.4	9.2	6.5	9.7	14.3	12.7	13.8	6.3	10.0	12.7	6.0	5.4
White	Mean	197.9	205.3	201.7	196.7	193.4	200.1	207.8	199.9	199.9	204.1	189.7	198.3
	SD	21.4	28.6	23.1	24.5	33.7	31.8	34.7	26.6	22.1	28.6	15.5	19.3
	SEM	8.6	9.4	6.3	9.7	15.4	13.3	14.6	6.6	10.0	12.9	5.8	5.5
African American	Mean	180.2	185.8	183.2	191.4	177.8	182.3	184.8	184.5	183.5	181.1	184.4	183.6
	SD	20.9	24.6	20.8	23.4	28.3	27.5	29.8	22.9	21.1	24.8	16.0	17.5
	SEM	8.4	8.0	5.7	9.3	12.9	11.5	12.5	5.7	9.5	11.2	6.0	4.9
Hispanic	Mean	177.0	185.6	181.5	186.0	179.0	184.7	185.2	184.1	186.7	185.2	187.1	186.7
	SD	20.9	24.6	21.0	22.9	27.9	26.8	28.0	22.1	20.3	24.7	15.3	17.0
	SEM	8.4	8.0	5.7	9.1	12.7	11.2	11.7	5.5	9.1	11.1	5.8	4.8
Special Education	Mean	171.3	175.2	173.7	175.9	165.6	172.8	172.3	172.3	178.3	174.9	180.6	178.3
	SD	25.4	27.3	24.7	24.0	25.7	24.9	28.6	21.9	23.1	26.7	16.4	19.3
	SEM	10.2	8.9	6.7	9.5	11.7	10.4	12.0	5.4	10.4	12.0	6.2	5.5
LEP	Mean	169.0	178.3	173.9	179.0	172.9	178.7	178.1	177.5	183.8	180.6	188.0	184.6
	SD	20.9	23.4	20.5	21.0	25.6	24.4	25.7	19.7	20.7	23.9	16.1	17.1
	SEM	8.4	7.7	5.6	8.3	11.7	10.2	10.8	4.9	9.3	10.8	6.1	4.8
Economically Disadvantaged	Mean	176.7	184.1	180.7	185.3	175.5	181.2	183.1	181.8	184.6	183.1	184.9	184.7
	SD	21.3	24.2	20.9	22.3	27.1	25.8	28.0	21.3	20.5	24.0	14.9	16.6
	SEM	8.6	7.9	5.7	8.8	12.4	10.8	11.7	5.3	9.2	10.8	5.6	4.7

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

Table 7.1 (continued)
Standard Score Summary Statistics
Fall 2007 Texas Representative Sample by Groups

Grade 5 By Groups		Reading V	Reading RC	Reading RT	Mathematics M1	Mathematics M2	Mathematics M3	Mathematics MT+	Science SC
Texas Sample	Mean	197.6	207.4	202.6	205.2	207.5	201.2	205.2	210.7
	SD	28.2	28.8	26.4	25.8	33.5	22.5	23.7	32.6
	SEM	11.5	9.4	7.2	10.9	15.2	8.8	6.7	14.8
White	Mean	211.6	220.9	216.2	215.1	220.1	204.2	213.5	225.5
	SD	25.6	27.1	24.3	24.7	33.0	22.6	23.2	32.1
	SEM	10.5	8.8	6.6	10.4	15.0	8.9	6.5	14.6
African American	Mean	192.8	200.6	197.0	196.9	195.4	196.5	197.2	200.4
	SD	25.4	25.6	23.0	24.0	29.7	22.3	21.9	28.5
	SEM	10.4	8.3	6.3	10.1	13.5	8.8	6.2	13.0
Hispanic	Mean	188.2	198.6	193.6	199.5	200.8	199.8	200.6	201.7
	SD	26.4	26.9	24.5	24.2	31.6	21.8	22.3	29.6
	SEM	10.8	8.8	6.7	10.2	14.3	8.6	6.3	13.5
Special Education	Mean	178.5	183.4	180.9	187.9	186.2	186.6	187.3	194.8
	SD	26.7	28.0	25.1	22.8	28.6	21.5	20.6	26.9
	SEM	10.9	9.1	6.8	9.6	13.0	8.4	5.8	12.2
LEP	Mean	175.6	183.9	180.0	193.1	191.6	196.5	194.3	188.5
	SD	25.4	26.0	23.6	24.0	30.5	22.0	22.0	26.4
	SEM	10.4	8.5	6.4	10.1	13.8	8.6	6.2	12.0
Economically Disadvantaged	Mean	188.9	198.5	193.9	198.7	198.9	198.1	199.3	200.9
	SD	25.8	26.9	24.2	24.1	30.8	22.1	22.0	28.7
	SEM	10.5	8.8	6.6	10.2	14.0	8.7	6.2	13.1

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

Table 7.1 (continued)
Standard Score Summary Statistics
Fall 2007 Texas Representative Sample by Groups

Grade 6 By Groups		Reading V	Reading RC	Reading RT	Mathematics M1	Mathematics M2	Mathematics M3	Mathematics MT+
Texas Sample	Mean	209.7	214.9	212.8	217.5	216.9	214.5	216.8
	SD	28.3	33.7	29.1	28.7	33.0	22.6	24.4
	SEM	10.5	11.2	7.6	10.1	15.7	9.4	6.7
White	Mean	224.1	230.4	227.6	228.6	229.2	217.9	225.8
	SD	27.2	33.8	28.6	28.0	34.4	23.8	25.2
	SEM	10.1	11.2	7.5	9.9	16.3	9.9	6.9
African American	Mean	203.8	205.7	205.1	207.9	206.8	210.6	208.9
	SD	25.5	29.9	25.4	26.3	29.2	20.8	21.3
	SEM	9.5	9.9	6.7	9.3	13.9	8.6	5.9
Hispanic	Mean	199.7	205.1	202.7	210.9	209.6	212.2	211.4
	SD	24.9	29.7	25.3	26.7	29.3	21.2	21.8
	SEM	9.2	9.9	6.6	9.4	13.9	8.8	6.0
Special Education	Mean	192.3	189.3	191.3	196.2	192.3	198.0	195.7
	SD	24.9	28.8	25.1	25.4	29.3	21.2	21.5
	SEM	9.2	9.6	6.6	8.9	13.9	8.8	5.9
LEP	Mean	188.3	187.7	188.6	201.3	197.7	208.1	202.8
	SD	22.0	24.7	21.3	25.9	27.7	21.3	20.8
	SEM	8.2	8.2	5.6	9.1	13.1	8.8	5.7
Economically Disadvantaged	Mean	199.4	203.9	202.0	210.1	208.3	211.3	210.4
	SD	24.7	29.3	24.8	26.5	29.8	21.6	21.9
	SEM	9.2	9.7	6.5	9.3	14.1	9.0	6.0

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

Table 7.1 (continued)
Standard Score Summary Statistics
Fall 2007 Texas Representative Sample by Groups

Grade 7 By Groups		Reading			Language					Mathematics			
		V	RC	RT	L1	L2	L3	L4	LT	M1	M2	M3	MT+
Texas Sample	Mean	223.4	229.0	226.4	231.3	218.5	228.9	231.1	227.9	236.7	233.0	225.5	232.1
	SD	30.0	35.7	31.0	31.3	41.2	42.6	45.1	35.2	28.0	37.6	29.8	28.1
	SEM	10.5	10.4	7.4	9.2	15.2	13.6	15.6	6.8	9.2	15.0	12.2	7.1
White	Mean	239.2	245.1	242.3	237.3	231.9	242.7	248.8	240.6	247.4	247.5	231.4	242.4
	SD	26.6	34.2	28.7	31.5	42.3	44.0	45.3	35.8	27.4	37.6	30.6	28.5
	SEM	9.3	10.0	6.9	9.3	15.6	14.1	15.6	6.9	9.0	15.0	12.5	7.2
African American	Mean	215.7	215.8	216.1	229.4	209.5	217.1	216.3	218.6	224.0	218.8	217.6	220.5
	SD	27.9	33.3	28.5	31.2	38.3	38.9	41.2	32.2	26.3	33.1	28.1	25.3
	SEM	9.7	9.7	6.8	9.2	14.1	12.4	14.2	6.2	8.7	13.2	11.5	6.4
Hispanic	Mean	212.0	218.7	215.6	225.6	208.5	219.2	219.8	218.7	230.6	224.2	221.8	225.8
	SD	27.0	31.9	27.3	29.3	36.8	38.0	40.3	31.1	25.4	34.1	27.9	25.0
	SEM	9.4	9.3	6.5	8.6	13.6	12.1	13.9	6.0	8.4	13.6	11.4	6.3
Special Education	Mean	199.8	196.1	198.2	201.8	187.5	193.8	191.0	194.0	209.2	202.4	204.1	205.5
	SD	25.8	27.8	24.3	23.5	29.2	27.8	30.5	21.6	23.8	27.2	24.2	19.6
	SEM	9.0	8.1	5.8	6.9	10.8	8.9	10.5	4.2	7.9	10.9	9.9	5.0
LEP	Mean	192.1	197.0	194.8	209.2	189.9	201.1	193.5	198.9	217.4	206.1	217.2	214.0
	SD	24.6	27.4	23.4	25.1	29.5	29.8	33.3	24.2	23.6	30.4	27.2	22.4
	SEM	8.6	8.0	5.6	7.4	10.9	9.5	11.5	4.7	7.8	12.2	11.1	5.7
Economically Disadvantaged	Mean	209.8	215.2	212.6	224.1	206.8	216.4	215.8	216.3	227.4	220.7	218.9	222.6
	SD	27.2	32.0	27.5	29.5	37.2	37.1	39.9	30.8	25.9	33.9	27.8	25.0
	SEM	9.5	9.3	6.6	8.7	13.7	11.8	13.8	5.9	8.6	13.6	11.4	6.3

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

Table 7.1 (continued)
Standard Score Summary Statistics
Fall 2007 Texas Representative Sample by Groups

Grade 8 By Groups		Reading V	Reading RC	Reading RT	Mathematics M1	Mathematics M2	Mathematics M3	Mathematics MT+	Social Studies SS	Science SC
Texas Sample	Mean	233.9	241.0	237.7	246.9	244.4	237.7	243.5	232.0	241.5
	SD	30.0	39.4	32.6	31.6	41.0	36.5	31.7	37.2	37.7
	SEM	10.3	11.4	7.6	11.8	16.3	15.0	8.2	13.3	14.1
White	Mean	247.3	257.0	252.3	257.6	258.2	247.3	255.1	244.1	254.0
	SD	27.5	37.8	30.7	30.6	42.0	37.3	32.0	37.7	38.5
	SEM	9.4	10.9	7.1	11.4	16.7	15.3	8.3	13.4	14.4
African American	Mean	230.7	232.2	231.6	239.9	236.1	231.7	236.6	225.3	233.3
	SD	27.6	36.6	29.8	29.2	37.1	33.7	28.4	34.7	34.7
	SEM	9.4	10.5	6.9	10.9	14.7	13.9	7.4	12.4	12.9
Hispanic	Mean	223.4	229.9	226.9	239.6	235.0	231.0	235.6	223.3	232.3
	SD	28.2	36.8	30.0	30.1	37.5	34.3	28.9	34.2	34.2
	SEM	9.6	10.6	7.0	11.3	14.9	14.1	7.5	12.2	12.8
Special Education	Mean	208.0	205.5	207.1	217.1	209.8	212.9	213.5	206.5	217.1
	SD	27.2	32.4	27.0	28.5	31.4	30.6	24.9	29.2	31.0
	SEM	9.3	9.3	6.3	10.7	12.5	12.6	6.4	10.4	11.6
LEP	Mean	212.5	213.4	213.2	229.4	223.0	226.2	226.4	214.1	223.5
	SD	28.6	36.5	29.7	30.7	35.2	33.8	28.3	33.0	31.6
	SEM	9.8	10.5	6.9	11.5	14.0	13.9	7.3	11.8	11.8
Economically Disadvantaged	Mean	222.5	228.3	225.7	238.0	233.1	230.3	234.3	220.5	230.8
	SD	28.7	37.0	30.3	29.9	37.3	34.7	28.9	33.4	34.9
	SEM	9.8	10.7	7.0	11.2	14.8	14.3	7.5	11.9	13.0

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

Table 7.1 (continued)
Standard Score Summary Statistics
Fall 2007 Texas Representative Sample by Groups

Grade 9 By Groups		Reading			Mathematics		
		V	RC	RT	M	M3	MT+
Texas Sample	Mean	240.8	244.1	242.5	255.7	248.0	253.3
	SD	31.9	43.1	35.2	36.8	35.4	33.5
	SEM	9.6	12.3	7.8	12.1	14.5	7.6
White	Mean	258.7	267.1	263.3	273.8	261.2	269.9
	SD	27.5	41.1	32.0	34.1	35.3	31.7
	SEM	8.3	11.8	7.1	11.3	14.4	7.2
African American	Mean	237.0	232.0	234.8	243.7	240.6	242.8
	SD	28.1	37.1	29.9	32.1	33.0	29.4
	SEM	8.4	10.6	6.6	10.6	13.5	6.6
Hispanic	Mean	227.8	228.1	228.2	244.0	239.2	242.9
	SD	29.5	37.4	30.7	33.7	32.2	30.1
	SEM	8.9	10.7	6.8	11.1	13.2	6.8
Special Education	Mean	216.7	211.4	214.3	225.4	221.7	224.3
	SD	30.7	33.9	29.1	29.0	29.1	24.9
	SEM	9.2	9.7	6.4	9.6	11.9	5.6
LEP	Mean	209.1	205.8	207.9	227.8	227.0	228.2
	SD	28.1	31.3	26.7	30.1	30.9	26.6
	SEM	8.4	9.0	5.9	9.9	12.6	6.0
Economically Disadvantaged	Mean	228.8	228.2	228.5	244.5	239.0	242.8
	SD	29.3	36.3	29.9	32.9	32.0	29.2
	SEM	8.8	10.4	6.6	10.9	13.1	6.6

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

Table 7.1 (continued)
Standard Score Summary Statistics
Fall 2007 Texas Representative Sample by Groups

Grade 10 By Groups		Reading V	Reading RC	Reading RT	Language L1	Language RW	Mathematics M	Mathematics M3	Mathematics MT+	Social Studies SS	Science SC
Texas Sample	Mean	250.2	253.0	251.8	258.1	256.7	257.6	248.2	254.8	246.0	248.9
	SD	35.2	46.5	38.1	36.5	43.5	40.5	35.1	35.5	46.3	48.2
	SEM	11.4	13.2	8.8	12.7	12.1	13.9	13.2	7.9	15.2	18.5
White	Mean	263.2	269.9	266.7	265.0	271.5	270.4	256.0	266.0	260.2	264.3
	SD	33.3	47.0	37.5	37.3	43.7	40.6	35.7	35.7	47.0	49.0
	SEM	10.8	13.3	8.6	13.0	12.2	13.9	13.4	8.0	15.4	18.8
African American	Mean	239.1	239.2	239.4	252.7	243.1	241.2	237.3	240.1	231.1	231.0
	SD	33.7	40.6	33.9	34.6	39.4	35.7	32.8	31.4	41.5	41.5
	SEM	10.9	11.5	7.8	12.0	11.0	12.2	12.3	7.0	13.6	15.9
Hispanic	Mean	240.1	239.2	239.8	251.3	245.1	249.0	242.7	247.3	234.4	237.0
	SD	32.9	41.0	34.0	33.8	39.1	37.2	32.3	32.0	41.6	43.1
	SEM	10.7	11.6	7.8	11.8	10.9	12.7	12.1	7.2	13.7	16.5
Special Education	Mean	220.7	220.2	220.8	228.6	220.6	227.2	222.9	226.2	219.0	221.3
	SD	33.3	36.3	31.7	29.4	32.8	34.1	29.1	29.2	34.3	37.2
	SEM	10.8	10.3	7.3	10.2	9.2	11.7	10.9	6.5	11.3	14.3
LEP	Mean	223.4	218.4	220.9	235.6	223.6	234.8	233.8	234.6	216.6	220.4
	SD	34.2	37.6	32.9	32.2	34.8	35.1	30.9	29.8	35.2	39.0
	SEM	11.1	10.6	7.6	11.2	9.7	12.0	11.6	6.7	11.6	15.0
Economically Disadvantaged	Mean	240.7	243.4	242.3	251.9	247.2	249.6	241.5	247.2	235.7	238.2
	SD	33.5	41.1	34.6	34.5	38.8	37.3	32.2	32.3	42.3	43.9
	SEM	10.9	11.6	8.0	12.0	10.8	12.8	12.1	7.2	13.9	16.8

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

Table 7.1 (continued)
Standard Score Summary Statistics
Fall 2007 Texas Representative Sample by Groups

Grade 11 By Groups		Reading V	Reading RC	Reading RT	Language L1	Language RW	Mathematics M	Mathematics M3	Mathematics MT+	Social Studies SS	Science SC
Texas Sample	Mean	261.7	264.3	263.3	270.2	271.0	271.1	264.0	269.4	258.5	264.0
	SD	36.2	47.5	38.8	36.7	44.1	40.6	33.7	35.2	47.7	47.1
	SEM	11.7	13.2	8.8	12.4	11.9	14.0	11.1	7.5	15.0	15.0
White	Mean	274.6	279.4	277.3	276.7	283.4	282.9	271.5	279.9	274.4	278.2
	SD	35.3	47.5	38.3	37.7	44.7	41.1	35.6	36.3	48.1	48.4
	SEM	11.4	13.2	8.7	12.8	12.1	14.2	11.7	7.7	15.1	15.5
African American	Mean	250.0	248.3	250.0	265.4	257.5	252.6	253.6	253.9	241.0	248.6
	SD	33.1	43.2	34.4	34.0	40.2	36.0	29.0	29.8	41.2	40.8
	SEM	10.7	12.0	7.8	11.5	10.9	12.4	9.5	6.3	13.0	13.0
Hispanic	Mean	251.2	253.3	252.4	263.7	261.8	264.4	258.8	263.0	246.8	253.4
	SD	33.3	43.6	35.4	34.3	40.3	37.3	30.8	31.8	43.8	43.1
	SEM	10.7	12.1	8.1	11.6	10.9	12.9	10.1	6.7	13.8	13.8
Special Education	Mean	227.9	227.3	227.4	236.2	230.2	239.3	237.3	239.1	226.0	235.4
	SD	33.6	39.2	32.7	30.5	33.2	33.5	28.1	28.0	36.1	38.3
	SEM	10.8	10.9	7.5	10.3	9.0	11.6	9.2	5.9	11.4	12.2
LEP	Mean	231.1	234.1	232.9	244.9	240.5	253.3	249.8	252.5	225.9	240.4
	SD	35.5	39.1	33.8	31.8	33.4	34.8	27.0	29.2	38.4	40.3
	SEM	11.4	10.8	7.7	10.8	9.0	12.0	8.9	6.2	12.1	12.9
nomically lvantaged	Mean	250.1	251.9	251.3	261.7	259.8	262.0	258.1	261.1	244.3	251.6
	SD	33.0	43.1	34.8	34.0	40.0	37.3	31.5	31.9	42.8	42.9
	SEM	10.6	12.0	7.9	11.5	10.8	12.9	10.4	6.8	13.5	13.7

Note. V = vocabulary; RC = reading comprehension; RT = reading total; L1 = spelling; L2 = capitalization; L3 = punctuation; RW = revising written materials; L4 = usage/expression; LT = language total; M1 = concepts & estimation; M2 = problem solving and data interpretation; M = *ITED* concepts & problem solving; M3 = computation; MT+ = math total; SS = social studies; SC = science

Figures 7.1 through 7.4 depict the growth trajectories of mean standard scores (SS) in Reading and Math across grade levels. Figures 7.1 and 7.3 show growth trajectories for Whites, African American, and Hispanics. Figures 7.2 and 7.4 show growth trajectories for special-education, LEP, and economically disadvantaged students.

Figure 7.1
Reading Total Standard Score by Race and Grade
Fall 2007 Texas Representative Sample

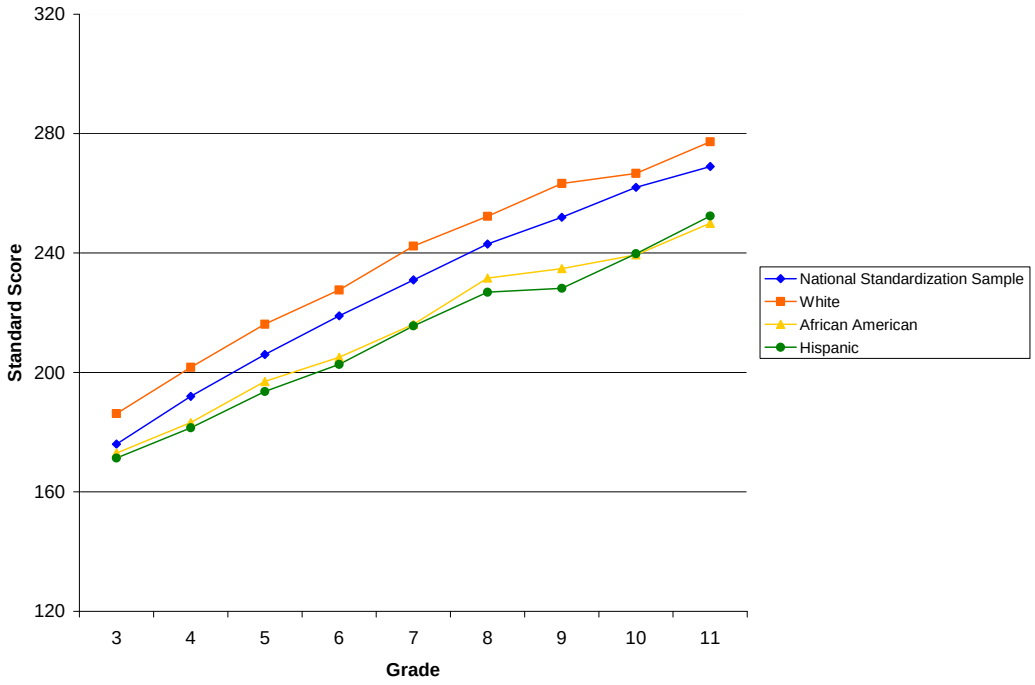


Figure 7.2
Reading Total Standard Score by Special Population and Grade
Fall 2007 Texas Representative Sample

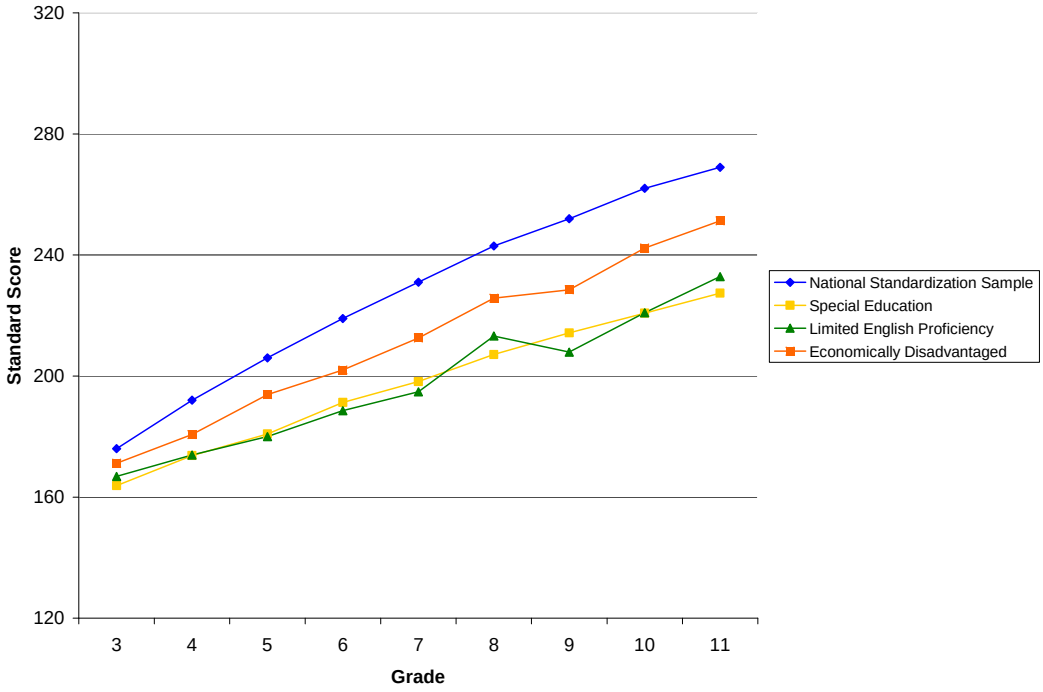


Figure 7.3
Math Total Standard Score by Race and Grade
Fall 2007 Texas Representative Sample

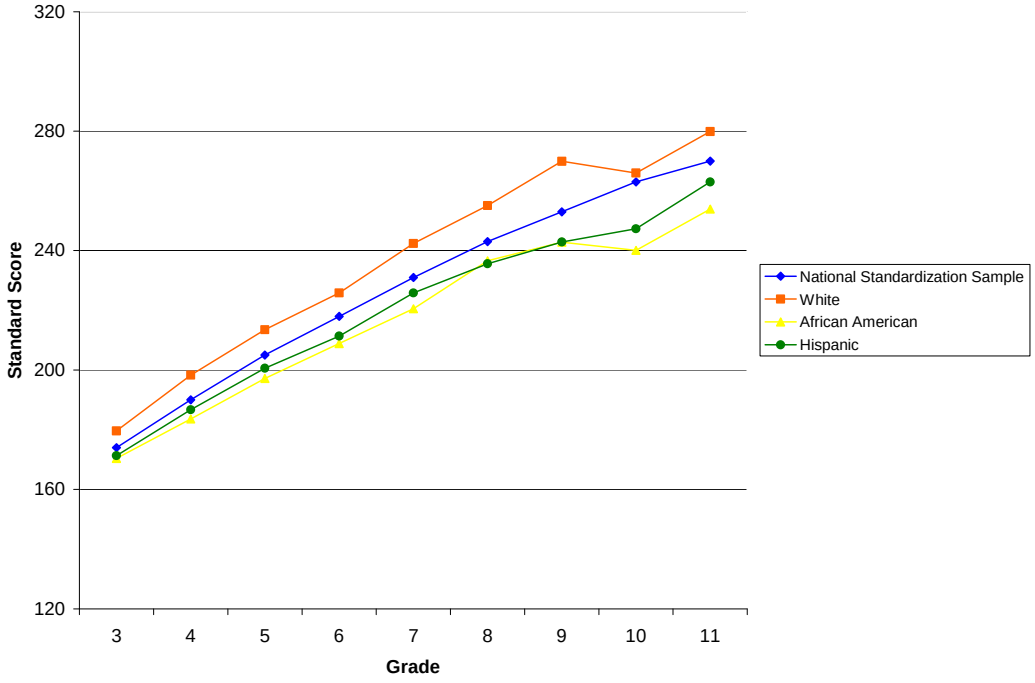


Figure 7.4
Math Total Standard Score by Special Population and Grade
Fall 2007 Texas Representative Sample