HumRRO Report No.
FR-04-26

HumRRO
Human Resources Research Organization

Relationships Among the Pennsylvania System of School Assessment (PSSA) Scores, SAT Scores, and Self-Reported High School Grades for the Classes of 2002 and 2003

Milton E. Koger
Arthur A. Thacker
Emily R. Dickinson

Human Resources Research Organization (HumRRO)
950 Breckenridge Lane, Suite 170
Louisville, KY 40207
Phone (502) 721-9045
FAX (502) 721-9983

Prepared for:

Central Susquehanna Intermediate Unit
P.O. Box 213
Lewisburg, PA 17837

Abstract

As part of Pennsylvania's examination of the Pennsylvania System of School Assessment (PSSA), Human Resources Research Organization (HumRRO) examined whether PSSA scores are appropriately related to other measures of educational achievement by investigating the relationship between students' performance on PSSA, SAT, and student-reported grade-point average and course grades obtained from the SAT's Student Descriptive Questionnaire (SDQ). Additionally, student demographic information obtained from the PSSA database was used to examine demographic groups of interest as designated by the No Child Left Behind (NCLB) Act of 2001. However, because students self-select to take the SAT as part of their college application process, the student population examined in this report consists predominantly of the higher achieving students (about the upper two-thirds) attending public schools in Pennsylvania.

The data presented in this report suggest the following main points:

- Students who do well on any of the assessments tend to do well on all of the assessments—PSSA, SAT, GPA, or course grades.
- When schools' means are calculated using only the SAT-taking population of students, schools with high scores on SAT also have high PSSA scores. GPA and course grades are also related to both PSSA and SAT although not as strongly as the relationship between the PSSA and SAT.

PSSA exhibits strong convergent validity coefficients. Correlations are very high between PSSA and SAT. PSSA is positively correlated with other measures of student achievement including course grades and GPA. Gains on PSSA are reflective of changes on SAT at the school level. All these data provide strong evidence in support of PSSA as a valid measure of student achievement.

There are, however, a few clarifications and qualifications to these general conclusions, but none that diminish the basic findings. These clarifications and qualifications include the following:

Changes in scores. Both mean PSSA and SAT scores experienced an increase. Changes between the two assessments are positively correlated. Schools that have gained on one assessment tend to have also gained on the other assessment. However, the same relationship is also true for those schools that have decreased.

PSSA and SAT are tied to different content domains, use differently formatted items, and were designed to serve very different purposes. However, it is clear from the data that students who tend to perform well on the PSSA can also be expected to perform well on the SAT and vice-versa. It is clear from the data that schools that perform well on one test can be expected to perform well on the other. It is also clear that schools that improve on one test can be expected to improve on the other.

Discriminant validity. At both the student level and the school level, the different assessments of mathematics achievement are more highly related to each other than to assessments of other subjects. Students with high mathematics scores on one assessment will tend to do well on all other assessments, but that tendency is most pronounced for other mathematics assessments. The same holds for school scores. There is a similar differentiation on

reading/verbal assessments at both the student and school level, but not as pronounced as in mathematics.

GPA shows positive relationships with both PSSA and SAT assessments at the student and school level, but not as pronounced as the relationship between the two assessments. This is interpreted as being due to differences in schools' grading practices. Grading practices apparently differ sufficiently to reduce school-level associations with PSSA and SAT scores.

Neither gender, race, socioeconomic status, nor limited English proficiency appear to influence the PSSA scores any more than would be expected based on observed differences for SAT scores. In other words, PSSA items are not injecting any unexpected gender, racial/ethnic, socioeconomic status, or limited English proficiency bias.

Relationships Among the Pennsylvania System of School Assessment (PSSA) Scores, SAT Scores, and Self-Reported High School Grades for the Classes of 2002 and 2003

Table of Contents

List of Tables

List of Figures

Relationships Among the Pennsylvania System of School Assessment (PSSA) Scores, SAT Scores, and Self-Reported High School Grades for the Classes of 2002 and 2003

Introduction

As part of Pennsylvania's System of School Assessment (PSSA), students in Grades 5, 8, and 11 take tests in reading and mathematics. Students in Grades 6, 9, and 11 are assessed in writing. The annual PSSA is a standards-based criterion-referenced assessment used to measure a student's attainment of academic standards while also determining the degree to which school programs enable students to attain proficiency standards. PSSA results are produced at student and school levels. Student scores, which are provided to their respective schools, can be used diagnostically to identify students in need of additional educational opportunities. School scores are provided for schools and districts to use for curriculum and instruction improvement discussions and planning (www.pde.state.pa.us/a_and_t/site).

PSSA tests students' abilities in relation to academic standards adopted in 1999. The standards identify what students should know and be able to do within each subject area at each designated grade level. PSSA test items are linked to the standards and PSSA scores are used to stratify student performance within the standards. Students receive designations of Below Basic, Basic, Proficient, or Advanced, depending on how they score in each tested subject. These proficiency levels are determined using cut scores on the PSSA measurement scale. Cut scores were determined using the Bookmark (Lewis, Mitzel & Green, 1996) and Borderline Groups (Livingston & Zieky, 1978) standards-setting procedures.

Because PSSA serves as an assessment for individual students and for schools, it is configured using common and matrix items. Common items are administered to all students and are used to create all student-level measures. Matrix items are administered by form such that each student only takes a portion of the matrix items. There are typically 10-12 matrix forms, spiraled within classrooms to ensure that a random and representative sample of students receives each form. The matrix items add to the overall content coverage of the PSSA and allow for better diagnostic data to be produced at school and district levels.

In addition to the PSSA, many students also take the SAT test as a requirement for admission to Pennsylvania colleges and universities. SAT is designed to be predictive of college performance. Most Pennsylvania students who plan to attend college take the SAT 1 version of the test, which includes a verbal and a mathematics component. Both PSSA and SAT test data includes student demographic information, allowing students to be matched by name and date of birth and allowing the examination of test performance in relation to various demographic factors.

The quality of an assessment is typically characterized by its reliability and validity. The usual measure of reliability is an indication of how similar a student's scores on an assessment would be if student took the test multiple times, or test-retest reliability. Reliability is largely concerned with the consistency of an assessment. Reliability coefficients are provided in the PSSA technical manuals produced by Data Recognition Corporation (DRC) each year (Mead & Melby, 2002; Mead & Melby, 2003; and Mead, Smith, & Swanlund, 2003). PSSA test-retest reliabilities ranged from 0.93 to 0.94 for math and from 0.92 to 0.94 for reading for the full set of items (common + matrix) in 2002. They were slightly lower for the common items only, 0.92 for

math for all grades and ranging from 0.88 to 0.91 for reading (Estimates were very similar in 2001 and 2003). This reduction in reliability for the common items reflects that reliability statistics are influenced to a great extent by test length. The common section of the PSSA ranges from a minimum of 60 score points (for Grade 5 reading) to 80 or more score points (for all other grade/subjects). This relatively large number of items helps account for PSSA's very high reliability estimates.

So, we know that according to traditional reliability statistics, PSSA is a reliable measurement instrument. This report is concerned with the validity of the PSSA. Simply put, does the PSSA measure what it purports to measure? This is not as simple a question as it seems. One way of investigating validity is to compute convergent validity coefficients. Convergent validity coefficients are measures of the relationship between two separate tests of student ability for the same subject matter. They are correlations between students' performance on the two tests. Often, convergent validity statistics are used to ensure that multiple forms from a single assessment are measuring essentially equivalent content. When convergent validity coefficients are calculated for that purpose, higher correlations are considered better than lower correlations. When we correlate PSSA scores with scores from SAT, however, the interpretation of the coefficients requires additional explanation.

PSSA tests student ability related to content that is specific to Pennsylvania. The academic standards outline the content that Pennsylvania has collectively decided is essential for students in its public schools. The extent to which another test measures content that is different from Pennsylvania's will limit the strength of the correlation between the two assessments. In fact, if the correlation were very high, it would raise questions as to whether the two assessments were measuring anything different at all, and consequently whether both are necessary. PSSA is composed of multiple-choice and performance-task items. If another test uses only multiple-choice items, there may be method differences that reduce the strength of the correlations, as well. Still, all the tests analyzed as part of this research report scores for recognizable subjects, such as mathematics and reading. We expect student scores on the various tests to be related. We are left looking for what Hoffman (1998) refers to as "Goldilocks" correlations. Correlations between PSSA and SAT should fall in the not-too-high and not-too-low category.

Comparisons between PSSA and other tests can also provide insights into performance of various identified groups of interest. The federal No Child Left Behind (NCLB) Act of 2001 (NCLB, 2004) requires that each state measure the proficiency of its students as well as reduce gaps between traditionally lower-performing groups and their higher-performing peers. NCLB requires that states reduce gaps associated with race, socioeconomic status, language and disability. Comparisons of group performance on PSSA can determine if there is differential impact, a necessary but insufficient condition for bias, associated with the assessment. This report examines gender, race, language and socioeconomic status for differential impact. Due to inconsistencies in the manner in which students with disabilities are treated when taking the assessments, disability is not considered.

Description of Data

Student-level PSSA files from the administrations in 2001, 2002 and 2003 were provided by DRC. HumRRO downloaded the data from a secure file transfer protocol (FTP) site. Files included student responses and scores for reading and mathematics as well as demographic information. Files also identified students by name, school, and district. In most cases, student

birthdates were also included in the files. No student-level records were shared with the Pennsylvania Department of Education (PDE). Files included student-level identifiers for the expressed purpose of matching PSSA results with results from other assessments. Once matching was complete, student names and birthdates were purged from all working files. All PSSA files were provided as text files. Text files were converted to SAS databases prior to analyses or merging with other files.

College Board provided state-level SAT data from administrations conducted in 2001, 2002, and 2003. College Board provided only data from students enrolled in public schools. Private school students are not considered under the jurisdiction of the Pennsylvania Department of Education, so identifiers were not available for those students, even if their schools opted to take the PSSA. Consequently, the number of students and mean scores reported in this study do not match the numbers and means provided on College Board's web site for Pennsylvania.

Descriptive Statistics

PSSA

Tables 1-3 present a summary of the data included in the PSSA files. Statistics represent scale scores for reading and mathematics and are the same scale scores reported to students. They are based only on the common set of items to which all students respond. The number of cases presented in the first column is the total number of students in the file. Not all of those students had data in the scale-score fields, so n-counts are smaller than the number of cases. Scale score minimums were limited to 700 for 2002 and 2003.

Table 1. Summary Data from PSSA 2001

Grade		Mathematics (Non-Missing)	Reading (Non-Missing)
Grade 5	N	142,858	142,810
(Cases = 145,440)	Mean	1,311.42	1,312.14
	S.D.	190.65	202.68
	Minimum	595	381
	Maximum	2,212	2,506
Grade 8	N	139,322	139,073
(Cases = 143,119)	Mean	1,310.46	1,304.56
	S.D.	200.35	210.06
	Minimum	373	100
	Maximum	2,201	2,214
Grade 11	N	116,137	115,566
(Cases = 122,332)	Mean	1,304.11	1,296.87
	S.D.	220.57	217.57
	Minimum	100	100
	Maximum	2,481	2,088

Table 2. Summary Data from 2002 PSSA

Grade		Mathematics (Non-Missing)	Reading (Non-Missing)
Grade 5	N	143,159	143,046
(Cases =145,688)	Mean	1,315.19	1,316.39
	S.D.	208.32	205.13
	Minimum	700	700
	Maximum	2409	2630
Grade 8	N	141,910	141,801
(Cases =146,127)	Mean	1,315.78	1,309.70
	S.D.	199.62	218.87
	Minimum	700	700
	Maximum	2,293	2,345
Grade 11	N	123,550	123,234
(Cases =129,475)	Mean	1,314.47	1,312.86
	S.D.	228.29	215.13
	Minimum	700	700
	Maximum	2,437	2,362

Table 3. Summary Data from 2003 PSSA

Grade Level		Mathematics (Non-Missing)	Reading (Non-Missing)
Grade 5	N	141,402	141,530
(Cases =143,502)	Mean	1,338.13	1,331.60
	S.D.	207.98	224.59
	Minimum	700	700
	Maximum	2,325	2,444
Grade 8	N	144,964	145,302
(Cases =148,760)	Mean	1,319.92	1,336.97
	S.D.	209.04	227.05
	Minimum	719	700
	Maximum	2,228	2,666
Grade 11	N	126,941	127,427
(Cases =133,168)	Mean	1,314.78	1,315.61
	S.D.	214.78	235.39
	Minimum	700	700
	Maximum	2,238	2,355

SAT Data

Table 4 contains descriptive statistics for SAT data. College Board provided SAT data for 2001, 2002, and 2003. Note that the number of students taking the SAT is considerably fewer than those taking the Grade 11 PSSA. Students typically take the SAT in preparation for post-

secondary education. Students with no plans for continuing their education past high school may not take the SAT. The minimum and maximum scores on SAT, for both verbal and mathematics tests, was 200 and 800, respectively, for all years.

Table 4. Summary Data from SAT for 2001 through 2003

SAT Year		Verbal	Mathematics
2001	N	75,696	75,696
	Mean	497.34	499.18
	S.D.	103.50	109.61
2002	N	76,924	76,924
	Mean	494.68	500.06
	S.D.	104.08	110.73
2003	N	79,149	79,149
	Mean	497.32	502.96
	S.D.	103.25	110.20

Matching Data

All comparison test data had to be matched to PSSA data in order to perform correlations and other calculations. The SAT data provided by the College Board consisted of separate annual data files for Pennsylvania public school students graduating from 2001 to 2003. PSSA data files consisted of data files for 11th grade students from 2001 to 2003. SAT data files were merged with PSSA data files to produce data files for students graduating in 2003 (2003 SAT file merged with 2002 PSSA 11th grade file) and 2002 (2002 SAT file merged with the 2001 PSSA 11th grade file).

Rules were established to ensure the consistency of merging data. Each merge attempt resulted in three files—successfully merged student data, unmerged students from File 1, and unmerged students from File 2. Each successive merge attempt was made using only students from the unmerged files. The successfully merged data was then appended to create the final working files.

The data file for students graduating in 2003 was obtained by exactly matching student names and birthdates in the two files. This was done in four merge cycles. For the first cycle, we merged on exact matches of student last name, first name, middle initial, and birth date. For the second cycle, we merged on student last name, first name, and birth date. For the third cycle, we merged on last name, first name truncated to the first four letters, and birth date. For the fourth cycle, we merged on last name truncated to the first four letters, first name truncated to the first four letters, and birth date.

Because of an apparent error in the birth date for one of the files for students graduating in 2002, birth date could not be used. As a result, we merged files for this class in two cycles. For the first cycle, we merged on last name, first name, and middle initial. For the second cycle, we merged on last name and first name. A more detailed explanation of this is found in the following paragraph.

Table 5 presents the proportion of students retained in the final PSSA/SAT data set, reported by graduation year. Following the final merge cycle, slightly more than 90% of the original cases in the SAT data file were retained. The successful merging of the files was

hampered by inconsistent reporting of students' names across the tests. Student or coding errors/differences created inconsistencies between the two tests (e.g., O BRIEN versus O'BRIEN, or TOM versus THOMAS). Additionally, missing birth date or invalid birth date (e.g., 13th month, 32nd day of 2001 for a 2003 high school graduate) by a student or in the coding process on one of the tests may also have caused a portion of students' files not to merge. Also of note is the fact that the 2001 and 2002 PSSA files were missing more than 10,000 birth dates in the two files combined. Additionally, for the PSSA 2002 and the SAT 2003 data files, there appeared to be an error in the birth date file that resulted in virtually no matches between the files when birth date was included. As a result, the two files went through only two merge cycles which did not include birth date.

Table 5. Percentage of Students Retained in File Merge for Graduation Years 2003 and 2002

Year	Merge Cycle	Students in SAT File	Percent of SAT File
2003	Beginning File	79,095	100%
	1st Merge *(last name, first name, middle initial, birth date*	60,044	75.9%
	2nd Merge *(last name, first name, birth date)*	69,416	87.8%
	3rd Merge *(last name, truncated first name, birth date)*	70,831	89.6%
	4th Merge *(truncated last name, truncated first name, birth date)*	71,511	90.4%
2002	Beginning File	76,925	100%
	1st Merge *(last name, first name, middle initial)*	60,317	78.41%
	2nd Merge *(last name, first name)*	69,325	90.12%

Note: For 2002, there was a discrepancy in the birth date information between the 2001 PSSA data file for 11th graders and the 2002 SAT data file. As a result, birth date information could not be used in conducting the merge for that year.

An additional analysis was conducted to verify that students retained in the final data set did not differ significantly on SAT scores from those whose data failed to merge. Table 6 presents the means, standard deviations, and numbers of cases, for matched and all Pennsylvania public school students for 2002 and 2003. Mean scores in 2003 for students whose data were merged were approximately 3.0 scale score points higher than mean scores for all students and about 30 points higher than for unmerged students. The mean for merged students was about 2.6 scale score points higher than the mean for all students in 2002 for both tests. The difference in means between merged students and unmerged students also was about 30 points in 2002. Data includes students attending public schools in Pennsylvania only. While this difference is a concern, it is consistent between the two years. Correlations also were conducted on each cycle during the merge to check that correlations were approximately the same from cycle to cycle. Correlations remained approximately the same for each merge cycle. Results of these checks are not contained in this report.

Table 6. SAT Descriptive Statistics for Graduating Classes of 2002 and 2003 with Matched, Unmatched, and All Pennsylvania Public School Students

		Graduating Class of 2002			Graduating Class of 2003		
		Matched	Unmatched	Total	Matched	Unmatched	Total
Verbal	Mean	497.31	470.68	494.68	500.33	468.92	497.32
	S.D.	102.66	113.48	104.08	101.82	111.91	103.25
	N	69,325	7,600	76,925	71,511	7,584	79,095
Math	Mean	502.67	476.19	500.06	506.04	473.92	502.96
	S.D.	109.43	119.30	110.73	109.08	116.33	110.20
	N	69,325	7,600	76,925	71,511	7,584	79,095

Note: Student data are for students attending public schools in Pennsylvania only. No private school student data is included.

An additional analysis was conducted to compare PSSA scores between students whose data matched, whose data failed to match, and with the total file. Table 7 presents the means, standard deviations, and numbers of cases for Pennsylvania public school students for 2002 and 2003. As expected, the mean scores for merged students for both tests in both years are higher than for unmerged student scores and the state total. Students taking the SAT are generally college-bound students (the merged file) and typically score higher than non-college-bound students (most of the students in the unmerged file) on academically-oriented assessments. Since the total file is the combination of the two files—merged and unmerged—the total file mean should fall between the means for the two files.

Table 7. PSSA Descriptive Statistics for Graduating Classes of 2002 and 2003 with Matched Students, Unmatched Students, and All Pennsylvania Students

		Graduating Class of 2002 (11^{th} Grade PSSA for 2001)			Graduating Class of 2003 (11^{th} Grade PSSA for 2002)		
		Matched	Unmatched	Total	Matched	Unmatched	Total
Reading	Mean	1382.50	1174.32	1,296.89	1406.30	1186.18	1,312.87
	S.D.	179.02	209.03	217.56	168.56	206.45	215.13
	N	68,039	47,523	115,562	70,926	52,306	123,232
Math	Mean	1392.83	1177.96	1,304.13	1414.79	1179.02	1,314.48
	S.D.	200.78	182.75	220.57	206.26	181.92	228.28
	N	68,189	47,943	116,132	70,981	52,566	123,547

Relationships Among Measures of Student-Level Performance

This section presents relationships among PSSA and other measures of student performance—SAT scores, grade-point average (GPA), and course grades—by examining correlations among the measures and graphical portrayals of the relationships. GPAs and course

grades were obtained from the optional Student Descriptive Questionnaire (SDQ) which students are asked to complete as part of the SAT assessment process. HumRRO had specifically requested certain demographic information obtained by the SDQ when we requested the SAT data files. The use of self-reported data undoubtedly involves some inaccuracies. However, Cassady (2001) found students' self-reported grades to be a reliable measure of actual grades.

Correlations[1]

Pearson correlations were calculated for the PSSA-SAT merged files for both graduating classes. The tables differentiate between the correlations among the content areas within each of the different assessments from the correlations between the different kinds of assessments (Campbell & Fiske, 1959), thus allowing for the examination of the following relationships:

- The same content area within different achievement measures, or convergent validity coefficients (Campbell & Fiske, 1959). (These correlations are in bold and underlined)
- Different content areas within the same achievement measures. (These correlations are in italics.)
- Different content areas within different achievement measures. (These correlations are in bold, but not underlined.)

In correlation tables of this type, the expectation is for the highest correlations to be between different measures of the same content. Then, because of similarities in test-taking strategies or other method effects, the next highest correlations are typically those between different content, but measured by the same method of assessment. Correlations between different content areas within different measures should be the lowest in the table.

As the tables show, correlations are positive and above 0.30. This means that students who do well on any one measure of any content also tend to do well on all measures and in all content areas. In a recent study of school-level assessment scores, Sicoly (2002) discussed the existence of a general cognitive factor that "cuts across content areas." If such a "g-factor" exists, then it would be expected that students with high ability would score well on any test, regardless of the content. Correlations presented in these tables suggest that Pennsylvania students who exhibit high ability in one content area can be expected to perform well in other content areas.

The results are shown in Table 8 for the Class of 2002, Table 9 for the Class of 2003, and Table 10 for the two-year averaged[2] correlation for PSSA, SAT, GPA, and course grades. The correlations for both 2002 and 2003 and the two-year average show a strong correlation between the math and reading/verbal component on each assessment ($r = 0.686$ for PSSA and $r = 0.737$ for SAT for the two-year average) and between the math components on the two assessments and the reading/verbal assessments on the two assessments ($r = 0.856$ for math and $r = 0.742$ for reading/verbal for the two years averaged).

Correlations for student-reported scores tended to be lower than correlations between the two assessments. Correlations between PSSA and SAT to student-reported GPAs were lower, but were still approximately 0.5 ($r = 0.501$ for PSSA reading, $r = 0.539$ for PSSA math, $r = 0.491$ for SAT verbal, and $r = 0.525$ for SAT math). Correlations between GPA and both English

[1] Given the extremely large sample sizes, tests of statistical significance are irrelevant. All reported relationships are statistically significant; that is, unlikely to be due to chance. Therefore, the report focuses on interpretation of the results.

[2] The technical r to z correction was used to obtain the averaged correlations.

and math course grades were high (r = 0.708 for English and r = 0.669 for math). Correlations for student-reported course grades and the matching component of the two assessments were also fairly high. English was approximately 0.45 and math was approximately 0.530 (r = 0.443 for English course grades to PSSA reading, r = 0.446 for English course grades to SAT verbal, r = 0.534 for math course grades to PSSA math, and r = 0.529 for math course grades to SAT math for the two-year average). The lower correlations between course grades and the two assessments can be partially attributed to differences in courses taken and grading practices from teacher to teacher and school to school.

Table 8. PSSA, SAT, and Course Grade Correlations for Students in the Class of 2002

		PSSA		**SAT**		**Course Grades**		
Variable		Reading	Math	Verbal	Math	GPA (Total)	English	Math
PSSA	Reading	1.000						
	Math	*0.678*	1.000					
SAT	Verbal	**0.704**	**0.685**	1.000				
	Math	**0.593**	**0.846**	*0.735*	1.000			
Course Grades	GPA	0.501	0.539	0.491	0.525	1.000		
	English	**0.443**	**0.399**	**0.447**	**0.390**	0.708	1.000	
	Math	**0.361**	**0.528**	**0.354**	**0.527**	0.669	*0.449*	1.000

Table 9. PSSA, SAT, and Course Grades Correlations for Students in the Class of 2003

		PSSA		**SAT**		**Course Grades**		
Variable		Reading	Math	Verbal	Math	GPA (Total)	English	Math
PSSA	Reading	1.000						
	Math	*0.694*	1.000					
SAT	Verbal	**0.775**	**0.694**	1.000				
	Math	**0.663**	**0.865**	*0.739*	1.000			
Course Grades	GPA	0.520	0.553	0.496	0.535	1.000		
	English	**0.455**	**0.406**	**0.444**	**0.396**	0.704	1.000	
	Math	**0.380**	**0.540**	**0.363**	**0.531**	0.672	**0.454**	1.000

Table 10. PSSA, SAT, and Course Grades Correlations Averaged for the Classes of 2002 and 2003

Variable		PSSA		SAT		Course Grades		
		Reading	**Math**	**Verbal**	**Math**	**GPA (Total)**	**English**	**Math**
PSSA	Reading	1.000						
	Math	*0.686*	1.000					
SAT	Verbal	**0.742**	**0.690**	1.000				
	Math	**0.629**	**0.856**	*0.737*	1.000			
Course Grades	GPA	0.511	0.546	0.494	0.530	1.000		
	English	**0.449**	**0.403**	**0.446**	**0.393**	0.706	1.000	
	Math	**0.371**	**0.534**	**0.359**	**0.529**	0.671	*0.452*	1.000

Table 11 shows that student performance on both the SAT and PSSA increased from the Class of 2002 to the Class of 2003. For the SAT, the increase was 3.02 scale score points for the verbal portion of the assessment and 3.37 scale score points for the math portion of the assessment. For the PSSA, the increase was 23.80 scale score points for the verbal portion of the assessment and 21.96 scale score points for the math portion of the assessment.

Table 11. Changes in Mean Scale Scores Across Graduation Years

		Mean Scores Class of 2002	Mean Scores Class of 2003	Mean Score Difference
SAT	Verbal	497.31	500.33	+3.02
	Math	502.67	506.04	+3.37
PSSA	Reading	1,382.50	1,406.30	+23.80
	Math	1,392.83	1,414.79	+21.96

Illustration of Relationships

Figure 1 through Figure 16 present box plots that illustrate relationships between PSSA scores and other measures of student achievement. These graphs further illustrate the strong correlations between the two tests in both subject areas and other measures of student performance. Data are presented separately for the Classes of 2002 and 2003. The boxes and whiskers in each plot represent the distribution of scores versus the varying levels of the corresponding grouping of scores or performance measures (GPA or student course grades). Each box represents 50% of the students within each of the categories along the X-axis. The median is represented by the line in the box. The whiskers represent the spread of the distribution of students calculated at 1.5 times the length of the box. This spread should include approximately 99.3% of all students for the given category. Sample sizes within each category along the X-axis are noted.

SAT Verbal Score Versus PSSA Performance Level in Reading

Figure 1 and Figure 2 show the relationship between the SAT Verbal score on the y-axis and the 11th Grade PSSA Reading Performance Level on the x-axis for the Classes of 2002 and 2003. The graph shows a distinct stair-step, especially for the Basic to Advanced levels, for both classes. The number of students in each level also indicates that more of the higher-achieving students on the PSSA are taking the SAT than are lower-achieving students.

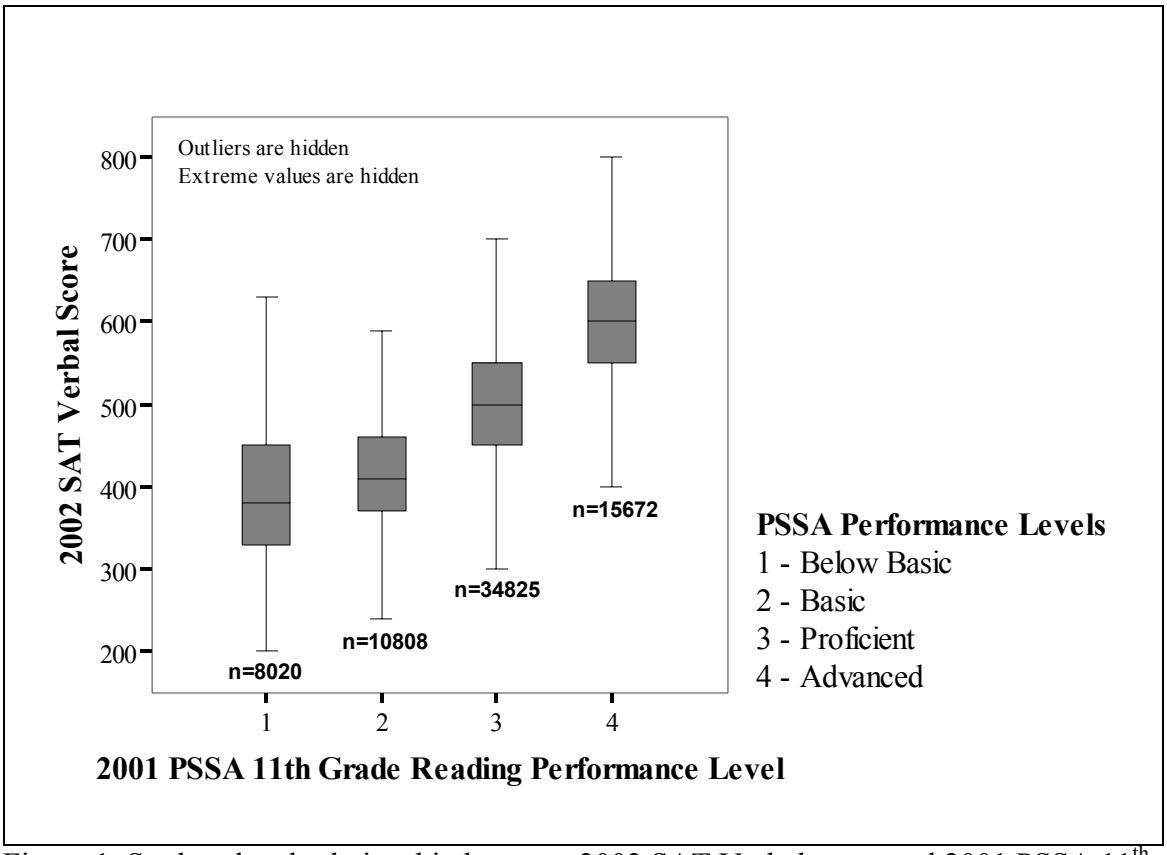

Figure 1. Student-level relationship between 2002 SAT Verbal score and 2001 PSSA 11th grade reading performance level for the Class of 2002.

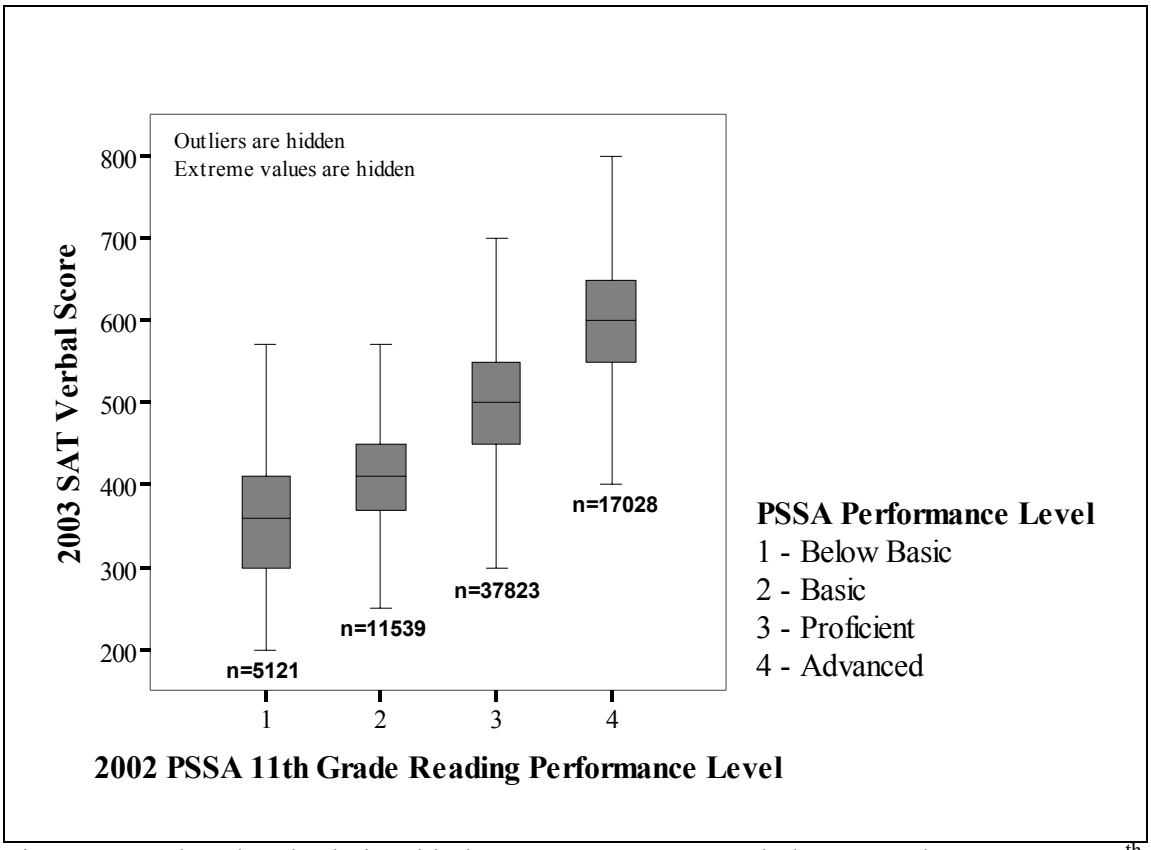

Figure 2. Student-level relationship between 2003 SAT verbal score and 2002 PSSA 11th grade reading performance level for the Class of 2003.

PSSA Reading Scale Score Versus SAT Verbal Score by Quintile

Figure 3 and Figure 4 show the relationship between the PSSA reading scale score on the y-axis and the SAT verbal score by quintile on the x-axis for the Classes of 2002 and 2003. The graph shows a distinct stair-step for each quintile in both grades. Two sets of cut scores are also illustrated on the graphs. The graph's legend lists the cut scores for each quintile. The cut scores were chosen to place approximately 20% of the students taking the SAT in each quintile. Because of the number of students with a score equal to the cut score, the number of students in each quintile is not the same. The cut points for each quintile therefore vary slightly between the two classes. The second set of cut scores is represented by the lines that have been added to show the 11th grade PSSA reading performance levels. This illustrates that most students in the top four quintiles, or top 80% of those students taking the SAT, are performing at the Proficient or Advanced level on the 11th grade PSSA reading assessment.

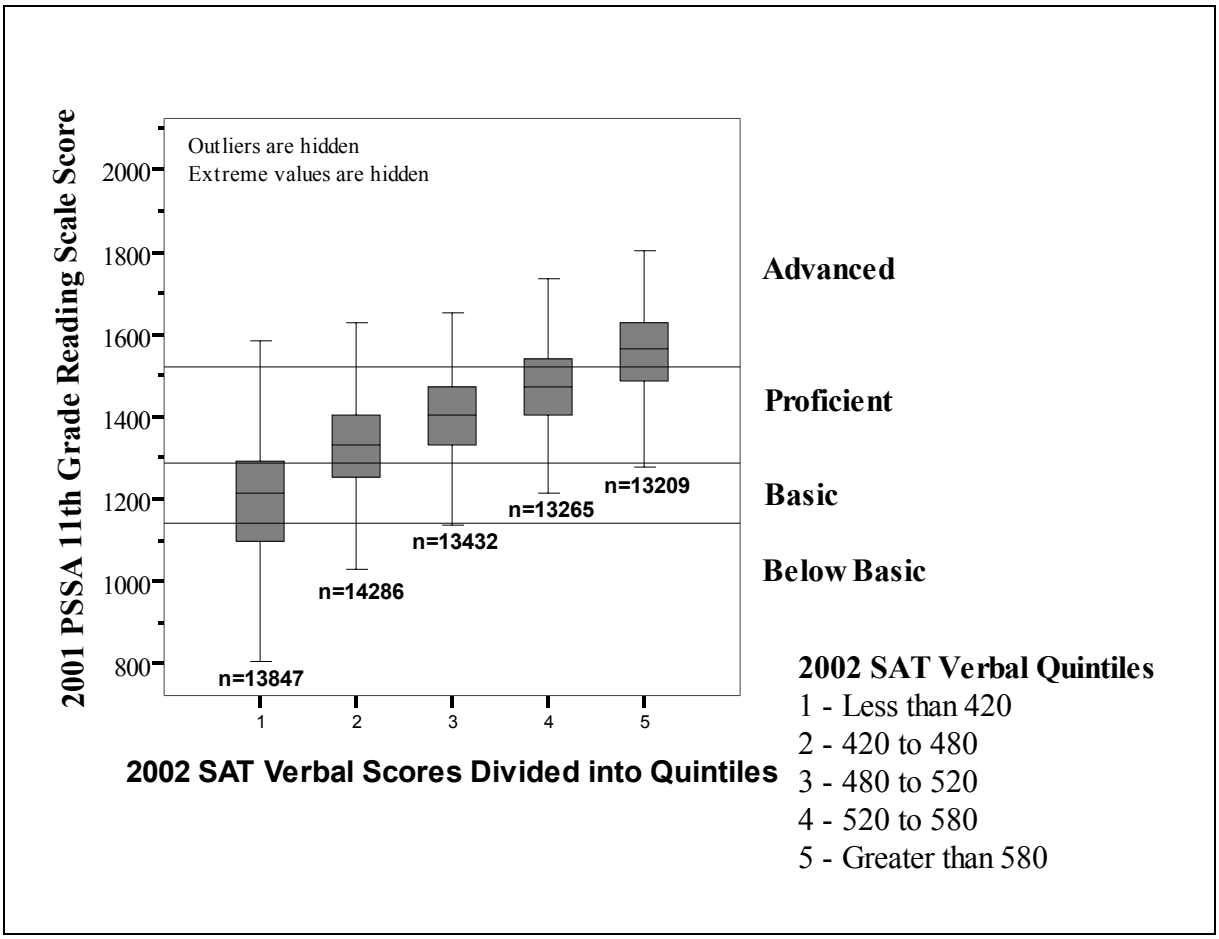

Figure 3. Student-level relationship between 2001 PSSA reading scale scores and 2002 SAT verbal scores divided into quintiles for the Class of 2002.

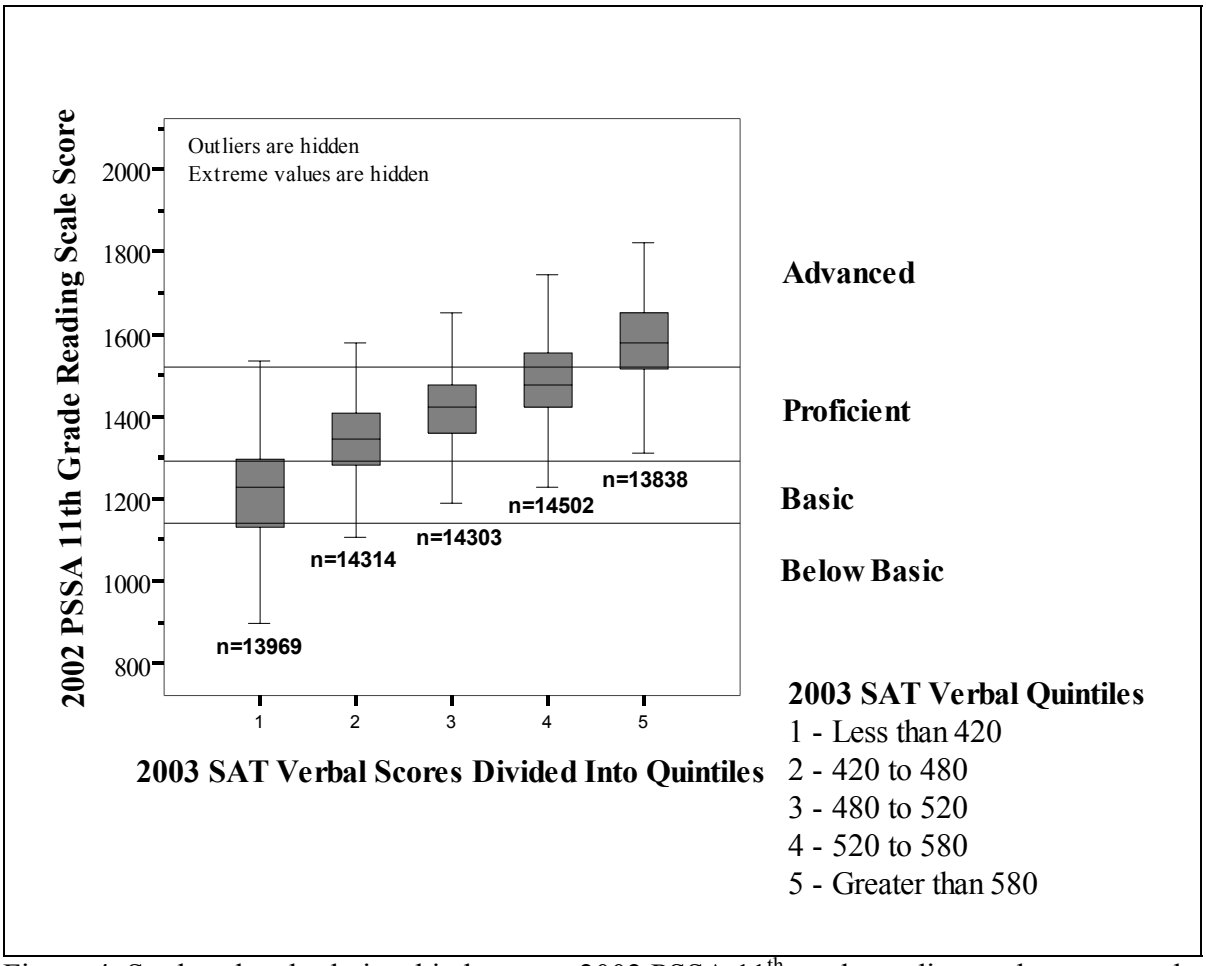

Figure 4. Student-level relationship between 2002 PSSA 11th grade reading scale scores and 2003 SAT verbal score divided into quintiles for the Class of 2003.

SAT Math Score Versus PSSA Performance Level in Math

Figure 5 and Figure 6 show the relationship between the SAT math score on the y-axis and the 11th grade PSSA math performance level on the x-axis for the Classes of 2002 and 2003. The graph shows a distinct stair-step for all levels in both grades. The number of students in each level also indicates that more of the higher-achieving students are taking the SAT than are lower-achieving students.

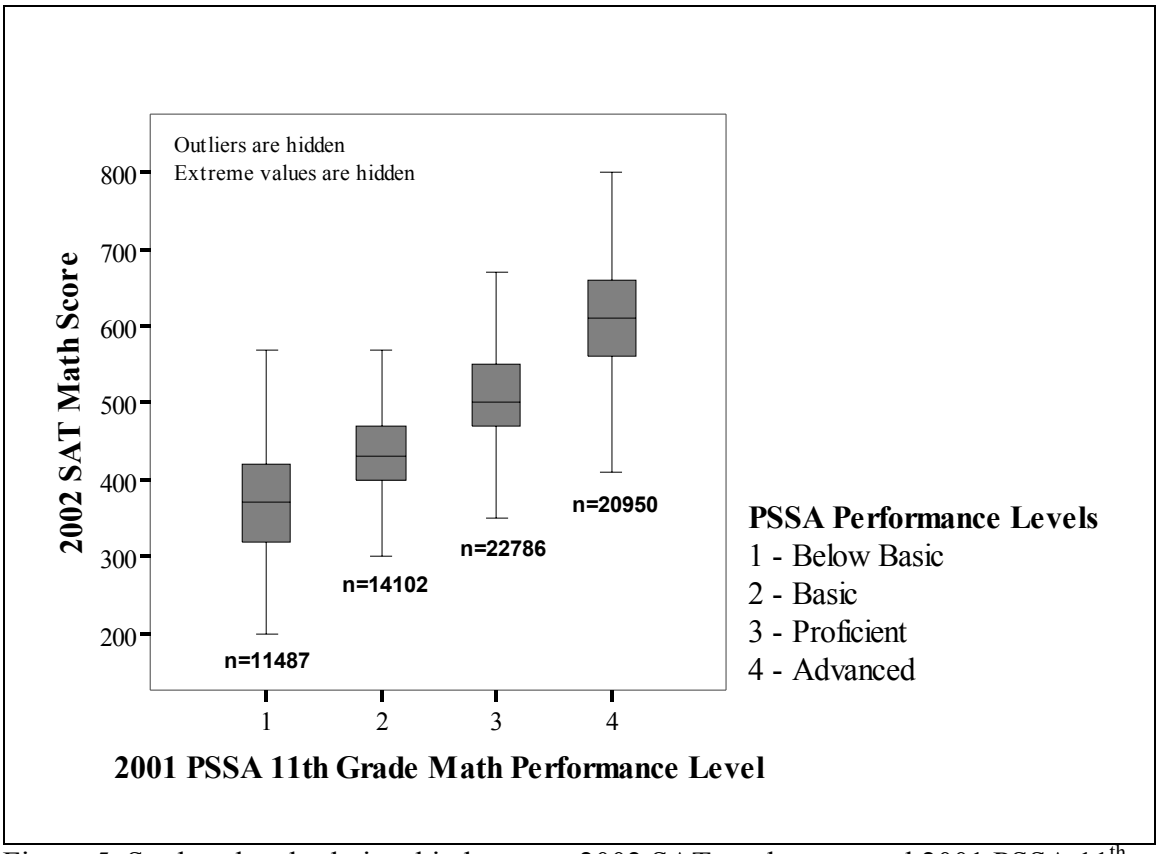

Figure 5. Student-level relationship between 2002 SAT math score and 2001 PSSA 11th grade math performance level for the Class of 2002.

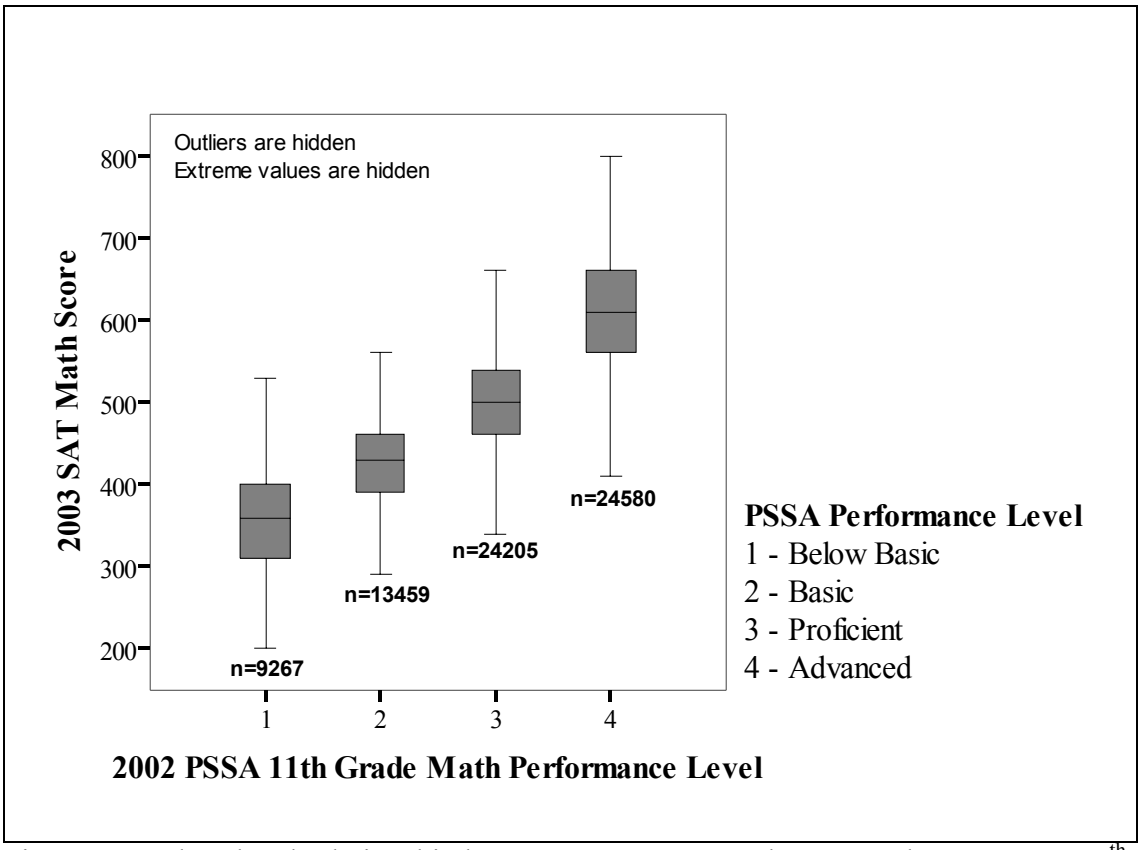

Figure 6. Student-level relationship between 2003 SAT math score and 2002 PSSA 11th grade math performance level for the Class of 2003.

PSSA Math Scale Score versus SAT Math Score by Quintile

Figure 7 and Figure 8 show the relationship between the 11th grade PSSA math scale score on the y-axis and the SAT math score by quintile on the x-axis for the Classes of 2002 and 2003. The graph shows a distinct stair-step for each quintile in both grades. Two sets of cut scores are also illustrated on the graph. The graph's legend lists the cut scores for each quintile. The cut scores were chosen to place approximately 20% of the students taking the SAT in each quintile. Because of the number of students with a score equal to the cut score, the number of students in each quintile is not the same. The cut points for each quintile therefore vary slightly between the two classes. The second set of cut scores is represented by the lines that have been added to show the for 11th grade PSSA math performance levels. This figure clearly illustrates that more than 75% of students in the each of the top three quintiles, or top 60% of those students taking the SAT, are performing at the Proficient or Advanced level on the 11th Grade PSSA math assessment. Additionally, more than 75% of students taking the SAT in the top quintile and more than 50% of students in the second quintile are scoring in the Advanced performance level on the 11th grade PSSA math assessment.

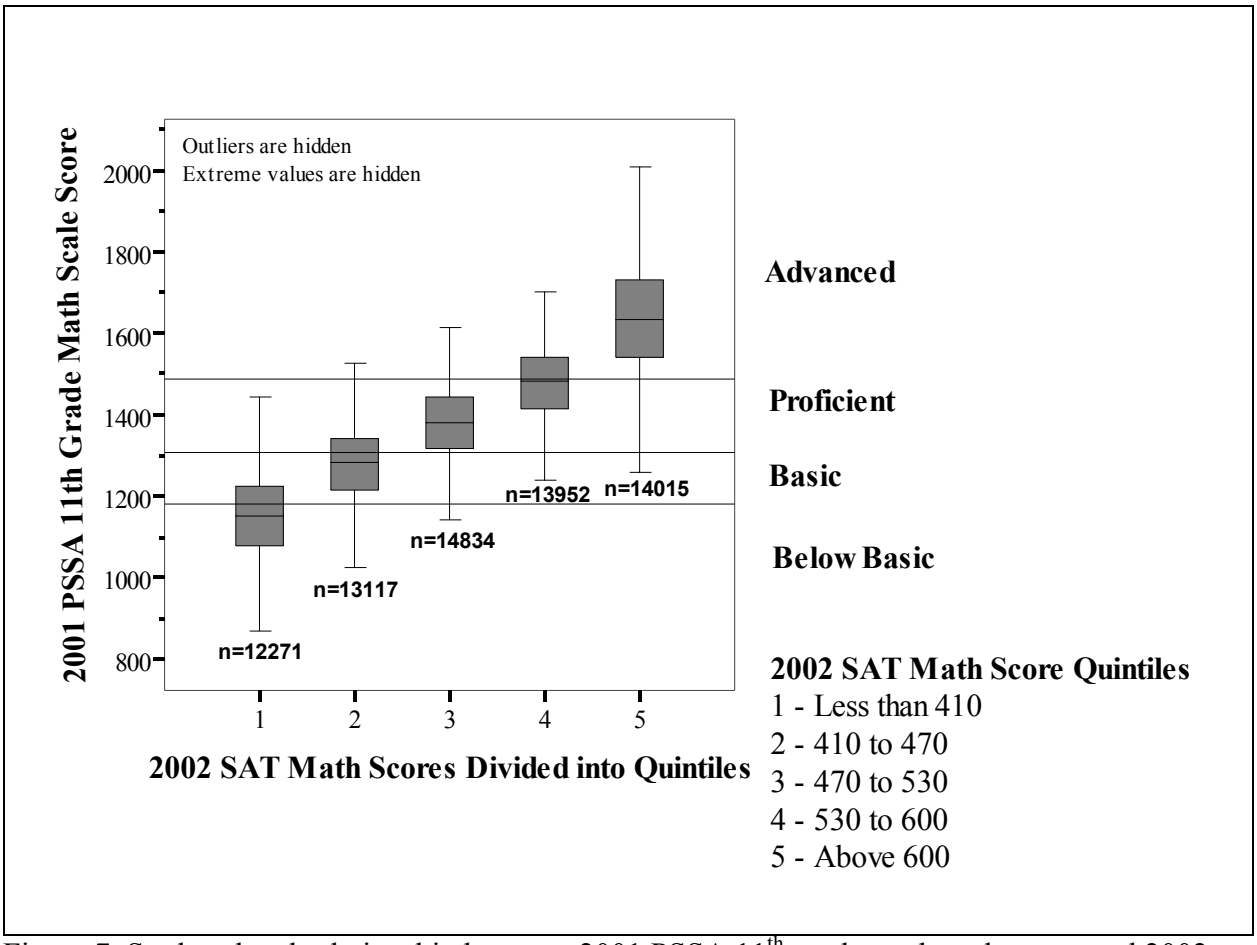

Figure 7. Student-level relationship between 2001 PSSA 11th grade math scale score and 2002 SAT math score divided into quintiles for the Class of 2002.

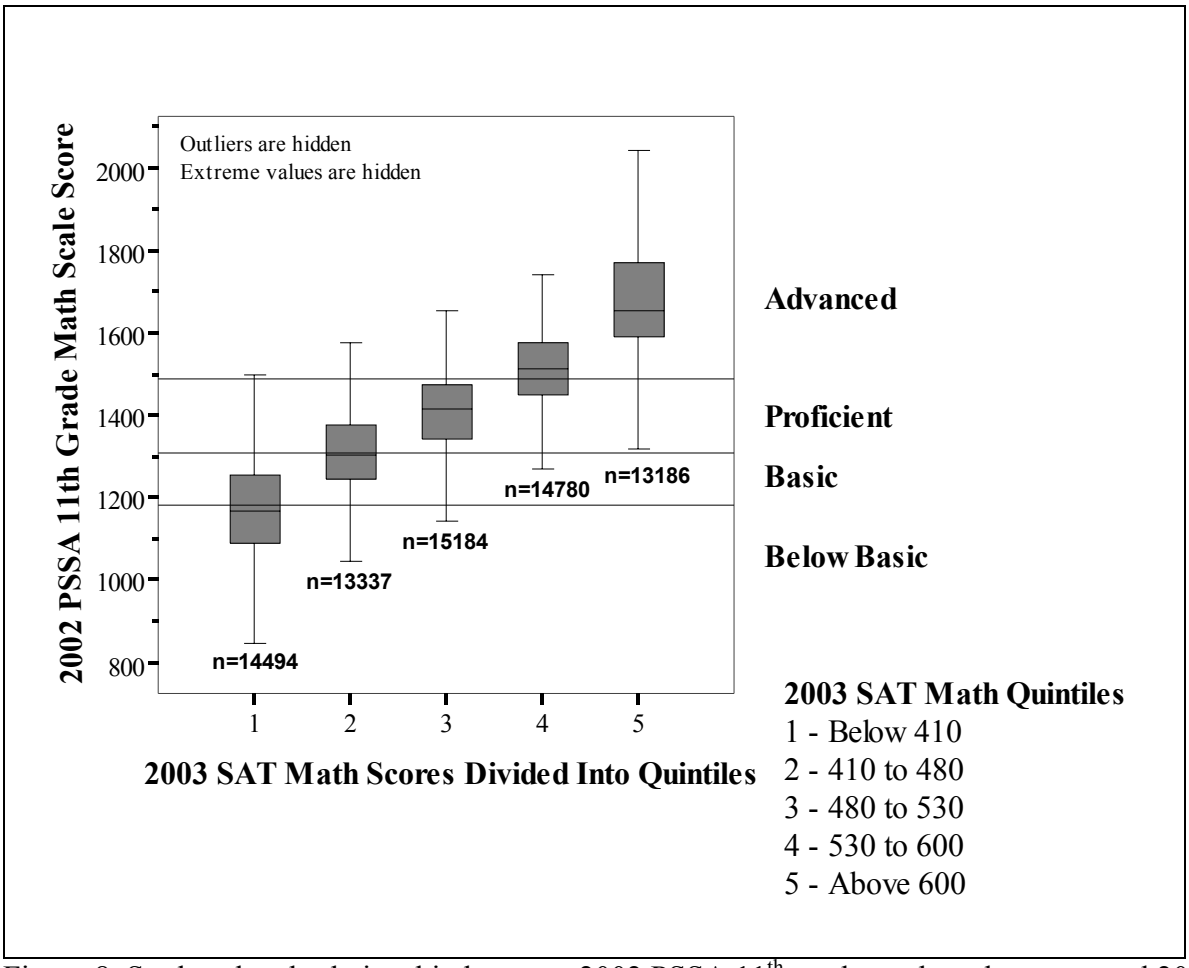

Figure 8. Student-level relationship between 2002 PSSA 11th grade math scale scores and 2003 SAT math scores divided into quintiles for the Class of 2003.

PSSA Reading Scale Score Versus SAT Reported Average Grade in English

Figure 9 and Figure 10 illustrate the relationship between the PSSA Reading scale scores and the grades students reported in the optional Student Descriptive Questionnaire (SDQ) portion of the SAT. The SDQ asked students to report their average grade in English. The available responses to this question were: 0 – Failing, 1 – Passing, 2 – Fair, 3 – Good, or 4 – Excellent. The category "No response" indicates that the student did not respond to the question. In 2002, 8,305 students did not respond to the English question. In 2003, 6,170 students did not respond. In 2002, 13 students responded that their average English grade was "Failing." In 2003, 6 students had this answer. Students who did not respond and those who responded that their average grade was "Failing" were not included in the graphs. As can be seen in Figure 10, scale scores in PSSA Reading tend to be higher for students reporting better performance in English coursework. The preponderance of students reporting either a "Good" or an "Excellent" English average scored in the Proficient or Advanced PSSA performance levels. Additionally, almost 90% of the students reported having either a "Good" or "Excellent" average in English courses. As previously noted, self-reported data undoubtedly involves some inaccuracies; however, Cassady (2001) found students' self-reported grades to be a reliable measure of actual grades.

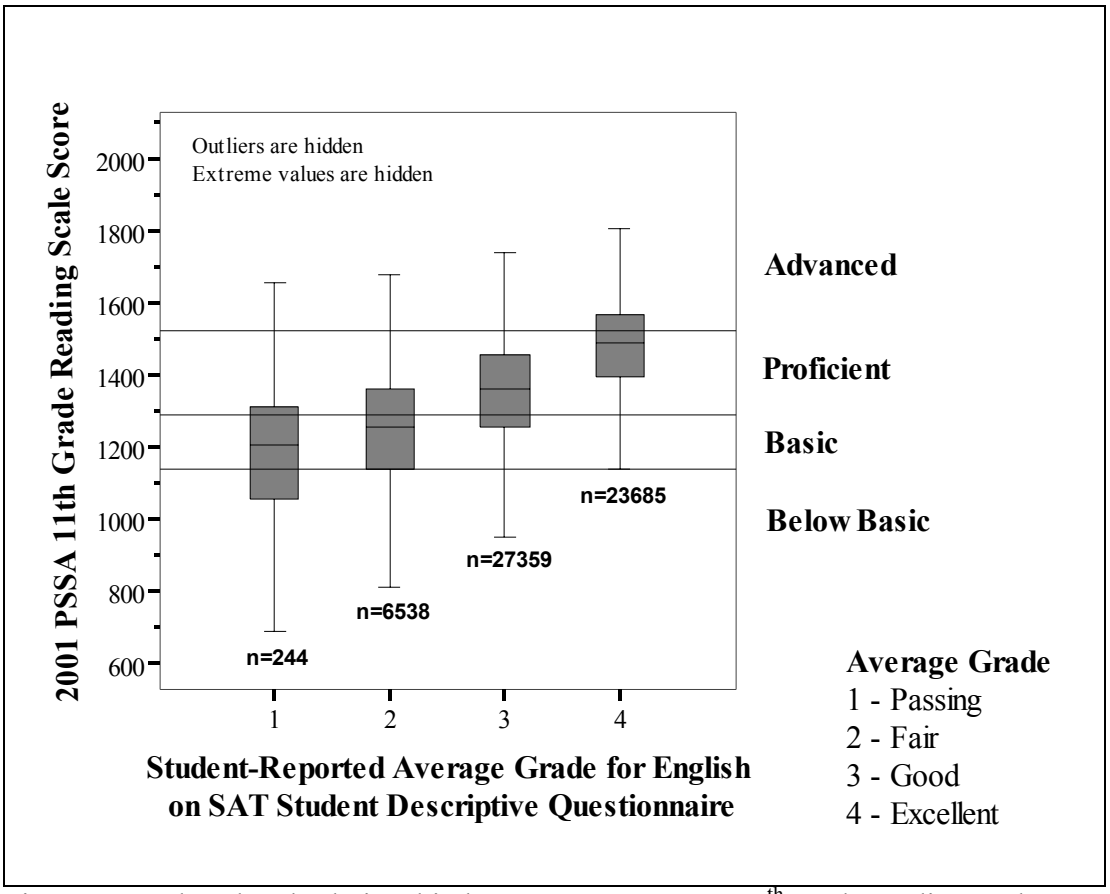

Figure 9. Student-level relationship between 2001 PSSA 11th grade reading scale score and student-reported average English course grades on the Student Descriptive Questionnaire (SDQ) of the SAT for the Class of 2002.

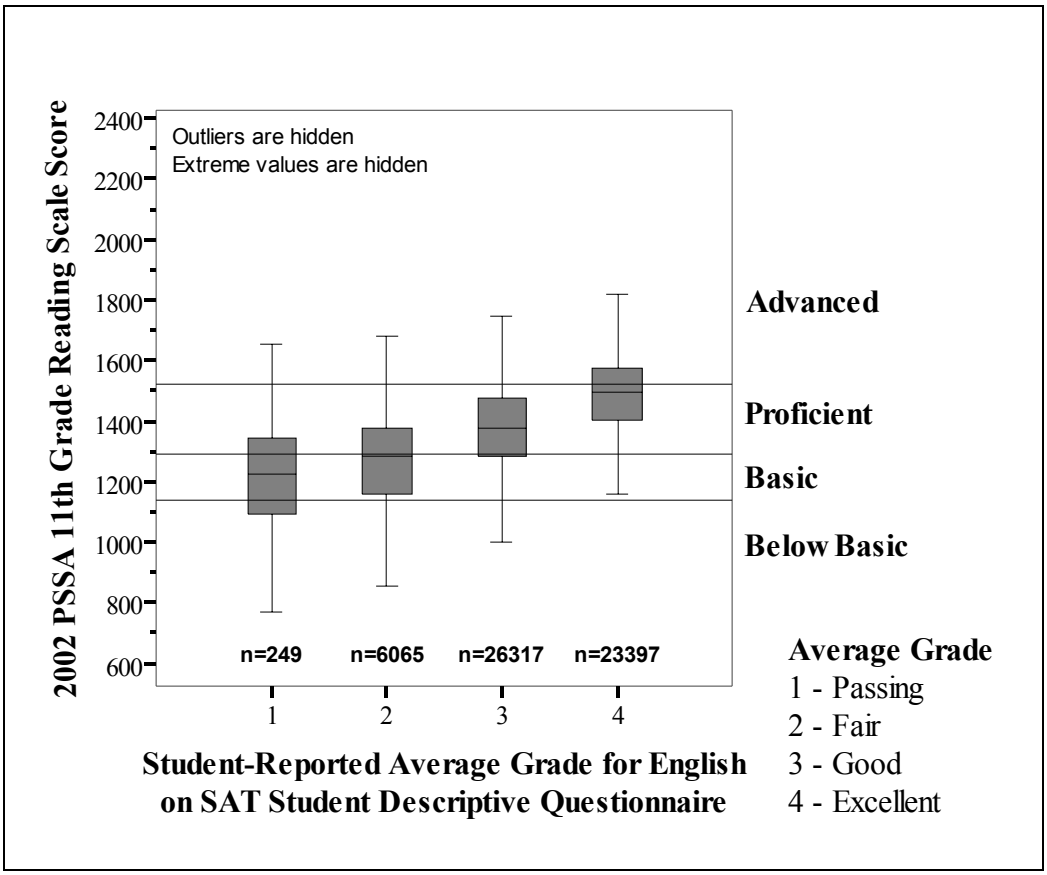

Figure 10. Student-level relationship between 2002 PSSA 11th grade reading scale score and student-reported English course grades on the SAT Student Descriptive Questionnaire (SDQ) for the Class of 2003.

PSSA Math Scale Score Versus SAT Reported Average Grade in Math

Figure 11 and Figure 12 illustrate the relationship between the PSSA 11th grade math scale scores and the grades students reported in the optional Student Descriptive Questionnaire (SDQ) portion of the SAT which students are asked to complete as part of registering for the SAT. The SDQ asked students to report their average grade in math. The available responses to this question were: 0 – Failing, 1 – Passing, 2 – Fair, 3 – Good, or 4 – Excellent. The category "No response" indicates that the student did not respond to the question. In 2002, 8,523 students did not respond to the math average grade question. In 2003, 6,316 students did not respond. In 2002, 41 students responded that their average math grade was "Failing." In 2003, 47 students had this answer. Students who did not respond and those who responded that their average grade was "Failing" were not included in the graphs.

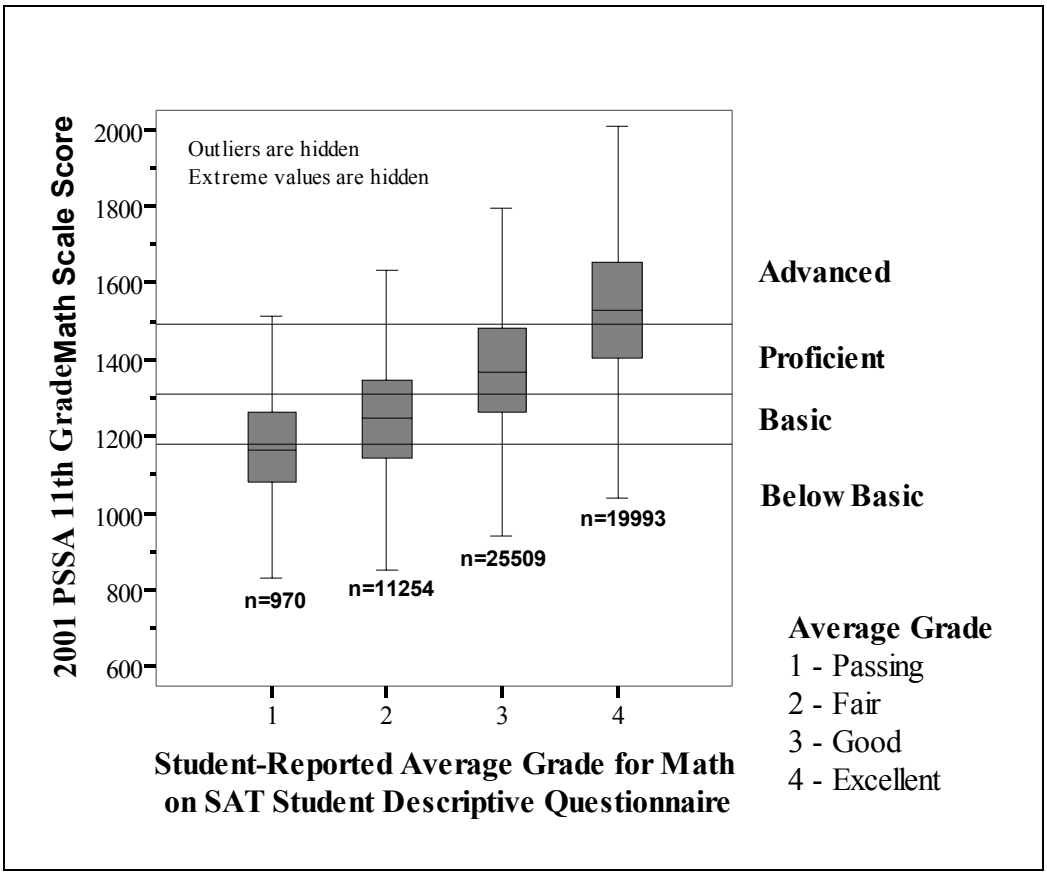

Figure 11. Student-level relationship between 2001 PSSA 11th grade math scale score and student-reported average math course grades in the Student Descriptive Questionnaire (SDQ) of the SAT for the Class of 2002.

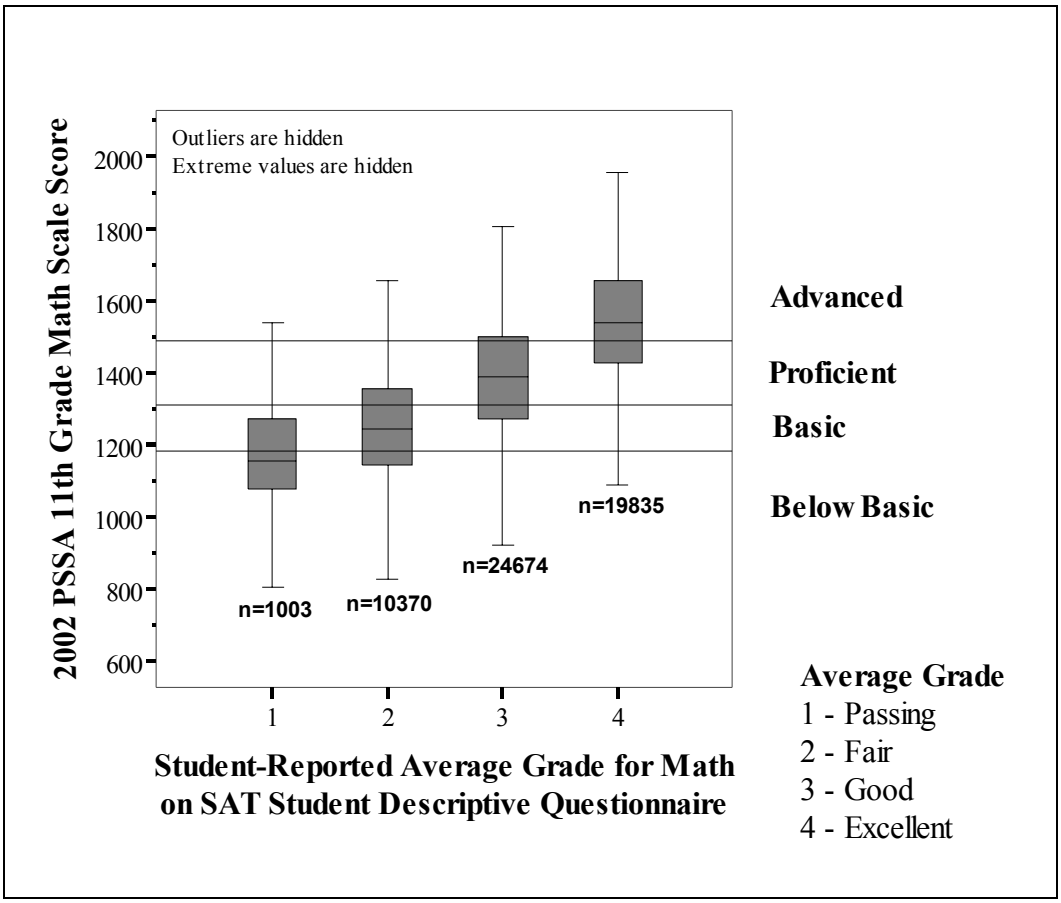

Figure 12. Student-level relationship between 2002 PSSA 11th grade math scale score and student-reported average math course grade in the SAT Student Descriptive Questionnaire (SDQ) for the Class of 2003.

PSSA Scale Scores Versus SAT Reported Grade-Point Average

Figure 13 through Figure 16 illustrate relationships between PSSA scale scores and overall grade-point averages (GPAs). The GPAs are from the SAT's SDQ. Students were asked to provide their GPA on a basic 4-point scale with pluses and minuses around each point (A+ = 4.3, A = 4.0, A- = 3.7, etc.). Approximately 87% (60,283 of 69,325) of students in the Class of 2002 completed the GPA question in the SDQ and 83% (59,258 of 71,511) of students in the Class of 2003 completed the question. The mean student-reported GPA was 3.33 (SD = 0.64) for the Class of 2002 and 3.35 (SD = 0.63) for the Class of 2003. The correlation for the 2001 PSSA 11th grade reading component with the 2002 GPA was 0.501. For the 2001 PSSA 11th grade math component, the correlation was 0.539. For the 2002 PSSA, the correlations were 0.520 in Reading and 0.553 in Math. The box-plots for PSSA scale scores to GPA for both years and both tests performed as expected—lower scores for lower GPAs and higher scores for higher GPAs. Note that the median and boxes align fairly well with the appropriate performance levels—A students in the Advanced and Proficient level, B students generally in the Proficient level, C students in the low Proficient and Basic levels, and D students in the Basic level. Note also that the number of students decreases sharply when the student-reported GPA drops below 3.0.

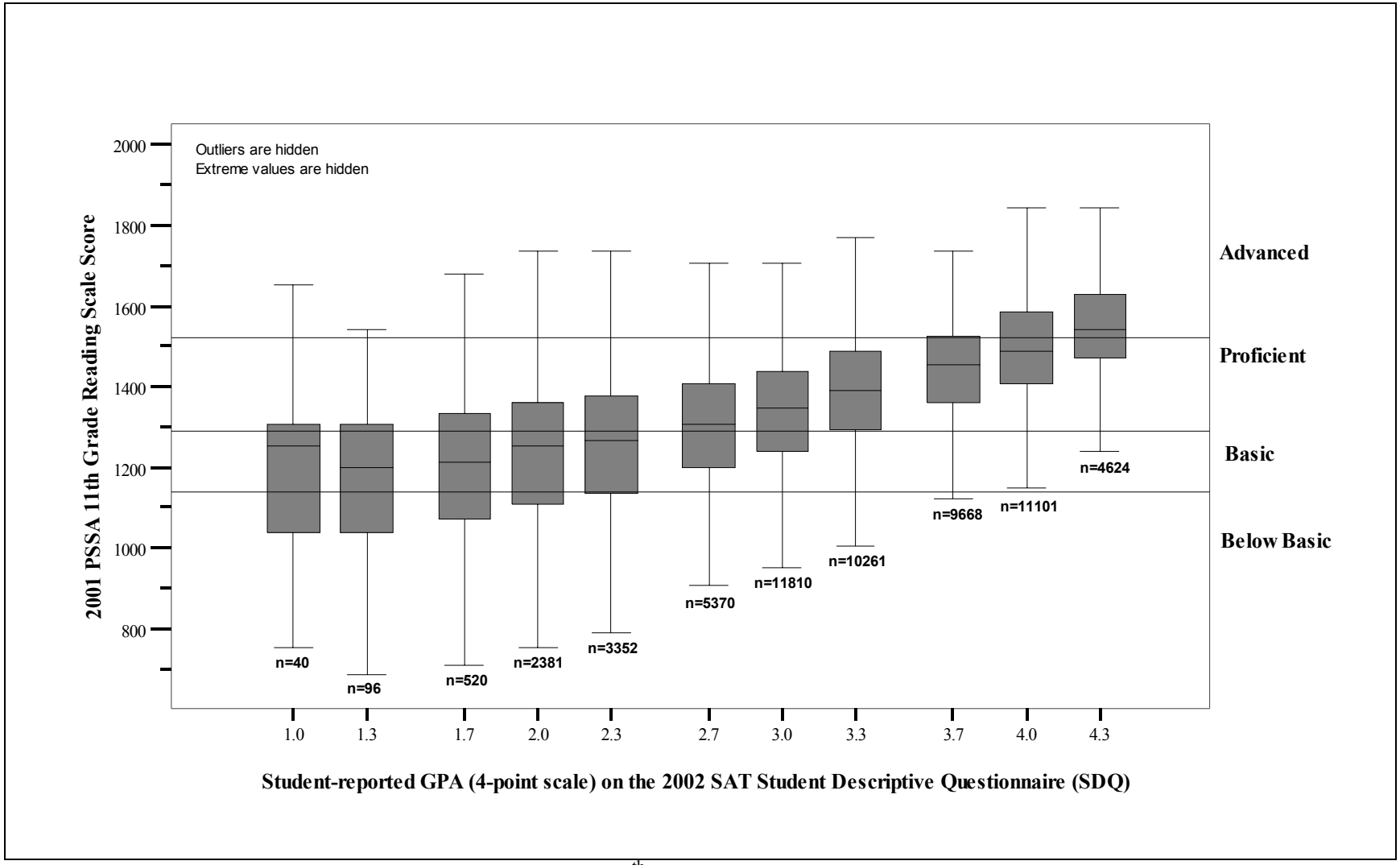

Figure 13. Student-level relationship between 2001 PSSA 11th grade reading scale score and student-reported grade-point average (GPA) from the SAT's Student Descriptive Questionnaire (SDQ) for the Class of 2002.

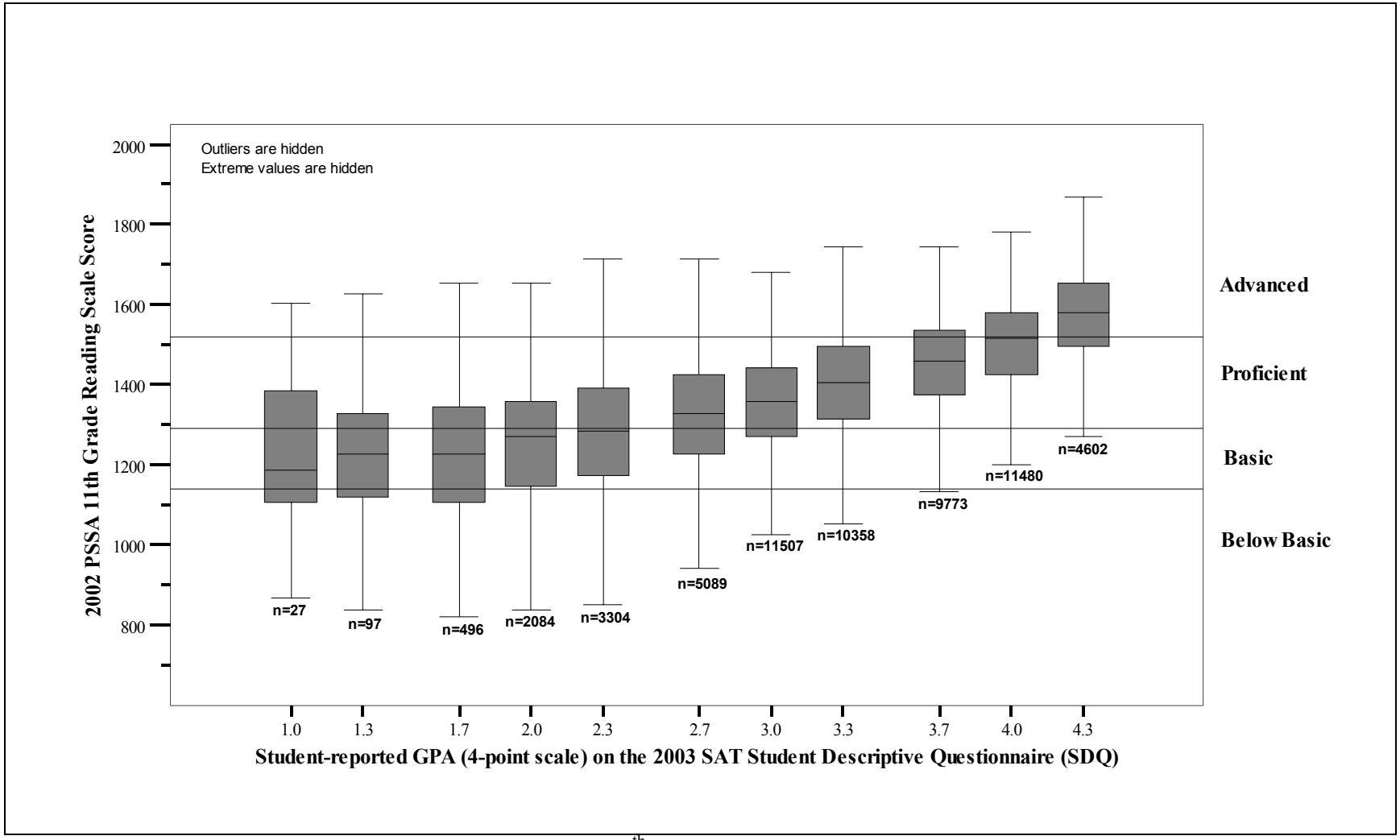

Figure 14. Student-level relationship between 2002 PSSA 11th Grade reading scale score and student-reported grade-point average (GPA) from the SAT's Student Descriptive Questionnaire (SDQ) for the Class of 2003.

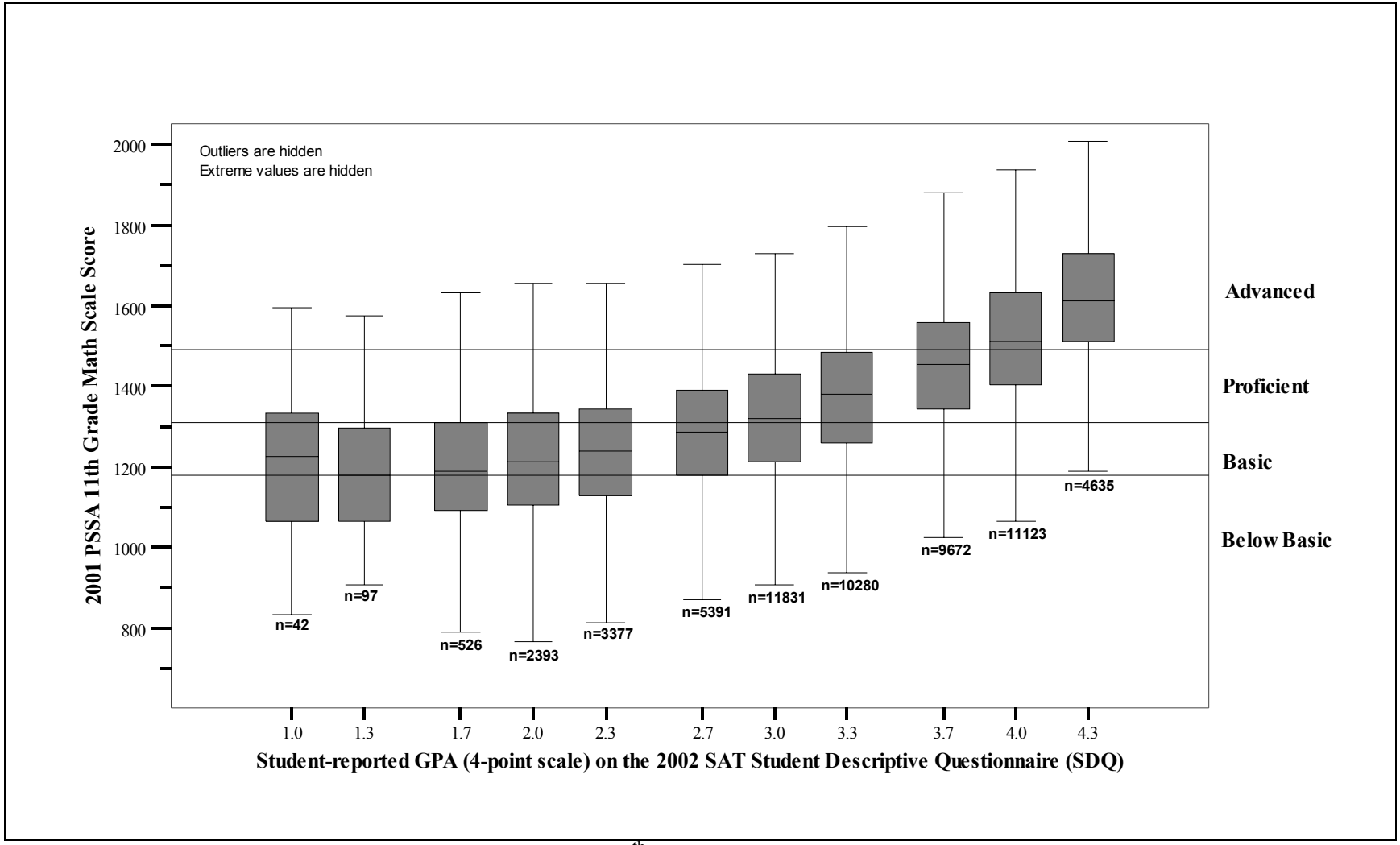

Figure 15. Student-level relationship between 2001 PSSA 11th grade math scale score and student-reported grade-point average (GPA) from the SAT's Student Descriptive Questionnaire (SDQ) for the Class of 2002.

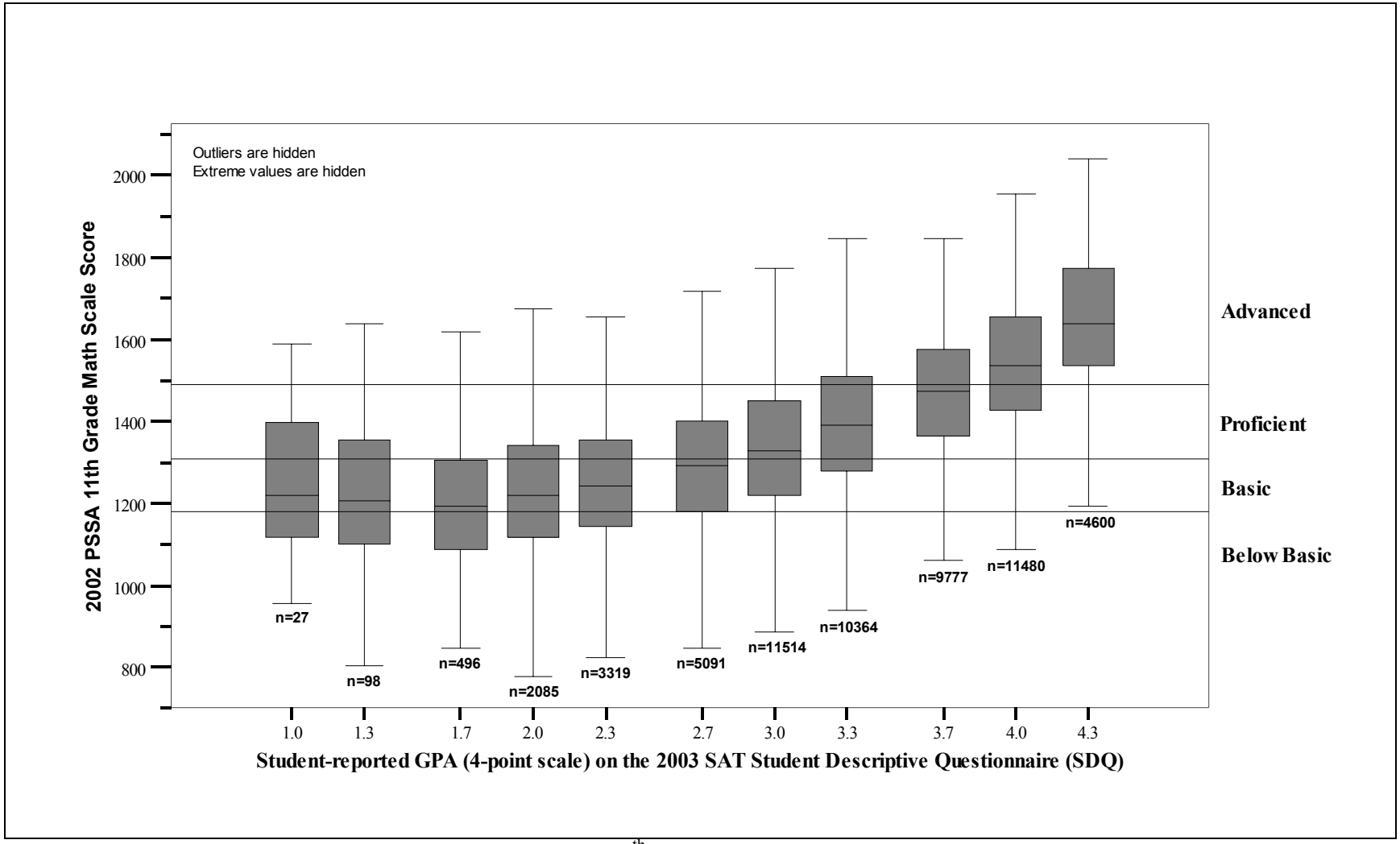

Figure 16. Student-level relationship between 2001 PSSA 11th grade math scale score and student-reported grade-point average (GPA) from the SAT's Student Descriptive Questionnaire (SDQ) for the class of 2002.

Differences in Performance for Demographic Groupings

Analyses were conducted to compare performance on the PSSA and SAT among students from varying backgrounds. Average test performance tends to differ for students of differing genders, ethnic groups, and various measures associated with socioeconomic status. The important question for judging bias on the PSSA is not whether specific groups show mean differences, but whether those differences are larger for the PSSA than the differences observed on other measures of student achievement.

Differences between measures of achievement can be captured in several ways. This report will present two measures of mean differences. First, effect size statistics will be calculated for the differences between major categories of students (Coe, 2002). The magnitude of the effect size statistics is less important than the difference between the effect sizes for PSSA compared to the other measures. However, measuring the magnitude of the gaps for NCLB-defined groups is instructive and important, since all states are expected to make progress toward closing those gaps (No Child Left Behind, 2004).

Effect size statistics can be interpreted as the average percentile standing of the average treated (or experimental) participant relative to the average untreated (or control) participant. An effect size of 0.0 indicates that the mean of the treated group is at the 50th percentile of the control group. An effect size of 0.8 indicates that the mean of the treated group is at the 79th percentile of the control group. An effect size of 1.7 indicates that the mean of the treated group is at the 95.5th percentile of the control group. Cohen (1988) hesitantly defined effect sizes as "small, d=.2," "medium, d=.5," and "large, d=.8." He went on to caution against offering conventional operational definitions such as these in the diverse field of behavioral science.

The second measure used to capture differences between measures of achievement was regression analysis. Regression analysis was used to examine the extent to which PSSA differences in performance for various demographic groupings are similar to those found in SAT scores. For each of the demographic groupings, we calculated a regression equation to predict PSSA scores based on the matching SAT content score. A second equation was created which added the specific demographic characteristic of interest. If PSSA scores exhibit greater differences based on that demographic characteristic, that characteristic will have a significant weight and there will be a meaningful increase in the prediction of the PSSA component score.

Gender Differences

Table 12 presents performance means for males and females for the PSSA only. This table provides information from the larger state population to use in comparison with the matched PSSA and SAT sample data. The SAT-taking population can be expected to post higher means. For all gender tables, data is coded such that positive numbers indicate higher mean scores for males; negative numbers indicate higher mean scores for females. For the full population of PSSA takers, the pattern of performance between males and females is consistent. Males outscore females, but only to a very small degree, on the PSSA mathematics test for all three included years. Females outscored males to a higher, but still small degree, on the PSSA reading test. The differences between years are small and inconsistent. This table shows that the gap in performance for both PSSA mathematics and reading between males and females is small and has not changed dramatically between 2001 and 2003.

Table 12. Gender Differences for PSSA (Unmatched State Data)

Year	Subject	Gender	Mean Scale Score	Standard Deviation	Number of Cases	Effect Size *d*
2001	Mathematics	Male	1,309.54	228.70	56,380	
		Female	1,302.88	211.70	55,496	0.030
	Reading	Male	1,271.07	227.31	56,091	
		Female	1,327.19	201.67	55,280	-0.261
2002	Mathematics	Male	1,324.75	238.13	60,777	
		Female	1,306.79	216.94	59,849	0.079
	Reading	Male	1,294.78	225.36	60,686	
		Female	1,334.43	200.34	59,636	-0.186
2003	Mathematics	Male	1,325.55	221.98	63,542	
		Female	1,304.34	206.80	61,849	0.099
	Reading	Male	1,293.08	246.74	63,840	
		Female	1,339.74	220.03	62,020	-0.200

Table 13 and Table 14 present data for the matched sample of students taking both PSSA and SAT, by gender. Scale scores for both assessments, as well as students' self-reported GPA, are included in the tables. The numbers of males versus females in the state student population is similar, typically with a few more males than females per grade. However, many more females take the SAT tests, suggesting that more females than males aspire to attend college. Males outscored females by a small margin on the PSSA and the SAT mathematics tests. This finding is consistent with scoring patterns on SAT and ACT reported by Willingham and Cole (1997). They suggest that, as a rule, females tend to avoid taking mathematics and science classes. In an apparent inconsistency, females outscored males on PSSA reading, but not on SAT verbal. Scores for SAT verbal were nearly identical for males and females. Willingham and Cole found gender differences favoring females in writing skills. The PSSA's performance-task items add a writing component that was not present in the SAT for the Classes of 2002 and 2003. However, SAT will include a student-written essay beginning in spring 2005. This study found that females tended to have higher self-reported GPAs than males.

Table 13. Gender Differences for Class of 2002 Matched Sample (2001 PSSA and 2002 SAT)

Performance Measure	Subject	Gender	Mean	Standard Deviation	Number of Cases	Effect Size *d*
PSSA	Mathematics	Male	1,419.25	205.36	29,403	
		Female	1,375.03	194.08	36,530	0.221
	Reading	Male	1,371.62	187.61	29,345	
		Female	1,394.20	169.21	36,451	-0.126
SAT	Mathematics	Male	522.73	110.76	29,787	
		Female	488.21	105.27	37,042	0.319
	Verbal	Male	500.81	103.33	29,787	
		Female	495.82	101.26	37,042	0.049
GPA		Male	3.243	.656	25,323	
		Female	3.404	.609	32,857	-0.255

Table 14. Gender Differences for Class of 2003 Matched Sample (2001 PSSA and 2002 SAT)

Performance Measure	Subject	Gender	Mean Score	Standard Deviation	Number of Cases	Effect Size *d*
PSSA	Mathematics	Male	1,450.23	210.49	31,338	
		Female	1,386.50	197.90	38,459	0.312
	Reading	Male	1,405.70	173.18	31,328	
		Female	1,407.15	164.27	38,418	-0.009
SAT	Mathematics	Male	525.96	110.67	31,557	
		Female	489.97	104.90	38,745	0.327
	Verbal	Male	505.24	102.35	31,557	
		Female	496.23	100.97	38,745	0.089
GPA		Male	3.265	.653	25,387	
		Female	3.415	.600	32,925	-0.239

Table 15 and Table 16 show regression results for adding gender as a predictor variable to the model for estimating PSSA scores based on SAT scores. For both PSSA reading and math, the regression weights for gender are negligible and the changes in R-square are essentially nonexistent. Gender is coded such that the positive weight means that females tend to have higher reading and math scores than would be expected from gender differences in SAT verbal.

Table 15. Regressions Results Showing the Adjusted Strength of Gender on Predicting 2001 PSSA 11th Grade Scale Scores Based on 2002 SAT Scores

PSSA Component	Standardized Coefficient			
	SAT Control	Gender	R^2	Change in R^2 due to Gender
Reading	0.704		0.495	
Reading	0.706	+0.081	0.502	+0.006
Math	0.847		0.717	
Math	0.851	+0.025	0.717	+0.001

Table 16. Regressions Results Showing the Adjusted Strength of Gender on Predicting 2002 PSSA 11th Grade Scale Scores Based on 2003 SAT Scores

PSSA Component	Standardized Coefficient			
	SAT Control	Gender	R^2	Change in R^2 due to Gender
Reading	0.775		0.601	
Reading	0.777	+0.038	0.602	+0.001
Math	0.865		0.749	
Math	0.863	-0.012	0.749	0.000

Ethnic or Racial Differences

Table 17 presents performance means for the four major ethnic or racial groups[3] (White, Black, Hispanic/Latino, and Asian) for the PSSA only. For these analyses, the racial/ethnic categories reported with the PSSA data were used. It should be noted that starting in 2003, the reported racial/ethnic categories were changed from those reported in 2001 and 2002. In 2003, the category of "Mixed" was no longer included. As a result, those students previously selecting "Mixed" were forced to select one of the other categories. For the full population of PSSA takers, the pattern of performance between these four groups is fairly consistent. In mathematics, Asians outscore all groups, Whites outscore Hispanics/Latinos and Blacks, and Hispanics/Latinos score between Blacks and Whites. For Reading, Asians and Whites score approximately the same, and Blacks score almost one standard deviation lower than Whites. However, the score for Hispanics/Latinos has shown a decrease over the three years of approximately 30 points. This decrease has dropped the Hispanic/Latino group to below the mean Black group score. The number of Hispanic/Latinos taking the PSSA has increased approximately 40%, Blacks approximately 65%, and Whites approximately 10% during that time, as well.

Table 17. Ethnic/Racial Differences for PSSA (Unmatched State Data)

Year	Subject	Measure	White	Black	Hispanic/ Latino	Asian
2001	Mathematics	Mean	1,324.81	1,135.96	1,167.57	1,408.20
		S.D.	215.25	170.93	188.73	246.57
		N	92,185	8,274	2,954	2,489
		Effect Size		0.972	0.777	-0.360
	Reading	Mean	1,316.44	1,151.58	1,170.25	1,329.27
		S.D.	210.66	203.93	211.04	217.07
		N	91,964	8,117	2,924	2,470
		Effect Size		0.795	0.693	-0.060
2002	Mathematics	Mean	1,339.35	1,140.90	1,164.38	1,434.30
		S.D.	220.98	176.26	190.41	267.57
		N	97,659	10,783	3,534	2,659
		Effect Size		0.993	0.848	-0.387
	Reading	Mean	1,339.17	1,148.87	1,161.93	1,337.11
		S.D.	203.51	200.70	211.55	229.14
		N	97,428	10,715	3,535	2,655
		Effect Size		0.942	0.854	0.010
2003	Mathematics	Mean	1,340.59	1,146.19	1,163.89	1,420.13
		S.D.	208.63	161.32	174.73	236.62
		N	104,064	13,669	4,106	2,865
		Effect Size		1.042	0.903	-0.357
	Reading	Mean	1,344.88	1,147.96	1,140.52	1,349.43
		S.D.	224.11	219.40	228.48	247.43
		N	104,271	13,871	4,158	2,866
				0.888	0.903	-0.019

[3] The racial/ethnic analyses were limited to these four groups because of the limited number of students in the other groups.

Table 18 and Table 19 present data for the matched sample of students taking both PSSA and SAT by ethnicity. Scale scores for both assessments, as well as students' self-reported GPA, are included in the tables. In 2002, Asians (72.6%) took the SAT tests at a much higher rate than any other group, followed by Whites (60.8%), Blacks (41.9%), and Hispanics/Latinos (36.4%). The order remained the same in 2003, although all groups but Asians had a slightly lower percentage in 2002 than in 2003. This suggests that more Asians and Whites aspire to attend college than do Blacks or Hispanic/Latinos. Scores on the PSSA, SAT and GPA show the same order for the matched sample as for the total state sample for the PSSA. Comparing the effect size for the PSSA versus the effect size for the SAT shows that the effect size for most groups in the Class of 2002 and components is approximately the same for the PSSA and the SAT. The largest difference for Blacks is between the PSSA reading and the SAT verbal components (PSSA reading – 0.883 and SAT verbal – 1.011). Note also that the effect size for GPA is approximately half the effect size for the PSSA and SAT for Blacks and Hispanics. However, for Asians, the GPA effect size is more similar to the effect size for math than it is for the SAT verbal or PSSA reading components.

Table 18. Ethnic/Racial Differences for the Class of 2002 Matched Sample (2001 PSSA 11th Grade and 2002 SAT)

Performance Measure	Subject	Measure	White	Black	Hispanic/ Latino	Asian
PSSA	Mathematics	Mean	1,408.09	1,208.09	1,285.29	1,462.56
		S.D.	192.72	177.59	194.83	228.12
		N	56,063	3,735	1,076	1,808
		Effect Size		1.079	0.634	-0.258
	Reading	Mean	1,396.58	1,236.58	1,291.12	1,380.40
		S.D.	170.04	191.80	191.99	191.83
		N	56,017	3,678	1,064	1,807
		Effect Size		0.883	0.582	0.089
SAT	Mathematics	Mean	510.41	402.86	443.32	548.20
		S.D.	104.09	102.68	110.82	130.77
		N	56,693	3,872	1,096	1,838
		Effect Size		1.040	0.624	-0.320
	Verbal	Mean	504.76	406.19	443.78	495.47
		S.D.	96.70	98.36	109.64	133.08
		N	56,693	3,872	1,096	1,838
		Effect Size		1.011	0.590	0.080
GPA		Mean	3.364	2.940	3.126	3.491
		S.D.	0.621	0.667	0.640	0.615
		N	49,617	3,337	871	1,616
		Effect Size		0.658	0.377	-0.205

Table 19. Ethnic/Racial Differences for the Class of 2003 for Matched Sample (2002 PSSA 11th Grade and 2003 SAT)

Performance Measure	Subject	Measure	White	Black	Hispanic/ Latino	Asian
PSSA	Mathematics	Mean	1,429.99	1,227.66	1,290.30	1,502.21
		S.D.	196.59	180.97	198.34	248.69
		N	59,076	4,516	1,178	1,951
		Effect Size		1.071	0.707	-0.322
	Reading	Mean	1,421.57	1,250.12	1,296.10	1,397.76
		S.D.	157.46	176.22	187.63	202.64
		N	59,038	4,500	1,184	1,951
		Effect Size		1.026	0.724	0.131
SAT	Mathematics	Mean	514.40	399.99	436.26	557.39
		S.D.	102.95	99.97	109.67	128.31
		N	59,443	4,573	1,206	1,970
		Effect Size		1.128	0.735	-0.370
	Verbal	Mean	508.34	405.68	438.51	501.66
		S.D.	95.39	99.13	107.45	131.63
		N	59,443	4,573	1,206	1,970
		Effect Size		1.055	0.687	0.058
GPA	Mathematics	Mean	3.385	2.933	3.175	3.538
		S.D.	0.610	0.666	0.641	0.595
		N	49,511	3,799	935	1,608
		Effect Size		0.708	0.336	-0.254

Table 20 through Table 23 show regression results for adding race/ethnicity (White versus Black and White versus Hispanic/Latino) as a predictor variable to the model for estimating PSSA scores based on SAT scores. For both PSSA reading and math, the regression weights for race/ethnicity are negligible and the changes in R-square are essentially nonexistent. Race/ethnicity is coded such that the negative weight means that Blacks or Hispanics/Latinos tend to have lower reading and math scores than would be expected from ethnicity differences in SAT verbal.

Table 20. Regressions Results Showing the Adjusted Strength of Being Black on Predicting 2001 PSSA 11th Grade Scale Scores Based on 2002 SAT Scores

	Standardized Coefficient			
PSSA Component	SAT Control	Black	R^2	Change in R^2 due to being Black
Reading	0.699		0.489	
Reading	0.686	-0.059	0.493	+0.003
Math	0.843		0.711	
Math	0.833	-.043	0.713	+0.002

Table 21. Regressions Results Showing the Adjusted Strength of Being Black on Predicting 2002 PSSA 11th Grade Scale Scores Based on 2003 SAT Scores

	Standardized Coefficient			
PSSA Component	SAT Control	Black	R^2	Change in R^2 due to being Black
Reading	0.769		0.591	
Reading	0.750	-0.068	0.595	0.004
Math	0.863		0.745	
Math	0.857	-0.021	0.746	0.000

Table 22. Regressions Results Showing the Adjusted Strength of Being Hispanic/Latino on Predicting 2001 PSSA 11th Grade Scale Scores Based on 2002 SAT Scores

	Standardized Coefficient			
PSSA Component	SAT Control	Hispanic/Latino	R^2	Change in R^2 due to being Hispanic/Latino
Reading	0.687		0.472	
Reading	0.685	-0.026	0.473	+0.001
Math	0.837		0.700	
Math	0.836	-0.014	0.700	0.000

Table 23. Regressions Results Showing the Adjusted Strength of Being Hispanic/Latino on Predicting 2002 PSSA 11th Grade Scale Scores Based on 2003 SAT Scores

	Standardized Coefficient			
PSSA Component	SAT Control	Hispanic/Latino	R^2	Change in R^2 due to being. Hispanic/Latino
Reading	0.754		0.568	
Reading	0.050	-0.034	0.569	+0.001
Math	0.857		0.735	
Math	0.856	-0.009	0.735	0.000

Economically Disadvantaged Differences

Table 24 presents performance means for those students identified as economically disadvantaged for the PSSA only. This table provides information from the larger state population to use in comparison with the matched PSSA and SAT sample data. The SAT-taking population can be expected to post higher means. For the full population of PSSA takers, the pattern of performance between males and females is consistent. Noneconomically disadvantaged students outscore economically disadvantaged student to a large degree on all components and all years except 2001 when it was a medium degree. The number of students identified more than doubled between 2001 and 2002 and increased an additional 40% in 2003.

Table 24. Differences for Economically Disadvantaged Students for PSSA (Unmatched State Data)

Year	Subject	Economically Disadvantaged	Mean Scale Score	Standard Deviation	Number of Cases	Effect Size *d*
2001	Mathematics	Yes	1,186.47	189.77	8,805	
		No	1,313.77	220.14	107,332	0.619
	Reading	Yes	1,189.18	211.42	8,702	
		No	1,305.64	215.70	106,864	0.545
2002	Mathematics	Yes	1,166.04	189.13	17,253	
		No	1,338.57	224.97	106,297	0.830
	Reading	Yes	1,163.80	209.14	17,199	
		No	1,337.04	206.17	106,035	0.834
2003	Mathematics	Yes	1,182.11	179.97	24,022	
		No	1,345.75	210.63	102,919	0.835
	Reading	Yes	1,167.47	228.17	24,290	
		No	1,350.50	223.19	103,137	0.811

Table 25. Differences for Economically Disadvantaged Students for Matched Sample for the Class of 2002 (2001 PSSA 11th Grade and 2002 SAT)

Performance Measure	Subject	Economically Disadvantaged	Mean Score	Standard Deviation	Number of Cases	Effect Size *d*
PSSA	Mathematics	Yes	1,305.98	188.30	2,893	
		No	1,396.68	200.44	65,296	0.466
	Reading	Yes	1,317.68	183.36	2,863	
		No	1,385.35	178.29	65,176	0.374
SAT	Mathematics	Yes	451.85	103.99	2,926	
		No	504.91	109.12	66,399	0.498
	Verbal	Yes	446.94	101.31	2,926	
		No	499.53	102.15	66,399	0.517
GPA		Yes	3.195	0.661	2,561	
		No	3.334	0.636	57,722	0.214

Table 26. Differences for Economically Disadvantaged Students for Matched Sample for the Class of 2003 (2002 PSSA 11th Grade and 2003 SAT)

Performance Measure	Subject	Economically Disadvantaged	Mean Score	Standard Deviation	Number of Cases	Effect Size *d*
PSSA	Mathematics	Yes	1,277.41	194.74	5,822	
		No	1,427.07	202.78	65,159	0.753
	Reading	Yes	1,281.69	184.98	5,812	
		No	1,417.42	162.43	65,114	0.780
SAT	Mathematics	Yes	424.12	107.46	5,907	
		No	513.41	106.16	65,604	0.836
	Verbal	Yes	421.00	104.91	5,907	
		No	507.48	98.45	65,604	0.850
GPA		Yes	3.116	0.680	4,854	
		No	3.369	0.620	54,404	0.389

Table 27. Regressions Results Showing the Adjusted Strength of Being Economically Disadvantaged on Predicting 2001 PSSA 11th Grade Scale Scores Based on 2002 SAT Scores

PSSA Component	Standardized Coefficient			
	SAT Control	Economically Disadvantaged	R^2	Change in R^2 due to being Economically Disadvantaged
Reading	0.704		0.495	
Reading	0.703	-0.004	0.495	0.000
Math	0.846		0.715	
Math	0.845	-0.009	0.715	0.000

Table 28. Regressions Results Showing the Adjusted Strength of Being Economically Disadvantaged on Predicting 2002 PSSA 11th Grade Scale Scores Based on 2003 SAT Scores

PSSA Component	Standardized Coefficient			
	SAT Control	Economically Disadvantaged	R^2	Change in R^2 due to being Economically Disadvantaged
Reading	0.775		0.601	
Reading	0.765	-.044	0.602	0.002
Math	0.865		0.749	
Math	0.864	-0.005	0.749	0.000

Annual Family Income Differences

Camara and Schmidt (1999) point out that family income is related to performance on tests such as the SAT I. They also point out that this finding is not unique to the SAT I or college admissions tests. Terenzini, Cabrera, and Bernal (2001) also examined various differences in educational outcomes associated with varying levels of socioeconomic status (SES). This report

examines differences associated with family income in this section. The following section shows differences in student performance related to parental education level. Terenzini, Cabrera, and Bernal (2001) and Camara and Schmidt (1999) point out differences between income, wealth, and SES.

Because annual family income data comes from the SDQ of the SAT, only matched student data can be presented showing relationships with family income. Table 29 (Class of 2002) and Table 30 (Class of 2003) show student performance on the PSSA, SAT, and GPA versus four levels of family income. As the table clearly shows, student scores increase on all performance measures as income goes up. However, the effect size for GPA is about half of the effect size for either of the assessments. The effect size for the PSSA is similar, although slightly less, than the effect size for the SAT for both components and both years.

Table 29. Annual Family Income Differences for Matched Sample for the Class of 2002 (2001 PSSA 11th Grade and 2002 SAT)

Performance Measure	Subject	Measure	Income Level			
			< $25,000	$25,000 to $40,000	$40,000 to $70,000	> $70,000
PSSA	Math	Mean	1,299.00	1,353.81	1,397.02	1,450.45
		S.D.	191.08	183.61	188.99	200.32
		N	6,658	8,709	16,365	14,902
		Effect Size		-0.293	-0.516	-0.774
	Reading	Mean	1,310.11	1,362.10	1,388.48	1,420.93
		S.D.	184.66	171.65	171.10	169.21
		N	6,594	8,693	16,348	14,867
		Effect Size		-0.292	-0.440	-0.626
SAT	Math	Mean	446.65	477.66	502.72	538.16
		S.D.	105.94	100.13	101.36	105.80
		N	6,822	8,856	16,549	15,122
		Effect Size		-0.301	-0.541	-0.864
	Verbal	Mean	444.69	476.19	497.96	528.25
		S.D.	101.77	95.46	94.85	98.15
		N	6,822	8,856	16,549	15,122
		Effect Size		-0.319	-0.542	-0.836
GPA		Mean	3.173	3.262	3.344	3.412
		S.D.	0.656	0.636	0.631	0.628
		N	6,596	8,607	16,141	14,726
		Effect Size		-0.138	-0.266	-0.372

Table 30. Annual Family Income Differences for Matched Sample for the Class of 2003 (2002 PSSA 11th Grade and 2003 SAT)

Performance Measure	Subject	Measure	Income Level			
			< $25,000	**$25,000 - $40,000**	**$40,000 - $70,000**	**> $70,000**
PSSA	Math	Mean	1,315.77	1,363.93	1,409.24	1,470.93
		S.D.	194.01	190.51	193.19	205.60
		N	6,230	7,958	14,963	14,295
		Effect Size		-0.250	-0.483	-0.776
	Reading	Mean	1,326.98	1,373.11	1,405.72	1,445.69
		S.D.	175.76	162.01	159.19	159.65
		N	6,232	7,948	14,958	14,284
		Effect Size		-0.273	-0.470	-0.707
SAT	Math	Mean	448.71	476.89	500.75	539.65
		S.D.	105.62	101.40	100.89	105.67
		N	6,304	8,010	15,045	14,388
		Effect Size		-0.272	-0.504	-0.861
	Verbal	Mean	447.93	475.81	496.53	529.05
		S.D.	103.45	95.78	93.54	96.89
		N	6,304	8,010	15,045	14,388
		Effect Size		-0.280	-0.493	-0.809
GPA		Mean	3.188	3.279	3.357	3.428
		S.D.	0.652	0.642	0.623	0.617
		N	6,115	7,832	14,659	14,000
		Effect Size		-0.141	-0.265	-0.378

Parent Educational Differences

Camara and Schmidt (1999) also looked at the impact of parent education on student performance in standardized testing. Their study showed that parental education is related to performance on the SAT I. Higher parental education is associated with higher scores on the SAT I. Table 31 and Table 32 show the means, standard deviations, number of students, and effect size for various levels of student-reported mother's education for the PSSA, SAT, and GPA. The analyses indicate that the effect size for both assessments and GPA are larger as the mother's educational level increases. The effect size for PSSA is not as large as the effect size for SAT. The effect size for GPA is about half of the effect size for PSSA or SAT. Table 33 and Table 34 provide the same information regarding the father's education level. The results follow the same pattern as that for mother's education level.

Table 31. Mother's Education Level Differences for Matched Sample for the Class of 2002 (2001 PSSA 11th Grade and 2002 SAT)

	Subject	Measure	< High School	High School	Some College	BA/BS Degree	Graduate Degree
PSSA	Math	Mean	1,288.55	1,359.08	1,382.12	1,457.89	1,480.95
		S.D.	187.85	185.42	191.61	199.82	206.88
		N	2,196	18,737	16,804	12,094	6,259
		Effect Size		-.378	-0.493	-0.873	-0.974
	Reading	Mean	1,290.39	1,358.83	1,378.03	1,433.60	1,447.35
		S.D.	178.86	171.34	174.41	167.33	169.81
		N	2,189	18,696	16,762	12,078	6,245
		Effect Size		-0.377	-0.495	-0.883	-0.985
SAT	Math	Mean	437.95	480.49	495.74	541.33	556.39
		S.D.	104.76	100.21	104.29	105.47	109.24
		N	2,247	18,986	17,075	12,291	6,357
		Effect Size		-0.415	-0.553	-0.983	-1.107
	Verbal	Mean	426.14	474.05	492.77	535.12	550.72
		S.D.	99.85	92.91	96.35	98.56	103.00
		N	2,247	18,986	17,075	12,291	6,357
		Effect Size		-0.497	-0.679	-1.099	-1.228
GPA		Mean	3.142	3.267	3.289	3.456	3.504
		S.D.	0.653	0.630	0.640	0.618	0.601
		N	2,145	18,341	16,424	11,914	6,150
		Effect Size		-0.195	-0.227	-0.494	-0.577

Table 32. Mother's Education Level Differences for Matched Sample for the Class of 2003 (2002 PSSA 11th Grade and 2003 SAT)

	Subject	Measure	< High School	High School	Some College	BA/BS Degree	Graduate Degree
PSSA	Math	Mean	1,297.23	1,373.23	1,400.38	1,480.27	1,501.40
		S.D.	195.13	187.46	198.65	204.74	224.94
		N	2,072	17,801	16,646	11,406	5,961
		Effect Size		-0.397	-0.524	-0.915	-0.970
	Reading	Mean	1,290.22	1,377.51	1,397.99	1,457.48	1,470.71
		S.D.	180.92	160.42	162.41	159.51	162.72
		N	2,066	17,787	16,638	11,406	5,954
		Effect Size		-0.511	-0.627	-0.981	-1.059
SAT	Math	Mean	436.31	481.06	497.38	544.11	556.41
		S.D.	104.94	100.46	103.87	104.67	113.05
		N	2,095	17,922	16,756	11,496	5,992
		Effect Size		-0.436	-0.585	-1.029	-1.101
	Verbal	Mean	425.49	475.92	494.32	536.18	550.50
		S.D.	101.24	92.54	95.47	97.95	105.22
		N	2,095	17,922	16,756	11,496	5,992
		Effect Size		-0.520	-0.700	-1.111	-1.211
GPA		Mean	3.148	3.285	3.307	3.475	3.510
		S.D.	0.646	0.622	0.636	0.608	0.599
		N	2,003	17,344	16,134	11,088	5,797
		Effect Size		-0.216	-0.248	-0.521	-0.581

Table 33. Father's Education Level Differences for Matched Sample for the Class of 2002 (2001 PSSA 11th Grade and 2002 SAT)

	Subject	Measure	< High School	High School	Some College	BA/BS Degree	Graduate Degree
PSSA	Math	Mean	1,289.08	1,350.41	1,383.68	1,448.43	1,500.72
		S.D.	184.43	182.51	187.21	196.45	207.36
		N	2,850	17,616	14,573	12,461	7,835
		Effect Size		-0.334	-0.509	-0.836	-1.079
	Reading	Mean	1,295.71	1,354.53	1,379.85	1,423.52	1,459.94
		S.D.	181.91	170.66	172.91	167.98	166.61
		N	2,289	17,558	14,539	12,454	7,838
		Effect Size		-0.333	-0.474	-0.730	-0.942
SAT	Math	Mean	439.58	475.10	496.26	536.51	567.59
		S.D.	102.29	99.26	101.16	103.68	107.27
		N	2,915	17,845	14,800	12,636	7,980
		Effect Size		-0.352	-0.557	-0.941	-1.22
	Verbal	Mean	432.45	470.58	492.10	528.30	561.14
		S.D.	97.78	92.12	94.75	95.71	101.28
		N	2,915	17,845	14,800	12,636	7,980
		Effect Size		-0.401	-0.620	-0.991	-1.293
GPA		Mean	3.136	3.247	3.301	3.432	3.548
		S.D.	0.637	0.634	0.633	0.618	0.594
		N	2,784	17,269	14,263	12,198	7,723
		Effect Size		-0.175	-0.260	-0.472	-0.669

Table 34. Father's Education Level Differences for Matched Sample for the Class of 2003 (2002 PSSA 11th Grade and 2003 SAT)

	Subject	Measure	< High School	High School	Some College	BA/BS Degree	Graduate Degree
PSSA	Math	Mean	1,311.97	1,365.33	1,400.22	1,471.83	1,525.03
		S.D.	191.07	187.33	191.63	202.47	220.47
		N	2,766	16,981	14,422	11,633	7,272
		Effect Size		-0.282	-0.461	-0.812	-1.033
	Reading	Mean	1,309.55	1,372.22	1,400.37	1,447.65	1,487.52
		S.D.	177.15	159.01	159.80	159.90	160.14
		N	2,761	16,968	14,412	11,629	7,265
		Effect Size		-0.372	-0.538	-0.818	-1.054
SAT	Math	Mean	445.05	475.69	497.72	539.11	571.15
		S.D.	102.30	99.38	100.23	103.52	110.09
		N	2,803	17,087	14,508	11,709	7,314
		Effect Size		-0.304	-0.520	-0.914	-1.187
	Verbal	Mean	437.01	472.24	494.99	529.20	562.96
		S.D.	98.87	91.95	93.10	95.80	102.94
		N	2,803	17,087	14,508	11,709	7,314
		Effect Size		-0.369	-0.604	-0.947	-1.248
GPA		Mean	3.139	3.272	3.316	3.454	3.558
		S.D.	0.639	0.624	0.632	0.606	0.582
		N	2,696	16,545	14,014	11,281	7,046
		Effect Size		-0.211	-0.279	-0.506	-0.686

Limited English Proficiency Differences

Students identified as having limited English proficiency (LEP) would be expected to have more difficulty taking a test in English than those not identified. As a result of these difficulties, LEP students would be expected to have lower scores than non-LEP students. Table 36 shows the mean, standard deviation, number of students, and the effect size for the entire state on the PSSA in 2001 and 2002. As expected, LEP students have lower scores than non-LEP students on both components and effect sizes are generally large. However, the math component effect size is only about half of the effect size for reading.

Table 35. Limited English Proficiency Differences for 11th Grade PSSA (Unmatched State Data)

Year	Subject	Limited English Proficiency	Mean Scale Score	Standard Deviation	Number of Cases	Effect Size *d*
2001	Math	Yes	1,153.12	189.77	492	
		No	1,304.76	220.47	115,645	0.737
	Reading	Yes	1,062.16	153.38	473	
		No	1,297.84	217.27	115,093	1.253
2002	Math	Yes	1,165.60	202.54	844	
		No	1,315.50	228.12	122,706	0.695
	Reading	Yes	1,026.18	161.38	840	
		No	1,314.83	214.13	122,394	1.522

Table 36 and Table 37 show the mean, standard deviation, number of cases, and effect size for the matched sample of students in the Classes of 2002 and 2003. The effect sizes are similar to those in the state sample for the PSSA and SAT—very larger in verbal/reading and medium in math. The effect size is slightly larger for PSSA than SAT in math, but slightly smaller for PSSA than SAT in reading/verbal. However, the GPA for LEP students is actually slightly higher than the GPA for non-LEP students.

Table 36. Limited English Proficiency Differences for Matched Sample for the Class of 2002 (2001 PSSA 11th Grade and 2002 SAT)

Performance Measure	Subject	Limited English Proficiency	Mean Score	Standard Deviation	Number of Cases	Effect Size *d*
PSSA	Math	Yes	1,259.74	196.61	134	
		No	1,393.10	200.70	68,055	0.671
	Reading	Yes	1,137.45	146.49	132	
		No	1,382.97	178.76	67,907	1.502
SAT	Math	Yes	428.60	130.45	136	
		No	502.82	109.34	69,189	0.617
	Verbal	Yes	333.53	98.07	136	
		No	497.64	102.41	69,189	1.637
GPA		Yes	3.393	0.617	111	
		No	3.328	0.637	60,172	-0.104

Table 37. Limited English Proficiency Differences for Matched Sample for the Class of 2003 (2002 PSSA 11th Grade and 2003 SAT)

Performance Measure	Subject	Limited English Proficiency	Mean Score	Standard Deviation	Number of Cases	Effect Size *d*
PSSA	Mathematics	Yes	1,268.15	202.57	289	
		No	1,415.39	206.07	70,692	0.721
	Reading	Yes	1,093.86	158.55	288	
		No	1,407.57	167.41	70,638	1.924
SAT	Mathematics	Yes	415.64	121.59	291	
		No	506.41	108.87	71,220	0.787
	Verbal	Yes	325.26	87.75	291	
		No	501.05	101.26	71,220	1.855
GPA		Yes	3.358	.585	226	
		No	3.348	.629	59,032	-0.008

The regression analyses results (Table 38 and Table 39) show that there are negligible changes in R square for predicting PSSA from SAT scores when adding LEP to the equation for both reading and math for both classes.

Table 38. Regressions Results Showing the Adjusted Strength of Being Limited English Proficient on Predicting 2001 PSSA 11th Grade Scale Scores Based on 2002 SAT Scores

PSSA Component	Standardized Coefficient			
	SAT Control	Limited English Proficiency	R^2	Change in R^2 due to Limited English Proficiency
Reading	0.704		0.495	
Reading	0.703	-0.011	0.495	0.000
Math	0.846		0.715	
Math	0.845	-0.004	0.715	0.000

Table 39. Regressions Results Showing the Adjusted Strength of Being Limited English Proficient on Predicting 2002 PSSA 11th Grade Scale Scores Based on 2003 SAT Scores

PSSA Component	Standardized Coefficient			
	SAT Control	Limited English Proficiency	R^2	Change in R^2 due to Limited English Proficiency
Reading	0.775		0.601	
Reading	0.771	-0.033	0.602	+0.001
Math	0.865		0.749	
Math	0.865	0.000	0.749	0.000

School-Level Scores

Means were calculated by school for all scale scores and grade indicators to analyze school-level results. School-level descriptive statistics are presented for the Classes of 2002 and 2003 for the PSSA (Table 40), SAT (Table 41), and GPA and course grades (Table 42). Only schools with at least 10 students with merged PSSA and SAT scores are included in the analyses. School-level means showed an increase on both tests, both test components, GPA, and course grades from the Class of 2002 to the Class of 2003. The change in scoring procedures from the 2001 PSSA to the 2002 PSSA (instituting a minimum score of 700 for the 2002 PSSA) is problematic. However, there were only 9 students in this college-bound sample in the Class of 2003 who scored a 700 on the 2002 PSSA reading component and no students who scored a 700 on the math component. The change in course grades was extremely small.

Table 40. PSSA School-Level Descriptive Statistics for the Classes of 2002 and 2003

Subject	Measure	Class of 2002	Class of 2003
Reading	Mean	1369.73	1393.16
	S.D.	74.35	70.88
	N	586*	586
Math	Mean	1369.66	1393.33
	S.D.	83.35	84.96
	N	586	586

* Limited to schools with at least 10 students with merged PSSA and SAT scores

Table 41. SAT School-Level Descriptive Statistics for the Classes of 2002 and 2003

Subject	Measure	Class of 2002	Class of 2003
Verbal	Mean	485.78	489.07
	S.D.	42.07	42.51
	N	586*	586
Math	Mean	488.47	491.59
	S.D.	45.00	46.01
	N	586	586

* Limited to schools with at least 10 students with merged PSSA and SAT scores

Table 42. School-Level Descriptive Statistics GPA and Course Grades for the Classes of 2002 and 2003

Subject	Measure	Class of 2002	Class of 2003
GPA	Mean	3.34	3.46
	S.D.	0.19	0.19
	N	586*	586
English	Mean	3.29	3.31
	S.D.	0.20	0.21
	N	586	586
Math	Mean	3.11	3.12
	S.D.	0.21	0.22
	N	586	586

* Limited to schools with at least 10 students with merged PSSA and SAT scores

Average Within-Year School-Level Correlations

Table 43 and Table 44 present correlations among these school-level scores. Table 45 presents correlations for the average across the two years.[4] As Hoffman (1998) pointed out, there is a tendency for correlations among school scores to be higher than correlations among student-level scores (Class of 2002 - Table 8, Class of 2003 - Table 9, and average for the Classes of 2002 and 2003 - Table 10). This is the case, except for those correlations involving GPA and course-grades (except for PSSA reading to math course grade and GPA to English course grade where school-level correlations are higher).

As Hoffman (1998) pointed out, the lower school-level correlations for GPA suggests that grading scales are not equivalent across schools. There is a tendency for different schools to grade differently. As a result, at the school level of analysis there is a weaker relationship between GPA and either PSSA or ACT scores. When students' scores are averaged at the school level, observed differences are related not only to differing levels of student achievement, but to the particular school's grading practices. However, the school-level relationship between GPA and course grades remains strong for both PSSA and SAT.

[4] Average correlations were technically corrected with Fisher's *r* to *z* transformations.

Table 43. School-Level PSSA, SAT, and Course Grade Correlations for the Class of 2002

		PSSA		SAT		Course Grades		
Variable		Reading	Math	Verbal	Math	GPA (Total)	English	Math
PSSA	Reading	1.000						
	Math	*0.861*	1.000					
SAT	Verbal	**0.818**	**0.883**	1.000				
	Math	**0.773**	**0.935**	*0.941*	1.000			
Course Grades	GPA	0.413	0.393	0.371	0.376	1.000		
	English	**0.411**	**0.391**	**0.359**	**0.362**	0.778	1.000	
	Math	**0.423**	**0.454**	**0.416**	**0.450**	0.719	*0.542*	1.000

* Limited to schools with at least 10 students with merged PSSA and SAT scores

Table 44. School-Level PSSA, SAT, and Course Grades Correlations for the Class of 2003

		PSSA		SAT		Course Grades		
Variable		Reading	Math	Verbal	Math	GPA (Total)	English	Math
PSSA	Reading	1.000						
	Math	*0.870*	1.000					
SAT	Verbal	**0.905**	**0.849**	1.000				
	Math	**0.868**	**0.915**	*0.943*	1.000			
Course Grades	GPA	0.491	0.443	0.439	0.425	1.000		
	English	**0.442**	**0.394**	**0.402**	**0.394**	0.780	1.000	
	Math	**0.501**	**0.489**	**0.484**	**0.486**	0.784	*0.615*	1.000

* Limited to schools with at least 10 students with merged PSSA and SAT scores

Table 45. School-Level PSSA, SAT, and Course Grades Correlation Averages for the Class of 2002 and 2003

Variable		PSSA		SAT		Course Grades		
		Reading	Math	Verbal	Math	GPA (Total)	English	Math
PSSA	Reading	1.000						
	Math	*0.866*	1.000					
SAT	Verbal	**0.868**	**0.867**	1.000				
	Math	**0.826**	**0.926**	*0.942*	1.000			
Course Grades	GPA	0.453	0.418	0.406	0.401	1.000		
	English	**0.427**	**0.393**	**0.381**	**0.378**	0.779	1.000	
	Math	**0.463**	**0.472**	**0.451**	**0.468**	0.753	*0.580*	1.000

* Limited to schools with at least 10 students with merged PSSA and SAT scores

Table 46 shows that the mean school-level scores for both PSSA 11th grade reading and SAT verbal components of each assessment increased from the Class of 2002 to the Class of 2003. For the SAT, the mean school-level change on the verbal component was +3.30 scale score points. For the PSSA, the mean school-level scale score change on the 11th grade reading component was +23.43 scale score points. Table 47 shows the fairly strong relationship between the differences in mean school-level scores on the SAT verbal component and the PSSA reading component for the Classes of 2002 and 2003. This relationship is graphically portrayed in Figure 17. Table 48, Table 49, and Figure 18 show the same information for the math components of the PSSA and SAT. As would be expected, the relationship between the math components is stronger than for the reading/verbal components.

Table 46. Difference in Mean School-Level SAT Verbal and PSSA Reading Scores From the Classes of 2002 and 2003

Difference in Mean School-Level Score From the Class of 2002 and Class of 2003	Mean	Standard Deviation	Number of Schools*
SAT Verbal	+3.30	16.73	586
PSSA Reading	+23.43	46.60	586

* Limited to schools with at least 10 students with merged PSSA and SAT scores

Table 47. Correlation of Difference in School-Level Mean SAT Verbal Score and PSSA Reading Score Between the Classes of 2002 and 2003

Difference in Mean School-Level Score From the Class of 2002 and Class of 2003	SAT Verbal	PSSA Reading
SAT Verbal	1	
PSSA Reading	.462	1

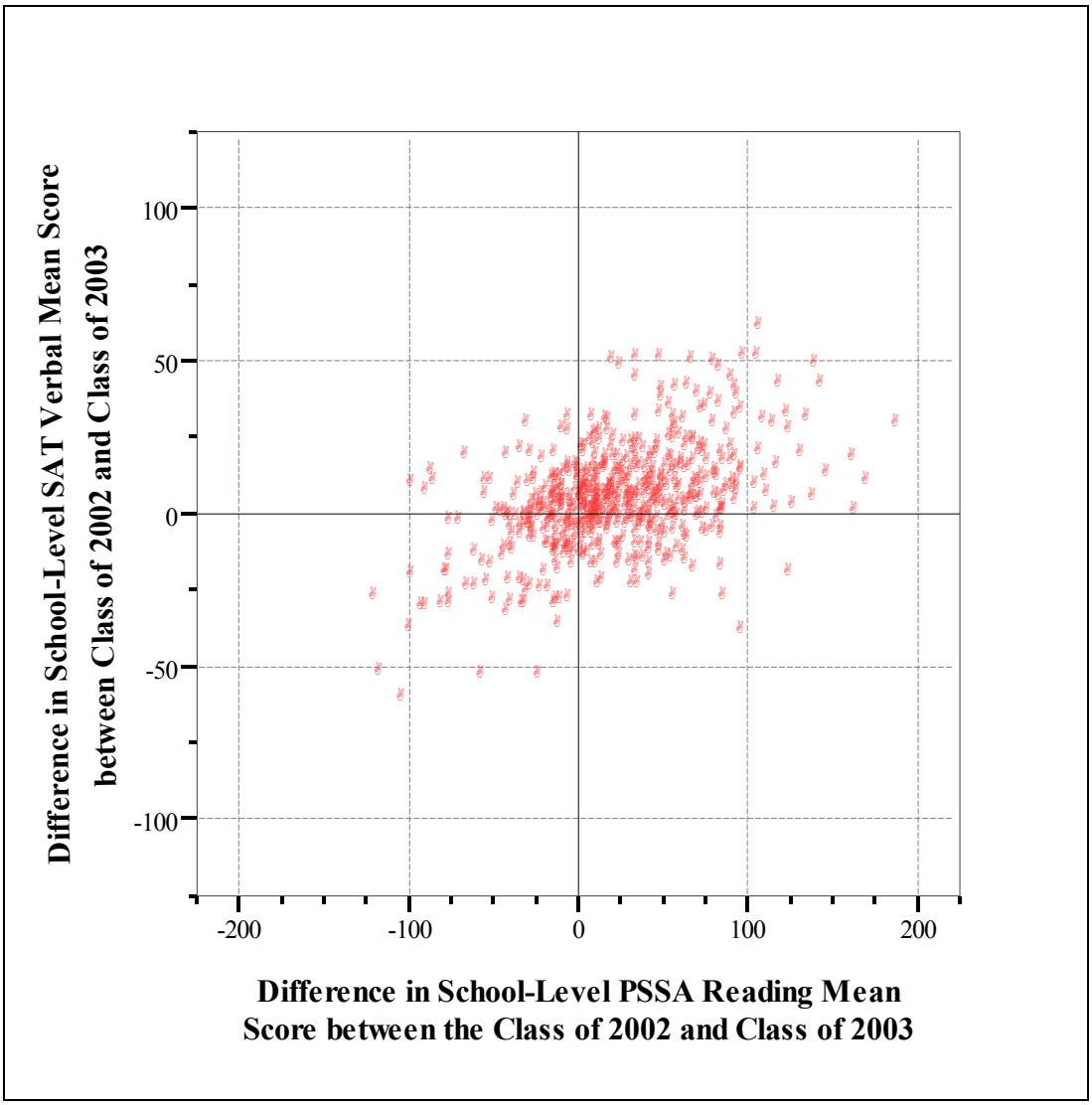

Figure 17. Scatterplot of difference in school-level SAT verbal mean scores and school-level PSSA 11th grade reading mean scores between the Class of 2002 and Class of 2003.

Table 48. Difference in Mean School-Level SAT Math and PSSA Math Scores From the Class of 2002 and Class of 2003

Difference in Mean School-Level Math Score From the Class of 2002 and Class of 2003	Mean	Standard Deviation	Number of Schools
SAT	+3.13	18.42	586
PSSA	+23.67	46.87	586

* Limited to schools with at least 10 students with merged PSSA and SAT scores

Table 49. Correlation of Difference in School-Level Mean SAT Math Score and PSSA Math Score Between the Classes of 2002 and 2003

Difference in Mean School-Level Math Score From the Class of 2002 and Class of 2003	SAT	PSSA
SAT	1	
PSSA	.644	1

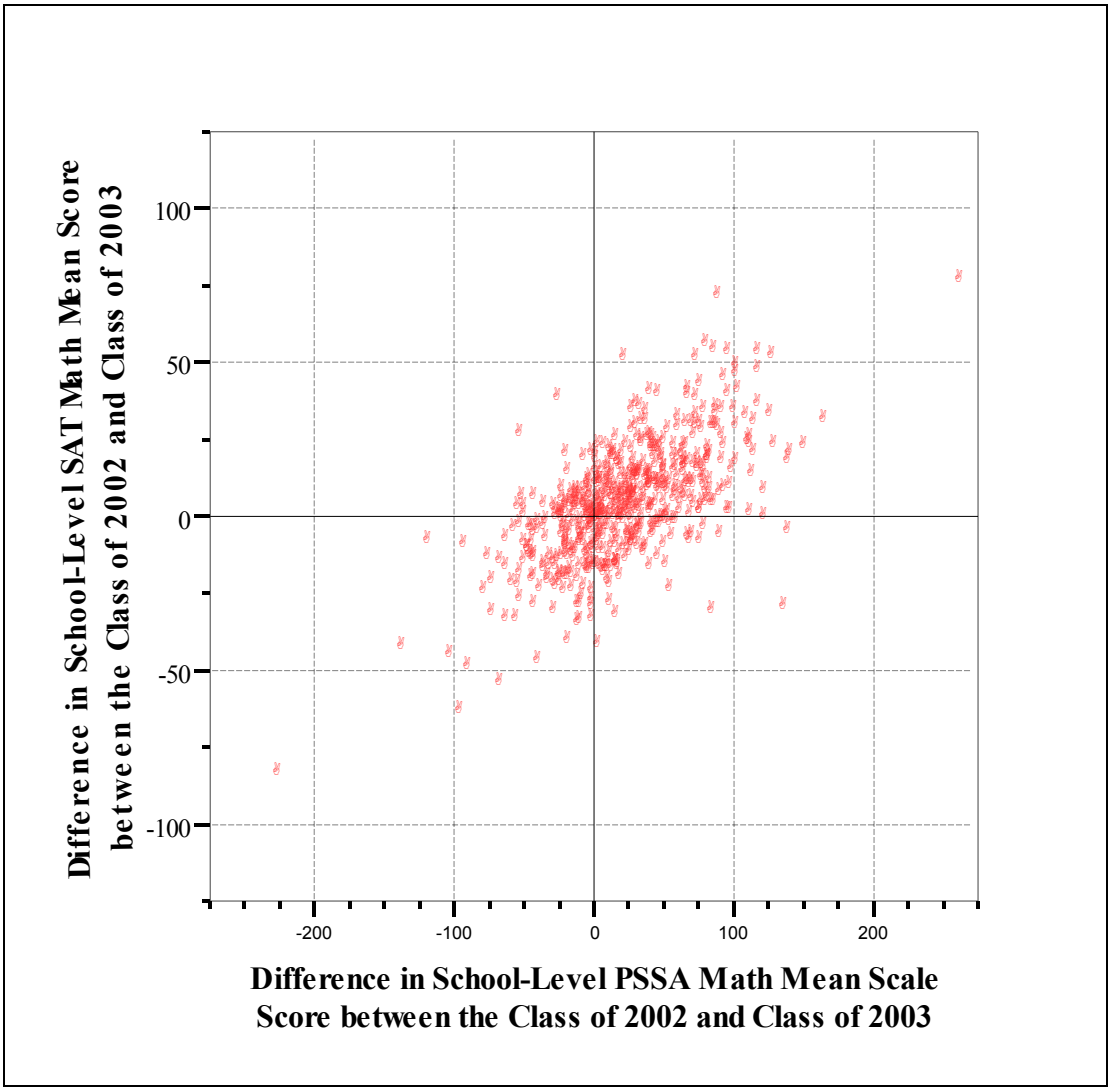

Figure 18. Scatterplot of differences in school-level mean SAT math score and school-level mean PSSA math scores between the Classes of 2002 and 2003.

Pooled, Within-School Correlations

The above discussion alluded to school-level factors, such as differences in grading practices, which may be impacting observed correlations. Figure 19 depicts an extreme version of such a situation. Schools A, B, and C are clearly different in where their students lie with respect to PSSA or SAT scores, but they are identical in terms of the grades their students

receive. Also, within each school, there is a relationship between grades and either PSSA scores or SAT scores. However, if a correlation were calculated using students from all schools, without regard to school-level characteristics, it would be zero.

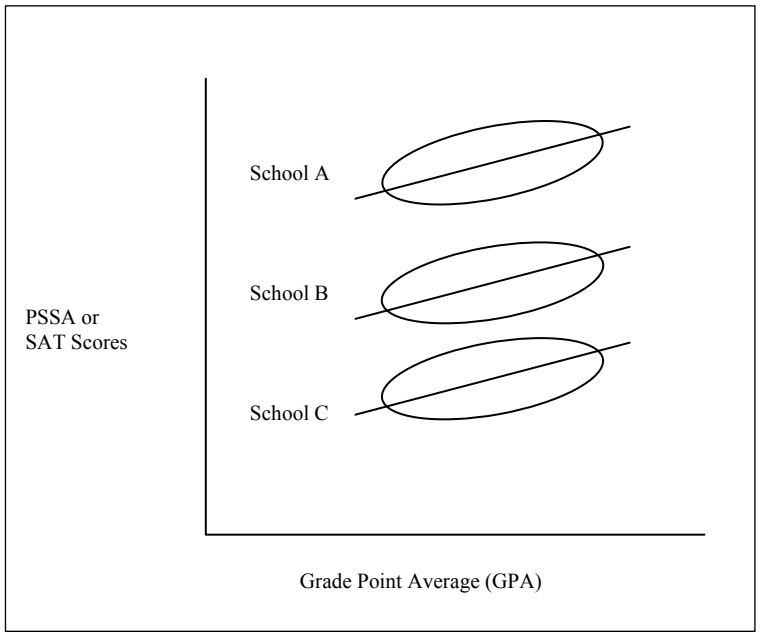

Figure 19. Extreme hypothetical illustration of how mean differences among schools on PSSA or SAT scores that are not mirrored by differences in GPA can impact correlations. (Ellipses represent students' scores within each school).

Pooled, within-school correlations are used to estimate relationships between measures calculated within schools and then averaged across all schools. These pooled, within-school correlations were calculated for each of the two years of data (Table 50 for the Class of 2002 and Table 51 for the Class of 2003), and then averaged across the two years (Table 52).[5] The resulting average pooled, within-school correlations are presented in Table 52.

[5] Again, the technical *r* to *z* correction was used.

Table 50. Pooled, Within-School Correlation for the Class of 2002

Variable		PSSA		SAT		Course Grades		
		Reading	Math	Verbal	Math	GPA (Total)	English	Math
PSSA	Reading	1.000						
	Math	*0.648*	1.000					
SAT	Verbal	**0.693**	**0.650**	1.000				
	Math	**0.569**	**0.832**	*0.695*	1.000			
Course Grades	GPA	0.528	0.573	0.516	0.559	1.000		
	English	**0.461**	**0.412**	**0.465**	**0.402**	0.703	1.000	
	Math	**0.361**	**0.542**	**0.349**	**0.541**	0.668	*0.443*	1.000

* Limited to schools with at least 10 students with merged PSSA and SAT scores

Table 51. Pooled, Within-School Correlation for the Class of 2003

Variable		PSSA		SAT		Course Grades		
		Reading	Math	Verbal	Math	GPA (Total)	English	Math
PSSA	Reading	1.000						
	Math	*0.702*	1.000					
SAT	Verbal	**0.757**	**0.629**	1.000				
	Math	**0.666**	**0.857**	*0.662*	1.000			
Course Grades	GPA	0.512	0.564	0.534	0.679	1.000		
	English	**0.457**	**0.405**	**0.466**	**0.414**	0.698	1.000	
	Math	**0.350**	**0.541**	**0.369**	**0.548**	0.667	*0.443*	1.000

* Limited to schools with at least 10 students with merged PSSA and SAT scores

Table 52. Pooled, Within-School Correlation Averaged for Classes of 2002 and 2003

		PSSA		SAT		Course Grades		
Variable		Reading	Math	Verbal	Math	GPA (Total)	English	Math
PSSA	Reading	1.000						
	Math	*0.676*	1.000					
SAT	Verbal	**0.727**	**0.640**	1.000				
	Math	**0.620**	**0.845**	*0.679*	1.000			
Course Grades	GPA	0.520	0.569	0.525	0.569	1.000		
	English	**0.459**	**0.409**	**0.466**	**0.408**	0.701	1.000	
	Math	**0.356**	**0.542**	**0.359**	**0.546**	0.668	*0.443*	1.000

* Limited to schools with at least 10 students with merged PSSA and SAT scores

If differences between schools' grading procedures are substantial, and if within schools, students with higher grades also have higher PSSA scores, then the correlations in Table 52 will be higher than student correlations in Table 10 (student-level correlations averaged for the Classes of 2002 and 2003). The difference between the GPA correlations in Table 10 and Table 52, however, are only marginally different (see Table 53). In nearly all cases, the within-school correlations are higher. However, the range is relatively small—0.009 to 0.039 for GPA, 0.176 to 0.219 for English course grades, and 0.147 to 0.237 for math course grades. Thus, grading differences among schools do appear to exist, but the effect is small for students' GPA, but larger for both English and math course grades especially when compared to the matching assessment component (i.e., English course grade to PSSA reading and SAT verbal, math course grade to PSSA math and SAT math).

Table 53. Differences Between Student-Level Correlations and Pooled Within-School Correlations Averaged for the Classes of 2002 and 2003

GPA or Course Grade to Assessment	Student-Level Correlations	Pooled, Within-School Correlations	Difference in Correlations
GPA to PSSA Reading	0.511	0.520	0.009
GPA to PSSA Math	0.546	0.569	0.023
GPA to SAT Verbal	0.494	0.525	0.031
GPA to SAT Math	0.530	0.569	0.039
English to PSSA Reading	0.240	0.459	0.219
English to PSSA Math	0.226	0.409	0.183
English to SAT Verbal	0.256	0.466	0.210
English to SAT Math	0.232	0.408	0.176
Math to PSSA Reading	0.202	0.356	0.154
Math to PSSA Math	0.305	0.542	0.237
Math to SAT Verbal	0.212	0.359	0.147
Math to SAT Math	0.313	0.546	0.233

Comparison of the correlations in Table 52 to Table 10 can also be used to test a hypothesis about the relationship between PSSA and SAT performance. It is suggested above that some schools might be working more successfully toward improving PSSA performance than others. Again, if these efforts are only tangentially affecting SAT scores, then we might expect mean differences among schools on PSSA scores that are not mirrored by differences on SAT. Within schools, however, we can still expect students who do better on PSSA to also have a tendency to do better on SAT. Figure 20 provides an extreme illustration of this potential effect. Comparing correlations between PSSA and SAT from Table 52 to Table 10 reveals no such trend (see Table 54). The differences are all small (-0.020 to +0.009), suggesting that there are no significant differences among schools in the average PSSA scores of their students that are not mirrored by differences in students' SAT scores. This conclusion is consistent with the earlier conclusion from the individual level data. That is, no strong methodological differences in skills requirements are evidenced.

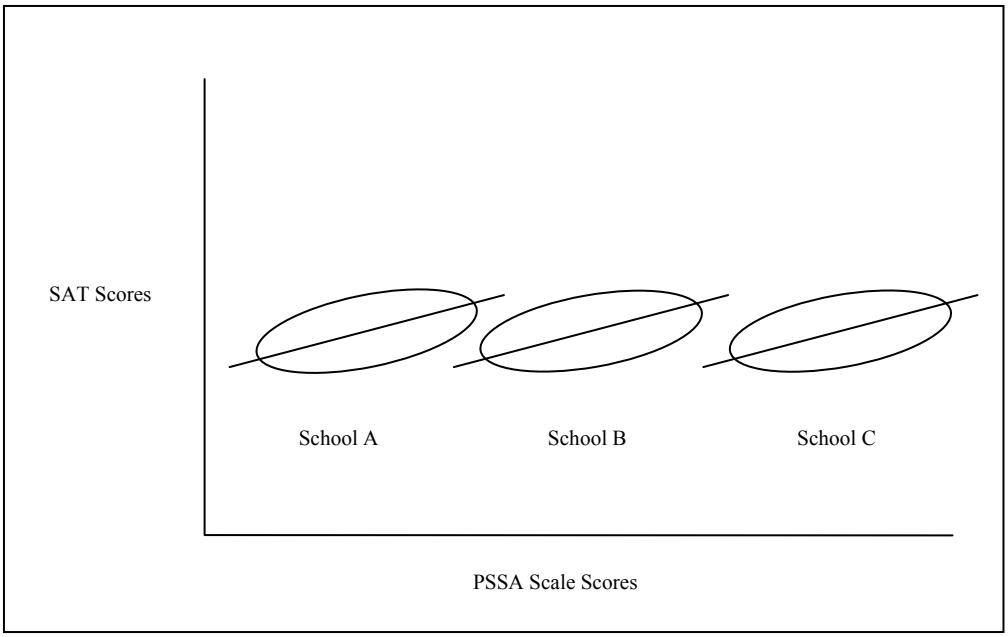

Figure 20. Extreme hypothetical illustration of how mean differences among schools on PSSA scores that are not mirrored by differences on SAT scores can impact correlations. (Ellipses represent students' scores within each school.)

Table 54. Differences Between SAT and PSSA Components for Student-Level Correlations and Pooled Within-School Correlations Averaged for the Classes of 2002 and 2003

GPA or Course Grade to Assessment	Student-Level Correlations	Pooled, Within-School Correlations	Difference in Correlations
PSSA Reading to SAT Verbal	0.742	0.727	-0.015
PSSA Reading to SAT Math	0.690	0.640	-0.020
PSSA Math to SAT Verbal	0.629	0.620	+0.009
PSSA Math to SAT Math	0.856	0.845	-0.009

A third comparison can also be made with the pooled, within-school correlations. This concerns the intercorrelations within each assessment format. The pooled, within-school PSSA intercorrelation (Table 52) shows a very slight decrease from student-level PSSA correlations (Table 10), while the pooled, within-school SAT intercorrelation (Table 52) shows a larger increase from the student-level SAT intercorrelations (Table 10). It is possible that this is due to an increased homogeneity for the pooled, within-school results. However, Figure 21 presents an exaggerated illustration of another manner in which this difference might be obtained. This pattern would yield higher correlations for all students than for students within a school and a higher correlation of school-level scores (school-level PSSA intercorrelation is 0.866, school-level SAT intercorrelation is 0.942 [both from Table 45]) than for student-level scores (student-level PSSA is 0.686, student-level SAT intercorrelation is 0.737 [both from Table 10]). Although small (see Table 55), this is the pattern of differences in correlations obtained for both PSSA intercorrelations and for SAT intercorrelations. Note that the difference between the PSSA components is smaller than the difference between the SAT components. The obtained pattern,

like that in Figure 21 but much less exaggerated, could result from differences between schools in the population of their students, from differences in instruction that impact PSSA and SAT test scores, or both.

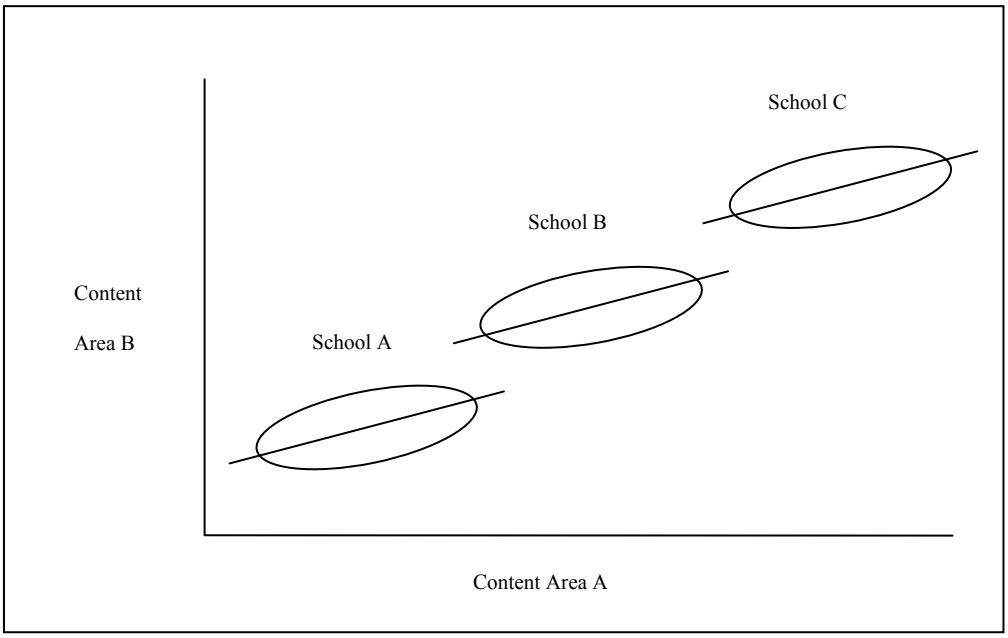

Figure 21. Exaggerated illustration of how mean differences among schools can increase overall correlations. (Ellipses represent students' scores within each school.)

Table 55. Internal Differences between SAT and PSSA Components for Student-Level Correlations and Pooled, Within-School Correlations Averaged for the Classes of 2002 and 2003

GPA or Course Grade to Assessment	Student-Level Correlations	Pooled, With-School Correlations	Difference in Correlations
PSSA Reading to PSSA Math	0.686	0.676	-0.010
SAT Verbal to SAT Math	0.737	0.845	0.108

Summary and Conclusion

The purpose of this report is to explore the validity questions: "Are PSSA scores appropriately related to other measures of educational achievement?"

The data presented in this report suggest the following main points:

- Students who do well on any of the assessments tend to do well on all of the assessments—PSSA, SAT, GPA, or course grades.
- When schools' means are calculated using only the SAT-taking population of students, schools with high scores on SAT also have high PSSA scores. GPA and course grades are also related to both PSSA and SAT although not as strongly as the relationship between the PSSA and SAT.

PSSA exhibits strong convergent validity coefficients. Correlations are very high between PSSA and SAT. PSSA is positively correlated with other measures of student achievement including course grades and GPA. Gains on PSSA are reflective of changes on SAT at the school level. All these data provide strong evidence in support of PSSA as a valid measure of student achievement.

There are, however, a few clarifications and qualifications to these general conclusions, but none that diminish the basic findings. These clarifications and qualifications include the following:

Changes in scores. Both mean PSSA and SAT scores experienced an increase. Changes between the two assessments are positively correlated. Schools that have gained on one assessment tend to have also gained on the other assessment. However, the same relationship is also true for those schools that have decreased.

PSSA and SAT are tied to different content domains, use differently formatted items, and were designed to serve very different purposes. However, it is clear from the data that students who tend to perform well on the PSSA can also be expected to perform well on the SAT and vice-versa. It is clear from the data that schools that perform well on one test can be expected to perform well on the other. It is also clear that schools that improve on one test can be expected to improve on the other.

Discriminant validity. At the student level and the school level, the different assessments of mathematics achievement are more highly related to each other than to assessments of other subjects. Students with high mathematics scores on one assessment will tend to do well on all other assessments, but that tendency is most pronounced for other mathematics assessments. The same holds for school scores. There is a similar differentiation on reading/verbal assessments at both the student and school level, but not as pronounced as in mathematics.

GPA shows positive relationships with both PSSA and SAT assessments at the student and school level, but not as pronounced as the relationship between the two assessments. This is interpreted as being due to differences in schools' grading practices. Grading practices apparently differ sufficiently to reduce school-level associations with PSSA and SAT scores.

Neither gender, race, socioeconomic status, nor limited English proficiency appear to influence the PSSA scores any more than would be expected based on observed differences for SAT scores. In other words, PSSA items are not injecting any unexpected gender, racial/ethnic, socioeconomic status, or limited English proficiency bias.

References

Camara, W.J. & Schmidt, A.E. (1999). *Group differences in standardized testing and social stratification*. (College Board Report No. 99-5). New York: College Entrance Examination Board.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56(2), 81-105.*

Cassady, J.C. (2001). Self-reported GPA and SAT: A Methodological Note. *Practical Assessment, Research & Evaluation, 7(12).*

Coe, Robert. (2002). *It's the effect size, stupid: What effect size is and why it's important.* Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England. Retrieved from http://www.leeds.ac.uk/educol/documents/00002182.htm

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed), Hillsdale, NJ: Erlbaum.

Hoffman, R. G., (1998). *Relationships among Kentucky's open-response assessments, ACT scores, and students' self-reported high school grades.* (HumRRO Report FR-WATSD-98-27), Radcliff, KY: Human Resources Research Organization.

Lewis, D.M., Mitzel, H.C., & Green, D.R. (1996). Standard setting: A bookmark approach. Symposium presented at the Council of Chief State School Officers Conference on Large –Scale Assessment, Phoenix, AZ.

Livingston, S.A. & Zieky, M.J. (1978). *Manual for setting standards on the Basic Skills Assessment Test*. Princeton, NJ: Educational Testing Service.

Mead, R. & Melby, K. (2003, February). *Technical analysis: Pennsylvania System of School Assessment, 2003 reading and mathematics PSSA*. Data Recognition Corporation.

Mead, R. & Melby, K. (2002, April). *Technical analysis: Pennsylvania System of School Assessment, 2001 reading and mathematics PSSA.* Data Recognition Corporation.

Mead, R., Smith, R.M., & Swanlund, A. (2003). *Technical analysis: 2003 Pennsylvania System of School Assessment, mathematics and reading.* Data Recognition Corporation.

NCLB. (2004). http://www.ed.gov/nclb/landing.jtml.

Sicoly, Fiore (2002). What do school-level scores from large-scale assessments really measure? *Educational Measurement: Issues and Practice, 21(4), 17-26.*

Terenzini, P.T., Cabrera, A.F., & Bernal, E.M. (2001). *Swimming against the tide: The poor in American higher education.* (Research Report No. 2001-1). New York: College Entrance Examination Board.

Willingham, W.W., & Cole, N.S. (1997). *Gender and Fair Assessment*. Hillsdale, NJ: Erlbaum.