OPPOSITE OBJECTIVES, SIMILAR SCORES: THE MCAS AND SAT

An Interactive Qualifying Project Report

Submitted to the Faculty

Of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

By

Nathaniel Freedman

Professor John M. Wilkes, Major Advisor

May 3, 2007

Table of Contents

Abstract

The MCAS is a criterion-based achievement test created by the state of Massachusetts to measure public school students' mastery of the academic materials in the Curriculum Frameworks. Previous WPI students' studies of urban public high school students found that scores on the pilot MCAS correlated strongly to scores on the SAT, an aptitude test, and that both tests were biased towards certain MBTI types. Beginning with the Class of 2003, public high school sophomores are required to pass the MCAS to graduate. This study showed that the prior findings for the experimental test were also true for the official test taken by the Class of 2003. It also found that the SAT was equal to or greater than grades in magnitude of correlation with the MCAS. Further research should be conducted to study this important issue.

Introduction

The MCAS is a criterion-based achievement test created by the state of Massachusetts. The SAT is a test intended to measure aptitude, and used by many colleges to assist in admissions. However, previous studies of Worcester public high school students by Worcester Polytechnic Institute students have found a strong correlation between the scores on the SAT and scores on pilot versions of the MCAS. They also found that a personality type measure known to correlate with SAT scores, the S-N preference of the MBTI, also had the same correlation with MCAS scores, but did not correlate with subject grades.

Since the MCAS didn't have official status and the curriculum the curriculum the students took to prepare for it had not met the standards set by the state, the validity of this set of findings were unclear. This study is the first to examine the SAT and official MCAS for Worcester public high school students in the class of 2003, the first class to complete certified Math and English curricula and take a test they were required to pass to graduate from high school. The results of previous WPI studies of Worcester high schools were found to also apply to the students studied from the Class of 2003 who took the official MCAS. Academic data are also used in the current analysis.

It was determined that SAT scores are in general at least as predictive of MCAS scores as grades. Also, while the MCAS Math test appears relatively homogeneous with regards to the SAT and academic achievement, major differences between the multiple choice and non-multiple choice subsections of the MCAS ELA test were found. An unexpected finding was that grades and MCAS scores had weak to insignificant correlations for many students who took less difficult courses. These results raise

questions about the validity of the MCAS as an unbiased achievement test, the relationship of the SAT to academic achievement, and the quality of the curricula in Worcester public high schools. Further studies of this subject should be conducted to determine how universal these findings are for Massachusetts public schools; whether changes to the curriculum, MCAS, and SAT since 2003 have changed the situation today; and to better understand the relationship between academic achievement, aptitude, the MCAS, and the SAT.

Background

Massachusetts Comprehensive Assesment System (MCAS)

The MCAS is a standardized test created by the state of Massachusetts to test students' mastery of the materials specified by the learning standards of the Massachusetts Curriculum Framework for a given subject area such as English/language arts or mathematics. The Curriculum Framework is the state standard for curricula in public schools, giving minimum requirements for skills and knowledge to be learned by high school graduates. The Curriculum Framework was established by the Education Reform Act of 1993. It contains standards for English language arts, mathematics, science and technology/engineering, history and social science, art, and comprehensive health. However, not all of these subjects are currently tested by the MCAS. Schools must submit their curricula to the state Department of Education for approval under the Curriculum Framework. (Massachusetts curriculum frameworks)

The MCAS results are used by the state to assess academic achievement at the district, school, and student level. They were originally intended to be administered to students in public schools in 4th, 8th, and 10th grade. Due to teachers' concerns about the time devoted to testing, the tests for 4th- and 8th-grade students were spread out over multiple grades. Beginning with the class of 2003, all 10th grade students in public high schools are required to pass the MCAS for English language arts (ELA) and Math in order to graduate (the Science and Technology/Engineering MCAS test has been added to the graduation requirement starting with the class of 2010). Before then the MCAS was administered three times (10th grade for the Class of 2000, 2001, 2002, and the 8th grade test to the Class of 2002) on an experimental basis, both because the test was not

fully developed and curricula that met the standards of the Curriculum Framework had not been implemented statewide long enough for students to have been taught most of the knowledge and skills tested by the MCAS. Students who fail the MCAS have five more opportunities in the next two years to pass the sophomore test in order to be able to graduate. If schools have consistently low MCAS scores, the state will intervene in the school to improve teaching and administration. (About MCAS)

The MCAS unambiguously has the form of a criterion-based achievement test. The ELA and math tests have four different question formats. Both tests have multiple choice and open response questions. The ELA test has a composition section in which students must write an essay in response to a prompt, and the math test has a short answer section that requires a brief response to a math problem. The multiple choice section makes up half of the total score. Each subject test is scored separately and given a scaled score between 200 and 280. The scaled scores are assigned to different performance levels. 200 - 219 is “failing” (F), 220 - 239 is “needs improvement” (NI), 240 - 259 is “proficient” (P), and 260 - 280 is “advanced” (A). (About MCAS)

Goldsmith (2002) reviewed the 2001 MCAS ELA, taken by the class of 2003. In reaction to improved scores on the 2001 MCAS compared to previous years, Goldsmith wrote, “Celebration of improvements in student performance on the 2001 MCAS and preoccupation with helping students pass the test have obscured the absence of an evaluation of the MCAS tests and their probable effects on teaching and learning in Massachusetts classrooms.” She analyzed sample questions from the 2001 ELA test at each grade levels for how challenging the test was and how well it evaluated students’ knowledge. Goldsmith found that the composition sections on almost all tests were near

the limit of the writing skills expected of most students at that age. She also criticized the composition on the 10th grade ELA test for asking students to write about a specific theme in the context of any book they had read. Without a specific work associated with the prompt, this could result in students “bending the prompt to the work” or vice versa. Goldsmith wrote that having the students choose a book they have read would also unfairly disadvantage students who did not have a diverse reading background.

Goldsmith recommended that more questions testing basic skills should be added to the test, as it did not do much to ensure that “prerequisite learning has been mastered.” In particular, she faulted the 10th grade test for having few questions about grammar and vocabulary. By subject, 28% of the 2001 test was composition, 64% was literature analysis, and 8% was vocabulary and grammar.

Scholastic Aptitude Test (SAT)

The SAT is a standardized test taken by most students applying to undergraduate colleges, and many college admission departments use the scores to guide admission decisions. The SAT was first used as a college admissions test in 1933 for scholarship selection by Harvard University. By 1948 it was the primary college admissions test for college applicants nationwide. The test was derived from an IQ test used by the U.S. Army in World War I, and it was originally intended to measure solely the aptitude of college applicants, independent of academic performance. High school students typically take the test in their senior year. (A brief history of the SAT)

In 2003 the test consisted of two sections and was known as the SAT I: Reasoning Test. The Verbal section only had multiple choice questions on critical reading, sentence completion, and analogies. The Mathematics section was mainly multiple choice questions on geometry, arithmetic, and algebra, with some questions requiring a numerical answer instead of a multiple choice answer. Students' scores on each section were scaled to a range of 200 - 800, with 800 the highest score possible for a section. (Frequently asked questions)

The SAT is scored by a method known as formula scoring. Correct multiple choice answers add one point to the raw score, but incorrect answers result in a penalty of one-quarter of a point. With 5 multiple choice answers available, random guessing is not likely to improve the test-takers' formula corrected score. However, a controlled experiment by Higham (2007) that used excerpts from the 1997 SAT I found that the assumptions for the use of formula scoring are not true for the SAT. Using a statistical analysis known as type-2 signal detection theory, Higham determined that test-takers

could guess correct answers at a rate better than random chance. However, the test-takers did not usually guess enough to have the best possible score. This contradicts the basis for using formula scoring, which is intended to discourage guessing by preventing it from improving the corrected score. He also found that if test-takers thought the penalty was -4 instead of -0.25, their mean formula score was improved more by being forced to answer all of the questions than test-taker who were not misinformed that the penalty was -4. By the assumptions of formula scoring, neither group should improve their formula corrected score by answering all questions if they did not know the correct answer. This implied that not only was formula scoring inadequate to correct for guessing, but that how much the score could be improved by answering more questions was sensitive to the incentive of test-takers to avoid wrong answers.

In 2006, the SAT was changed significantly. Analogy problems were removed from the Verbal section and more critical reading problems were added. The math section was also altered, with the addition of more advanced mathematics problems and the removal of some other problems. The most important change was the addition of a writing section, which consists of a writing prompt that students have 25 minutes to write an essay on, and a multiple choice section on writing skills. The name of the test was also changed from SAT I: Reasoning Test to SAT Reasoning Test. (Frequently asked questions)

The Preliminary SAT (PSAT) is a standardized test typically taken by high school juniors. The test is both a practice for the SAT and it is used to select students to receive National Merit Scholarships. It is similar to the SAT in format and subjects, except that it had a writing section before the SAT did. However, it did not have a written essay, and as

of 2006 it still does not. PSAT scores are reported on a scale of 20 - 80 for each of the three sections, and a composite score that is the sum of all three section scores. (PSAT/NMSQT) There is evidence linking the PSAT to academic achievement, as Milewski & Sawtell (2006) found that composite scores on the PSAT had a correlation of 0.52 to the self-reported GPA of high school students from a nation-wide sample. Ansley et al. found a 0.604 Spearman's rank-order correlation between math PSAT scores and average math grade for Worcester public high school student from the classes of 1998 and 1999. They also found a 0.484 Spearman's rank-order correlation between verbal PSAT scores and average English grade for the same students.

Myers-Briggs Type Indicator (MBTI)

The MBTI is a personality test created by Katherine Briggs and Isabel Myers. It is based on Carl Jung's theory of psychological type. Jung theorized that there are two opposite ways in which people perceive, sensing, and intuition, and two opposite processes people use to make judgments, feeling and thinking. Jung termed judging or perceiving the world around a person "extraversion," and judging or perceiving of a person's own emotions and ideas "introversion." Jung theorized that people have a preference for using one type of perceiving and one type of judging, and preferred to use either extraversion or introversion. These preferences are not exclusive, but are the stronger and more often used process. (Myers, 1993)

The MBTI uses these ideas to index the preferences of people, adding a fourth preference, judging or perceiving, to represent whether people prefer to use judging or perceiving processes in extraversion. The combination of these preferences results in sixteen different personality types. The MBTI is a measure of each preference for determining personality type. The preferences are assigned letters: extraversion - introversion (E-I), sensing -intuition (S-N), thinking - feeling (T-F), and judging - perceiving (J-P). Someone whose preferences are extraversion, sensing, feeling, and perceiving would be a ESFP type. (Myers, 1993)

The T-F dimension is not usually connected to learning, because it is a judging preference while most education involves perception or problem-solving without an emotional or social context. Thinking types logically analyze situations in terms of principles and criteria to determine what to do, while feeling types look at how decisions will affect the people involved to choose a course of action. The T-F dimension is

relevant to what teaching style is best for interacting with students, as feeling types prefer teachers who show that they care how well students are doing, and thinking types want to be sure that rewards for good performance are fairly distributed. They also seem to prefer courses that do not have a social context, because they feel it adds unnecessary complexity instead of relevance, as opposed to feeling types, who are more engaged with classroom learning when they are in courses that involve discussion about issues in society. The J-P dimension describes how someone deals with organization, with J-types seeking an orderly and planned task environment, and P-types preferring flexibility to keep options open to gather more information. J-types tend to put more effort into homework and projects than P-types, while P-types prefer classes and assignment that are not strictly structured and allow the student to have an active role in learning. (Lawrence, 2004)

The two perceiving preferences, sensing and intuition are the most important for education. Sensing types prefer to focus on details to understand something, while intuiting types will try to find patterns and make informed guesses to understand the concept as a whole. S-types prefer to learn by looking at the application or example before the theory, while N-types prefer the opposite. N-type students may do better on multiple choice tests because they are better at figuring out which choices are incorrect from context and presentation when they do not know which choice is correct. (Lawrence, 2004)

A study of college students by Schurr and Ruble (1986) found that students who were intuition or introverted types scored higher on the SAT verbal. The same study found that judging types had a higher average college GPA than perceiving types despite

having similar scores on aptitude tests. A study by Schurr, Ruble, and Henriksen (1988) found that intuition and sometimes perceiving were associated with higher SAT scores on both sections.

Past WPI Studies

One of the earliest studies by WPI students of standardized tests in Worcester public schools was by Kingsland et al. in 1995. They found that N-types had significantly better scores on the SAT than S-types. The ETS made some alterations to the test the next year. A 1997 study by Batey et al. made the same finding for a later graduating class in Worcester taking the modified SAT. A 1998 full-scale study by Pieper et al. that was supported and endorsed by the Worcester public school system of the juniors in the Class of 1997 found that N-types also outscored S-types on the PSAT in Worcester public schools. The study was of a large proportion of the Worcester public school system because the administration strongly encouraged students in the Class of 1997 to take the PSAT to prepare for the SAT. In 2001, Stambaugh et al. found that students in the Worcester public schools from the class of 2000 and students in the class of 2000 and 1999 from Fitchburg also had a significant N-type advantage in SAT scores. A study of Leicester students by Mangenot et al. also found an N-type advantage on the SAT.

The first study by WPI students to compare the MCAS and SAT/PSAT was by Stambaugh et al. They studied the Fitchburg high school class of 1999 and 2000, and the Worcester public high school class of 2000. The study found in Fitchburg a 0.63 correlation between the MCAS ELA score and the SAT Verbal score, and a 0.53 correlation between the MCAS Math score and SAT Math score. For the Worcester class of 2000, a 0.71 correlation was found between the Verbal and ELA scores, and the math test scores had a 0.83 correlation. Such strong correlations were not expected, as the MCAS is intended to measure academic achievement while the SAT is intended to measure aptitude. The tests should not have strong correlations to each other if they are

measuring different qualities, which led to the hypotheses that both tests are actually aptitude tests or both tests are actually achievement tests.

However, a major issue with the validity of these findings was that the MCAS did not have official status at the time. This was because curricula meeting the standards of the Curriculum Framework had not been implemented long enough, and students were not required to pass the MCAS to be able to graduate. As a result of this, failure rates on the MCAS for sophomores exceeded 50% statewide, with failure rates in Worcester and Fitchburg approximately 2/3. Without having been taught most of the material covered by the MCAS tests, the students' academic achievement is not really being evaluated.

The study by Ansley et al. in 2006 was supposed to overcome this issue by comparing SAT and MCAS scores of the Worcester public schools class of 2003, the first class for which the MCAS ELA and mathematics test were required for graduation. However, they were unable to obtain SAT or PSAT data in time to carry out that study, so they instead compared the class of 2003 MCAS scores to grades and the MBTI. They found that there was an N-type advantage in MCAS scores, as there was for the SAT in prior studies. This advantage was present even after controlling for grades and course difficulty.

The current study is the first at WPI to directly compare SAT and MCAS scores for the class of 2003.

Statistical Analysis

Section 1: About The Data

The MCAS, MBTI, and academic data in this sample were provided by Worcester public high schools for a previous study by Poland et al. The SAT data was provided by the Worcester public high schools in 2006, and this is the first study at WPI to use their data. The student cases in this data sample have varying levels of data completeness, with three categories of data that are present or absent. The categories are SAT score data; MCAS score data; and MBTI and academics data. The main subset of the sample used is 225 cases which have all three categories of data, which will be referred to as the tertiary sample. This subset is part of a much larger subset that contains all cases that have MCAS, MBTI, and academics data. This larger subset has 784 cases, and is used for some analyses in this study. It will be referred to as the secondary sample. Approximately 1300 cases of student data available for this project had MCAS data, which is likely the entire Class of 2003 that took the MCAS as sophomores. Retest scores were not available for the students who failed the MCAS on the first testing.

	Frequency	Percent
Accelerated Learning	11	4.9
Burncoat	85	37.8
Doherty	73	32.4
North High	17	7.6
South High	39	17.3
Total	225	100.0

Table 1.1: Students' School for Tertiary Sample

Table 1.1 shows the composition by high school for the students in the subset of the sample that has complete data. The information for high school attended for this table comes from the high school code on the SAT form.

		Frequency	Percent
	ALL School	40	5.1
	Burncoat	159	20.3
	Doherty	249	31.8
	North	104	13.3
	South	78	9.9
	Voc. Tech.	153	19.5
	Total	783	99.9
	Missing	1	.1
Total		784	100.0

Table 1.2: Students' School for Secondary Sample

The information for high school attended for Table 1.2 comes from the prior MCAS/MBTI data set. It is available for more students in the sample than the SAT high school code, but conflicts with that identifier for several students. "Voc Tech." stands for Worcester Technical High School (formerly Worcester Vocational High School), and "ALL School" stands for Accelerated Learning Laboratory.

Table 1.3 and Table 1.4 show the distribution of MCAS performance levels by school.

			ELA Performance Level				
			F	NI	P	A	Total
School Code w/Name	ALL School	Count	14	11	11	3	39
		% within School Code w/Name	35.9%	28.2%	28.2%	7.7%	100.0%
	Burncoat	Count	40	55	56	8	159
		% within School Code w/Name	25.2%	34.6%	35.2%	5.0%	100.0%
	Doherty	Count	44	80	95	29	248
		% within School Code w/Name	17.7%	32.3%	38.3%	11.7%	100.0%
	North	Count	18	50	32	3	103
		% within School Code w/Name	17.5%	48.5%	31.1%	2.9%	100.0%
	South	Count	10	39	24	4	77
		% within School Code w/Name	13.0%	50.6%	31.2%	5.2%	100.0%
	Voc. Tech.	Count	64	71	18	0	153
		% within School Code w/Name	41.8%	46.4%	11.8%	.0%	100.0%
Total		Count	190	306	236	47	779
		% within School Code w/Name	24.4%	39.3%	30.3%	6.0%	100.0%

Table 1.3: MCAS ELA Performance Level by School (Secondary Sample)

F: Failing (200 - 219)
NI: Needs Improvement (220 – 239)
P: Proficient (240 – 259)
A: Advanced (260 – 280)

			Math Performance Level				
			F	NI	P	A	Total
School Code w/Name	ALL School	Count	14	22	2	2	40
		% within School Code w/Name	35.0%	55.0%	5.0%	5.0%	100.0%
	Burncoat	Count	49	55	45	10	159
		% within School Code w/Name	30.8%	34.6%	28.3%	6.3%	100.0%
	Doherty	Count	50	76	72	51	249
		% within School Code w/Name	20.1%	30.5%	28.9%	20.5%	100.0%
	North	Count	32	44	21	7	104
		% within School Code w/Name	30.8%	42.3%	20.2%	6.7%	100.0%
	South	Count	17	32	20	9	78
		% within School Code w/Name	21.8%	41.0%	25.6%	11.5%	100.0%
	Voc. Tech.	Count	72	74	7	0	153
		% within School Code w/Name	47.1%	48.4%	4.6%	.0%	100.0%
Total		Count	234	303	167	79	783
		% within School Code w/Name	29.9%	38.7%	21.3%	10.1%	100.0%

Table 1.4: MCAS Math Performance Level by School (Secondary Sample)

The secondary sample is not completely representative of the entire sample of MCAS scores, as grades and MBTI data was not available for a disproportionate number of the students who failed the MCAS in 10th grade. In the entire data set of MCAS scores for class of 2006 Worcester public high school students, 32.5% failed the ELA section, and 39.8% failed the math section. However, the probable self-selection of students taking the SAT surpasses this problem for the tertiary sample, as very few students who failed the MCAS in 10th grade took the SAT relative to students who passed the MCAS. Thus, students who failed the MCAS are not adequately represented in the tertiary sample.

Even so, average scores on the 2001 MCAS were clearly better than on previous administrations of the MCAS in Worcester. It appears that development of the curricula and requiring students to pass the test resulted in the failure rate dropping to a level that does not create overwhelming concerns about the validity of the MCAS as an achievement test. The high failure rate and questions of validity were a major issue for interpreting the results in past studies by WPI students, but it appears that this study does not have the same problem. The proportion of students who failed the test is unacceptably high from an educational perspective, but it is not extraordinarily high from a research perspective.

			ELA Performance Level				Total
			F	NI	P	A	
High School	Accelerated Learning	Count	3	3	3	2	11
		% within High School	27.3%	27.3%	27.3%	18.2%	100.0%
	Burncoat	Count	11	28	40	6	85
		% within High School	12.9%	32.9%	47.1%	7.1%	100.0%
	Doherty	Count	1	13	42	17	73
		% within High School	1.4%	17.8%	57.5%	23.3%	100.0%
	North High	Count	2	4	11	0	17
		% within High School	11.8%	23.5%	64.7%	.0%	100.0%
	South High	Count	2	17	16	4	39
		% within High School	5.1%	43.6%	41.0%	10.3%	100.0%
Total		Count	19	65	112	29	225
		% within High School	8.4%	28.9%	49.8%	12.9%	100.0%

Table 1.5: MCAS ELA Performance Level by School (Tertiary Sample)

			Math Performance Level				Total
			F	NI	P	A	
High School	Accelerated Learning	Count	3	8	0	0	11
		% within High School	27.3%	72.7%	.0%	.0%	100.0%
	Burncoat	Count	14	29	33	9	85
		% within High School	16.5%	34.1%	38.8%	10.6%	100.0%
	Doherty	Count	3	14	29	27	73
		% within High School	4.1%	19.2%	39.7%	37.0%	100.0%
	North High	Count	0	8	7	2	17
		% within High School	.0%	47.1%	41.2%	11.8%	100.0%
	South High	Count	2	14	15	8	39
		% within High School	5.1%	35.9%	38.5%	20.5%	100.0%
Total		Count	22	73	84	46	225
		% within High School	9.8%	32.4%	37.3%	20.4%	100.0%

Table 1.6: MCAS Math Performance Level by School (Tertiary Sample)

High School [From SAT Code]		Verbal SAT Score	Math SAT Score
Accelerated Learning	Mean	374.	341.
	N	11	11
Burncoat	Mean	467.	481.
	N	85	85
Doherty	Mean	509.	509.
	N	73	73
North High	Mean	468.	478.
	N	17	17
South High	Mean	468.	481.
	N	39	39
Total	Mean	476.	483.
	N	225	225

Table 1.7: Mean SAT Score by School (Tertiary Sample)

The 225-student tertiary sample used in this study has a gender imbalance, as 65.8% of the sample is female. This imbalance is not present in the secondary sample. However, the only significant difference in scores between genders within the tertiary

sample is a mean difference of 41 points between males and females on the SAT Math, with higher scores for males. A small significant advantage for females for English average grade was also found.

MBTI intuitive-types are overrepresented in the tertiary sample, with 51.6% of the sample intuitive, compared to 30-35% in the general population. This is also present in the MCAS/MBTI subset, though to a somewhat lesser degree. Perceiving types are also somewhat over represented, with about 60% in the secondary and tertiary sample.

The academics data in this sample is only for the students' freshman and sophomore years. The junior and senior courses and grades are not included; therefore it is not accurate to compare the grades data directly to the SAT scores. Since the MCAS is taken in the sophomore year, MCAS scores and grades are a measure of the same part of the student's secondary education, and are appropriate for direct comparison.

A previous IQP used the transcript data from the freshman and sophomore year to calculate average grades and average course levels for English and math courses. Courses were assigned course level 2 if they were an "honors" course, course level 3 if they were a "college" course, and course level 4 if they were a "general" or "technical" course. Course level 1 represents Advanced Placement courses, but no students in the sample took AP math or English courses in their freshman or sophomore years. The average course levels by student were originally decimal values, but this study rounded them to integers to make it easier to use them as categories.

The following tables show the distribution of average course levels by school for the secondary and tertiary sample. The students of the tertiary sample mainly have average course levels of 2, but the secondary sample has a nearly even distribution

between level 2 and level 3. Almost no students from the small proportion with average course level of 4 are in the tertiary sample. Worcester Technical is not represented in the SAT sample.

		English Average Course Level			
		2	3	4	Total
	ALL School	0	40	0	40
	Burncoat	91	51	17	159
	Doherty	139	89	20	248
	North	51	48	4	103
	South	52	24	2	78
	Voc. Tech	25	87	41	153
Total		358	339	84	781

Table 1.8: Students by Average English Course Level and HS. (Secondary Sample)

		Math Average Course Level			
		2	3	4	Total
	ALL School	0	40	0	40
	Burncoat	88	55	16	159
	Doherty	130	97	21	248
	North	42	59	3	104
	South	53	22	3	78
	Voc. Tech.	22	79	52	153
Total		335	352	95	782

Table 1.9: Students by Average Math Course Level and HS. (Secondary Sample)

		English Average Course Level			
		2	3	4	Total
High School [From SAT Code]	Accelerated Learning	0	11	0	11
	Burncoat	67	13	5	85
	Doherty	59	14	0	73
	North High	14	2	0	16
	South High	35	4	0	39
Total		175	44	5	224

Table 1.10: Students by Average English Course Level and HS. (Tertiary Sample)

		Math Average Course Level			
		2	3	4	Total
High School [From SAT Code]	Accelerated Learning	0	11	0	11
	Burncoat	65	15	5	85
	Doherty	60	13	0	73
	North High	14	3	0	17
	South High	35	4	0	39
Total		174	46	5	225

Table 1.11: Students by Average Math Course Level and HS. (Tertiary Sample)

The changes in proportions of average course level between the samples are not likely to be random. It is probable that it is due to self-selection of students in higher level courses taking the SAT because they are more likely to apply to college than students in lower level courses. The proportion of students between schools also changes from the secondary to tertiary sample, and it is not clear if this is random or not. Regardless, the only schools that can be analyzed singly with accuracy are Burncoat, Doherty, and South High.

Because of these factors, the tertiary sample is only representative of students with higher-level academic transcripts who plan to attend college, and then likely only students from public urban high schools. Analysis of the SAT will look mainly at students with an average course level of 2. It is likely that a future study using PSAT scores instead of SAT scores will overcome many of the problems with the representativeness of the tertiary sample and grades.

Many of the MCAS scores used in this report are raw scores, not scaled scores. As detailed in Appendix A, scaled scores are not a scale statistic, as equal differences between scaled scores are not always equal differences in raw score. The distributions of scaled scores are also often bimodal. Raw score is appropriate to use as a replacement for

scaled score because it is rank-ordered identically to scaled scores, and is unimodally distributed.

Section 2: Correlation between MCAS and SAT Scores

Bivariate correlations were computed for corresponding subjects of the MCAS and SAT using the tertiary sample. Parametric and non-parametric correlations were used to determine whether the relationship was linear.

		Verbal SAT Score	MCAS English - Raw Score
Verbal SAT Score	Pearson Correlation	1	.747(**)
	Sig. (2-tailed)		.000
	N	225	225
MCAS ELA - Raw Score	Pearson Correlation	.747(**)	1
	Sig. (2-tailed)	.000	
	N	225	225

** Correlation is significant at the 0.01 level (2-tailed).

			Verbal SAT Score	MCAS English - Raw Score
Spearman's rho	Verbal SAT Score	Correlation Coefficient	1.000	.725(**)
		Sig. (2-tailed)	.	.000
		N	225	225
	MCAS ELA - Raw Score	Correlation Coefficient	.725(**)	1.000
		Sig. (2-tailed)	.000	.
		N	225	225

** Correlation is significant at the 0.01 level (2-tailed).

		Math SAT Score	MCAS Math - Raw Score
Math SAT Score	Pearson Correlation	1	.813(**)
	Sig. (2-tailed)		.000
	N	225	225
MCAS Math - Raw Score	Pearson Correlation	.813(**)	1
	Sig. (2-tailed)	.000	
	N	225	225

** Correlation is significant at the 0.01 level (2-tailed).

			Math SAT Score	MCAS Math - Raw Score
Spearman's rho	Math SAT Score	Correlation Coefficient	1.000	.811(**)
		Sig. (2-tailed)	.	.000
		N	225	225
	MCAS Math - Raw Score	Correlation Coefficient	.811(**)	1.000
		Sig. (2-tailed)	.000	.
		N	225	225

** Correlation is significant at the 0.01 level (2-tailed).

Tables 2.1 – 2.4: Parametric and Non-Parametric Correlations Between MCAS and SAT scores. (Tertiary Sample)

There was not a large difference between the Spearman and Pearson correlation coefficient, suggesting that the correlation between the MCAS scores and SAT scores is linear. Examining scatterplots confirms this. The analysis shows that the correlations between the MCAS score and SAT score is significant and very strong, explaining 56% of the variance for the ELA and Verbal sections and 66% of the variance for the Math sections. The 99% confidence interval for the math scores was (0.745, 0.864). The 99% confidence interval for the ELA and verbal score was (0.660, 0.814). The confidence intervals were calculated using the Fisher's z' transformation.

The following scatterplots show the relation between MCAS and SAT scores.

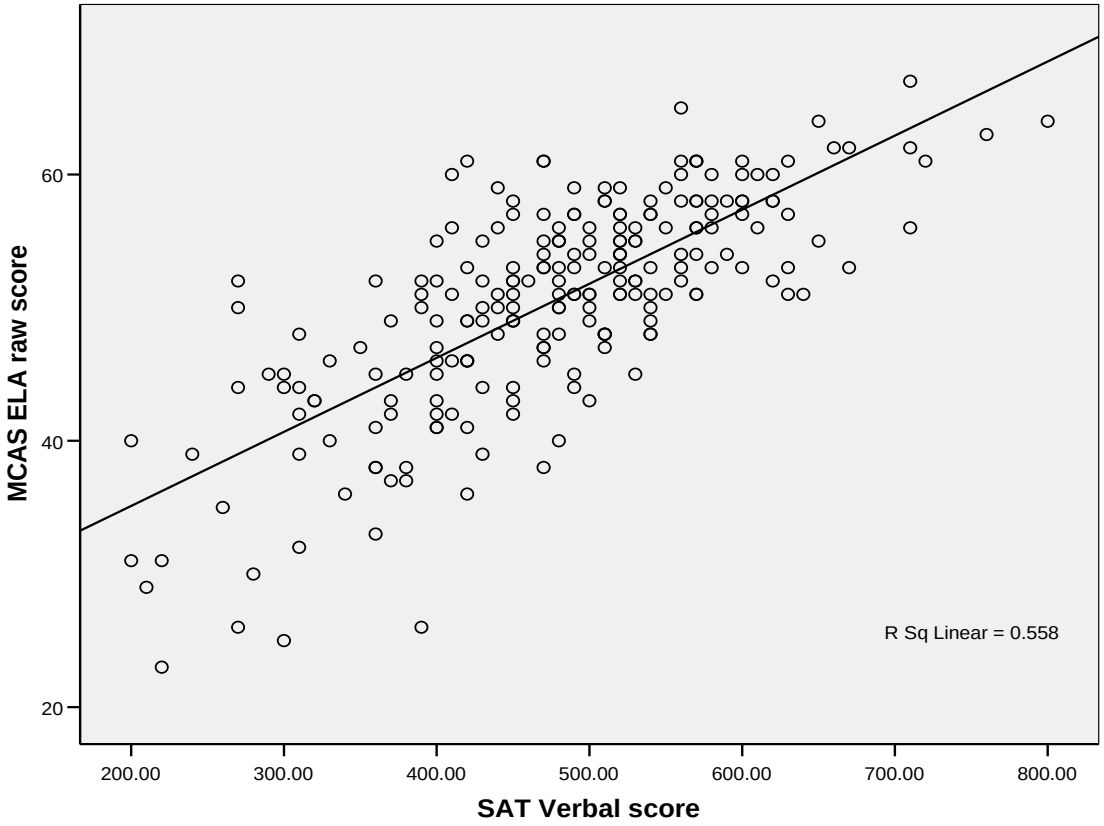


Figure 1.1: SAT Verbal Scores vs. MCAS ELA Raw Scores (Tertiary Sample).

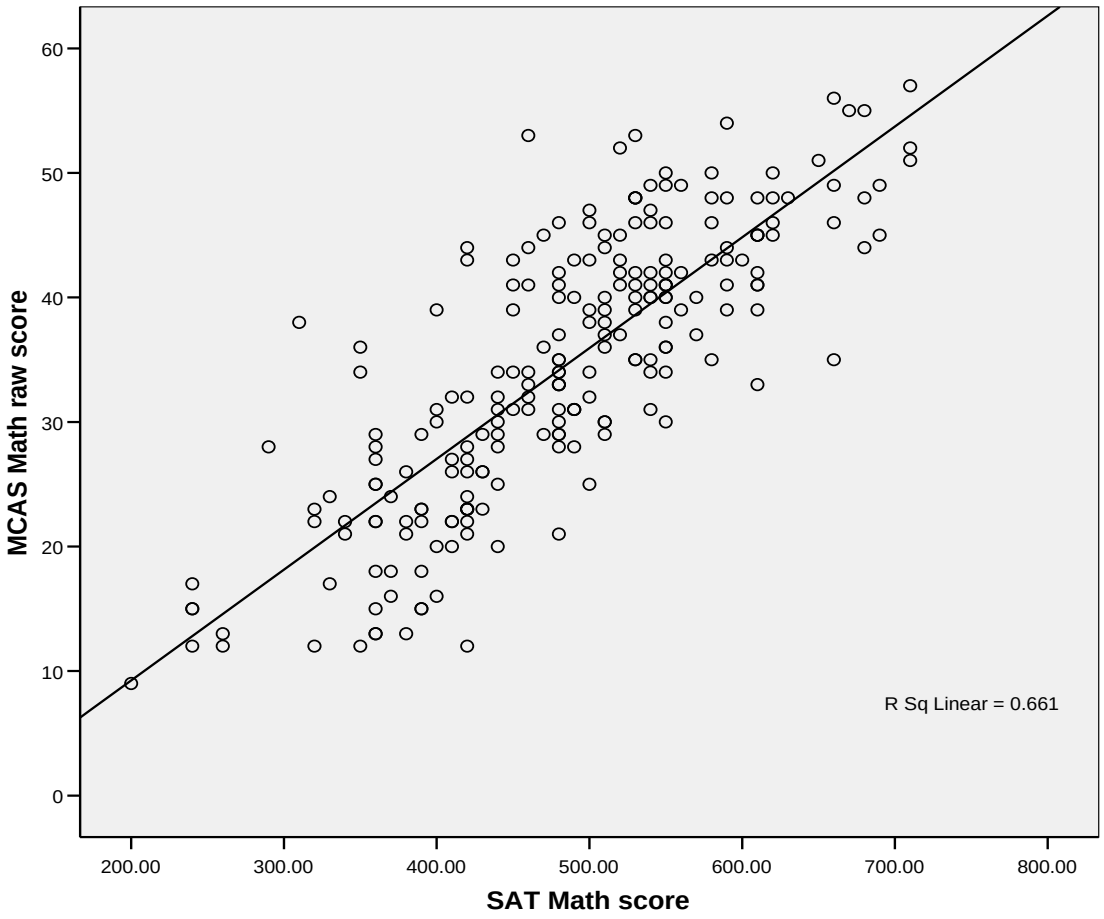


Figure 1.2: SAT Math Scores vs. MCAS Math Raw Scores (Tertiary Sample)

Section 3: MBTI Similarities Between Tests

Ansley et al. (2006) previously analyzed the sensing – intuition MBTI type difference in scores for the MCAS taken by the Class of 2003, but similar analysis had not been done for the SAT scores of the same students prior to this study. The table below has the means for the types in the tertiary sample.

Sensing - Intuition		MCAS Math - Raw Score	MCAS ELA - Raw Score	Math SAT Score	Verbal SAT Score
N	Mean	37.	53.	502.	516.
	N	116	116	116	116
	Std. Deviation	10.	6.	89.	96.
S	Mean	32.	48.	463.	434.
	N	109	109	109	109
	Std. Deviation	12.	9.	108.	105.
Total	Mean	34.	51.	483.	476.
	N	225	225	225	225
	Std. Deviation	11.	8.	100.	108.

Table 3.1: Mean Scores for Sensing and Intuition MBTI Types (Tertiary Sample).

Intuition-type students have higher mean scores on both the MCAS and SAT. Two-sided t-tests showed the difference in scores to be significant at the 0.01 level. For comparison, here are the means for the MCAS from the secondary sample.

Sensing - Intuition		MCAS Math - Raw Score	MCAS English - Raw Score
N	Mean	29.	48.
	N	364	363
	Std. Deviation	12.	9.
S	Mean	25.	43.
	N	418	415
	Std. Deviation	11.	9.
Total	Mean	27.	45.
	N	782	778
	Std. Deviation	12.	9.

Table 3.2: Mean Scores for Sensing and Intuition MBTI Types (Secondary Sample).

T-tests also showed these differences to be significant at the 0.01 level.

An exception to this pattern was found. There was no significant N-type advantage on the math tests for students at Doherty in the tertiary sample. Below are the means for students at Doherty.

Sensing - Intuition		MCAS Math - Raw Score	Math SAT Score
N	Mean	41.	514.
	N	43	43
	Std. Deviation	10.	102.
S	Mean	39.	501.
	N	30	30
	Std. Deviation	10.	107.
Total	Mean	40.	509.
	N	73	73
	Std. Deviation	10.	103.

Table 3.3: Mean Scores for Sensing and Intuition MBTI Types, Doherty Only (Tertiary Sample).

T-tests showed the differences to have p-values both greater than 0.4, meaning that greater than a 40% of all pairs of random samples from the two populations will have a difference that is at least as large as the one in this sample if the population means are equal. The 99% confidence interval for the mean difference of the MCAS math scores was (-4.2, 8.2) and the 99% confidence interval for Math SAT scores was (-53.2, 79.3). To see if the lack of S-N type difference for the MCAS math raw score was also present in all Doherty students in the secondary sample, the means were compared for that sample.

MCAS Math - Raw Score

Sensing - Intuition	Mean	N	Std. Deviation
N	33.	129	13.
S	31.	120	12.
Total	32.	249	13.

Table 3.4: Mean Scores for Sensing and Intuition MBTI Types, Doherty Only (Secondary Sample).

T-tests showed the difference to have a p-value of 0.17. The 99% confidence interval for the mean difference was (-1.9, 6.3). These confidence intervals include the difference in means found for the entire sample, which indicates that the possibility that this is a Type II error, resulting from insufficient statistical power, should not be ruled out. Therefore, no assumptions were made about a relation or lack of relation between S-N type and test scores for Doherty. The positive result on the t-tests for the rest of the schools were still regarded as evidence that N-type students in general have better scores on the MCAS and SAT, with the exception of students at Doherty for the math tests only.

Ansley et al. found no significant or large S-N non-parametric correlation with grades in their study using the MCAS, MBTI, and academics data set. To show that the tertiary sample has the same property, the non-parametric correlation of grades and S-N in the tertiary sample was tested. The strongest correlation was the $r_s = .14$ correlation of N-type with higher grades for English only, significant only at a 0.05 level.

			Sensing - Intuition	math average grade	english average grade
Spearman's rho	Sensing - Intuition	Correlation Coefficient	1.000	-.093	-.143(*)
		Sig. (2-tailed)	.	.163	.032
		N	225	225	224

* Correlation is significant at the 0.05 level (2-tailed).

Table 3.5: Non-Parametric Correlations Between S-N Type and Average Grades. (Tertiary Sample)

Section 4: Correlation Between MCAS Score and Grades

The correlation between average English grades and MCAS ELA scores was 0.387 for the entire secondary sample. The correlation for average math grades and MCAS math score for the same sample was 0.537. Both were significant at the 0.01 level. It was hypothesized that the correlation between average grades and MCAS score would be confounded by differences in academic difficulty between schools and courses, and that the correlation would be greater when students were analyzed separated by school and average course level. The secondary sample was used to test this. It was found that this was generally true only for students in the highest level courses, with average course level 2. This group is composed of students who mainly took honors-level courses in math or English.

School	Correlation of average math grade and MCAS math raw score (# of students)	Correlation of average English grade and MCAS ELA score (# of students)
Burncoat	0.554** (88)	0.528** (91)
Doherty	0.704** (130)	0.574** (139)
North	0.320 (42)	0.412** (51)
South	0.755** (53)	0.630** (52)
Voc. Tech.	0.622** (22)	0.137 (25)

** Correlation is significant at the 0.01 level (2-tailed).

Table 4.1: Correlation of Average Grade and MCAS Score for Average Subject Level 2 (Secondary Sample).

It cannot be ruled out that there was a real correlation between math grades and MCAS math scores for North High School of similar magnitude to the correlation for the other schools, as the sample for that test had only 42 students, and the 99% confidence interval was (-0.08, 0.631). Worcester Technical may have had a real correlation for the English grade and ELA score equal to or greater than the correlation for the entire

secondary sample, as the small sample size of 25 students resulted in a 99% confidence interval of (-0.389, 0.596).

For Math and English, it was found that there were generally no 0.01 significant correlations between average subject grade and the raw MCAS score of the same subject for students in lower course levels. The only exceptions in this study were students in average math course level 3 at Doherty and Accelerated Learning Laboratory, and students at Worcester Technical in average math course level 3 and 4. Those correlations were 0.629, 0.514, 0.388, and 0.379, respectively. The sample sizes for those groups were, respectively, 97, 40, 79, and, 52 students.

Many samples tested for the schools at lower course level were small, so 99% two-sided confidence intervals were calculated to see if correlations equal to or greater than that found for the entire population were possible, in order to rule out the possibility of a practically significant Type II error.

Doherty and Worcester Technical were the only schools with confidence intervals that were below the correlation of the entire population for average English course level 3, with upper limits of 0.335 for 89 students sampled and 0.348 for 87 students sampled respectively. For average English course level 4, Doherty had an upper limit of 0.353 for 20 students, and Worcester Technical had an upper limit of 0.230 for 41 students. Therefore it can only be concluded for students in sub-honors level English curricula at Doherty and Worcester Technical that there was not an English grades-MCAS ELA correlation equal to or greater than that found for the entire secondary sample.

Of the schools lacking a significant correlation for math grades and scores for average math course level 3 and 4, only Burncoat for average math course level 3 had

confidence intervals below the population correlation, with an upper limit of 0.521 for the 55 students sampled. Therefore only Burncoat can be concluded with certainty to lack a stronger correlation between math grades and MCAS math scores for students in average math course level 3 than for all students in the secondary sample.

Scatter-plots of grades and raw MCAS score for different course levels and schools can be found in Appendix B.

Section 5: Multiple Regression of MCAS Score With Grades and SAT Score

It was hypothesized that the SAT and grades each had some independent correlation to the MCAS score while having limited correlation to each other, so multiple regressions were done for the demographics of students in the groups controlled for curriculum difficulty as described in Section 4, for the three largest control groups: Burncoat, Doherty, and South High. The regression was done step-wise using an F-test at the 0.01 level to determine inclusion of an additional independent variable. The regression details for the different schools can be found in Appendix C. Tolerances for the regressors, average subject grade and SAT subject score, were not low, so multicollinearity was not a problem in regression. Below are the part correlations for average grades and SAT score for each school, and the R^2 for each model.

School	Part correlation – average English grade	Part correlation – SAT Verbal score	R^2
Burncoat	0.339	0.380	0.568
Doherty	0.316	0.523	0.590
South High	0.337	0.419	0.565

Table 5.1: Part Correlations and R^2 for Regression Models of MCAS ELA Raw Score.

School	Part correlation – average math grade	Part correlation – SAT Math score	R^2
Burncoat^	0.159	0.628	0.718
Doherty	0.466	0.321	0.687
South High	0.312	0.360	0.745

^The p-value for inclusion of average math grade was 0.022, these values comes from accepting entry at a 0.05 level of significance instead of 0.01. The zero-order correlation for SAT Math score was 0.832.

Table 5.2: Part Correlations and R^2 for Regression Models of MCAS Math Raw Score.

The part correlations, also known as semi-partial correlations, indicate the correlation of the regressor to the dependent variable after removing the correlation of the regressor to all other regressors in the model. The square of the part correlation is equal to the change in variance explained when the predictor is added to the model. When there is moderate correlation between the regressors, as is the case for average grade and SAT score, part correlations are more appropriate to use than the simple bivariate correlations to compare the correlations of regressors to the dependent variable.

The part correlation of SAT score was greater than that of average grade for almost all models. Even accounting for sampling variation, this implies that the correlation of SAT scores and MCAS scores that was independent of average grades was likely to be equal to or greater than the independent correlation for average grades and MCAS scores.

Section 6: Analysis of Subsections

The scores on the multiple choice subsection of the English MCAS were strongly correlated to the scores on the subsections of the Verbal SAT. The same was true of the multiple choice subsections of the Math MCAS and Math SAT.

		SAT Verbal Critical Reading Raw Score	SAT Verbal Analogies Raw Score	SAT Verbal Sentence Completion Raw Score	MCAS ELA MC - Raw Score
SAT Verbal Critical Reading Raw Score	Pearson Correlation	1	.711(**)	.720(**)	.690(**)
	Sig. (2-tailed)		.000	.000	.000
	N	225	225	225	225
SAT Verbal Analogies Raw Score	Pearson Correlation	.711(**)	1	.712(**)	.627(**)
	Sig. (2-tailed)	.000		.000	.000
	N	225	225	225	225
SAT Verbal Sentence Completion Raw Score	Pearson Correlation	.720(**)	.712(**)	1	.675(**)
	Sig. (2-tailed)	.000	.000		.000
	N	225	225	225	225
MCAS ELA MC - Raw Score	Pearson Correlation	.690(**)	.627(**)	.675(**)	1
	Sig. (2-tailed)	.000	.000	.000	
	N	225	225	225	225

** Correlation is significant at the 0.01 level (2-tailed).

Table 6.1: Correlation of English MCAS Multiple Choice Subsection to Verbal SAT Subsections (Tertiary Sample).

		SAT Math Arithmetic and Algebraic Reasoning Raw Score	SAT Math Geometric Reasoning Raw Score	MCAS Math MC - Raw Score
SAT Math Arithmetic and Algebraic Reasoning Raw Score	Pearson Correlation	1	.813(**)	.747(**)
	Sig. (2-tailed)		.000	.000
	N	225	225	225
SAT Math Geometric Reasoning Raw Score	Pearson Correlation	.813(**)	1	.663(**)
	Sig. (2-tailed)	.000		.000
	N	225	225	225
MCAS Math MC - Raw Score	Pearson Correlation	.747(**)	.663(**)	1
	Sig. (2-tailed)	.000	.000	
	N	225	225	225

** Correlation is significant at the 0.01 level (2-tailed).

Table 6.2: Correlation of Math MCAS Multiple Choice Subsection to Math SAT Subsections (Tertiary Sample).

These correlations are as large as or larger than the correlations of the multiple choice subsections' raw score to the other subsections' raw score from the subject. An exception is the MCAS math open ended section, which also has a significant 0.768 correlation to the SAT Math score.

		MCAS Math Short Answer - Raw Score	MCAS Math Open Ended - Raw Score	MCAS Math MC - Raw Score
MCAS Math Short Answer - Raw Score	Pearson Correlation	1	.608(**)	.655(**)
	Sig. (2-tailed)		.000	.000
	N	225	225	225
MCAS Math Open Ended - Raw Score	Pearson Correlation	.608(**)	1	.742(**)
	Sig. (2-tailed)	.000		.000
	N	225	225	225
MCAS Math MC - Raw Score	Pearson Correlation	.655(**)	.742(**)	1
	Sig. (2-tailed)	.000	.000	
	N	225	225	225

** Correlation is significant at the 0.01 level (2-tailed).

Table 6.3: Correlations of Math MCAS Subsections (Tertiary Sample)

		ELA MC - Raw Score	ELA Open Ended - Raw Score	MCAS Composition - Raw Score	MCAS Writing Prompt - Raw Score
ELA MC - Raw Score	Pearson Correlation	1	.621(**)	.501(**)	.454(**)
	Sig. (2-tailed)		.000	.000	.000
	N	225	225	225	225
ELA Open Ended - Raw Score	Pearson Correlation	.621(**)	1	.452(**)	.508(**)
	Sig. (2-tailed)	.000		.000	.000
	N	225	225	225	225
MCAS Composition - Raw Score	Pearson Correlation	.501(**)	.452(**)	1	.688(**)
	Sig. (2-tailed)	.000	.000		.000
	N	225	225	225	225
MCAS Writing Prompt - Raw Score	Pearson Correlation	.454(**)	.508(**)	.688(**)	1
	Sig. (2-tailed)	.000	.000	.000	
	N	225	225	225	225

** Correlation is significant at the 0.01 level (2-tailed).

Table 6.4: Correlations of ELA MCAS Subsections (Tertiary Sample)

The sum of the raw scores of the non-multiple choice subsections of each MCAS section still had significant correlations to the corresponding SAT section. The correlations were still strong, with the math non-multiple choice subsections having a 0.78 Pearson's correlation coefficient with the Math SAT section, and the ELA section and English SAT having a Pearson's correlation coefficient of 0.60.

Structure detection was done on the tertiary sample for average course level 2 using principal axis factoring. The variables chosen were all the subsections of both tests and the average grades for both subjects. Average course level was fixed at 2 to control for curriculum differences that affect grades-MCAS relation as described in Section 4. Three factors had eigenvalues greater than 1 and explained 69% of the variance. A Scree Plot showed that these three factors were the only major factors resulting from the structure detection, with the variance explained by each factor much greater than that of the unused factors. They were selected for extraction, and 9% of the variance explained was lost during extraction because of correlation only between individual variables. The three factors were rotated with the varimax method, and after rotation the first two factors each explained about 20% of the variance and the third factor explained 16% of the variance.

	Factor		
	1	2	3
SAT Verbal Critical Reading Raw Score	.216	.748	.367
SAT Verbal Analogies Raw Score	.256	.724	.158
SAT Verbal Sentence Completion Raw Score	.151	.781	.177
SAT Math Arithmetic and Algebraic Reasoning Raw Score	.735	.497	.151
SAT Math Geometric Reasoning Raw Score	.666	.446	-.014
MCAS Math Short Answer - Raw Score	.619	.021	.215
MCAS Math Open Ended - Raw Score	.754	.193	.296
MCAS Math MC - Raw Score	.716	.354	.146
MCAS Writing Prompt - Raw Score	.180	.121	.632
MCAS Composition - Raw Score	.169	.250	.613
MCAS ELA Open Ended - Raw Score	.082	.185	.675
MCAS ELA MC - Raw Score	.257	.617	.391
English average grade	.372	.187	.649
math average grade	.659	.125	.339

Table 6.5: Structure Detection of Subsections and Grades

Factor 1 is mainly linked to the math subsections and math average grade. Factor 2 is linked to the ELA multiple choice subsections for both tests, with weaker linkage to the math multiple choice subsections. Factor 3 is linked to the essay and short answer subsections of the ELA MCAS, as well as the English average grade.

A result of this analysis was that the math open-ended and multiple choice subsections on the MCAS Math test are very similar in score and relation to the SAT Math. All three of those scores are also correlated strongly to average math grade. Another result was a dissimilar finding for the MCAS ELA multiple choice and non-multiple choice subsections. The ELA multiple choice subsection is strongly correlated to

SAT score, while the essay and open-ended subsections are strongly correlated to average English grade. This suggests that the MCAS ELA test is actually two different tests; a multiple choice test that is similar to the SAT Verbal, and an essay and open-response test that is more correlated to English grades than to the SAT Verbal.

Conclusions

This study confirmed the hypothesis that scores on the official MCAS would correlate strongly with SAT scores. In addition, it showed that one psychological preference associated with SAT scores, the S-N measure of the MBTI, is also associated in the same way with MCAS scores. Another important finding of this study was that the magnitude of the correlation of grades to MCAS score was less than the magnitude of the correlation of MCAS scores with SAT scores. It was also determined that format similarities between the MCAS and SAT may be a major source of their correlation, but do not comprehensively account for the similarity in scores.

The strength of correlation between MCAS scores and SAT scores alone is worthy of note given the contrasting objectives of the tests. Overall, 56% of the variance in the ELA MCAS score was explained by the Verbal SAT score, and 66% of the Math MCAS score variation was explained by the Math SAT score. Furthermore, N-type students, as determined with the MBTI, have better scores on both tests than S-types, while no such advantage was found for average subject grade for the first two years of student's high school transcript in either subject. This is one point supporting the possibility that at least some of the similarity in scores is due to properties or factors inherent to the tests and test-takers and unrelated to differences in academic achievement.

It may be hypothesized that the intuitive advantage is connected to the violation of the assumptions of formal scoring found by Higham. N-type test-takers may be better at eliminating alternative answers and may be more prone to guess answers than S-types. However, an advantage on multiple choice questions cannot fully explain the correlation

of MCAS and SAT scores, as comparing the score on all non-multiple choice MCAS problems to the score on the multiple choice SAT still indicates strong correlations, especially for the MCAS Math open-ended subsection.

A hypothesis explaining the broad correlation of the MCAS Math with the SAT Math is that the tests have the same background. Both are primarily tests of subjects usually taught by the end of 10th grade: basic algebra and high school geometry. Students who did well in math classes teaching those subjects are likely to do well on the SAT Math and MCAS Math. If this hypothesis is correct, one consequence would be that the SAT Math and MCAS Math have a stronger correlation to average math grade for grades 9 and 10 than for grades 11 and 12, as both tests are hypothetically related to student's skill in basic high school mathematics instead of advanced high school mathematics.

Directly comparing the correlation of SAT scores and average grades to MCAS scores was informative on the nature of the relation between the three. The strong correlation of the SAT with the MCAS would be of less concern if grades had a stronger correlation with the MCAS. However, for the groups that had strong correlation between the MCAS and grades, the students in honors-level courses, multiple regression showed that, after controlling for the correlation between average grades and SAT score, SAT scores typically had a stronger correlation with the MCAS scores than the correlation of average grades with MCAS scores. Even if the independent correlations to the MCAS of grades and SAT scores are only equal, this supports the hypothesis that the MCAS-SAT correlation is larger in scope than a shared relation to academic performance. This implies that the 2001 MCAS was partially an aptitude test. However, the correlation of SAT score and grades is not evidence against the pre-revision SAT also having elements

of an achievement test. It could show a relation between academic achievement and SAT score, if the correlation between grades and the SAT is not just due to greater aptitude resulting in higher grades. Given this triangle of relationships, it is difficult to determine what cognitive and academic abilities and skills are the causal factors for the scores on each test and academic grades, and the relative magnitude of those effects.

The incidental finding that MCAS scores were not likely to be at least moderately correlated to average grades for most students in college and vocational-technical courses was unexpected, and should be a major concern for the Worcester public schools. It was initially hypothesized that this was due to a problem with the students taking the courses; that some students may be learning the material well but not making an effort to do well on homework or in-class tests. If this was the case, it would be expected that students with higher grades in the lower level of courses would have less variance and higher scores on the MCAS than students with the lowest grades in the same level of courses. This was not the case, as shown in Appendix B.

Another hypothesis is that the curricula of the less challenging courses did not meet the standards of the curriculum framework, so students could achieve high grades while not learning enough of the material tested on the MCAS to score well. If this is true, school administrations should use similar analyses to identify classes that are not teaching students enough for them to perform well on the MCAS. They should then revise the curricula of those courses to meet the standards of the curriculum framework, which should result in the students with high grades in these courses consistently doing well on the MCAS.

The possibility that the lack of relationship to grades is a problem with the MCAS should not be excluded. Goldsmith (2002) criticized the MCAS ELA test in this study for unrealistic expectations of student ability by age. If the test actually is connected to how advanced the abilities of the students are, then the expected pattern would be for students in more advanced courses who presumably have more advanced abilities to have a strong correlation between academic performance and MCAS score. On the other hand, the success of students in less advanced courses would depend on how advanced their ability is for their age, possibly more than their academic performance. Therefore, a clear relation between grades and MCAS score would not exists for students in less advanced courses. This is what is observed, particularly for the ELA test criticized by Goldsmith.

A possible source of procedural error, rather than statistical error, in the analysis involving grades and MCAS scores is that average subject grades, controlled for average curriculum difficulty and school as in this study, may be far from an optimal measure of academic achievement. If this is the case, a better measure of academic achievement could have a much higher correlation with MCAS score. The part correlation of SAT scores in the multiple regression of MCAS score, with this measure as the other predictor, might also be close enough to zero to not be important, if the correlation between the MCAS and SAT was entirely due to the correlation of the SAT and academic achievement, a correlation indicated by numerous studies. A caveat to this conjectured error is that this study and prior research has found no practically significant N-type advantage for high school grades, while an N-type advantage for SAT and MCAS scores has been found. That research is strong evidence for at least a weak connection between the SAT and the MCAS that has nothing to do with academic achievement.

Lastly, the factor analysis does point towards an overall theory of the relationship of grades and tests, and the skills and abilities responsible for scores and grades. The variance explained is only 60%, so other factors or random errors still have a large effect on the individual grades and scores, but the structure detection does indicate that there are three principal areas of tests and academics. First, there is a general math factor that is related to scores on all math tests and grades on math courses. There is an English multiple choice factor that is related to the SAT Verbal and MCAS ELA multiple choice section, and a factor that is related to the non-multiple choice sections of the MCAS ELA and English grades. The general math factor connecting the SAT Math score and math grades is consistent with the fact that the SAT Math test is based on math teaching up to the 10^{th} grade.

The most important results of this study are that in 2001 the official MCAS test had some biases in scores that were unrelated to academic achievement, and that there was a correlation to the SAT score, a test that is meant to measure aptitude. The correlation is not fully explained by the grades of those students, and the SAT had a relation to the MCAS that was at least as important as grades. Even more troubling was the lack of relationship between MCAS score and course grades for many students in college and technical courses. In 2001, the MCAS still needed improvement to be a fair test of academic achievement and reduce the possible linkage to aptitude, and much work was still needed to improve the lower-level high school courses in urban schools to meet the standards of the curriculum framework. Whether that work has been done is a question to be answered by future studies.

Areas For Future Research

One of the two major limitations of this study is that it exclusively studied urban schools in a single city. Also, none of the schools had a high average level of performance on the MCAS, with over 30% of the Class of 2003 in the entire system failing the test on their first try (after retakes and appeals of the test, only 11% of the Class of 2003 in Worcester public high schools failed to graduate because they failed the MCAS). A study of similar design sampling rural or suburban schools and schools with better performance on the MCAS would have an important role in determining how typical the SAT-MCAS correlation is. Also, it would allow comparisons with the MCAS-grades correlations in this study. Schools with higher scores on the MCAS may have greater correlations between grades and MCAS score, and could tell whether the lack of correlation for students in sub-honors level courses at various schools is unique to Worcester, or if it is common statewide.

The other major limitation on this study is that self-selection by students taking the SAT restricts the academic diversity of students taking the SAT in Worcester public high schools. For this reason, only the students in the highest academic level of courses in math and English were studied in great detail. A study that used the PSAT, which is taken by most juniors in some schools, instead of the SAT, would provide analogous data for students in less challenging academic curricula. Also, in non-urban high schools a higher proportion of students take the SAT. A question that I think needs to be answered is if the MCAS still has a strong correlation to SAT scores for students who are not in the most challenging level of courses offered. Two reasons for this are that most students in Massachusetts public high schools take “college” courses instead of honors or AP

courses, and a correlation between the MCAS and PSAT scores for this group could give conclusive evidence for the SAT's relation to the MCAS being independent of academic achievement, since no correlation between grades and MCAS scores has been found for most students in this group by this study of Worcester public high schools. Studies of high schools with better MCAS performance should take into consideration that the difficulty of courses at those schools may differ greatly from that of Worcester high school courses with the same name. Groups of students in an average course level in those schools may not be directly comparable to groups of students in the same average course level in Worcester high schools.

This study had a few deficiencies in the data available. It is estimated that nearly 300 more students in the Class of 2003 in Worcester public schools took the SAT, and finding that data could improve and increase the number of regression models predicting the MCAS score. More SAT data for Doherty could help determine whether the lack of relation between the test scores and S-N preference is a finding of the null hypothesis or a Type II error. Also, MBTI and academic data was not available for many students who failed the MCAS. What relevance this has to the current study cannot be known without better data coverage in the future. Lastly, there was not academic data for enough students to have sufficient statistical power to conclusively identify all the course levels that did or did not have a strong correlation between the MCAS and grades for each school, because the best way to control for differences in course difficulty and grading is by separating the students by average course level and school.

The Worcester Class of 2006 and onwards should be studied in a similar manner as this study to see if more recent MCAS tests still have an N-type advantage, if students

in lower course levels still do not have a strong correlation between MCAS score and grade, and if the revised SAT, with an essay section, correlates strongly to the MCAS. It is my hypothesis that the addition of an essay to the Verbal section will result in even more correlation of the MCAS ELA to the SAT Verbal. The correlation between grades and MCAS scores for students in lower course levels should improve over time due to the courses being improved to decrease the failure rate of students taking the MCAS, if the lack of correlation is caused by lower level curricula failing to meet the Curriculum Framework standards,.

Appendix A

The histograms below are the distribution of all scaled MCAS scores for each subject in the sample for Burncoat and Doherty, with normal curve. “512” is the code for Doherty, and “503” represents Burncoat.

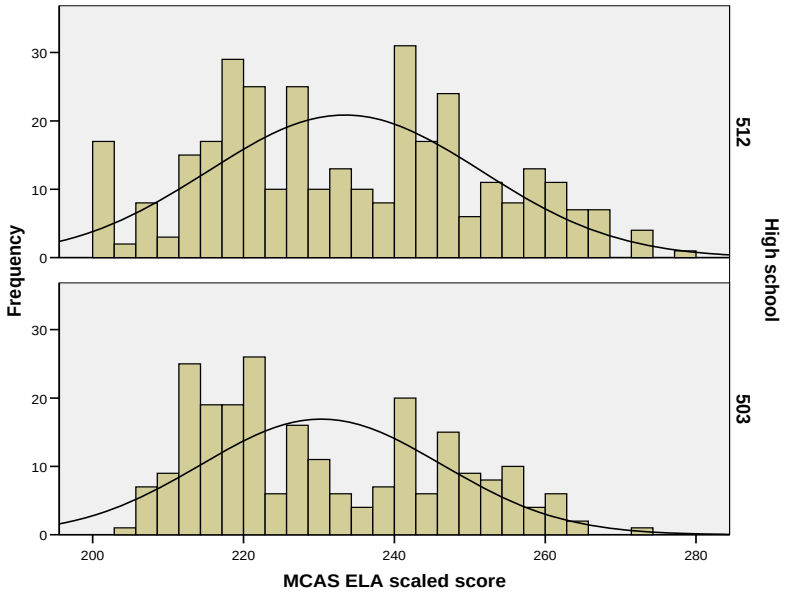


Figure A.1: Histogram of MCAS ELA Scaled Score for Doherty (512) and Burncoat (503).

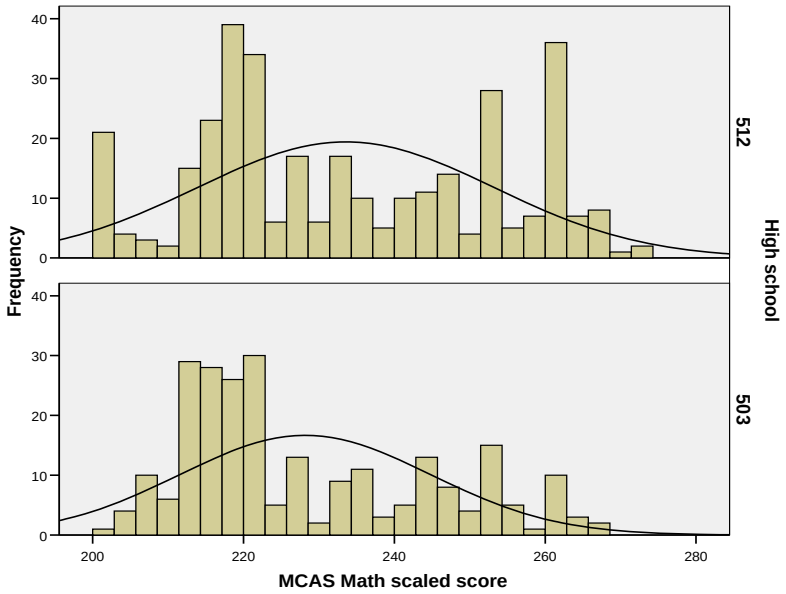


Figure A.2: Histogram of MCAS Math Scaled Score for Doherty (512) and Burncoat (503).

These histograms show all available MCAS raw scores for the same schools. The highest possible score on the ELA section is 72, and the highest possible score on the Math section is 60.

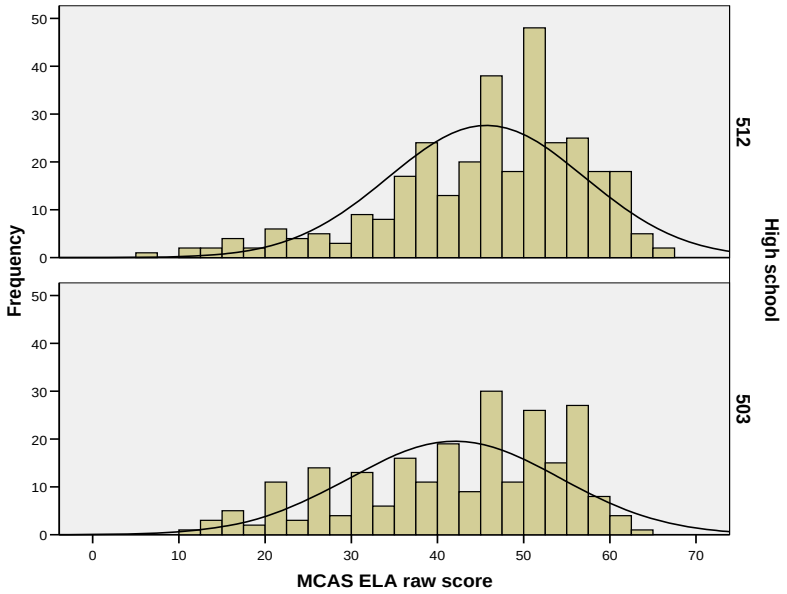


Figure A.3: Histogram of MCAS ELA Raw Score for Doherty (512) and Burncoat (503).

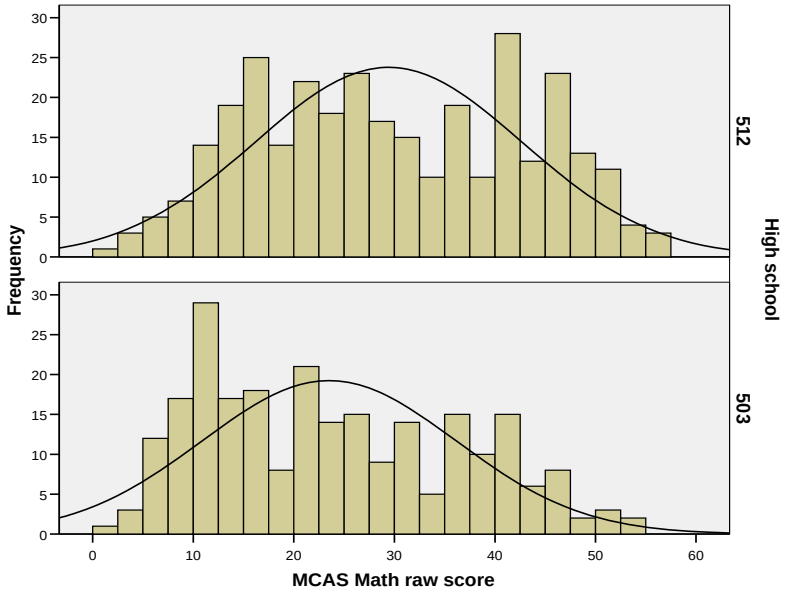


Figure A.4: Histogram of MCAS Math Raw Score for Doherty (512) and Burncoat (503).

The scatterplots below show raw score on the MCAS vs. scaled score for all scores.

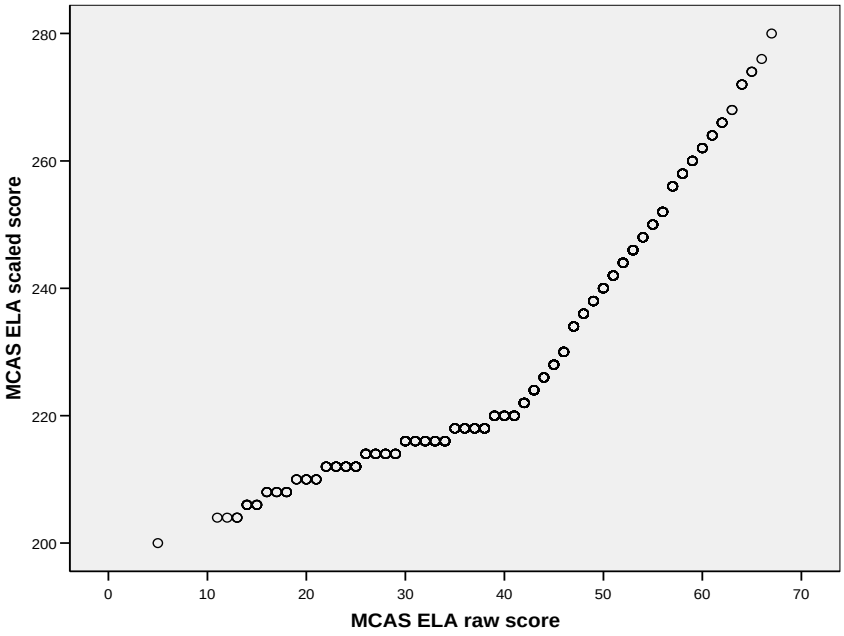


Figure A.5: Scaling of MCAS ELA Raw Scores

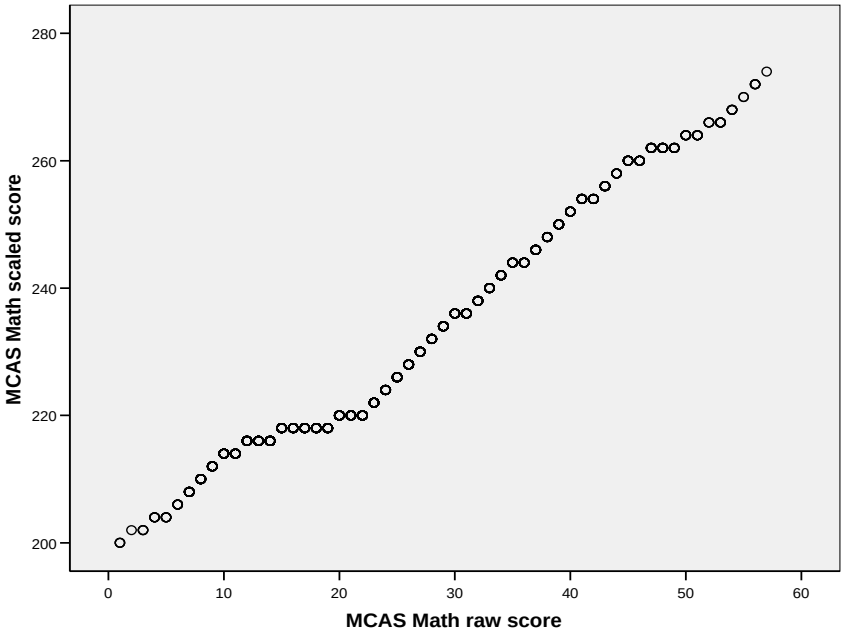


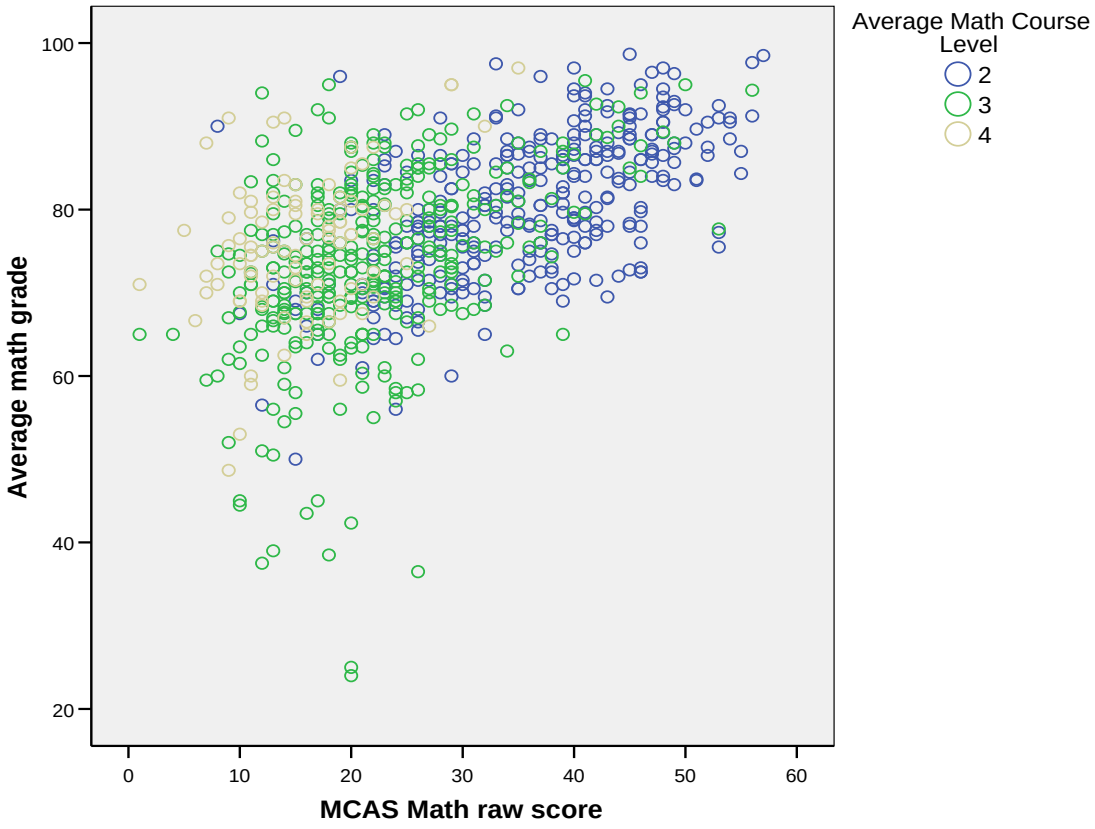
Figure A.6: Scaling of MCAS Math Raw Scores

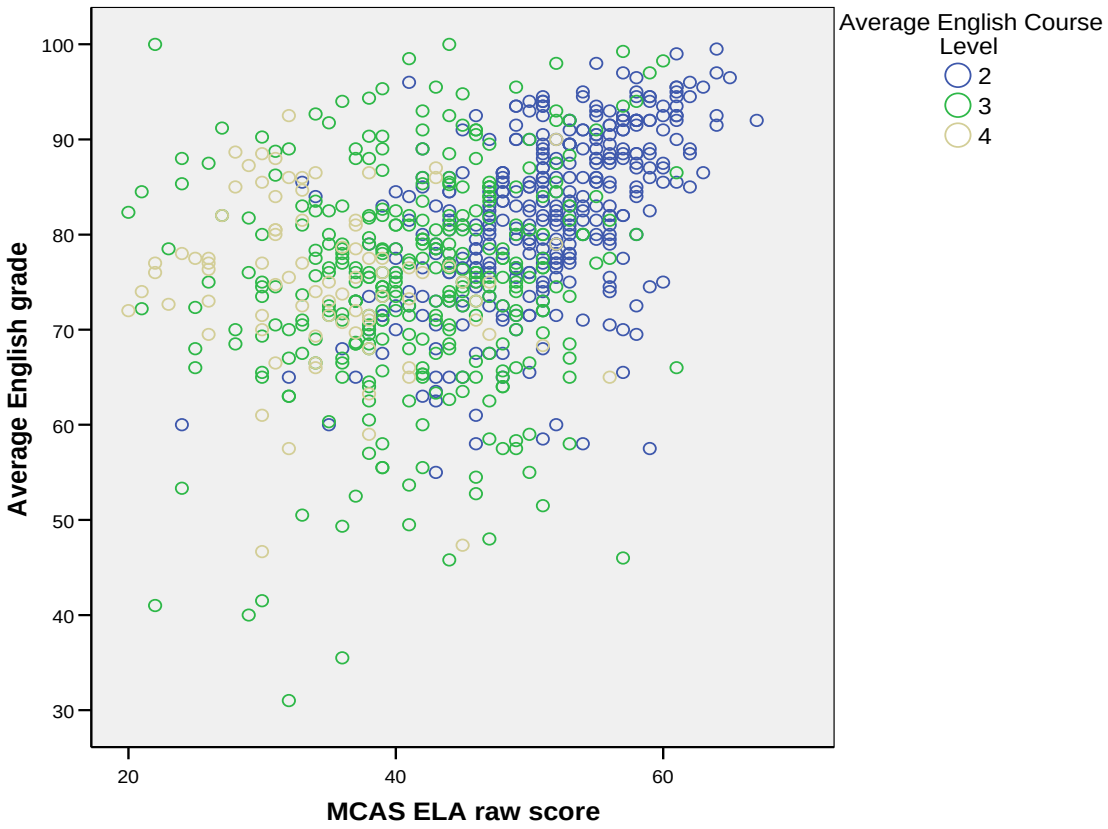
From the scatterplots, it is clear that the scaling of raw scores is not linear. The scaling is rank-ordered identically to the raw scores though, which means the points from the raw score are weighted equally. Combined, this means that scaled scores are not a scale variable because equal numeric differences in scaled scores do not represent an equal difference in performance on the MCAS.

Although there is still significant skew in the distribution of raw scores, their unimodality and scale property make them preferable to scaled scores for most statistical quantitative analysis. The large sample sizes of the analyses in this study help ensure that deviations from normality do not skew the distribution of sampling test statistics

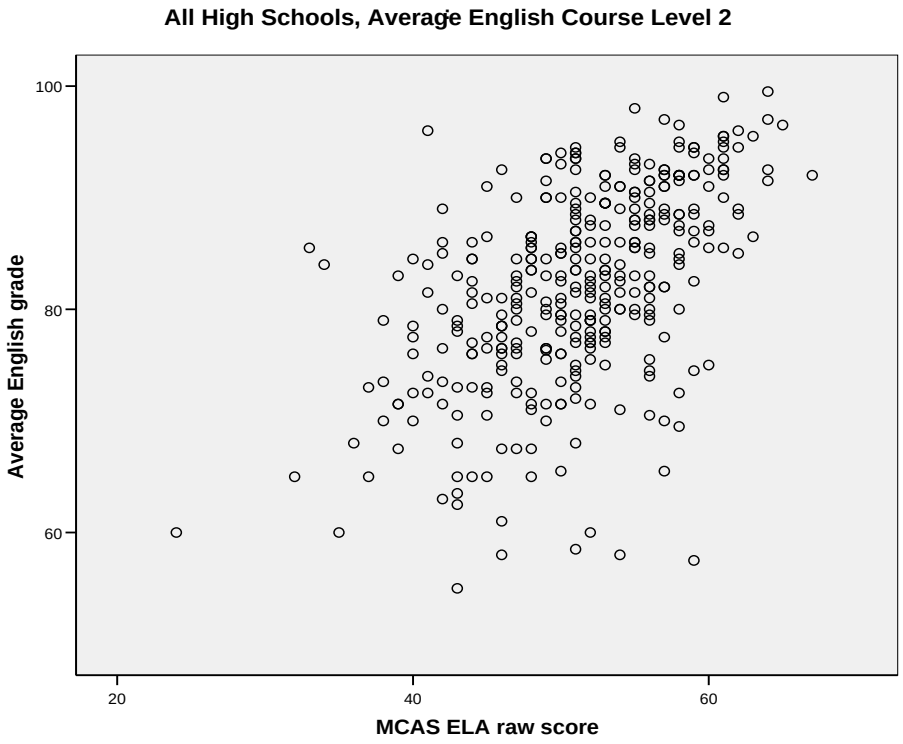
Appendix B

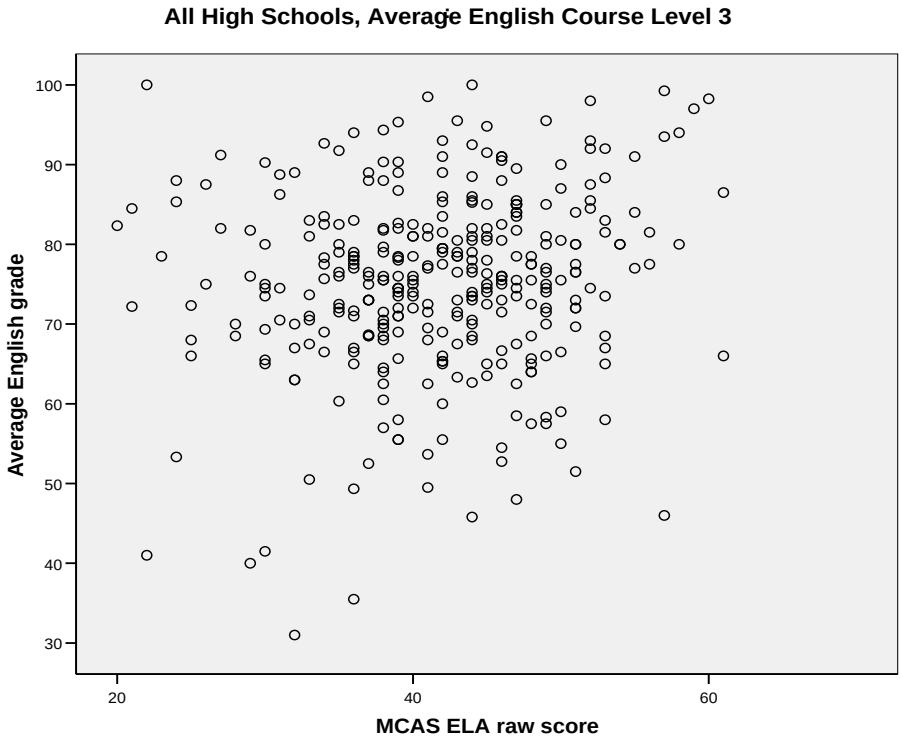
The scatterplot below shows the bivariate distribution of grades and MCAS raw scores for all students in the secondary sample.

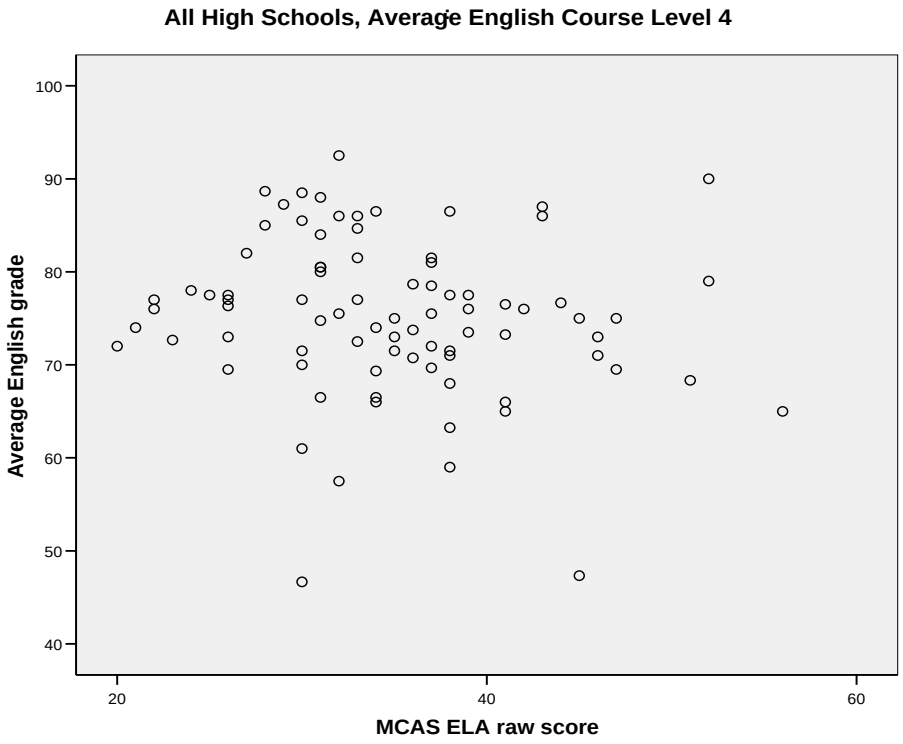


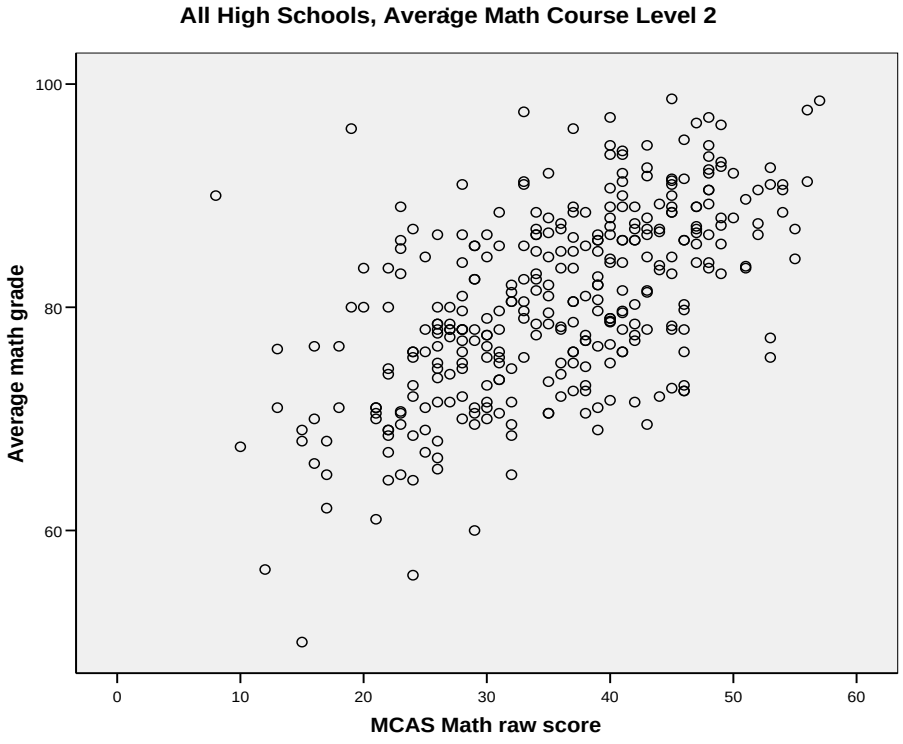

Average English Course Level
2
3
4
Average English grade
100
90
80
70
60
50
40
30
20
40
60
MCAS ELA raw score

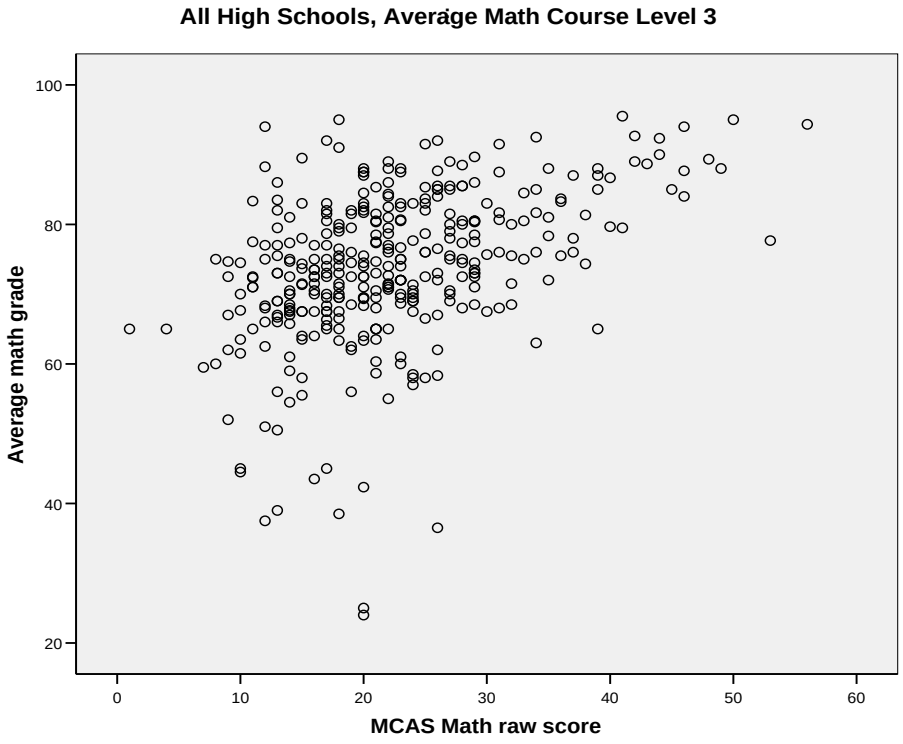
The following scatterplots show the bivariate distribution for each course level for all schools.

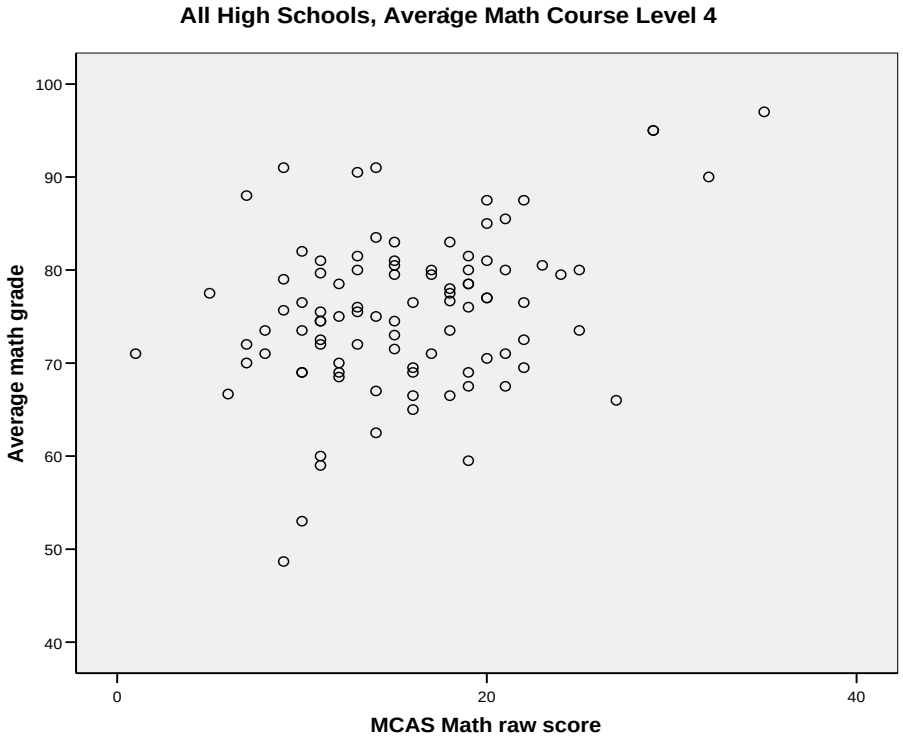




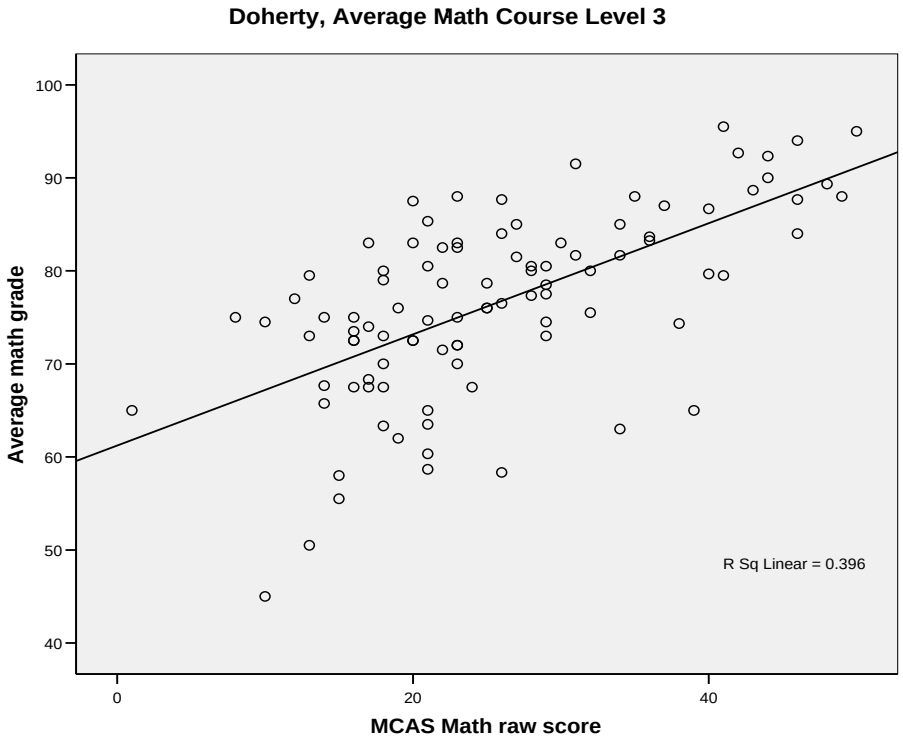

All High Schools, Average English Course Level 4
Average English grade
100
90
80
70
60
50
40
20
40
60
MCAS ELA raw score


All High Schools, Average Math Course Level 2
Average math grade
100
80
60
0
10
20
30
40
50
60
MCAS Math raw score

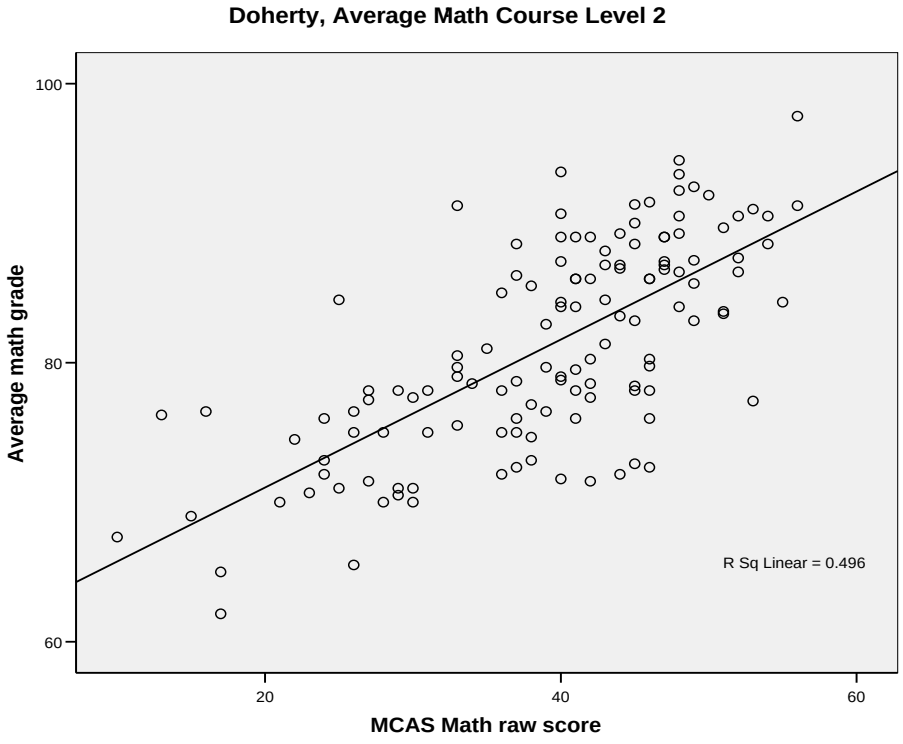


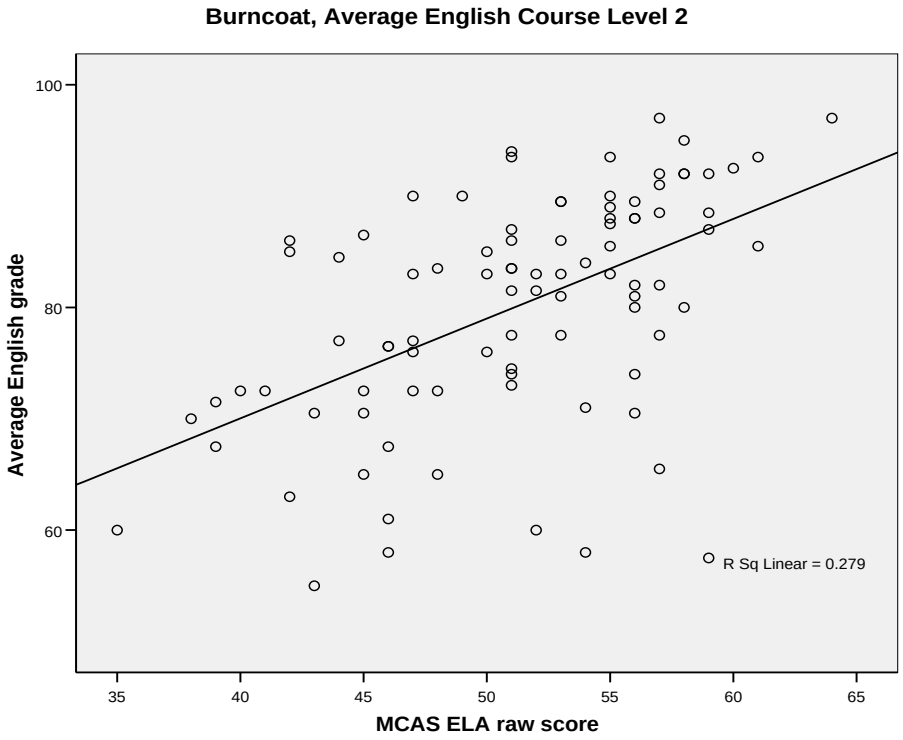


The points on the upper right for average math course level 3 are almost entirely from Doherty.

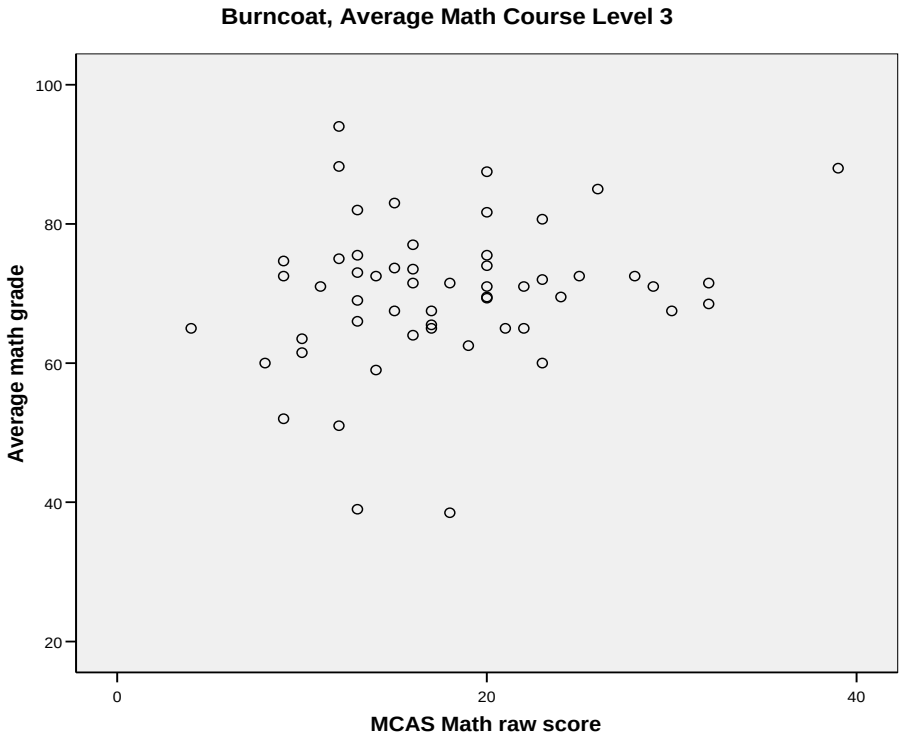


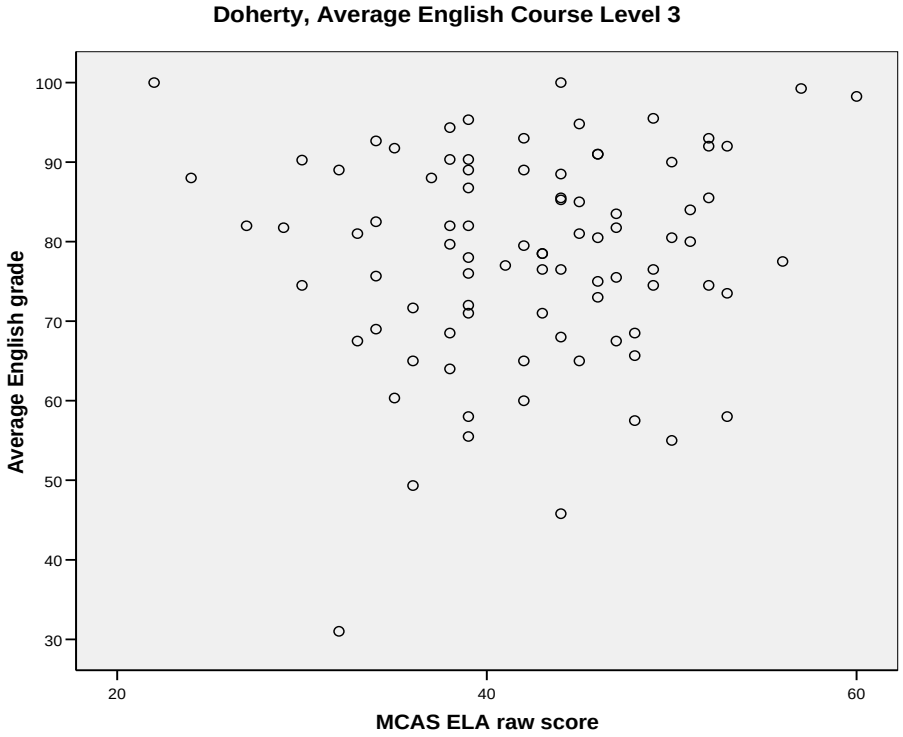
Below are examples of the grades-MCAS relation in the control groups that had significant correlations.



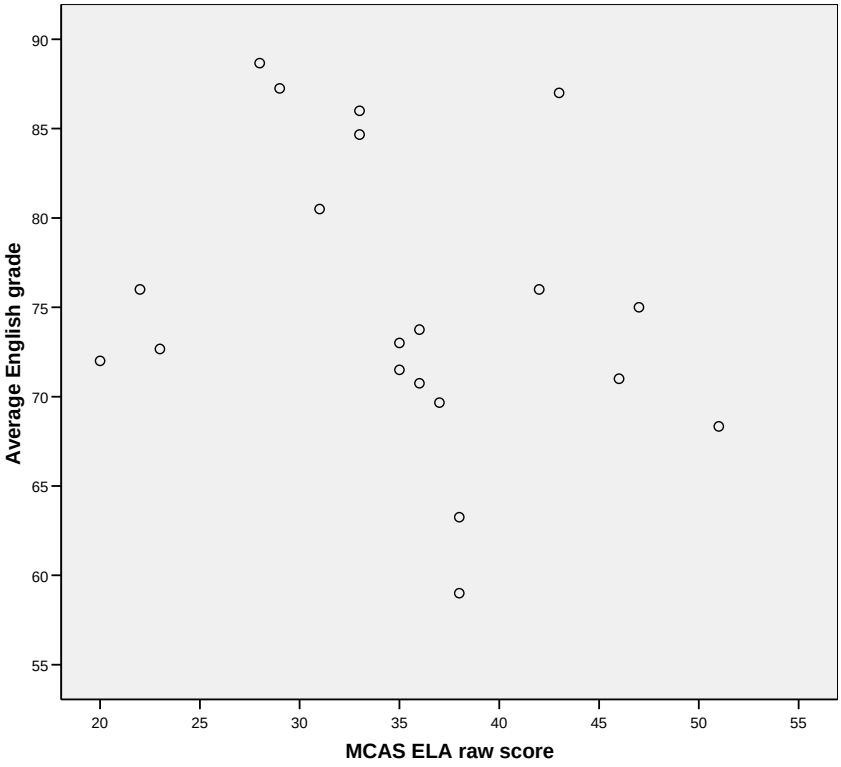


The following five graphs are of groups which did not have a significant correlation between grades and the MCAS, and were likely to lack a real correlation of practical importance. Note that there is no pattern of correlation with MCAS scores emerging as grades increase.

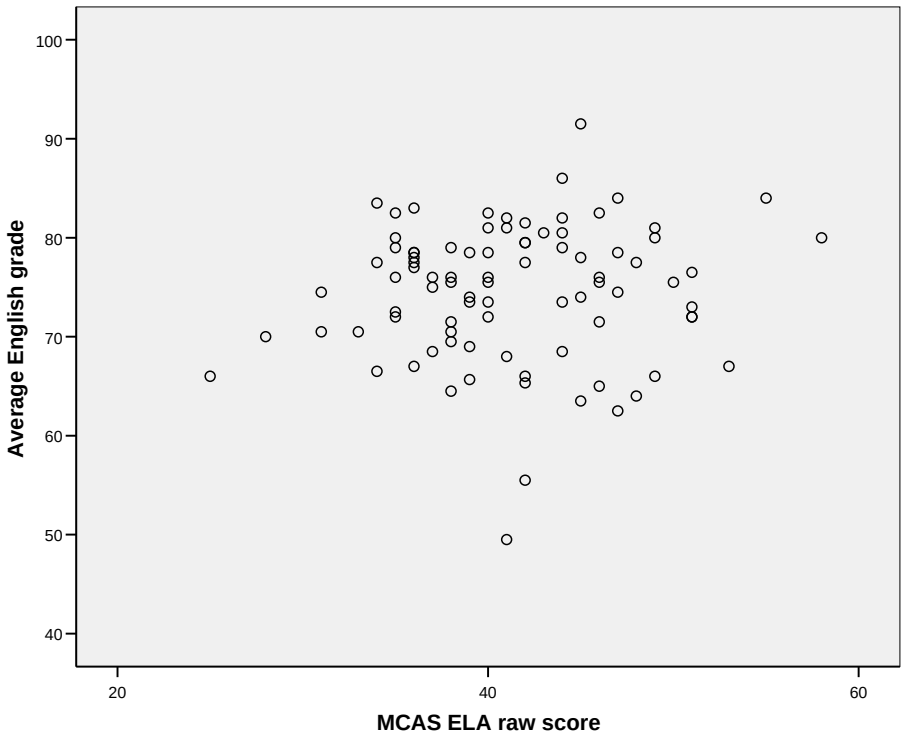

Burncoat, Average Math Course Level 3
Average math grade
100
80
60
40
20
0
20
40
MCAS Math raw score

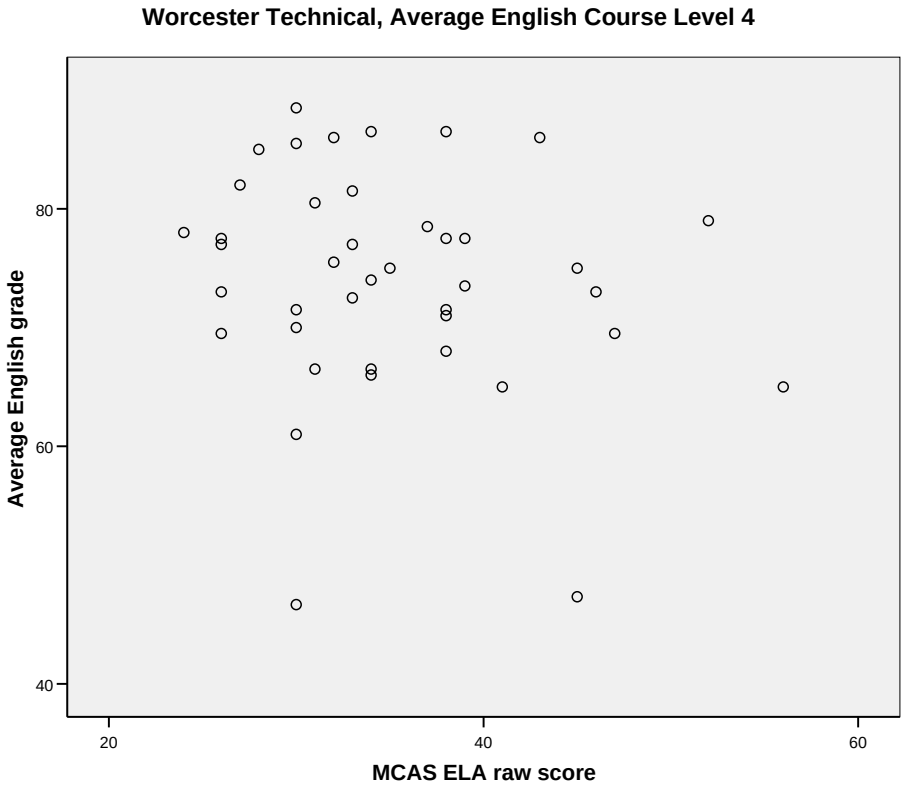

Doherty, Average English Course Level 3
Average English grade
100
90
80
70
60
50
40
30
20
40
60
MCAS ELA raw score

Doherty, Average English Course Level 4



Worcester Technical, Average English Course Level 3




Worcester Technical, Average English Course Level 4
Average English grade
80
60
40
20
40
60
MCAS ELA raw score

Appendix C

Below are the regression models for Burncoat, Doherty, and South High for the MCAS ELA and MCAS Math raw score.

MCAS ELA Score

Model Summary

High School [From SAT Code]	Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
Burncoat	1	.673(a)	.453	.445	4.646
	2	.754(b)	.568	.555	4.161
Doherty	1	.700(a)	.490	.481	3.633
	2	.768(b)	.590	.575	3.287
South High	1	.672(a)	.452	.435	5.065
	2	.752(b)	.565	.538	4.579

a Predictors: (Constant), Verbal SAT Score
b Predictors: (Constant), Verbal SAT Score, english average grade

Coefficients(a)

High School [From SAT Code]	Model		Unstandardized Coefficients		Standardized Coefficients			Correlations		
			B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
Burncoat	1	(Constant)	30.263	2.927		10.340	.000			
		Verbal SAT Score	.042	.006	.673	7.337	.000	.673	.673	.673
	2	(Constant)	15.050	4.521		3.329	.001			
		Verbal SAT Score	.028	.006	.453	4.621	.000	.673	.500	.380
		english average grade	.269	.065	.404	4.129	.000	.651	.459	.339
Doherty	1	(Constant)	35.947	2.666		13.483	.000			
		Verbal SAT Score	.036	.005	.700	7.404	.000	.700	.700	.700
	2	(Constant)	17.796	5.474		3.251	.002			
		Verbal SAT Score	.030	.005	.567	6.109	.000	.700	.632	.523
		english average grade	.251	.068	.343	3.694	.001	.563	.443	.316
South High	1	(Constant)	20.697	5.750		3.599	.001			
		Verbal SAT Score	.062	.012	.672	5.213	.000	.672	.672	.672
	2	(Constant)	4.393	7.666		.573	.571			
		Verbal SAT Score	.044	.012	.481	3.592	.001	.672	.536	.419
		english average grade	.291	.100	.388	2.894	.007	.625	.455	.337

a Dependent Variable: MCAS ELA - Raw Score

High School [From SAT Code]	Model		Collinearity Statistics	
			Tolerance	VIF
Burncoat	1	(Constant)		
		Verbal SAT Score	1.000	1.000
	2	(Constant)		
		Verbal SAT Score	.703	1.422
		english average grade	.703	1.422
Doherty	1	(Constant)		
		Verbal SAT Score	1.000	1.000
	2	(Constant)		
		Verbal SAT Score	.849	1.177
		english average grade	.849	1.177
South High	1	(Constant)		
		Verbal SAT Score	1.000	1.000
	2	(Constant)		
		Verbal SAT Score	.757	1.321
		english average grade	.757	1.321

a Dependent Variable: MCAS ELA - Raw Score

MCAS Math Score

Model Summary

High School [From SAT Code]	Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
Burncoat	1	.832(a)	.693	.688	4.783
	2	.847(b)	.718	.709	4.619
Doherty	1	.764(c)	.584	.577	5.484
	2	.829(d)	.687	.676	4.798
South High	1	.805(a)	.648	.637	5.764
	2	.863(b)	.745	.729	4.981

a Predictors: (Constant), Math SAT Score
b Predictors: (Constant), Math SAT Score, math average grade
c Predictors: (Constant), math average grade
d Predictors: (Constant), math average grade, Math SAT Score

Coefficients(a)

High School [From SAT Code]	Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
			B	Std. Error	Beta			Zero-order	Partial	Part
Burncoat	1	(Constant)	-15.048	4.254		-3.537	.001			
		Math SAT Score	.098	.008	.832	11.913	.000	.832	.832	.832
	2	(Constant)	-24.421	5.722		-4.268	.000			
		Math SAT Score	.087	.009	.735	9.302	.000	.832	.763	.628
		math average grade^	.188	.080	.186	2.353	.022	.569	.286	.159
Doherty	1	(Constant)	-38.097	8.890		-4.286	.000			
		math average grade	.961	.106	.764	9.031	.000	.764	.764	.764
	2	(Constant)	-33.541	7.848		-4.274	.000			
		Math SAT Score	.033	.008	.383	4.334	.000	.686	.498	.321
		math average grade	.699	.111	.556	6.290	.000	.764	.640	.466
South High	1	(Constant)	-5.264	5.467		-.963	.343			
		Math SAT Score	.086	.011	.805	7.793	.000	.805	.805	.805
	2	(Constant)	-18.543	6.065		-3.058	.004			
		Math SAT Score	.053	.013	.502	4.030	.000	.805	.580	.360
		math average grade	.360	.103	.435	3.492	.001	.785	.525	.312

a Dependent Variable: MCAS Math - Raw Score

^ The addition of average math grade to the model for Burncoat was not significant at the 0.01 level. The regressor was added to the model with the significance limit at the 0.05 level.

High School [From SAT Code]	Model		Collinearity Statistics	
			Tolerance	VIF
Burncoat	1	(Constant)		
		Math SAT Score	1.000	1.000
	2	(Constant)		
		Math SAT Score	.729	1.373
		math average grade	.729	1.373
Doherty	1	(Constant)		
		math average grade	1.000	1.000
	2	(Constant)		
		Math SAT Score	.703	1.423
		math average grade	.703	1.423
South High	1	(Constant)		
		Math SAT Score	1.000	1.000
	2	(Constant)		
		Math SAT Score	.514	1.946
		math average grade	.514	1.946

a Dependent Variable: MCAS Math - Raw Score

Bibliography

A brief history of the SAT. Retrieved May 12, 2007, from PBS Web site: http://www.pbs.org/wgbh/pages/frontline/shows/sats/where/history.html

About MCAS. Retrieved May 12, 2007, from Massachusetts Department of Education Web site: http://www.doe.mass.edu/mcas/about1.html

Ansley, T., Pereira, G., & Concepcion, M. (2006). *Analyses of the Similarities in the SAT and MCAS.* Unpublished Interactive Qualifying Project Report. Worcester Polytechnic Institute.

Batey, D. et al. (1995). *Cognitive Bias in the Scholatic Aptitude Test*. Unpublished Interactive Qualifying Project Report. Worcester Polytechnic Institute.

Frequently asked questions. Retrieved May 12, 2007, from College Board Web site: http://www.collegeboard.com/about/news_info/sat/faqs.html

Goldsmith, S. S. (2002). MCAS circa 2001: As good as it should be?. *Journal of Education*. *183*, 68.

Higham, P. A. (2007). No special k! A signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology*. *136*, 1-22.

Kingsland, D. et al. (1995). *Learning Style Biases In The Scholastic Aptitude Test*. Unpublished Interactive Qualifying Project Report. Worcester Polytechnic Institute.

Lawrence, G. (2004). *Looking at type and learning styles.* Gainesville, FL: Center for Applications of Psychological Type, Inc.

Mangenot, G. C., & Tongol, M. L. (1998). *The Leicester SAT Experience*. Unpublished Interactive Qualifying Project Report. Worcester Polytechnic Institute.

Massachusetts curriculum frameworks. Retrieved May 12, 2007, from Massachusetts Department of Education Web site: http://www.doe.mass.edu/frameworks/

Milewski, G. B. & Sawtell, E. A. (2006). *Relationships between PSAT/NMSQT scores and academic achievement in high school*. New York, NY: College Board.

Myers, I. B. (1993). *Introduction to type.* Palo Alto, CA: Consulting Psychologists Press, Inc.

Poland, A. M., & Levasseur, K. M. (2003). *WPS MCAS Experience (Class of 2003): Implications for the High School Principals.* Unpublished Interactive Qualifying Project Report. Worcester Polytechnic Institute.

PSAT/NMSQT. Retrieved May 12, 2007, from College Board Web site: http://www.collegeboard.com/student/testing/psat/about.html

Schurr, K. T., & Ruble, V. E. (1986). The Myers-Briggs Type Indicator and first-year college achievement: A look beyond aptitude test results. *Journal of Psychological Type, 14,* 57-59.

Schurr, K. T., Ruble, V. E., & Henriksen, L. W. (1988). Relationships of Myers-Briggs Type Indicator personality characteristics and self-reported academic problems and skill ratings with scholastic aptitude scores. *Educational and Psychological Measurement, 48*, 187-196.

Stambaugh, J. L., & Fountain, K. J. (2001). *Comparative Analysis of Learning Styles and MCAS Scores in the Worcester and Fitchburg Public School Systems.*

Unpublished Interactive Qualifying Project Report. Worcester Polytechnic Institute.