HumRRO Report No.
FR-05-55

HumRRO
Human Resources Research Organization

Relationships Among Pennsylvania System of School Assessment (PSSA) Scores, University Proficiency Exam Scores, and College Course Grades in English and Math

Draft until released by the Pennsylvania Board of Education.

Andrea L. Sinclair
Arthur A. Thacker

Human Resources Research Organization (HumRRO)
10503 Timberwood Circle, Suite 101
Louisville, KY 40223
Phone (502) 339-9331
FAX (502) 339-9432

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

September, 2005

Abstract

As part of Pennsylvania's examination of the Pennsylvania System of School Assessment (PSSA), the Human Resources Research Organization (HumRRO) investigated whether PSSA scores are appropriately related to other measures of educational achievement. This report represents one part of that larger effort. For this report, HumRRO matched students' 11th grade data from the PSSA with data from three public universities' proficiency exams in Pennsylvania. These universities' proficiency exams are administered to incoming freshmen. Proficiency scores are then used to make English and math course recommendations. Scores were correlated between PSSA and proficiency exams in order to compute convergent validity coefficients. All comparison tests were moderately to highly correlated with PSSA (r = 0.40 to 0.75 for math and r = 0.59 to 0.64 for English multiple choice). The results from this study indicate that students who tend to perform well on the PSSA can also be expected to perform well on the university proficiency exams and vice-versa. These results further support the growing body of validity evidence for the PSSA (see also Koger, Thacker, & Dickinson, 2004; Thacker, Dickinson, & Koger, 2004).

We also examined the predictive validity of the PSSA and the proficiency exams by investigating their relationships with students' grade point averages (GPAs). Students' course GPAs in their first college English and math class generally show positive relationships with both the PSSA and the university proficiency exams, but these correlations are not as strong as the correlations between the PSSA and the proficiency exams. These weaker correlations are interpreted as being partially due to differences in teachers' grading practices, teaching styles and differences in course difficulty, and suggest that student effort may play an important role in predicting course GPA. It is important to note that the correlations between PSSA and course GPA were similar in magnitude to the correlations between the university proficiency exams and course GPA, thereby indicating that neither assessment is substantially better than the other at predicting course GPA.

RELATIONSHIPS AMONG PENNSYLVANIA SYSTEM OF SCHOOL ASSESSMENT (PSSA) SCORES, UNIVERSITY PROFICIENCY EXAM SCORES, AND COLLEGE COURSE GRADES IN ENGLISH AND MATH

TABLE OF CONTENTS

INTRODUCTION **1**
Description of Proficiency Exams 3
University #1 Proficiency Exams 3
University #2 Proficiency Exams 3
University #3 Proficiency Exams 4
Description of Data 4
PSSA Data 4
University #1 Data 6
University #2 Data 7
University #3 Data 7
Matching Data 8
Relationships Among Measures of Student-Level Performance 11
Convergent Validity 12
Predictive Validities 15
Illustration of Relationships 19
University #1 Illustrative Relationship with PSSA 20
University #2 Illustrative Relationship with PSSA 23
University #3 Illustrative Relationship with PSSA 26
Summary and Conclusion 31

List of Tables

Table 1. Descriptive Statistics from PSSA 2001 - 2003 5
Table 2. Summary Descriptive Statistics for University #1 6
Table 3. Summary Descriptive Statistics for University #2 7
Table 4. Summary Descriptive Statistics for University #3 7
Table 5. Percentage of Students Retained in Merged Files 8
Table 6. University #1 Descriptive Statistics for Matched, Unmatched, and Total Students 9
Table 7. University #2 Descriptive Statistics for Matched, Unmatched, and Total Students 10
Table 8. University #3 Descriptive Statistics for Matched, Unmatched, and Total Students 11
Table 9. Correlations for University #1 13
Table 10. Correlations for University #2 14
Table 11. Correlations for University #3 15
Table 12. English/Reading Predictive Validities for University #1 16
Table 13. Math Predictive Validities for University #1 17
Table 14. English/Reading Predictive Validities for University #2 18
Table 15. Math Predictive Validities for University #2 19
Table 16. University #1 PSSA Cut Score by College English Course 21
Table 17. University #1 PSSA Cut Score by College Math Course 22

Table 18. University #2 PSSA Cut Score by College Math Course 25
Table 19. University #3 PSSA Cut Score by College English Course 27
Table 20. University #3 PSSA Cut Score by College Math Course 29

List of Figures

Figure 1. Relationship between University #1 English Scores and PSSA Reading Scores. 20
Figure 2. Relationship between University #1 Math Scores and PSSA Math Scores. 22
Figure 3. Relationship between University #2 English Scores and PSSA Reading Scores. 24
Figure 4. Relationship between University #2 Math Scores and PSSA Math Scores. 25
Figure 5. Relationship between University #3 Writing Scores and PSSA Reading Scores. 26
Figure 6. Relationship between University #3 Algebra/Trig Scores and PSSA Math Scores. 28
Figure 7. Relationship between University #3 Calculus Scores and PSSA Math Scores. 30

RELATIONSHIPS AMONG PENNSYLVANIA SYSTEM OF SCHOOL ASSESSMENT (PSSA) SCORES, UNIVERSITY PROFICIENCY EXAM SCORES, AND COLLEGE COURSE GRADES IN ENGLISH AND MATH

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 content coverage of the PSSA and allow for better diagnostic data to be produced at school and district levels.

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 a 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 high reliability

estimates. Consequently, 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? 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. Prior investigations demonstrate convergent validity evidence for PSSA when correlated with commercially available norm-referenced and criterion-referenced assessments (Koger, Thacker & Dickinson, 2004; Thacker, Dickinson, & Koger, 2004). Same-subject correlations were highest for math, typically ranging from about .70 to about .90. For reading correlations were also quite high, although slightly weaker than for math, ranging from about 0.60 to about 0.80. Moreover, Koger et al. (2004) found that correlations are very high between PSSA and SAT (r = 0.78 for reading/verbal, and r = 0.87 for Math in 2003). They also found that PSSA is positively correlated with students' course grades and grade point average (GPA), although not as highly as with SAT (r = 0.46 to r = 0.55).

The purpose of this report is to provide additional convergent validity evidence for PSSA by correlating PSSA with common subjects on proficiency exams at three Pennsylvania universities and with students' GPA in their first college English and math course. High school graduates accepted for admission to these colleges take proficiency exams in English/reading and math. The results from these tests are used to determine students' appropriate starting levels in English and math course sequences (e.g., remedial level, regular beginning level, or advanced level courses).

For Pennsylvania students who were 11th graders in 2001 - 2003, we will correlate their 11th grade PSSA reading and math scores with their scores on the university proficiency exams. Because PSSA and university proficiency exams report scores for common subject areas (i.e., math and reading/English), we expect student scores on the various tests to be related. However, there are three main reasons why the strength of the correlations between the two types of assessments may be attenuated. First, PSSA tests student ability related to content that is specific to the state of Pennsylvania. The academic standards outline the content that Pennsylvania has collectively decided is essential for students in its public schools. The extent to which the universities' proficiency exams measure content that is different from PSSA's will limit the strength of the correlations between the assessments. Second, students attending college may represent a high performing subset of all students taking PSSA; therefore, the full range of PSSA scores may not be represented by this subset. Finally, differences in the format of the test items (e.g., multiple-choice vs. performance-task items) can also weaken the correlations between the two types assessments. If the correlations between the two types of assessments are very high, it will raise questions as to whether the assessments are measuring anything different, and consequently whether the universities' proficiency exams are necessary. If the correlations are very low, one might question whether the two types of assessment are measuring the same general construct at all (math or reading/English ability). Consequently, we are looking for what Hoffman (1998) refers to as "Goldilocks" correlations; that is, correlations between PSSA and the university proficiency exams should fall in the not-too-high and not-too-low category.

This report also compares the predictive validity of the PSSA and the proficiency exams in terms of their ability to predict college academic success. For the purpose of this report, college academic success is defined as college freshmen's GPA in their first English and math course. Because the two types of assessments are expected to be positively correlated, we also expect both assessments to explain some variance in college performance.

Description of Proficiency Exams

University #1 Proficiency Exams

University #1 proficiency exams were developed by the university's academic departments in cooperation with University Testing Services. The English proficiency exam measures students' competence in spelling, vocabulary, punctuation, grammar, and diction. Students are asked to identify misspelled words, synonyms, antonyms, and inaccurate punctuation and to evaluate the correctness of the grammar and diction of parts of sentences. There is no essay component to the English proficiency exam at University #1. The math proficiency exam measures students' knowledge of various algebra and trigonometry topics. This test is particularly important for students entering programs of study that require calculus, since a low score on this test indicates that a student is not ready for calculus and will be required to schedule courses in algebra and/or trigonometry. The math proficiency exam is targeted at assessing four areas: (1) basic math, (2) algebra, (3) business calculus, and (4) calculus.

University #2 Proficiency Exams

The English proficiency exam for University #2 consists of an essay component and a multiple choice component. For the essay, students are asked to write on one of two topics which are among four topics students receive prior to the exam. Before taking the exam, students are asked to review sample essays, to give some thought to each of the four potential topics provided, to make notes on the topic, and to bring the notes with them when they come to take the exam. On the day of the exam, the proctor designates two of the four topics from which the student may choose. Each essay topic quotes a writer's position on an issue. The exam instructs students to explain the writer's position, and then to express his/her own position about what the writer has said. Essays are read by at least two university instructors. To place the student properly in the First-Year Writing Program, graders need to see how well the student has read the passage, explained the writer's argument, and presented an argument in an organized way with relatively few grammatical errors. The multiple choice component of the English proficiency exam tests Sentence Structure skills, Conventions of Written English, and Reading Comprehension. The total multiple choice score is used in conjunction with the student's SAT verbal score and high school percentile to develop a base score. The essay score and the base score are then used to determine the English verdict, which designates the appropriate English course placement for the student. This study is interested in the validity of the proficiency exams only; consequently, students' SAT verbal scores and high school percentile ranks are not taken into consideration for the purposes of this study.

The math proficiency exam for University #2 is a three-part test. The first part measures elementary computational skills such as addition, subtraction, multiplication and division of whole numbers; decimals and fractions; and ratios. The second part consists of algebra problems, and the third part measures calculus readiness. Math course recommendations are based on students' scores on each of these three parts of the math proficiency exam.

University #3 Proficiency Exams

University #3 has a writing proficiency exam that is designed to measure students' ability to read and understand college-level material and to write a thoughtful, coherent response. Students are asked to respond to a single essay prompt. They read a short passage and write an essay in which they explain what the writer of the passage is saying about a particular issue and respond to that writer's view and the issue in terms of their own knowledge and experience. There is no multiple choice component to University #3's writing exam. The math proficiency exam consists of an algebra proficiency exam, a trigonometry proficiency exam, and a calculus proficiency exam. The algebra proficiency exam is a five-part exam designed to determine whether students have acquired the algebraic skills sufficient for a college-level quantitative course. The five-part exam covers problems ranging from elementary to advanced algebra, and covers the following material: algebra of rational numbers, operations with algebraic expressions, linear equations and inequalities, factoring and algebraic functions, exponents and radicals, graphing and distance, fractional and quadratic equations and inequalities, functions and their graphs, complex numbers, absolute values, and systems of equations and exponential and logarithmic functions. The trigonometry exam tests definitions of trigonometric functions, right angles, evaluation of special angles, related angles, radian measure, graphing, identities, laws of sines and cosines, trigonometric equations, and arc functions. Lastly, the calculus exam tests differential and integral calculus of algebraic and trigonometric functions, applications of the derived trigonometric functions and of the derivation related area, max-min problems, applications to graphing, and application of the integral to problems involving area, volumes and revolution, arc, length, and work.

Description of Data

PSSA 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 math as well as demographic information. 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 the university proficiency exams. Only 11th grade PSSA scores were used for the purposes of this report. Once matching was complete, student-level identifiers 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.

Table 1 presents a summary of the data included in the PSSA files. Statistics represent scale scores for reading and math 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. Descriptive Statistics from PSSA 2001 - 2003

		Math (Non-Missing)	Reading (Non-Missing)
PSSA 2001 Grade 11 (Cases = 122,332)	N	116,137	115,566
	Mean	1,304.11	1,296.87
	S.D.	220.57	217.57
	Minimum	100	100
	Maximum	2,481	2,088
PSSA 2002 Grade 11 (Cases =129,475)	N	123,550	123,234
	Mean	1,314.47	1,312.86
	S.D.	228.29	215.13
	Minimum	700	700
	Maximum	2,437	2,362
PSSA 2003 Grade 11 (Cases =133,168)	N	126,941	127,427
	Mean	1,314.78	1,315.61
	S.D.	214.78	235.39
	Minimum	700	700
	Maximum	2,238	2,355

University #1 Data

University #1 provided student-level data from the summer 1999 – fall 2004 administrations of the proficiency exams for all students who reported Pennsylvania home addresses. Table 2 contains descriptive statistics for the proficiency data. The scores on the proficiency exams represent raw scores. The initial database contained 46,224 cases. The database contained data for students with birth dates before 1983 and for students with missing birthdates. Because students with birth dates before 1983 were extremely unlikely to be 11th graders in 2001, 2002, or 2003, we deleted those cases[1], as well as those with missing birthdates. This same clean-up process was utilized for all three universities. After eliminating these, we were left with 28,546 cases. Not all of those students had data in the score fields, so n-counts are smaller than the number of cases.[2]

Table 2. Summary Descriptive Statistics for University #1

University #1 (cases = 28,546)	English	Basic Math	Algebra	Business Calculus	Calculus
N (non-missing)	27,510	28,530	28,526	28,514	28,513
Mean	25.43	13.40	12.80	10.30	12.48
S.D.	9.69	3.17	4.33	7.01	8.07
Minimum	0	0	0	0	0
Maximum	59	17	19	26	34

[1] By deleting students with birthdates prior to 1983, this also resulted in the deletion of 1999, 2000 and most of the 2001 proficiency exam data; students who took the college proficiency exams in these years would not be expected to be 11th graders in 2001 – 2003.

[2] Note that for all of the university databases, the number of students in the university databases is considerably less than the PSSA database. This is because only students admitted to the universities take the proficiency exams. Students with no plans for continuing their education past high school and students attending other universities or vocational institutions do not take the proficiency exams for that particular university.

University #2 Data

University #2 provided student-level data from the fall 2003 and fall 2004 administrations of the proficiency exams. Table 3 contains descriptive statistics for the proficiency data. The scores on the proficiency exam represent raw scores. The initial database contained proficiency exam data for 7,677 students. After cleaning the database in the same manner as described for University #1, there were 7,606. As we expected, a smaller percentage of students were deleted from the initial database provided by University #2 given that proficiency exam data was not provided for non-relevant years (i.e., 1999 – 2001).

Table 3. Summary Descriptive Statistics for University #2

University #2 (cases = 7,606)	English - Essay	English – Multiple Choice	Elementary Math	Algebra	Calculus
N (non-missing)	7,554	7,552	7,564	7,564	7, 564
Mean	6.96	86.61	15.84	11.10	2.63
S.D.	1.40	13.11	3.38	4.18	1.67
Minimum	2	22	3	0	0
Maximum	12	114	21	22	7

University #3 Data

University #3 provided proficiency exam data for all of the Pennsylvania-resident freshmen who took the tests from fall 1999 to fall 2004. Table 4 contains descriptive statistics for the proficiency data. As with the other two universities, the proficiency exam scores represent raw scores. University #3 provided individual databases for the writing exam, the algebra/trigonometry exam, and the calculus exam. The initial database for the writing exam contained 10,158 cases. After cleaning the database, 5,576 cases remained in the writing database. For the algebra/trigonometry database, the initial database contained 10,359 cases and the cleaned database contained 5,591 cases. The initial calculus database contained 1,034 cases and the cleaned database contained 651 cases.

Table 4. Summary Descriptive Statistics for University #3

University #3	Writing (cases = 5,576)	Algebra (cases = 5,591)	Trigonometry (cases = 5,591)	Calculus (cases = 651)
N (non-missing)	5,576	3,990	2,538	651
Mean	2.85	29.06	9.19	9.47
S.D.	0.41	7.05	3.44	3.96
Minimum	1	0	0	0
Maximum	5*	72	18	19

Note. *Out of 5,576 students, only two students received a score of "5" on the writing exam. There were no students who received a score of "4."

Matching Data

All comparison test data had to be matched to PSSA data in order to perform correlations and other calculations. The PSSA data file was merged with each university's data file(s). Students were matched on student-level indicators provided in each 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. Four merge attempts were made for each working file, and used less stringent merging criteria than the previous attempt. 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. Table 5 presents the proportion of students retained in each of the university data files. Approximately, 50% – 70% of the cases were retained across the universities.

Table 5. Percentage of Students Retained in Merged Files

Merges	No. Students in University File	Percent of University File
University #1		
Beginning Data File	28,546	100%
Final Data File with Successfully merged PSSA data	16,824	**58.94%**
University #2		
Beginning Data File	7, 606	100%
Final Data File with Successfully merged PSSA data	3,890	**51.14%**
University #3		
Beginning Data File – Writing	5,576	100%
Final Data File with Successfully merged PSSA data – Writing	3,451	**61.89%**
Beginning Data File – Algebra	5,591	100%
Final Data File with Successfully merged PSSA data – Algebra	3,990	**71.36%**
Beginning Data File – Trigonometry	2,538	100%
Final Data File with Successfully merged PSSA data – Trigonometry	1,255	**49.44%**
Beginning Data File – Calculus	651	100%
Final Data File with Successfully merged PSSA data – Calculus	427	**65.59%**

An additional analysis was conducted to verify that students' scores retained in the final data set did not meaningfully differ from those whose data failed to merge. Table 6, Table 7, and Table 8 present the means, standard deviations, and numbers of cases, for matched, unmatched and total data for Universities #1, #2, and #3, respectively. The means for matched students were generally never more than approximately 1 raw score point higher than the means for unmatched students, nor more than 0.50 raw points higher than the means for the total sample. The finding of higher means for the matched students is typical in investigations of this type (e.g., Koger, Thacker, & Dickinson, 2004).

Table 6. University #1 Descriptive Statistics for Matched, Unmatched, and Total Students

		University #1		
Subject		Matched	Unmatched	Total
English	Mean	25.85	25.74	25.80
	S.D.	9.67	9.71	9.69
	N*	16,215	11,312	27,510
Basic Math	Mean	13.57	13.15	13.40
	S.D.	3.02	3.34	3.17
	N	16,815	11,733	28,530
Algebra	Mean	13.00	12.52	12.80
	S.D.	4.26	4.42	4.33
	N	16,811	11,733	28,526
Business Calculus	Mean	10.66	9.78	10.30
	S.D.	7.07	6.90	7.01
	N	16,802	11,730	28,514
Calculus	Mean	12.90	11.89	12.48
	S.D.	8.16	7.90	8.07
	N	16,801	11,730	28,513

Note. Due to a few non-valid merges, the numbers of matched and unmatched cases do not exactly equal the number of total students. This is not surprising given that we did not have a unique identifier on which to match students.

* N = non-missing data.

Table 7. University #2 Descriptive Statistics for Matched, Unmatched, and Total Students

		University #2		
Subject		Matched	Unmatched	Total
English – Essay	Mean	6.97	6.96	6.96
	S.D.	1.41	1.40	1.40
	N*	3,875	3,682	7,554
English – Multiple Choice	Mean	86.72	86.50	86.61
	S.D.	13.12	13.09	13.11
	N	3,875	3,680	7,552
Elementary Math	Mean	15.86	15.82	15.84
	S.D.	3.43	3.33	3.38
	N	3,870	3,697	7,564
Algebra	Mean	11.29	10.90	11.10
	S.D.	4.25	4.10	4.18
	N	3,870	3,697	7,564
Calculus	Mean	2.70	2.57	2.63
	S.D.	1.70	1.64	1.67
	N	3,870	3,697	7,564

Note. Due to a few non-valid merges, the numbers of matched and unmatched cases do not exactly equal the number of total students. This is not surprising given that we did not have a unique identifier on which to match students.

* N = non-missing data.

Table 8. University #3 Descriptive Statistics for Matched, Unmatched, and Total Students

		University #3		
Subject		Matched	Unmatched	Total
Writing	Mean	2.84	2.87	2.85
	S.D.	0.40	0.42	0.41
	N*	3,451	2,128	5,576
Algebra	Mean	29.66	28.54	29.06
	S.D.	6.99	7.06	7.05
	N	1,838	2,153	3,990
Trigonometry	Mean	9.35	9.03	9.19
	S.D.	3.31	3.55	3.44
	N	1,255	1,284	2,538
Calculus	Mean	9.74	8.96	9.47
	S.D.	3.93	3.97	3.96
	N	427	224	651

Note. Due to a few non-valid merges, the numbers of matched and unmatched cases do not exactly equal the number of total students. This is not surprising given that we did not have a unique identifier on which to match students.
* N = non-missing data.

Relationships Among Measures of Student-Level Performance

This section presents relationships among PSSA and other measures of student performance by examining correlations among the various measures. The other measures of student performance investigated in this study are: (1) students' scores on the university proficiency exams, and (2) students' course GPAs in their first college English course and their first college math course.

Correlations are more informative than are the relative means of the different assessments. We cannot simply compare the PSSA means with the university means because each is on a different scale. The PSSA is on a transformed scale and the university proficiency exams use raw scores. Just as water's freezing point is either 0 or 32 degrees depending on whether one uses the Celsius or Fahrenheit temperature scale, student proficiency could be represented by essentially any number one chooses to define. It is true that water freezes at different numbers of degrees depending on the scale we choose, but if several temperatures are measured on each scale and correlated together, the correlation would be perfect. Similarly, if PSSA and the proficiency exams and course GPA's measure a particular content construct in much the same way, then we can expect a high correlation even though the tests are on different scales. If the correlation is low, we can surmise that the construct the tests are measuring is different. The correlation coefficient is always between -1 and +1, the closer the correlation is to +/-1, the closer to a perfect linear relationship. However, correlations are never perfect due to measurement error inherent in all measures. As a general rule of thumb, correlations less than .30

are generally considered weak, correlations .40 to .60 are generally considered moderate, and correlations above .60 are generally considered moderately strong to strong.

Convergent Validity

Pearson correlations[3] were calculated for the merged files. 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.

University #1 Correlations. The correlations for PSSA, the proficiency exams, and course GPAs for University #1 are shown in Table 9. Not surprisingly, the correlations between the four math scores on the proficiency exam are quite strong, particularly among algebra, business math and calculus (r = 0.76 to 0.94). The basic math score tended to have somewhat weaker correlations with the business math and calculus scores (r = 0.59 and 0.56, respectively). Interestingly, the correlations between the English exam and the math exams were similar in magnitude to the basic math—business math and basic math—calculus correlations (r = 0.54 to 0.53). The correlation between PSSA reading and math (r = 0.63) was slightly stronger than were the correlations between the English and math proficiency exams for University #1.

The convergent validity coefficient between PSSA reading and the English proficiency exam was moderately strong (r = 0.59). The convergent validity coefficients between PSSA math and the math proficiency exam tended to be stronger; this is a common finding in research of this type (e.g., Bacci, Koger, Hoffman, & Thacker, 2003; Sinclair & Thacker, 2004). The strongest math-to-math correlation was between PSSA math and the algebra proficiency exam (r = 0.75), and the weakest was between PSSA math and the basic math proficiency exam (r = 0.68), which was still quite strong. Finally, as expected, the correlations between different content areas on the different assessments were among the lowest (r = 0.45 to 0.57).

[3] 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.

Overall, the PSSA correlations and University #1 proficiency exam correlations are all positive and above 0.47. This indicates 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 this report suggest that Pennsylvania students who exhibit high ability in one content area can be expected to perform well in other content areas.

Correlations with course GPAs were considerably lower than the correlations between the PSSA and the proficiency exams. However, the correlations between course GPA and PSSA (r = 0.29 for reading/English and r = 0.27 for math) were similar to the correlations between course GPA and the proficiency exams (r = 0.26 for English and r = 0.19 for math).

Table 9. Correlations for University #1

Variable		PSSA		University #1					Course GPA	
		Read	Math	Engl.	Basic	Alge.	Buss.	Calc.	Engl.	Math
PSSA	Read	1.00								
	Math	*0.63*	1.00							
Univ. # 1	Engl.	**0.59**	**0.57**	1.00						
	Basic	**0.48**	**0.68**	*0.48*	1.00					
	Alge.	**0.51**	**0.75**	*0.53*	*0.80*	1.00				
	Buss.	**0.47**	**0.74**	*0.54*	*0.59*	*0.80*	1.00			
	Calc.	**0.45**	**0.73**	*0.53*	*0.56*	*0.76*	*0.94*	1.00		
Course GPA	Engl.	**0.29**	**0.25**	**0.26**	**0.23**	**0.27**	**0.25**	**0.24**	1.00	
	Math	**0.20**	**0.27**	**0.19**	**0.19**	**0.27**	**0.27**	**0.26**	*0.39*	1.00

University #2 Correlations. The correlations for PSSA, the proficiency exams, and course GPAs for University #2 are shown in Table 10. The correlations among the three math components of the math proficiency exam are among the highest (r = 0.47 to 0.65). The correlation between reading and math on the PSSA (r = 0.45) was stronger than the correlations between the English components (particularly the English essay) and the math components of the proficiency exams (r = 0.07 to 0.24).

The convergent validity coefficients between PSSA reading and the components of the English proficiency exam varied somewhat. The strongest correlation occurred between PSSA reading and the multiple choice component of the English proficiency exam (r = 0.64). The correlation between PSSA reading and the essay component of the proficiency exam was weaker (r = 0.37). It is not surprising that correlations with the essay component were weaker than correlations with other components of the proficiency exam. The multiple choice component of the English exam and the three multiple choice components of the math exam are likely measured on an interval scale where the difference between a score of 30 and 40, for example, is

essentially the same as the difference between a score of 50 and 60. The essay component is more likely to be measured on an ordinal scale where a score of "4" is better than a score "3," but the difference between a score of "3" and a score of "4" is not necessarily the same as the difference between a score of "5" and a score of "6." As a result, correlations with the essay component of the English exam are likely to be lower. Lastly, the convergent validity coefficients between PSSA math and the math proficiency exams were moderate (r = 0.51 to 0.58). Finally, as expected, the correlations between different content areas on the different assessments were among the lowest (r = 0.16 to 0.40).

As seen with University #1, correlations with course GPAs were considerably lower than the correlations between PSSA and the proficiency exams. However, it is important to note that the correlations between course GPA and PSSA (r = 0.17 for reading/English and r = 0.20 for math) were similar to the correlations between course GPA and the proficiency exams (r = 0.08 and 0.12 for English and r = 0.10 to 0.15 for math).

Table 10. Correlations for University #2

Variable		PSSA		University #2					Course GPA	
		Read	Math	Engl. Essay	Engl. M.C.	Elem. Math	Alge.	Calc.	Engl.	Math
PSSA	Read	1.00								
	Math	*0.45*	1.00							
Univ. #2	Engl. Essay	**0.37**	**0.19**	1.00						
	Engl. M.C.	**0.64**	**0.40**	*0.44*	1.00					
	Elem. Math	**0.21**	**0.58**	*0.09*	*0.24*	1.00				
	Alge.	**0.16**	**0.56**	*0.07*	*0.19*	*0.65*	1.00			
	Calc.	**0.19**	**0.51**	*0.10*	*0.20*	*0.47*	*0.61*	1.00		
Course GPA	Engl.	**0.17**	**0.08**	**0.08**	**0.12**	**0.02**	**0.04**	**0.05**	1.00	
	Math	**0.11**	**0.20**	**0.08**	**0.12**	**0.15**	**0.11**	**0.10**	*0.30*	1.00

University #3 Correlations. The correlations between PSSA and the proficiency exams for University #3 are shown in Table 11. University #3 did not provide data on students' GPAs in their first English course and first math course; consequently, correlations with these assessments are not available. Also, the students' scores on the proficiency exams for writing, algebra/trigonometry, and calculus were represented in separate databases. Consequently, no intercorrelations among these content areas are presented.

The convergent validity coefficients for PSSA reading and the writing proficiency exam were weak to moderate (r = 0.30). The writing proficiency exam for University #3 is an essay exam; consequently, the same explanation provided above for the weaker correlations with the essay component of University #2's English proficiency exam is likely to apply here. The convergent validity coefficients between PSSA math and the math proficiency exams were stronger with the strongest being between PSSA math and the algebra proficiency exam (r = 0.52).

Table 11. Correlations for University #3

Variable		PSSA		University #3			
		Reading	Math	Writing	Algebra	Trig.	Calculus
PSSA	Reading	1.00					
	Math	*0.43*	1.00				
Univ. # 3	Writing	**0.30**	**0.23**	1.00			
	Algebra	**0.22**	**0.52**	--	1.00		
	Trig.	**0.14**	**0.40**	--	*0.39*	1.00	
	Calculus	**0.17**	**0.42**	--	--	--	1.00

Predictive Validities

Next, we more closely investigated the predictive validity of the PSSA and the proficiency exams for Universities #1 and #2 by examining the correlations between PSSA and GPAs and the correlations between the proficiency exams and GPAs broken down by course level. Students' GPAs were categorized into course levels based on the actual courses taken by the students. Categorization into course level was not based on students' score(s) on the proficiency exams (although in most cases students took the course they were recommended to take). As a result, the students in each course level contain some students who were recommended to take a higher level course based on their proficiency exam score, and some students who were recommended to take a lower level course based on their proficiency exam score. This increases the amount of variance within course level categorizations and therefore should provide conservative estimates of the proficiency exams' predictive validity.

University #1 Predictive Validities. University #1 offers multiple English courses for its freshmen students. We divided the English classes into four levels with higher levels being more advanced than lower levels. The first level represents the Basic Writing Skills course. The second level represents the Writing Tutorial course. The third level represents the Rhetoric and Composition course, and the fourth level represents the Honors Freshmen Composition course. The majority of students in the database fit into one of these categories. A few other English courses are available to freshmen, but only a minority of students took those courses; therefore, they are not included in these analyses.

As displayed in Table 12, the correlations for the two assessments with English course GPA were relatively weak regardless of the level of the course. Because the two assessments yielded similar correlations with English course GPA, this suggests that one assessment is not substantially better than the other at predicting course GPA. Despite these similarities, PSSA did exhibit slightly stronger correlations for three of the four English course categories (r = 0.23 to 0.24 for PSSA vs. r = 0.07 to 0.22 for University #1 proficiency exam). Honors Composition was the only course for which the proficiency exam had a higher correlation than the PSSA reading exam, although this correlation was only 0.10.

Table 12. English/Reading Predictive Validities for University #1

GPAs by Course Category		University #1 English Exam	PSSA Reading
Level 1 Basic Writing Skills	r	0.07	0.23
	N	2,027	2,244
Level 2 Writing Tutorial	r	0.20	0.24
	N	483	502
Level 3 Rhetoric and Comp.	r	0.22	0.24
	N	10,803	10,965
Level 4 Honors Comp.	r	0.10	0.08
	N	1,158	1,152

Table 13 compares the correlations between University #1 scores on the math proficiency exam and math course GPA with the correlations between PSSA math and math course GPA. These correlations are broken down by course level. We divided the math classes into six levels with higher levels being more advanced than lower levels. Level 1 is composed of Developmental Mathematics, Elementary Geometry with Problem Solving, and Basic Skills, Level 2 represents Intermediate Algebra, Level 3 represents College Algebra I, Level 4 is composed of College Algebra II, Analytic Geometry, and Plane Trigonometry, Level 5 is composed of Techniques of Calculus I and Calculus with Analytic Geometry I, and Level 6 is composed of Techniques of Calculus II and Calculus with Analytic Geometry II. There were a few other math courses available, although few students were enrolled in those courses; consequently, they were excluded from these analyses. Overall, the magnitudes of the PSSA correlations with course GPA were similar to the proficiency exam correlations with course

GPA, thereby suggesting that one assessment is not substantially better than the other at predicting math course GPA.

Despite the similarities, there were some subtle differences between the two assessment types. As displayed in Table 13, the strongest correlation occurred between the algebra proficiency exam and GPA in the Level 3 course (College Algebra I) (r = 0.34). The weakest correlation occurred between the business proficiency exam and GPA in the Level 2 course (Intermediate Algebra) for which there was a slight negative relationship (r = -0.02). The business proficiency exam was the best predictor of GPA in the Level 5 courses (Calculus I) (r = 0.33). Interestingly, the algebra proficiency exam was a slightly better predictor of GPA in the Level 6 courses (Calculus II) (r = 0.24) than the calculus proficiency exam (r = 0.22). Level 1 and Level 2 courses tended to exhibit the lowest correlations with the two assessments.

Table 13. Math Predictive Validities for University #1

GPAs by Course Category		University #1 Basic Math	University #1 Algebra	University #1 Business Ma.	University #1 Calculus	PSSA Math
Level 1	r	0.08	0.18	0.08	0.09	0.21
Basic Math	N	398	398	396	395	389
Level 2	r	0.12	0.15	-0.02	-0.01	0.19
Intermediate Alge.	N	850	849	849	850	842
Level 3	r	0.20	0.34	0.28	0.26	0.29
College Algebra I	N	4,516	4,516	4,511	4,511	4,487
Level 4	r	0.06	0.16	0.22	0.16	0.15
College Algebra II	N	2,217	2,216	2,216	2,216	2,200
Level 5	r	0.10	0.25	0.33	0.30	0.24
Calculus I	N	4,042	4,042	4,041	4,041	4,028
Level 6	r	0.14	0.24	0.22	0.22	0.20
Calculus II	N	449	449	448	448	447

University #2 Predictive Validities. University #2 has a less fine-grained distinction among the English courses offered for its freshmen students than University #1. Depending on an incoming student's score on the English proficiency exam, the university recommends that the student take either the pre-college composition course or the standard freshmen composition course. If the student receives a writing verdict of "7," then the student receives an exemption from freshman composition[4].

Overall, the correlations for the two assessments types with English course GPA were relatively weak regardless of the level of the course (see Table 14). This was particularly true for

[4] Recall that "writing verdict" is based on a combination of the essay exam, the multiple choice exam, SAT verbal score, and high school percentile rank. Scores range from 1 – 7.

the essay component of the English proficiency exam, which is not surprising giving the constraints of its ordinal measurement scale. The multiple choice component of the English proficiency exam and the reading portion of the PSSA did a similar job at predicting students' GPAs in the Level 1 and Level 2 courses, although the proficiency exam was a slightly better predictor. Because the two assessments yielded similar correlations with English course GPA, this suggests that one assessment is not substantially better than the other at predicting course GPA.

Table 14. English/Reading Predictive Validities for University #2

GPAs by Course Category		University #2 English Essay	University #2 English M.C.	PSSA Reading
Level 1 Pre-College Composition Course	r	0.04	0.25	0.21
	N	740	740	736
Level 2 Standard Freshman Composition Course	r	0.10	0.24	0.17
	N	1774	1772	1768

Table 15 compares the correlations between University #2 scores on the math proficiency exam and math course GPA with the correlations between PSSA math and math course GPA. These correlations are broken down by course level. We divided the math classes into four levels with higher levels being more advanced than lower levels. Level 1 represents Pre-College Elementary Algebra, Level 2 represents Intermediate Algebra, Level 3 represents Calculus I, and Level 4 represents Calculus II. There were a few other math courses offered, although the vast majority of freshmen took one of these classes.

The strongest correlation occurred between the elementary math component of the university proficiency exam and students' GPAs in the Level 1 math course ($r = 0.42$). This seems logical in that it makes sense that a test of one's elementary math skills should be a reasonably good predictor of one's success in a basic math class. Interestingly, however, the strongest correlation between PSSA math and students' GPAs also occurred for the Level 1 math course ($r = 0.35$). That is, students' math scores on the PSSA were better able to predict students' GPAs in the basic math course than in any of the higher level math courses. Also among the highest correlations was the correlation between the algebra component of the university exam and students' GPAs in the highest level math course (Calculus II). Interestingly, the algebra component of the proficiency exam did a better job of predicting calculus GPA than the calculus component of the proficiency exam. This was true for both Level 3 (Calculus I) and Level 4 (Calculus II) math courses. Lastly, it is of note that the majority of freshmen took Level 2 math courses. PSSA math did a better job at predicting students' GPAs at this level than any component of the university proficiency exam, although this correlation was only weak to moderate ($r = 0.27$).

Table 15. Math Predictive Validities for University #2

GPAs by Course Category		University #2 Elementary Math	University #2 Algebra	University #2 Calculus	PSSA Math
Level 1 Pre-College Algebra	r	0.42	0.30	0.10	0.35
	N	286	286	286	283
Level 2 Intermediate Algebra	r	0.17	0.19	0.15	0.27
	N	1258	1258	1258	1259
Level 3 Calculus I	r	0.08	0.17	0.14	0.08
	N	338	338	338	339
Level 4 Calculus II	r	0.05	0.39	0.20	0.15
	N	171	171	171	171

Illustration of Relationships

Figures 1 to 7 present box plots that illustrate relationships between PSSA scores and university proficiency exam scores. These graphs further illustrate the relationships between university proficiency exams and PSSA in both reading and math. The boxes and whiskers in each plot represent the distribution of university scores versus the varying levels of the corresponding grouping of PSSA scores. Each box represents 50% of the students within each of the university 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 university category. Two sets of cut scores are illustrated in the figures. The grouping of scores on the x-axis corresponds to the cut scores on the proficiency exams for making course recommendations. The second set of cut scores is represented by the dashed horizontal lines that have been added to show the 11th grade PSSA performance levels (i.e., cut scores).[5] Sample sizes within each category along the X-axis are noted.[6] Tables 16 to 20 show the percent of students who scored at or above proficient on the PSSA (cut score of 1290 for reading and 1310 for math), and who took at least the standard level freshman course in English (e.g., Writing Composition) or the standard level freshmen course in math (e.g., College Algebra).

[5] On June 30, 2005 the Pennsylvania State Board of Education adopted a new set of cut scores. However, given that the current data are based on the former PSSA cut scores, those cut scores are provided in this report.

[6] University cut scores on the proficiency exams reflect the decision points for course recommendations; however, not all students followed those recommendations; this accounts for differences in samples sizes in the figures as compared to the sample sizes listed in the previous tables.

University #1 Illustrative Relationship with PSSA

English/Reading. Figure 1 shows the relationship between the PSSA reading score on the y-axis and the University #1 English score on the x-axis. The figure shows a distinct stair-step. Eleventh graders who scored in the Below Basic and Basic performance levels on PSSA reading tended to score in the bottom level of the English proficiency exam; these students were recommended to take the Basic Writing Skills course. The majority of 11th graders who scored in the proficient performance level on PSSA reading scored in Level 2 and Level 3 of the university exam, resulting in a recommendation to take either the Writing Tutorial course or the Rhetoric and Composition course. Finally, the 11th graders who scored in the Advanced performance level on PSSA reading tended to score in the top level on the proficiency exam; these students received recommendations to enroll in the Honors Composition course.

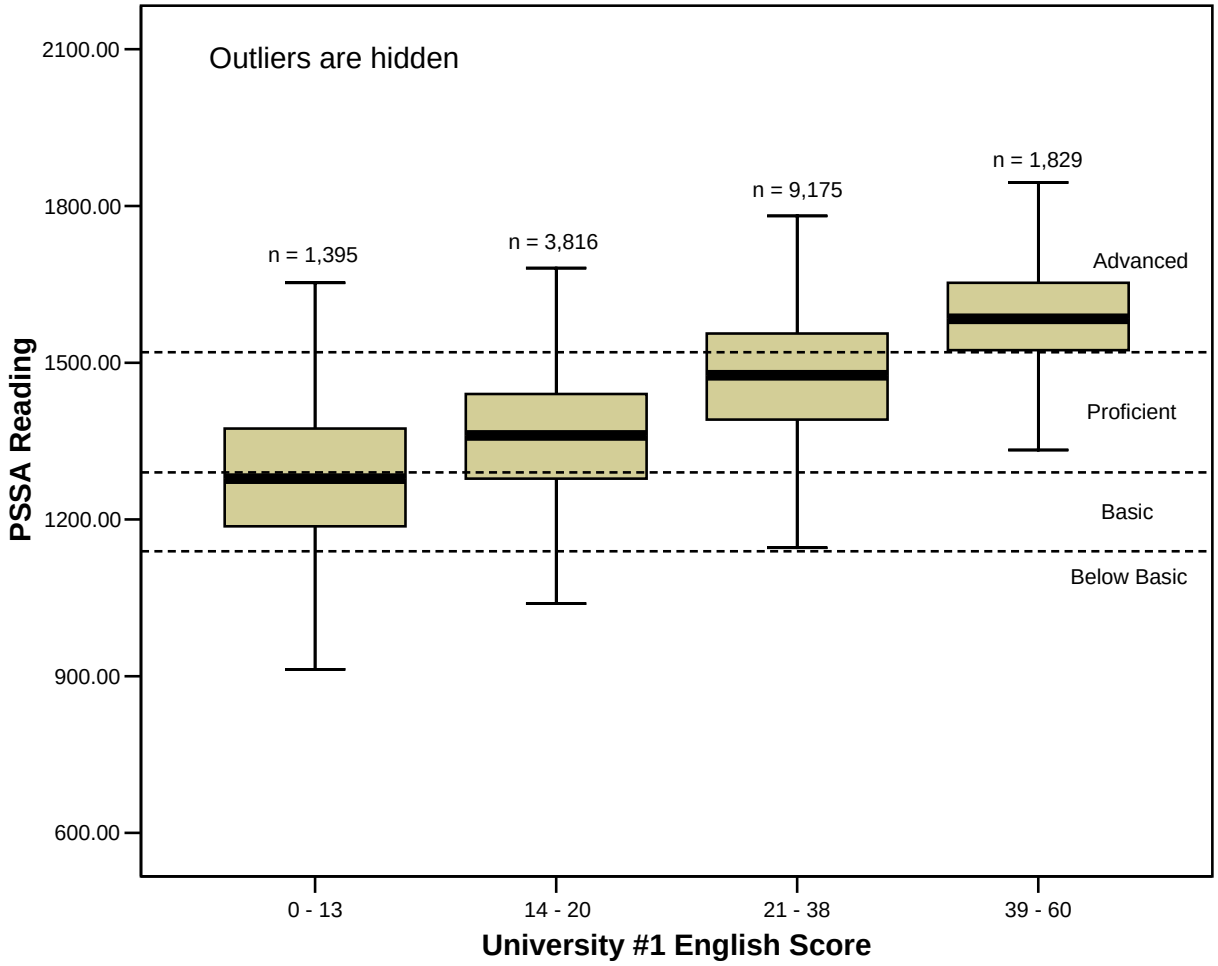

Figure 1. Relationship between University #1 English Scores and PSSA Reading Scores.

Table 16 demonstrates that 90.7% of students who scored at least 1290 (i.e., proficient or above) on their 11th grade PSSA reading assessment took at least the standard level freshmen composition course.

Table 16. University #1 PSSA Cut Score by College English Course

	PSSA English < 1290	PSSA English >= 1290
Basic English	1148 48.7%	1194 9.3%
Standard English and Above	1211 51.3%	11,610 90.7%

Math.

Figure *2* shows the relationship between the PSSA math score on the y-axis and the University #1 math scores on the x-axis. Once again, the figure shows a distinct stair-step. Eleventh graders who scored in the Below Basic and Basic performance levels on PSSA math tended to score in the two lowest levels of the math proficiency exam; these students received recommendations to take beginning level math courses. The majority of eleventh graders who scored in the proficient performance level on PSSA math scored in the middle two levels of the university math exam, and received recommendations to take intermediate level math courses. Finally, the eleventh graders who scored in the Advanced performance level on PSSA math tended to score in the two highest levels of the math proficiency exam; these students received recommendations to enroll in advanced math courses.

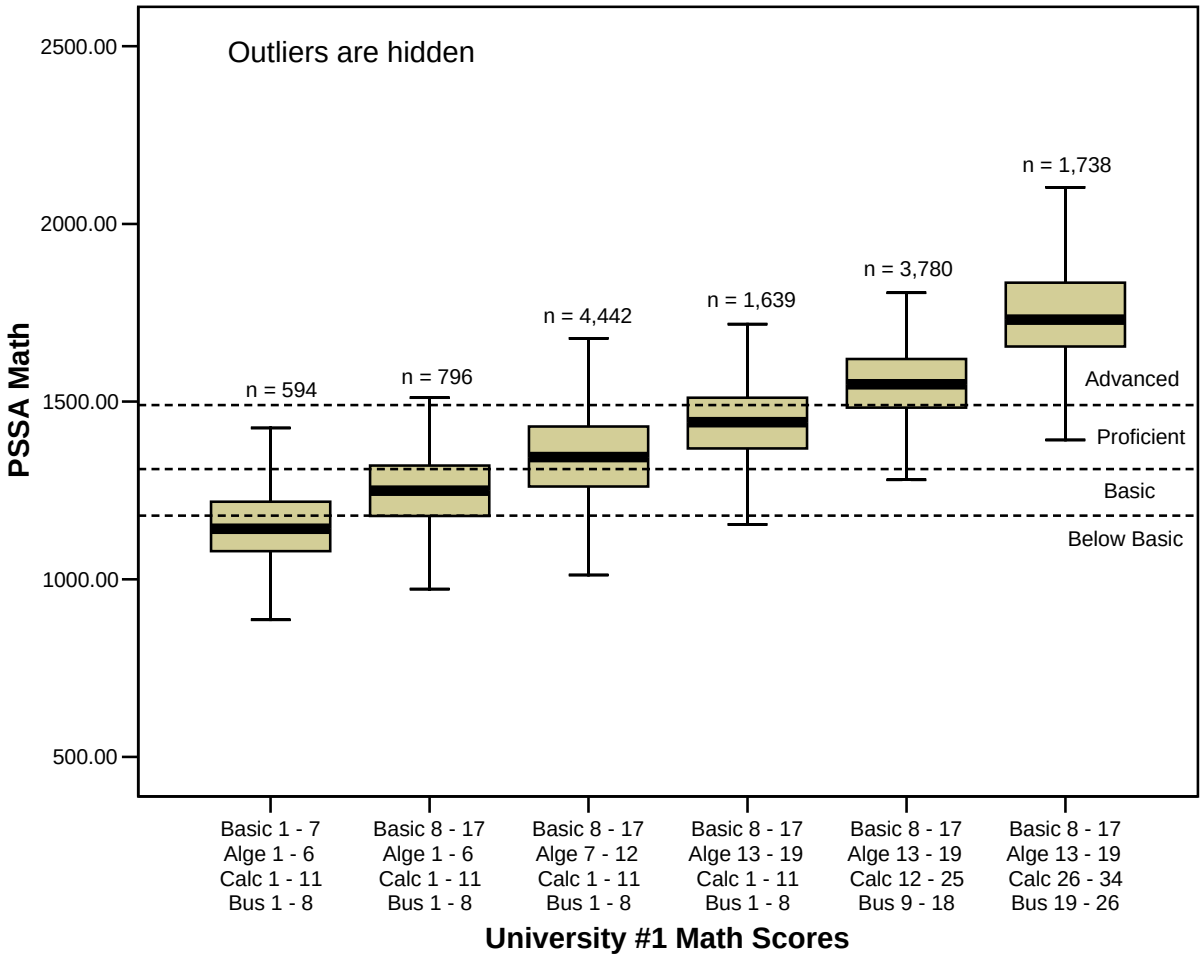

Figure 2. Relationship between University #1 Math Scores and PSSA Math Scores.

Table 17 demonstrates that 97.4% of 11th graders who scored at least 1310 on their PSSA math assessment took at least the standard level freshmen math course.

Table 17. University #1 PSSA Cut Score by College Math Course

	PSSA Math < 1310	PSSA Math >= 1310
Basic Math	1063 37.0%	283 2.6%
Standard Math and Above	1810 63.0%	10,424 97.4%

University #2 Illustrative Relationship with PSSA

English/Reading. Figure 3 shows the relationship between the PSSA reading score on the y-axis and the University #2 English multiple choice score on the x-axis. Recall that University #2 makes English course recommendations based on a combination of students' SAT verbal score, high school percentile rank, English essay exam, and the multiple choice component of the proficiency exam. This study is interested in the validity of the proficiency exams only and because of the constraints discussed regarding the essay exam (and because we don't know what weight is given to the essay in relation to the multiple choice test), only the multiple choice component of the proficiency exam is represented in Figure 3. Students' scores on the multiple choice exam ranged from 22 – 113, and were divided into quartiles[7]. The figure shows a distinct stair-step. Eleventh graders who scored in the Below Basic and Basic performance levels on PSSA reading tended to score in the bottom quartile of the English multiple choice exam. Students who scored in the Proficient performance level on PSSA tended to score in the 2nd and 3rd quartiles on the English multiple choice exam. Finally, students who scored in the Advanced performance level on the PSSA tended to score in the top quartile of the English multiple choice exam.

[7] Because course placement is not based solely on the multiple choice component of the English proficiency exam, the table representing the percent of students scoring at proficient or above on PSSA and taking the standard freshmen writing course is not provided for University #2.

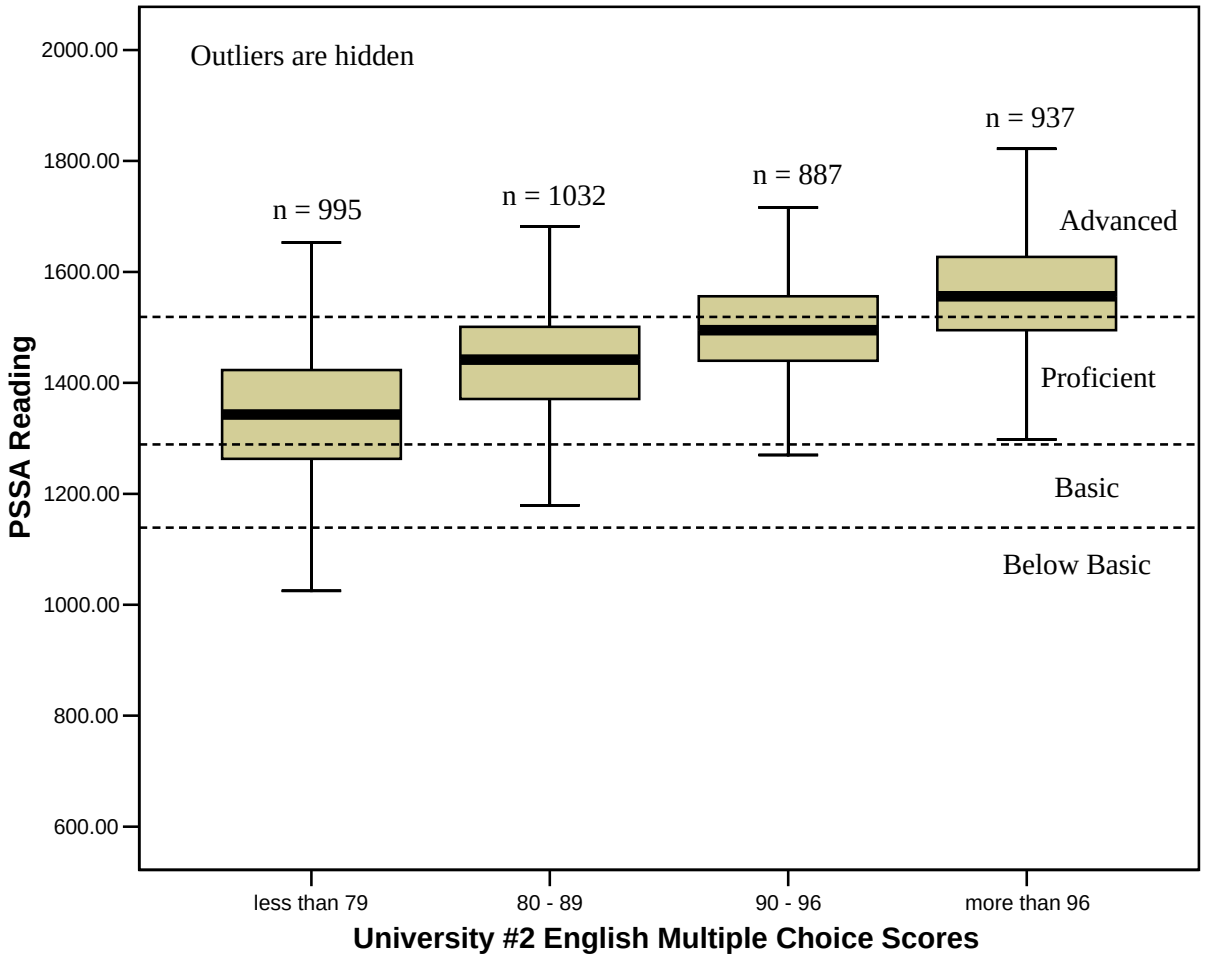

Figure 3. Relationship between University #2 English Scores and PSSA Reading Scores.

Math. Figure 4 shows the relationship between the PSSA math score on the y-axis and the University #2 math scores on the x-axis. Once again, the figure shows a distinct stair-step. Eleventh graders who scored in the Below Basic and Basic performance levels on PSSA math tended to score in the lowest level of the math proficiency exam; these students received recommendations to take pre-college algebra courses. The majority of eleventh graders who scored in the proficient performance level on PSSA math scored in the middle two levels of the university math exam, and received recommendations to take either intermediate algebra or Calculus I. Finally, the eleventh graders who scored in the Advanced performance level on PSSA math tended to score in the highest level of the math proficiency exam; these students received recommendations to enroll in Calculus II.

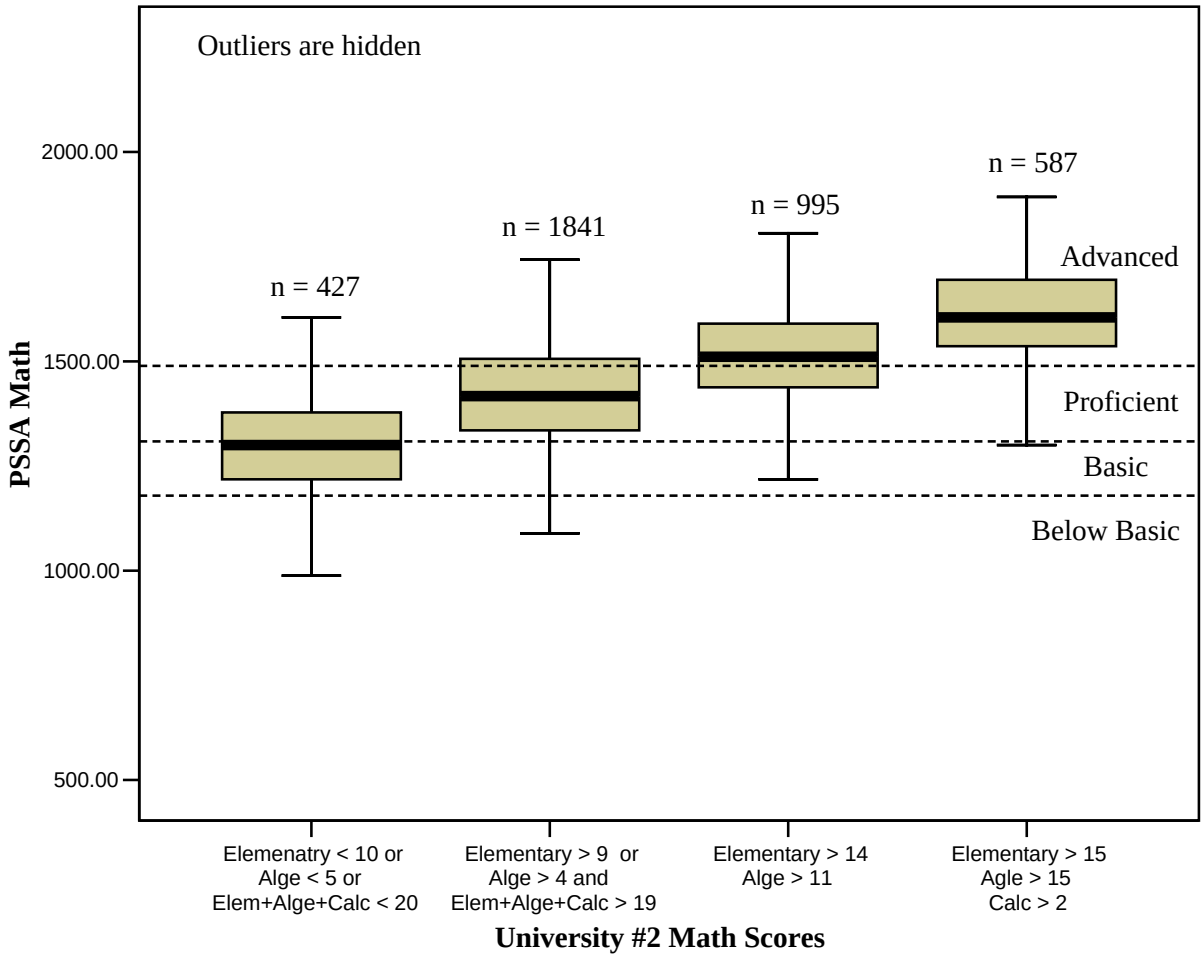

Figure 4. Relationship between University #2 Math Scores and PSSA Math Scores.

Table 18 demonstrates that 93.9% of 11th graders who scored at least 1310 on their PSSA math assessment took at least the standard level freshmen math course.

Table 18. University #2 PSSA Cut Score by College Math Course

	PSSA Math < 1310	PSSA Math >= 1310
Basic Math	231 36.2%	196 6.1%
Standard Math and Above	407 63.8%	3016 93.9%

University #3 Illustrative Relationship with PSSA

Writing/Reading. Figure 5 shows the relationship between the PSSA reading score on the y-axis and the University #3 writing score on the x-axis. Students received a writing score of "1," "2," or "3" on the university proficiency exam. A score of "1" corresponded to a recommendation to take an intensive workshop in composition. A score of "2" corresponded to a recommendation to take the standard workshop in composition. Finally, a score of "3" corresponded to a recommendation to bypass the workshops and take a seminar in composition. Figure 5 shows a distinct stair-step. Eleventh graders who scored in the Below Basic and Basic performance levels on PSSA reading and those who scored in the lower level of the Proficient performance level tended to receive a score of "1" on the writing proficiency exam. The majority of 11th graders who scored in the proficient performance level on PSSA reading received a score of "2" on the university writing exam. Finally, the 11th graders who scored in the upper level of the Proficient performance level and those who scored in the Advanced performance level on PSSA reading typically received a score of "2" or "3" on the university writing exam.

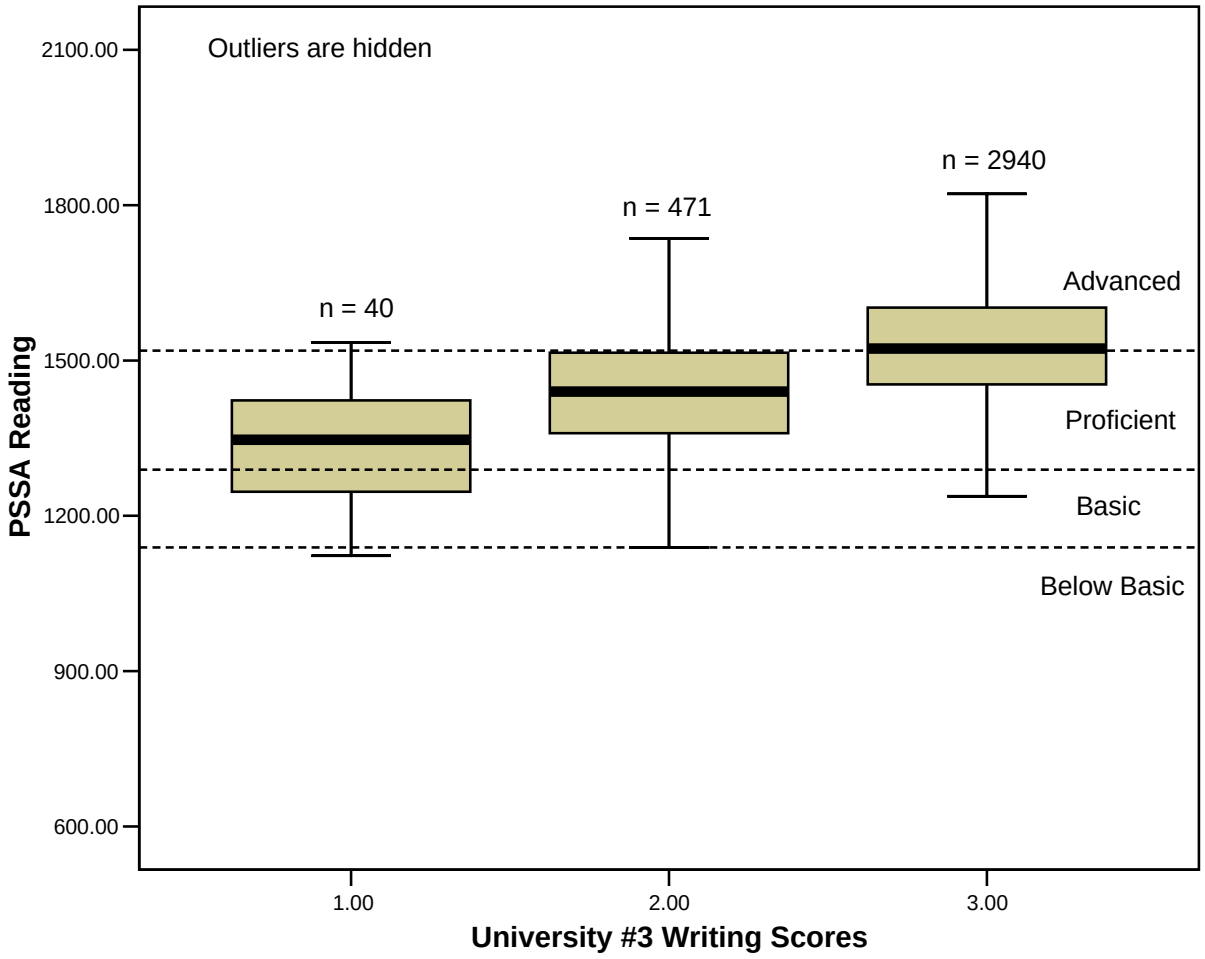

Figure 5. Relationship between University #3 Writing Scores and PSSA Reading Scores.

Table 19 demonstrates that 99.2% of students who scored at least 1290 (i.e., proficient or above) on their 11th grade PSSA reading assessment took at least the standard level freshmen composition course.

Table 19. University #3 PSSA Cut Score by College English Course

	PSSA English < 1290	PSSA English >= 1290
Basic English	14 11.6%	25 .8%
Standard English and Above	107 88.4%	3271 99.2%

Math. Recall that University #3 provided a separate data file for algebra/trigonometry proficiency scores and calculus proficiency scores. All incoming freshmen were required to take the algebra/trigonometry exam, but only students who had studied calculus in high school and who wished to exempt the first level of calculus in college were asked to take the calculus exam. Consequently, two separate math figures are presented for University #3.

Figure 6 shows the relationship between the PSSA math score on the y-axis and the University #3 algebra/trigonometry score on the x-axis. There were six parts to the algebra/trigonometry exam. The first five parts covered various algebraic skills, and the sixth part tested various trigonometric functions and equations. Based on a student's scores on these parts of the exam, the student was categorized into one of six math levels with lower levels representing basic math courses and higher levels representing advanced math courses. Once again, the figure shows a stair-step pattern. Eleventh graders who scored in the Below Basic and Basic performance levels on PSSA math tended to score in the two lowest levels of the algebra/trigonometry exam; these students received recommendations to take beginning level math courses. The majority of eleventh graders who scored in the proficient performance level on PSSA math scored in the middle two levels of the university exam, and received recommendations to take intermediate level math courses. Finally, the eleventh graders who scored in the Advanced performance level on PSSA math tended to score in the two highest levels of the math proficiency exam; these students received recommendations to enroll in advanced math courses (e.g., Calculus I).

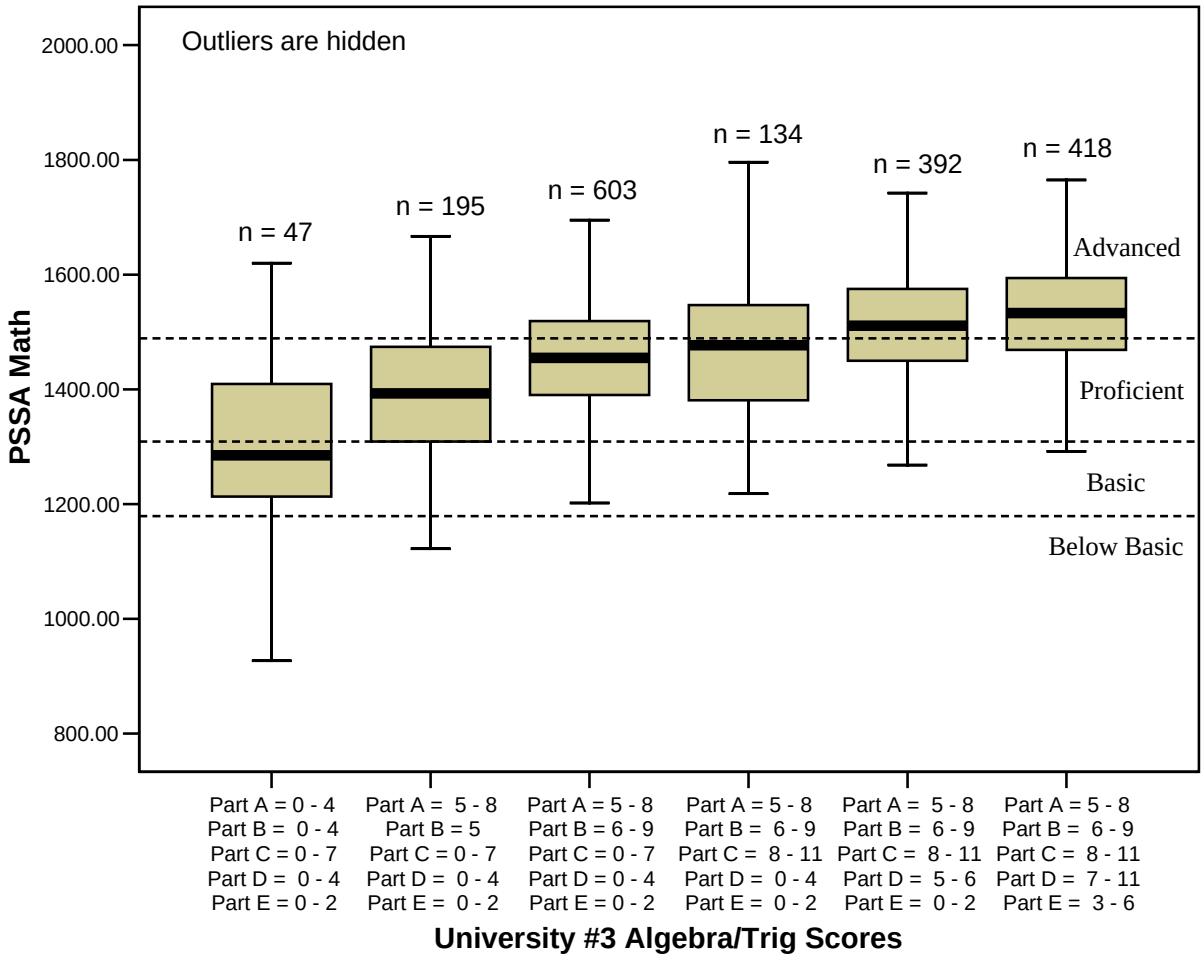

Figure 6. Relationship between University #3 Algebra/Trig Scores and PSSA Math Scores.

Table 20 demonstrates that 93.9% of 11th graders who scored at least 1310 on their PSSA math assessment took at least the standard level freshmen math course.

Table 20. University #3 PSSA Cut Score by College Math Course

	PSSA Math < 1310	PSSA Math >= 1310
Basic Math	75 50.7%	165 10.1%
Standard Math and Above	73 49.3%	1464 89.9%

Figure 7 shows the relationship between the PSSA math score on the y-axis and the University #3 calculus score on the x-axis. Calculus scores ranged from 1 to 18. If a student scored a 13 or lower on the exam, then the university deferred to the students' algebra/trigonometry score to make a course recommendation. Students who scored between 14 and 16 were exempt from Calculus I and were recommended to take Analytic Geometry and Calculus II. Students who scored a 17 or higher were recommended to take Honors Variable Calculus. Once again, the figure displays a distinct stair-step pattern. In this case, however, the majority of all students who took the calculus exam scored in the Advanced performance level on PSSA math. This is not surprising given that the calculus proficiency exam was intended for students who had taken calculus in high school. Nonetheless, the stair-step pattern indicates that those who scored in the lower third of the calculus exam had lower PSSA scores than those who scored in the middle third, who in turn had lower PSSA scores than those who scored in the upper third on the calculus exam.

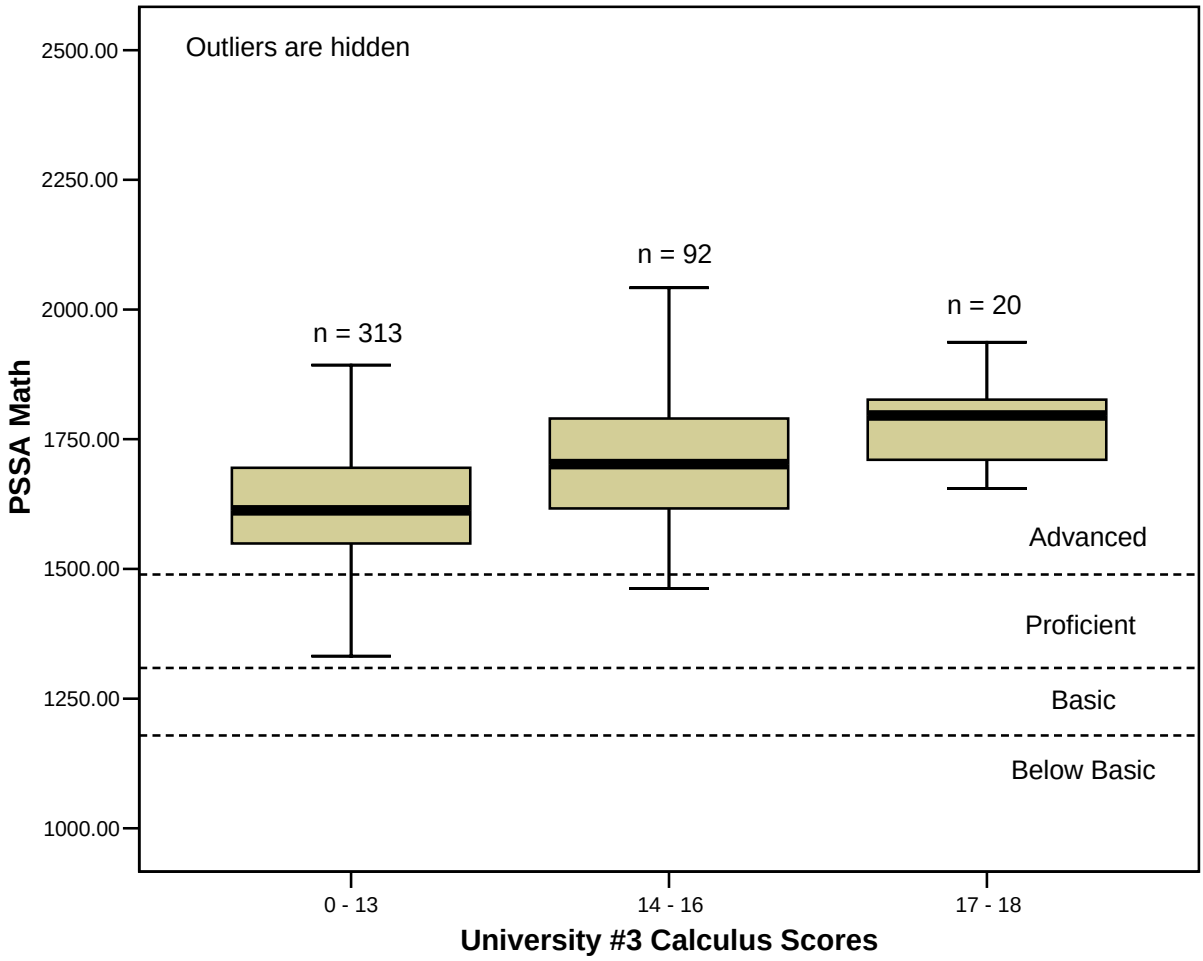

Figure 7. Relationship between University #3 Calculus Scores and PSSA Math Scores.

Summary and Conclusion

The primary purpose of this investigation was to determine the relationship between students' 11th grade PSSA scores and those same students' scores on their university proficiency exams in math and reading. There were subtle variations between universities; however, the same general pattern was observed among all three universities. For math, the results indicate that students' 11th grade PSSA math scores were moderately to highly correlated with their math scores on the universities' proficiency exams. University #1 had the highest convergent validity coefficients for math (r = 0.68 to 0.75), followed by University #2 (r = 0.51 to 0.58), and University #3 (r = 0.40 to 0.52). It is important to note that the sample sizes were largest for University #1 and smallest for University #3. For reading, there was more variability among universities. This was largely due to the format of the reading proficiency exams. University #2 included an essay component on its English exam and University #3's exam was comprised solely of a single essay prompt. These essay exams are likely scored using an ordinal scale, which attenuates the validity coefficient when correlated with PSSA's interval level data. This appears to be the case as the correlation between PSSA reading and University #2's essay exam was .37, and the correlation between PSSA reading and University #3's essay exam was .30. When the multiple choice English exams for University #1 and University #2 were correlated with students' scores on PSSA reading, the correlations were more in-line with what would be expected (r = 0.59 for University #1 and r = 0.64 for University #2).

The graphic illustrations of the relationships between PSSA scores and university proficiency scores further demonstrated that students who scored in the proficient and advanced levels on PSSA tended to be the students who scored high on the proficiency exams, and were thereby placed into advanced level courses (and vice-versa). Similarly, across all three universities at least 90% of students who scored at or above proficient on the PSSA took at least the standard level freshmen course in English/math. Overall, these results indicate that students who tend to perform well on the PSSA can also be expected to perform well on the university proficiency exams. Similarly, students who tend to perform poorly on the PSSA can be expected to perform poorly on the university proficiency exams.

Because the PSSA and the university proficiency exams report scores for common subject areas, we expected student scores on the various tests to be positively related. Overall, the above results support this expectation. The convergent validity coefficients meet the "Goldilocks" criterion; that is, the correlations between PSSA and the university proficiency exams fall in the not-too-high and not-too-low category. There are three main reasons we did not expect the correlations between PSSA and the university proficiency exams to be too high. First, PSSA tests student ability related to content that is specific to the state of Pennsylvania. The extent to which the universities' proficiency exams measure content that is different from PSSA's limits the strength of the correlation between the assessments. Second, students attending college represent a high performing subset of all students taking PSSA; therefore, the full range of PSSA scores may not be represented by this subset. This range restriction also attenuates the correlations. Third, differences in the format of the test items (e.g., multiple-choice vs. performance-task items) may also weaken the correlations between the two types assessments. Differences in the number of raw score points on the two assessments could also be another explanation for attenuated validity coefficients.

In addition to examining convergent validity evidence for the PSSA, we also examined the predictive validity of the PSSA and the university proficiency exams by investigating their relationships with students' GPAs. This type of analysis was only possible for University #1 and University #2 because course grade information was not available for University #3. For both universities, students' GPAs in their first college English course generally show positive relationships with PSSA reading scores and with university English proficiency exams. Similarly, students' GPAs in their first college math course generally show positive relationships with PSSA math scores and with both universities' math proficiency exams. In all cases, these correlations were only weak to moderate. These lower correlations were interpreted as being partially due to differences in teachers' grading practices, differences in teaching styles and differences in course difficulty. Lower correlations can also be partially attributed to range restriction on PSSA and range restriction in course grades. Lastly, these findings suggest that student effort may play an important role in predicting course GPA. It is important to note that the correlations between PSSA and course GPAs were similar in magnitude to the correlations between the universities' proficiency exams and course GPAs, thereby indicating that neither the PSSA nor the university proficiency exams is substantially better than the other at predicting course GPA. Together, these results suggest that PSSA scale scores/performance levels might serve as an alternate or additional means for students to demonstrate readiness for university courses in English and math.

References

Bacci, E. D., Koger, M. E., Hoffman, R. G. & Thacker A. A. (2003). *Relationships among Kentucky's Core Content Test, ACT scores, and students' self-reported high school grades for the classes of 2000 through 2002.* (HumRRO Report FR-03-19), Louisville, KY: Human Resources Research Organization.

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.*

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.

Koger, M.E., Thacker, A.A., & Dickinson, E.R. (2004). *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.* (HumRRO Report FR-04-26), Louisville, 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.*

Sinclair, A.L., & Thacker, A.A. (2004). *Relationships Between Students' Scores on KCCT and CTBS* (HumRRO Report FR-04-24). Alexandria, VA: Human Resources Research Organization.

Stroud, Robert E. (1995). Correlations between the metropolitan achievement tests, seventh edition, and the Ohio ninth-grade proficiency tests. Ohio. (ERIC Document Reproduction Service No. ED394992.

Thacker, A.A., & Hoffman, R.G. (1999). *Relationships between MCAS and SAT-9 for one district in Massachusetts* (HumRRO Report FR-WATSD-99-05). Alexandria, VA: Human Resources Research Organization.

Thacker, A.A., Dickinson, E.R., & Koger, M.E. (2004). *Relationships Among the Pennsylvania System of School Assessment (PSSA) and Other Commonly Administered Assessments.* (HumRRO Report FR-04-33), Louisville, KY: Human Resources Research Organization.

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