

01D028F

01D028I

Project Number: 51-JMW-ASPR

MCAS AND THE MBTI: EXPLORATIONS

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

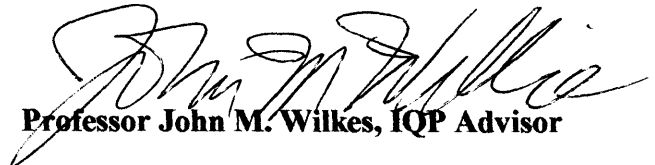
by



**Michael F. McElhaney**

Date: March 16, 2001

Approved:



**Professor John M. Wilkes, IQP Advisor**

1. education
2. testing
3. technology

<b>ABSTRACT.....</b>	<b>2</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>3</b>
<b>EXECUTIVE SUMMARY.....</b>	<b>4</b>
<b>1 INTRODUCTION.....</b>	<b>10</b>
<b>2 LITERATURE REVIEW.....</b>	<b>13</b>
2.1 MYERS-BRIGGS TYPE INDICATOR (MBTI).....	13
2.2 CENTER FOR APPLICATION OF PSYCHOLOGICAL TYPE (CAPT).....	16
2.3 STANDARDIZED TESTING (APTITUDE AND ACHEIVEMENT TESTS).....	17
2.4 MASSACHUSETTS CURRICULUM FRAMEWORK.....	20
2.5 ORIGINS OF THE MASSACHUSETTS COMPREHENSIVE ASSESSMENT SYSTEM.....	25
2.6 MCAS AND THE MBTI.....	30
2.6.1 <i>Public Debate</i> .....	32
2.7 SCHOLASTIC APTITUDE TEST (SAT).....	39
2.8 ACT ASSESSMENT™.....	41
2.9 OVERVIEW OF STATISTICS.....	43
2.10 PREVIOUS STUDIES AND FINDINGS.....	48
2.10.1 <i>Previous MCAS Studies</i> .....	48
2.10.2 <i>Previous ACT Studies</i> .....	49
2.10.3 <i>Previous SAT Studies</i> .....	50
2.10.4 <i>Average SAT Scores from Previous Studies</i> .....	53
<b>3 GENERAL HYPOTHESES.....</b>	<b>56</b>
<b>4 METHODOLOGY.....</b>	<b>58</b>
4.1 OBJECTIVES.....	58
4.2 CURRENT STATUS OF WORCESTER AND FITCHBURG MCAS.....	59
4.3 PARAMETERS OF STUDY.....	61
4.4 DATABASE CREATION.....	61
4.5 STATUS OF PROJECT/EXPLANATION OF FINDINGS.....	62
<b>9 PROJECT EVALUATION.....</b>	<b>64</b>
<b>BIBLIOGRAPHY.....</b>	<b>66</b>

## **ABSTRACT**

The Massachusetts Comprehensive Assessment System (MCAS) has caused a fiery debate among public school officials due to the fact that many students are at risk of failing this test. The MCAS takes effect starting this year as a high stakes state achievement test. It is supposed to improve standards and provide accountability tied to state-wide curriculums. Prior study of the SAT has revealed an “intuitive” advantage using MBTI data from the Worcester and Fitchburg Public school systems. This study seeks to find a similar cognitive bias on the MCAS. Also, if there is a bias, is it the same size and strength involving the same variables? In this socio-economic context, there might be a consequence of the MCAS relating so highly to the SAT in what the exams tell you. Finally, the study seeks to help identify those students “at risk” of failing the MCAS given the findings about what it is correlated with.

## **ACKNOWLEDGEMENTS**

The authors would like to thank Professor John M. Wilkes for his advising and mentoring throughout this project. Also, we would like to thank Patricia Mostue, Ph.D., of the Worcester Public School system, as well as her staff, for her unending advice and in providing us with the data we needed to complete this project. Similarly, we would like to thank Mr. Thomas Lamey and the Fitchburg Public School system, the WPI work-study students who aided in data entry, family, and friends.

## **EXECUTIVE SUMMARY**

In education, the standards of evaluation are very dependent on the school a student attends. Because of this, it is very difficult to gauge what and how much a student has been taught simply by looking at GPA or even individual grades. To try to solve this problem, the State of Massachusetts has mandated a statewide test. This test, the Massachusetts Comprehensive Assessment System (MCAS), is used as a measuring stick of sorts to determine if the schools systems across the state are effectively teaching a curriculum that meets the state's minimum standard for graduation as defined by the curriculum frameworks published by the State Department of Education.

The problem that has arisen out of this requirement is the fact that in some schools, especially city schools, large majorities of the student body are expected to fail the test. This can be seen in schools such as Worcester, Fitchburg, and Boston. In the Worcester and Fitchburg public schools, 66% of the students failed the Science and Technology portion of the practice round of the MCAS, and Boston had 71% failing. These were eighth grade students who took the MCAS in 1999, therefore making them part of the class of 2003, the class that must pass the test as 10<sup>th</sup> graders, or try again in 11<sup>th</sup> or 12<sup>th</sup> grade in order to graduate. When this information is combined with the fact that passing the test is required just to graduate high school, the result not to enter college is potentially devastating. If Worcester could only graduate a third of their students, the school system would be thrown into turmoil.

With this fact known, it becomes apparent that something must be done first to assure that the test is not biased, and if it is not, to help the teachers prepare their students more effectively for this exam. The purpose of this study is to examine the

effects of personality type (as determined by the MBTI) on MCAS results. If the MCAS performance of some types of learners is predictably weak and people with certain personality types are found to be at risk of under performing on the test, then a case for bias can be made. Also, a look at the MCAS (an achievement test) itself will be instructive. Our goal is to see if there is a correlation between it and the SAT/PSAT, which are considered aptitude tests. If it is found that there is a correlation between the two tests, then re-evaluation of the MCAS might be necessary. A double check using the ACT Plan (which is administered to 10<sup>th</sup> graders in the Fitchburg system) will also be undertaken. Though it too is a college admissions indicator, it is considered an achievement test.

This project will combine Worcester and Fitchburg High School students' MCAS scores, personal information, PSAT/SAT data, ACT Plan scores, GPA's, programs of study, and related materials with the results of their respective MBTI (a personality and learning styles indicator) to create a sizeable database. Using this database, we will be able to test our hypotheses and find out if the expected correlations actually exist. If the correlations are found to exist, then this database of MBTI information will prove very useful to the school system as a whole. It would allow teachers to identify students at risk earlier in their high school programs, and assist them in taking steps that are appropriate to help the students most likely to struggle with the MCAS. Ideally at the completion of this project, an expandable MBTI based database will have been constructed. This database will be given to the participating urban school systems in the hope of their continuing to maintain and update it as part of their efforts to cope with the MCAS crisis unfolding around them.

Certain parameters of the MBTI, especially the Sensing (S)/ Intuitive (N) and Judging (J)/ Perceiving (P) types, have been shown to somewhat predict how a student will perform on the SAT. It is best that the data is analyzed under a control variable, i.e. for those classmates in the same academic program. When this is not done, MBTI relationships can be exaggerated as in Pieper's WPS study, or can be obscured altogether, as in the Leicester study. This is due to the impact of differences in preparation and which learning types have the greatest access to the more challenging Advanced Placement and Honors courses in a given high school. Therefore, the correlations that will be made between learning style via the MBTI and the SAT should also be strong because the students are of the same class.

In relation to the MBTI and individual subject grades or overall GPA relationship, it is hypothesized that those with the higher GPA's will tend to be certain learning types, although it is unknown which ones at this point. Isabel Myers predicted that "judging" students would do better in class day-by-day, but the "perceivers" would outperform them on standardized tests like the PSAT. Students with a higher GPA are often encouraged to take more challenging courses, and therefore, they are more prepared for standardized examinations than other students. However, some decline, so as to keep their grades higher in a less challenging program. It is also thought that classroom performance in certain subjects, such as science and English, will be a predictor for performance on the same sections of the MCAS rather than general GPA predicting average scores.

Since both the ACT and MCAS are achievement tests, it is expected that certain learning types determined by the MBTI will be more prone to underperformance than others. This follows the logic that the ACT and MCAS are highly correlated as shown in

the literature review. Also, certain MBTI types have under performed on sections of standardized tests which have elements of achievement in them, indicating that certain MBTI types should be expected to perform similarly on the MCAS. While there has been little to no research in this area, the hypothesis has a foundation. Remember that the SAT has been shown to have elements of achievement in it. Also, previous research has shown correlations between the PSAT/SAT and ACT, as mentioned in section 2.10 of this report. Therefore, if certain learning styles by MBTI standards give some students an advantage on the sections of the PSAT/SAT that are supposedly “achievement sections”, the same MBTI types will give the same students an advantage on both the ACT and MCAS assessment exams. It was also shown before in section 2.10 that many sections of the MCAS and ACT show strong correlations. Therefore, performance on sections of one test should predict performance of similar sections on the other test.

Getting back to the main focus of the project, which is the effect of MBTI type on MCAS performance, it is expected that there will be certain types that will have advantages and others that will be at risk. In section 4.5 of this report, there are questions of correlation listed. It is expected that there will be significant correlations between most, if not all, of the relationships that are stated.

The study contained students who attended high school in the Worcester and Fitchburg public school systems. All of these students graduated from high school in the year 1999, 2000, or 2001. The MBTI was administered to them when they were in 10<sup>th</sup> or 11<sup>th</sup> grade, which was in 1997, 1998, or 1999 for most of them. The MCAS exam was also administered when they were sophomores. The number of students in the Worcester Public School system that took the MBTI and (presumably the MCAS) was about 1200, and there were 182 who took the MBTI (and presumably both exams) in the Fitchburg



school system. Therefore, the total number of students available for study from both school systems was 942. This number was large enough to provide a reliable study and give meaningful results. Of these 942 students, the database contained somewhere in the neighborhood of 750-850 students from both systems for whom we received MCAS and PSAT/SAT scores. A large portion of the data came from the Worcester Public School system, simply because there are four high schools there, and a significantly larger number of 10<sup>th</sup> graders for which there was MBTI data. Information for every student was not available. Hence, the database contained lapses and we took a loss of up to 200 students that took the MBTI.

Our results show that the Intuitive and Perceptive type advantages over Sensing and Judging types, respectively, form a unifying thread for the study of most aptitude and achievement tests studied thus far. This thread should be maintained. Replication studies which examine the correlation between high SAT, ACT, or MCAS scores, and these advantages year to year allow for the researcher to monitor changes in district performance, composition, and in the performance measure itself. The SAT changed in the early 1990's and changes in the MCAS are expected. Hence, a starting point for many measures, be they already studied for years or a brand new (test "X"), is to determine if there are advantages for any type, and how large and consistent they are. This sort of information provides the basis for regression analysis that may be done as part of a study, so it is important to monitor this relationship year to year for each measure available. Regression uses more than one independent variable to predict a dependent variable.

Overall, we found that there is no simple answer to the question of how to improve the MCAS for students at risk. While simplifying the test may increase certain student's

chances for passing, is that really the goal of the MCAS? The fact is the state feels that all high school graduates should have the knowledge tested in the MCAS. The reason why the MCAS is under such fire might not be because it is a poorly made test, but the first test that REALLY matters. The test itself might not have to change; all the schools and teachers might have to change to meet the test.

## **1 Introduction**

In education, the standards of evaluation are very dependent on the school a student attends. Because of this, it is very difficult to gauge what and how much a student has been taught simply by looking at GPA or even individual grades. To try to solve this problem, the State of Massachusetts has mandated a statewide test. This test, the Massachusetts Comprehensive Assessment System (MCAS), is used as a measuring stick of sorts to determine if the schools systems across the state are effectively teaching a curriculum that meets the state's minimum standard for graduation as defined by the curriculum frameworks published by the State Department of Education.

In an ideal world, the benefit of having this test would be looking at a few numbers can give one an idea of what the student has been taught, and how well the material has been mastered. This information would allow teachers to assess their level of success and adjust their teachings to meet the state's requirements if necessary. It would also allow the state to identify schools that are not routinely meeting the required minimum, and take an appropriate course of action to rectify the situation. However, this test is not just used to help assess the school system in general. The student must pass the test between 10<sup>th</sup> and 12<sup>th</sup> grade in order to graduate from high school with a diploma.

The problem that has arisen out of this requirement is the fact that in some schools, especially city schools, large majorities of the student body are expected to fail the test. This can be seen in schools such as Worcester, Fitchburg, and Boston. In the Worcester and Fitchburg public schools, 66% of the students failed the Science and Technology portion of the practice round of the MCAS, and Boston had 71% failing.

These were eighth grade students who took the MCAS in 1999, therefore making them part of the class of 2003, the class that must pass the test as 10<sup>th</sup> graders, or try again in 11<sup>th</sup> or 12<sup>th</sup> grade in order to graduate. When this information is combined with the fact that passing the test is required just to graduate high school, the result not to enter college is potentially devastating. If Worcester could only graduate a third of their students, the school system would be thrown into turmoil.

With this fact known, it becomes apparent that something must be done first to assure that the test is not biased, and if it is not, to help the teachers prepare their students more effectively for this exam. The purpose of this study is to examine the effects of personality type (as determined by the MBTI) on MCAS results. If the MCAS performance of some types of learners is predictably weak and people with certain personality types are found to be at risk of under performing on the test, then a case for bias can be made. Also, a look at the MCAS (an achievement test) itself will be instructive. Our goal is to see if there is a correlation between it and the SAT/PSAT, which are considered aptitude tests. If it is found that there is a correlation between the two tests, then re-evaluation of the MCAS might be necessary. A double check using the ACT Plan (which is administered to 10<sup>th</sup> graders in the Fitchburg system) will also be undertaken. Though it too is a college admissions indicator, it is considered an achievement test.

This project will combine Worcester and Fitchburg High School students' MCAS scores, personal information, PSAT/SAT data, ACT Plan scores, GPA's, programs of study, and related materials with the results of their respective MBTI (a personality and learning styles indicator) to create a sizeable database. Using this database, we will be able to test our hypotheses and find out if the expected correlations actually exist. If the

correlations are found to exist, then this database of MBTI information will prove very useful to the school system as a whole. It would allow teachers to identify students at risk earlier in their high school programs, and assist them in taking steps that are appropriate to help the students most likely to struggle with the MCAS. Ideally at the completion of this project, an expandable MBTI based database will have been constructed. This database will be given to the participating urban school systems in the hope of their continuing to maintain and update it as part of their efforts to cope with the MCAS crisis unfolding around them.

## **2 Literature Review**

### **2.1 Myers-Briggs Type Indicator (MBTI)**

The Myers-Briggs Type Indicator, which is commonly referred to as “The Indicator”, or simply as the MBTI, is a psychological tool developed by Isabel Briggs Myers and her mother, Katharine Cook Briggs. Their goal was to create a test that would reflect C. G. Jung’s research on understanding individual differences among people. Jung observed that human behavior is not random, but instead follows identifiable patterns that he felt derive from the structure of the human mind.

Myers and Briggs came upon Jung’s work in 1923 and thus, began their own effort, which consisted of nearly twenty years of “type watching.” After this extensive research, Myers decided that people might make better decisions concerning their careers and lives if they knew about Jung’s types. To that end, Myers began constructing a questionnaire to help assess an individual’s specific type. The MBTI was developed through several different iterations (over thirty years) as information was collected on thousands of people. At her death Isabel's forms F and G were the standard research instruments, and included the same core load of 100 items used to classify psychological type.

What the MBTI does is provide a description of a person’s preference in a set of four letters that define Psychological Type on four dimensions. Armed with this information about a person’s preferred, and sometimes habitual, way of dealing with situations involving data gathering or decision making, then approach to learning and likely reactions to various careers are able to be predicted and understood. It is also a way to study small group dynamics and leadership style. But knowing type does not just

help understand others. It also helps one understand one's own inclinations. The valuable area that MBTI data can be used to illuminate which concerns us here is its ability to identify a person's learning style without implying that one particular type is more able or intelligent than another. The MBTI deals with preferences, not abilities. It does not determine what skills or abilities someone will develop, but it does reveal the best channels by which to reach a learner. What it can do (for a teacher) is indicate circumstances under which given teaching and learning strategies are likely to be most effective by taking individual difference and preferences for handling information and coming to a decision into account.

The four dichotomous MBTI type preference dimensions are as follows:

Extrovert or Introvert

Sensing or Intuition

Thinking or Feeling

Judgment or Perception

By providing answers to multiple-choice questions in the indicator (Appendix B), each person shows that he or she is predisposed to either one or the other of these preferences. Following is a brief description of each of the MBTI's preferences assembled by Charles Martin, Ph.D., in Looking at Type: The Fundamentals, in which he indicated that Jung considered these differences in preference innate, rather than learned. What one learns later is how to function effectively when required to operate in one's less natural or preferred mode.

**Extroversion (E):**    *Key words:*  
                                  Outer world, people, action, breadth.  
                                  *Description:*

Extroverts are energized by active involvement in events. They are most excited when they are around people. They often find their understanding of a problem becomes easier if they can talk about it aloud.

Or

**Introversion (I):**

*Key words:*

Inner world, ideas, reflection, depth.

*Description:*

Introverts are excited when they are involved with ideas, images, and memories that are part of their inner world. They often prefer solitary activities or with one or two others whom they feel an affinity. Introverts truly like the idea of something often better than the something itself.

**Sensing (S):**

*Key words:*

Facts, details, experience, present.

*Description:*

People with a preference for sensing are more concerned with what is actual, present, current, and real. They are often good at seeing the practical applications of ideas. They learn best when they can first see the pragmatic side of what is being taught.

Or

**Intuition (I):**

*Key words:*

Symbols, pattern, theory, future.

*Description:*

People who have a preference for intuition would rather gain understanding through insight than hands-on experience. They like concepts, and they learn best when they have an impression of the overall idea first.

**Thinking (T):**

*Key words:*

Impersonal, truth, cool, thought-minded.

*Description:*

People of this type are concerned with the objective truth in a situation. They act based on truth or principle that is independent of what they or others might want.

Or

**Feeling (F):**

*Key words:*

Personal, value, warm, tender-hearted.

*Description:*



People of this preference are concerned with whether decisions and actions are worthwhile. They feel that they can make the best decisions by weighing what people care about.

**Judgment (J):**

*Key words:*

Structured, decided, organized, scheduled.

*Description:*

What the judging preference often looks like is that they prefer a planned or orderly way of life, and likes to bring life under control to the degree that is possible.

**Or**

**Perception (P):**

*Key words:*

Adaptable, spontaneous.

*Description:*

People with the perceiving preference are inclined to a more flexible and spontaneous way of life, like to understand and adapt to the world, and like to stay open for new experiences.

Once a person's predilection for each of the above preferences is found, then a four-letter type pattern (Appendix B), which is descriptive of the overall psychological type, is generated. With this four-letter pattern, it is possible to go into greater depth and analyze all sixteen different type combinations. However, this isn't germane to the fundamental understanding of the MBTI. The ultimate goal of the MBTI is to improve the quality of a person's life by helping them understand themselves and the people around them better in terms of how they prefer to handle data and come to a decision.

## **2.2 Center for Application of Psychological Type (CAPT)**

In 1975, Isabel Myers and Mary McCaulley founded the Center for Application of Psychological Type, known as the CAPT. This non-profit organization was created to

continue research on the MBTI and to provide training on type. The CAPT also publishes and distributes many different publications on type and the application of type in different fields, such as education, personal life, and the workplace. The CAPT's main focus is training people to be qualified in the use of the MBTI. Training takes 3-4 days, and costs roughly \$700 - \$900 dollars. Training seminars are held 2-3 times a month in a different U.S. City.

The Worcester Public schools became interested in the MBTI starting around 1995-1996, when evidence was gathered by several WPI student teams looking at data from the class of 1997, which connected their PSAT scores to their MBTI types. The three highest and three lowest of the 16 types scored 250 points apart, on average. Worcester made arrangements to profile each class using the MBTI during their 10<sup>th</sup> grade year, and paid to have a core group of about 50 school psychologists, guidance counselors, teachers, and administrators trained to the point of becoming qualified users of the MBTI. They would, in turn, help others in the future.

### **2.3 Standardized Testing (Aptitude and Achievement Tests)**

Objective testing, the use of objectively scored, paper-and-pencil tests of true-false and multiple-choice items, came into widespread use during World War II, but IQ testing began in World War I as the Army tried to identify the potential officers, assigning the rank and file by democratic means. At that time large numbers of recruits had to be tested for placement in armed forces positions. They needed to decide in a very short time whom would become pilots, cooks, infantrymen, officers, etc. The question of whether it is possible to use objective tests effectively to answer these kinds of questions

was found to be a qualified yes. The Army found that within limits, it was possible to use a paper and pencil test to predict a person's aptitude for things they had never even tried to do before. At least it was better than guessing or asking the recruits what they thought they could do best.

Standardized testing, which added rigidly defined procedures for administering the tests, came into use when the armed forces noticed that individuals from certain areas of the country were scoring much lower on the tests than individuals in other parts of the country. When this was looked into, it was found that the test was given under widely varying conditions. Some were given no instructions for taking the test and some were actually asked to take the test standing up, during their physical examinations. In order to alleviate these issues, standardized procedures to be followed by everyone were developed for administering the tests.

With the success of the armed forces achievement and aptitude tests, (the differences between them will be discussed below), it was only natural that this method would be used in schools to evaluate students. Aptitude tests (such as the SAT) were developed for use as college entrance exams, and achievement tests (such as Iowa) began to be used to measure grade-school performance. Most colleges at the time had a procedure to select eligible students that involved an extensive interview with the student, letters of recommendation and high-school performance information. Most selective colleges and universities also had their own admissions tests, which stressed Latin and the classics. Prep schools had sprung up to prepare students for the test of a given Ivy league university. Harvard proposed a College Board exam to replace this system and help identify working class students or unusual merit worthy of scholarship support. The objective, multiple-choice version of the SAT appeared during World War

II. When objective tests started to show up, many colleges wondered if they could use them effectively to predict a student's college success. The reason that virtually all colleges have gone over to using ACT and SAT test scores in addition to high school records of performance are not solely that they are more cost-effective and efficient than subjective interviews and past grades, although they are. No, they hoped that they would do a better job of predicting school success than the older, less efficient procedures and provide a level playing field (to be fairer than grades). By this measure they are a partial success in that the use of SAT scores and high school GPA is about 10% more predictive of freshman college grades than the use of H.S. GPA alone. Together they correlate 0.4 to 0.5 with freshman grades thus accounting for 16% to 25% of the variance in the early college grades.

The two kinds of tests mentioned above, achievement and aptitude, are fundamentally different. Standardized achievement tests are designed to measure how much a student has already learned about a subject. The results from these tests can be used to help teachers develop programs that suit students' achievement levels in each subject area, such as reading, math, language skills, spelling, or science to take them from where they are to where they need to be.

Standardized aptitude tests are an effort to measure students' ability to learn – their potential, or how well they are likely to do in future school work to which they are not yet exposed, as a general matter. Instead of measuring knowledge of subjects taught in school, these tests measure a broad range of abilities or skills that are considered important to success in school. They can be efforts to measure verbal ability, mechanical ability, creativity, clerical ability, or abstract reasoning. The SAT is designed to tap verbal and mathematical reasoning ability, which together are referred to as a general

(information reasoning) ability by the College Board. The results from aptitude tests help teachers to plan instruction that is appropriate (in terms of pace and presentation) for students, considered to be at different levels of aptitude as well as having different backgrounds. The ACT is different in that it focuses on where one is given the typical high school curriculum offered in the United States.

## **2.4 Massachusetts Curriculum Framework**

Equal opportunity for all students has been a focal point in the media and ongoing public debate over state and national education standards for the better part of 20 years. Even before that, there were racial equality battles being fought, which directly resulted in the integration of all races into the same public schools. This represented an early attempt to offer equality of opportunity to students regardless of their race.

In the past two decades concerns not only over racial equality within schools but the underlying inequalities between men and women and the social classes have become part of the SAT debate. As differences between schools and school districts serving the various ethnic groups and the more and less affluent communities were documented activists went to the legislature, the court and even the streets in search of redress. Many changes in the education system that we have today were efforts to achieve more equal opportunity. Affirmative Action Plans by race were only the most visible of these efforts that came to involve average SAT scores. Reverse discrimination cases began to reach the courts in which SAT scores were the primary evidence that a “stronger” student was discriminated against and did not get a college admission because they were of the wrong (white) race, given efforts to balance admissions along these lines, especially at publicly

funded state colleges and universities. The pressures of public debate involving social class differences came to a head in 1993 with the resolution of the McDuffy v. Robertson case in the Supreme Judicial Court of the Commonwealth of Massachusetts.

McDuffy v. Robertson represented a case of discrimination associated with social class. It was an outgrowth and extension of a case brought before the court system in 1978, *Webby v. Dukakis*, in which Webby claimed that townships within the Commonwealth were often unequally funded. Specifically, the poor school districts, possessing little income from property taxes, charged that the wealthier school districts had more money per student, thereby treating students differently depending on what school system they found themselves in. As the State contributed a smaller fraction of funding to school districts than did the local towns, this allowed for a wide range of per student spending, and hence monetary inequality under Commonwealth law.

Over the years various school systems, including poor metropolitan areas such as Worcester and Brockton, joined suit, evolving into the McDuffy case, which reached resolution in 1993. The Court decided in that case that the spending structure for the Commonwealth was inherently unequal and must be changed, with the details of a plan to achieve equity left to the legislature, as well as the allowance for tax increases associated with meeting the equal education standard. This opened the door for the rapid passage of new law to try to fix the system. The new laws, referred to the Education Reform Act of 1993, passed by the legislature and were signed by Governor Weld. They specified a new funding structure where the State contribution to education was raised from 30% to 50% for school districts. This was the central issue in the court case, but there were more sweeping changes in the dictates of the act.

The Supreme Judicial Court of Massachusetts laid down the concept of a Curriculum Framework in its decision when it delineated the seven areas a student would need to be proficient in after passing through the school systems of Massachusetts. The concept was simple. The only way to ensure that the equality sought by fixing financial inequity between school districts was being achieved was to measure the readiness of graduating students for the world somehow. These seven areas of knowledge were used to represent a minimum skill set that could be and should be delivered by an equitable statewide school system without vast discrepancies in resources between districts.

The Reform Act addressed these curriculum areas in a number of ways. The time allotted to the various core curriculum areas was increased greatly by the Act, so as to increase student preparedness in this area. Teacher certification requirements were strengthened, particularly in the sphere of core curriculum familiarity. Charter schools were added to the mix in some districts and while free from many generalized district regulations the core curriculum was set to be stressed therein as well. Lastly, the curriculum frameworks for the various content areas themselves were fleshed out.

The Curriculum Framework for the Commonwealth was set to include the following knowledge or skill areas: English/Language Arts; Mathematics; Science/Technology; History/Social Science; World Languages; Art (Culture); and Health. In order to give the new Curriculum some flesh on those bones, the Commission on the Common Core of Learning was formed, consisting of educators, administrators, and others considered likely to have valuable input. The Commission constructed a rough sketch of what students would need to know after passing through a K-12 school system. This sketch was done by 1994, and the Department of Education accepted the Common Core of Learning as the model for district education plans in the Commonwealth. By

1995 the first full implementation of the meat of the Framework, specifically in the areas of Mathematics, Science/Technology, Art, World Languages, and Health, was released to the schools.

Using this new set of guidelines was difficult for most districts due to the rapid changes that the Framework underwent in the immediate post 1995 period. The cause for this resided in the first revision document of the Framework itself, which was muddled within the learning processes of the Curriculum Framework rather than the knowledge and skill areas themselves. A 1996 Board of Education Chairman appointee named John Silber (Ex-President of Boston University) assessed that first revision and called for its overhaul using a new investigative commission composed of BOE members and curriculum specialists. These people were intended to focus on the intended outcome of K-12 education rather than the means to get there, as was the stipulation of the McDuffy court decision. Hence, the Curriculum Framework came to be a representation of the minimum standard knowledge base they felt a student graduating from a Massachusetts high school should possess in terms of content.

This commission, working under the direction of Silber (and his successor James Peyser) developed an achievement oriented educational environment. Opposition to this structure was mounted primarily by teachers and learning specialists who overwhelmingly held that focusing on knowledge rather than learning process keeps the students from effectively coping with new problems. Instead it limits them to a given pool of knowledge. This debate remains unresolved, though the court decision and the Reform Act itself do not explicitly call for or prohibit a focus on learning process rather than content outcomes.



Regardless of which position one takes on the debate over the philosophical nature of the Curriculum Frameworks the result is a seemingly easy way to measure the degree to which equity between the school systems has been achieved. If the Curriculum Framework is based on knowledge areas, it should be possible to define the statewide knowledge minimums in each area of the Framework. Once defined, it becomes a theoretically simple task to test the knowledge of students in these areas and not only assess their level of achievement individually but those of the individual schools and school districts as well. Here enters the Massachusetts Comprehensive Assessment System, a standardized test used to evaluate students and schools for their success in meeting the minimums of the Curriculum Framework areas being tested for. MCAS and its meaning are detailed below. For now assume that the MCAS is a tool designed to illustrate the relative competence of schools and districts in teaching their students the areas targeted by the Curriculum Framework – starting with English Language Arts, Science and Technology, and Mathematics. A first practice run of MCAS covering these fields was administered to 4<sup>th</sup>, 8<sup>th</sup>, and 10<sup>th</sup> graders at the end of the 1998-1999 school year. The Science and Technology test was considered experimental and subject to change at that point so attention focused on English Language Arts and Mathematics..

The importance of the MCAS in terms of the Framework becomes clear immediately. The standard performance expected per district in the Commonwealth reflects the minimum standard set forth in the Curriculum Framework and called for in the Reform Act. Taken to its logical conclusion, by assuring a minimum of knowledge requisite to survive in the world post graduation to all students in the Commonwealth the Reform Act fixes the inequity described in the McDuffy case. Hence, if the standard is not met, the cause of the failure must be assessed. This is where the fiercest debate

resides, specifically in the attribution of blame to an entity in the educational system. If the failure is in the exam, as is contended by opponents of MCAS and the Framework, then the test must be modified to more accurately reflect the Framework. If the failure is in the schools to teach the Curriculum completely, then the reasons must be examined and a remedy applied on the basis of the Reform Act. If the failure resides with the student, that student must be assisted in gaining the required knowledge before graduating from the system.

The interdependence of the Curriculum Framework, MCAS, and the school districts involved forms a tight, complex net of relationships. Sorting this web out involves more than just arguing philosophy. It depends on our understanding of how the Curriculum is assessed, and how students, as learning, thinking individuals, respond to this assessment. This requires knowledge of what MCAS is and its meaning is to us several years after its creation for these worthy purposes.

## **2.5 Origins of the Massachusetts Comprehensive Assessment System**

The Massachusetts Comprehensive Assessment System, or MCAS, is a standardized test designed to assess student and school performance in meeting the minimum knowledge required within the Curriculum Framework of the Department of Education. As specified in the section above on the Framework, the State needed a means to determine if the school districts across the state were affording students an equitable education. By defining the Curriculum Framework minimal knowledge base as an equitable education minimum, they were theoretically able to test for a student's preparedness in terms of the Framework as well as test the success of a school in

teaching the Framework to the student body as a whole. The test was to be administered to 4<sup>th</sup>, 8<sup>th</sup>, and 10<sup>th</sup> graders as an on-going assessment of their progress in understanding the Core Curriculum, but it was the 10<sup>th</sup> grade test that they had to pass. If they did not, the school would have 2 years to help them meet this standard in a maximum of 5 more attempts to pass the parts of the tests that were initially failed.

In order to be effective in molding the school districts to the Frameworks, certain responses were keyed to take effect should schools under perform on the test. The school would be studied to see if the cause was structural or due to the qualifications and efforts of personnel then within the school. If the cause was student related, or perhaps if the Curriculum needed more definition the staff in place would deal with it. The under performing students may or may not be flagged for special attention depending on their grade level. State funding may or may not be shifted in order to help assure later success. In brief, many things could result from weak performance on the MCAS by students. It is when the cause is determined to be the school, faculty, or school system that the remedy takes dramatic action such as the State taking over a bankrupt school, administration getting fired and curriculum being imported from elsewhere.

Not all of the 7 Curriculum Framework areas are presently tested on the MCAS, as the implementation of the Framework has been incremental throughout its history. By now – the third round of MCAS testing scheduled for May 2001 – the areas of the Framework incorporated into MCAS currently are Mathematics, Science/Technology, English Language Arts, and History/Social Science. The MCAS evaluation itself tests students in these areas according to different means of evaluation, specifically a mix of traditional bubble-filled standardized questions as well as write-in portions. The types of questions on the test are multiple choice, found in all sections, as well as short answer

write-ins for Mathematics portions of the test. There are open-ended questions for all sections of the test as well, which allow students to offer brief paragraph long answers to a myriad of questions. Lastly, in the English Language Arts section, there are writing prompt questions which allow for extensive student writing opportunities. Some involve creative problems, some are inspirational writing, some involve technical expository, etc.

The whole test comprises an enormous amount of material supposed to center on the content mandated by the Curriculum Framework. Therefore, it represents an incredibly ambitious achievement-style test reminiscent of a hybrid ACT/SAT. The MCAS multiple-choice sections are scored by machine, with all other parts being scored by teams of teachers and scoring professionals. Written answers are scored according to a preset scoring guide that all scoring teams refer to, and a range of points (1-4) are awarded on the basis of the student's answer. The scores awarded on the MCAS correspond to a range of points between 200 and 280, with the performance level cutoffs being every 20 points (hence at 220-239, 240-259, and over 260). Student performance itself is ranked below and within these levels as Failing (219 or less), Needs Improvement, Proficient, and Advanced, with the quality of the student's performance being from lowest to highest. A score of 245 for a given section, for instance, means that the student has performed proficiently on the section. In terms of the Framework, the student has learned that requisite portion of the Framework, and the school has provided that student with an equitable education under the Reform Act.

Different scores therefore mean different, important things. Students who are advanced have mastered the material required, and hence are receiving an equitable education. Students who are proficient, while not masters of the content area, are skilled

to the point of having received an equitable education. From the district standpoint, these two upper categories of students are considered total successes, with efforts targeted at assisting the lower two categories in understanding the Framework content.

The students who need improvement have met the minimum standard of the Framework, and hence the district has still succeeded in providing those students with an equitable education, and will not be subjected to punitive action but encouraged to attempt further improvement. Admittedly, the students are performing poorly, and large efforts should be made to improve their understanding. Lastly, the failing students do not understand enough material to meet the minimum standard required by the Framework and are considered unready to function in the broader society and world of work. This last has economic implications for the state that warrant intervention. From the district perspective, this can either be a school or student related problem. If a large number of failing students are in a school or district, it can be assumed that the school or district, for whatever reason, has been unsuccessful in teaching the Core Curriculum. If there are few students, then the failure may lie mostly with the individual student (or a certain type of student) who then needs special attention in order to gain an understanding of that core material.

This point of interpretation over the results is where the greatest debates lie, for particular schools (such as vocational schools), or particular students (such as students who primarily speak Spanish), may have considerable difficulty with the MCAS. This puts some districts in the difficult position of lacking any obvious remedy to the failing situation their students find themselves in. For the students' part, beginning with the class of 2003 graduating seniors will be required to have passed the MCAS between their

10<sup>th</sup> and 12<sup>th</sup> grade years, and for students who have a history of failing tests, the MCAS will prove a daunting requirement to meet.

Into this debate, at this point, enters the idea of learning styles as described in discussion of the MBTI. If the MCAS truly approximates an achievement-style test, then the types which get good grades due to consistent application and succeed at achievement tests will do so for the MCAS. Further, since achievement tests generally are considered to be content, not process oriented, there should be little bias overall toward any one type or certain type of learners. Conversely, if the MCAS is not a measure solely of content, but in some way takes into account process related thinking skills, then the types which perform well on it should appear to more closely resemble those who excel at aptitude tests such as the SAT. Either way this concept of psychological type and its relationship to test performance levels promises to be revealing. The S/N dimension of the MBTI seems to be a real factor in aptitude tests, some achievement tests, and other educational measures used nationally. If the MCAS is designed such that a few groups have the edge in success, then there are definite groups of students, differentiated by their MBTI types, which may be more at risk for failure on the MCAS. Does that mean that the educational system in general or a given school has served them badly, is biased against them, or that the test itself is inherently harder for some type of learners than others? It is this group of “at risk” students that represents an as of yet little considered source of tests failures that cognitively might not reflect on the school or their own efforts. Do these same students also show up badly on other measures such as grades, SAT performance, etc? The question of whether a “one size fits all” test is the appropriate way to assess educational outcomes is illuminated by the use of a learning style indicator with implications for career choice.

## 2.6 MCAS and the MBTI

As mentioned above, the test was initially given to fourth, eighth, and tenth graders. The possibility that certain areas might be tested at different grade levels to spread the burden was raised after the 2 experimental rounds of 1999 and 2000. To avoid the fatigue factor and let the students and teachers focus their preparations, there is likely to be a change in policy. For instance, the English Language Arts section has been administered to some third and seventh graders. The Mathematics section has been taken by sixth graders. Finally, the Science and Technology portion has been given to fifth graders, and will be omitted for fourth graders in the 2001 MCAS (<http://www.doe.edu.mass/mcas>). The reason for giving the examination to the fourth and eighth graders was to assess the performance of the school district in preparing the students according to the core curriculum before the students themselves face their high stakes moment at the end of the 10<sup>th</sup> grade. The tenth grade examination focuses more on the individual student, and measures his or her ability to pass the MCAS in order to graduate from high school.

The format of most of the questions that appear on the MCAS are similar to those that one would find on the SAT (Scholastic Aptitude Test) or PSAT (Preliminary Scholastic Aptitude Test). However, the MCAS designers stressed the need to balance multiple-choice with open-ended items and focus on achievement rather than reasoning. Hence, there are multiple-choice questions (all subject areas), short-answer questions (Mathematics only), open-response questions (all subject areas), and writing prompts (English Language Arts only). The results are reported in one of four ways for each

section and are used for several evaluations. These include using the results to make improvements in teaching and learning, to show school and district accountability, and individual student accountability. The four modes of reporting results from the MCAS and their respective scores are listed below with full descriptions:

**ADVANCED (260-280):** Student demonstrates a comprehensive and in-depth understanding of rigorous subject matter, and provides sophisticated solutions to complex problems.

**PROFICIENT (240-259):** Student demonstrates a solid understanding of challenging subject matter and solves a wide variety of problems.

**NEEDS IMPROVEMENT (220-239):** Student demonstrates a partial understanding of subject matter and solves some simple problems.

**FAILING (200-219):** Student demonstrates a minimal understanding of subject matter and does not even solve simple problems.

There has been some indication that previous tests such as the SAT and PSAT, which are aptitude tests, are easier for students of some learning styles as measured by the MBTI (Myers-Briggs Type Indicator) than others. However, so far we have found no research on how learning style can affect a student's performance on the MCAS, which is viewed as an assessment of achievement. The only basis for speculation is a comment based on research by Isabel Myers in the early 1960's at a number of suburban Philadelphia high schools. She reported then that intuitive students, especially INP's would be expected to excel at IQ style tests and seek out college preparation programs but that they might not do as well as the sensing and judging students when average grades based on day to day performance were assessed. This was her way of stressing that the INP students were not really smarter, just different in learning style as compared to the ESJ students. If learning style does affect MCAS performance, which personality types are most at risk of failure?



In addition, does the MCAS in fact test achievement, and not aptitude? Will the school systems' grades be a better predictor of success on the MCAS than scores on an aptitude test, like the PSAT?

### **2.6.1 Public Debate**

An important aspect of studying the MCAS that cannot be pushed aside by the analysis of the test itself is the public's opinion of the exam. Throughout the recent history of education reform in all states, public arguments for and against new policies have raged heated for decades, often reaching no resolution. MCAS thus far has proved itself to be good fodder for such arguments, with much talk and even mobilization of grass-roots organizations representing thousands of people who want their voices heard. This includes the Massachusetts Teachers Association.

One problem that has arisen out of the debate over the MCAS comes from the location of the primary opponents. A large percentage of the people that want to get rid of (or lower the importance of) the MCAS are in the suburbs, where the schools are generally in better shape than in the inner cities and the state is offering few new resources despite the threat of higher taxes. Public support for MCAS is strongest in the state's largest cities, where students "trapped" in failing schools and have the most to gain from a successful shift to higher standards. Unfortunately for MCAS supporters, there are more suburban districts than urban ones, so the suburbs have more of a voice than the urban districts. Further, the urban teachers are not in agreement with the urban voters and urban parents who now have a fear that their child, lacking a high school diploma, will not be able to go to college in a state university.

Besides the fact that passing the MCAS will eventually be mandatory to graduate from high school, the fact that it makes teachers radically change their curriculum has caused many people to shy away from it. Protesters say the MCAS has caused a dramatic shift in teaching priorities. Instead of teaching general themes and ideas, schools are trying to cram as many individual facts into students' heads as possible in order to increase test scores. This has caused many teachers to oppose the MCAS on philosophical grounds and in terms of educational practice. They have gone so far as boycotting it and refusing to administer it to their students. But, supporting the MCAS are many of the leaders of the state, and they feel that the MCAS is needed to put teeth into the state's education reform initiative and provide accountability for the use of state funds.

A recent addition to the debate has been the issue of retaking the test if a student fails it the first time. It was recently decided that students who fail the MCAS exam on their first try will have at least five more chances to pass it and can take retests only in areas they did not pass that don't include the toughest questions that would simply tire and discourage them. This is according to a plan approved by the state Board of Education. The plan is an effort to prepare for very high failure rates: forty-five percent of 10th-graders in the state failed English and 34 percent failed math on the 2000 test. Some say allowing retests that don't include the most difficult questions represents a retreat from high standards that the MCAS was intended to create, but most administrators feel that it is pointless to subject this population to the additional stress of a long exam involving a massive fatigue factor.

One problem that may arise because of the MCAS is policies to encourage the privatization of public education. Because there is already dissatisfaction with public

schools, very poor MCAS scores will further discredit the urban public schools. If this happens, the public humiliation may set the stage for them to be taken over by for-profit charter schools, and many feel that this is a problem. The Charter schools will be in a position to pick and choose their student bodies and then take funds from the public schools commensurate with each pupil they take. In the end, the public schools will have fewer resources to deal with the toughest remedial cases.

Interestingly enough many of the districts across the Commonwealth which originally supported education reform, and the Reform Act of 1993 (with its provision for a representative minimum curriculum that would represent equal education in all districts), now vehemently oppose MCAS. From school district and individual school standpoints the MCAS is increasingly viewed as an invalid measure of what the students know, or too different from the curriculum already in place for the school to rapidly adapt to it. Some schools are ill equipped to meet the minimums on all parts of the test, and some students who have special needs or other difficulties may not pass the test at all given its present level of difficulty. This would exclude them from higher education programs designed to accommodate them. Toss into this mix the factors surrounding the upcoming requirement for 10<sup>th</sup> graders to pass the MCAS prior to graduation and all the elements are in place for a lively public debate.

A prime opponent to the MCAS for some time has been the vocational/technical schools scattered throughout the state. These schools are focused upon teaching their students the skills of a trade rather than core academics. The students in these schools often are often not academically inclined, and typically do not go on to college. Instead, they go on to become carpenters, electricians, and other such blue-collar professionals. Indeed, employment rates for graduates of these schools are incredibly high compared

even to many universities, and drop out rates for the vocational schools tend to be low. They bridle at the idea that they do not prepare their students with the skills they need to live in our society. However, with the pressure now on to pass the MCAS in order to graduate, these schools face the specter of high dropout rates with disillusioned students leaving en masse for public schools geared to MCAS preparation after their students start failing the MCAS and lose the option to go on to higher education beyond community college.

The skill range of the vocational students is often focused on the trade that the student is pursuing for all 4 years of their high school education. With half the attention given to academics that would be found in a normal public high school, the students are especially likely to fail the MCAS. Some schools are seeking exemptions on this basis from the MCAS, while others are attempting to adapt. These adapting schools are toying with wide ranging solutions, like expanding the school day to allow for more academics. Some support restricting trade preparation to the last two years of high school so that the first two years of the vocational student's time will be spent on academics associated with passing the MCAS exam. Many call for changes in the MCAS, including additions of material, that would better measure the vocational students' practical and tangible skills and those of other students who are better with their hands than their heads. David Driscoll, the Education Commissioner for the Commonwealth, said that the "...standard is going to be the same for all, so we have an expectation that vocational schools will get these kids up over the bar, which I think has been properly set for graduation." Despite complaints by vocational schools, as per the concept of equity described by the court case which originated the Reform Act, no one school or district gets special treatments in

failing to provide the State mandated minimum standard of public education, just as no school may be denied the opportunity to do more than is required.

This seemingly immovable position of equity in testing has had a variety of effects, mostly forcing schools all over the State, of all socioeconomic backgrounds, to review the curriculum and scramble to meet the common testing requirements. Despite an overwhelming flow of bad news regularly coming from the assessment results, there has been good news from some districts on occasion. For instance in 1999, there were 23 schools that did not have one failing student on the MCAS, quite an amazing feat given the admitted challenge presented by the test. Many schools showed marked improvement, some showed a little nudge of improvement. Meanwhile, the overall numbers remain somewhat dismal, with tens of thousands of students likely to fail the 10<sup>th</sup> grade test this year (and hence need to retake it to graduate). Still, many observers are hopeful. They think the struggling districts can improve greatly in just a few years, and will be much the better for it while many currently successful districts may have difficulty hitting or maintaining Proficient or Advanced performance for years on end. Complacency is on the decline and attention has been focused on public education and resources provided to reform weak schools.

In any case, the performance from 1998 to the present on the test offers some insight into trends based on school-wide and system-wide scores and rankings about the validity of the MCAS. However, the data are insufficient to really make assertions. Given that this is still only the 6<sup>th</sup> year of incorporation of the Curriculum Framework into the school system, and many districts have yet to integrate much of the material focused on testing into their classes it is still very likely that the MCAS will not yet be testing the students on what they were actually taught since 4<sup>th</sup> grade. One thing that is

certain is that the MCAS and the Curriculum Framework minimums do not represent the maximum that will be demanded of students or districts.

As part of the Reform Act, districts are required to pursue ever higher standards of evaluation in the hope to make them more saleable in the State, National, and world markets as skilled graduates. As such, the minimum performance expected on the MCAS has been set to be an upward moving bar, set to new heights as old heights are reached and surpassed. The drive toward producing more skilled graduates is supposed to compensate for the risk of defining a statewide minimum that many students surpass, and would regress toward, should no driving force be given to them to achieve. Despite low scores in many areas of the State, Governor Cellucci insists that the State will "...hold the line on standards; I think that's been very clear." The fear that schools will focus on the low end students to such a degree that they don't challenge the high end students and serve the gifted young people increasingly poorly is very real.

With some schools failing overwhelmingly, and the bar set to move ever higher in the near future, many schools have come up with creative ways to deal with the exam. Many schools have restructured their curriculum to cover the new material at an earlier time, swapping the years material is taught or taking material not on the test out altogether. This approach dismays many teachers who see it as disrupting the fabric of the 4-year high school educational experience they are trying to provide for their students. Others see it as beneficial by reducing repetition in non-critical areas and focusing teachers in different grades to systematically decide who is supposed to cover what.

Some schools now offer tutoring, which is either during school hours or afterward, in areas specific to the MCAS. The concept is all too often to teach to the test,

rather than to master the Core Curriculum itself. Still in most places this represents rising standards and a more serious approach. Other schools have stressed team teaching, using two or more teachers to illustrate the unity and connections between subjects.

Educational integration that is sometimes left for the latter years of some universities is being incorporated at the high school level, allowing for algebra and geometry to be interwoven in an effort to develop the same thinking skills necessary to achieve on Science and Social Studies portions of the MCAS. The curriculum guidelines stress developing a mastery of processes of “inquiry” as much as endorsing the coverage of any particular concept on subject matter.

In the end the people of the Commonwealth will have to be the ones to decide what stays and what goes, for the MCAS and the Curriculum Frameworks. Lobbying groups are being formed to sway opinion at the State House to get the legislature to back down from using the MCAS as an individual graduation requirement, but rather to stress educational system level accountability. Efforts are being made to expand the numerous special education exemptions already allowed under the Department of Education's assessment rules. Still, other groups are arguing vehemently in support of MCAS, that despite its failings the test fulfills a useful function for the Commonwealth.

What may be necessary is for the public to step away from MCAS for a time, to step back and see if MCAS is a symptom rather than a cause of their concern over their children as students. The MCAS is an assessment of the equitable education provided to students across the State in terms of the Curriculum Framework, a Framework mandated by the Reform Act. The Act itself was a response to the court decision in favor of school districts which called for sweeping changes to make the education system across the State more equitable and homogenized. So perhaps it is time for the public debate to

concern itself with redressing a central question of education rather than perhaps compounding a folly that in another 15 years will be resolved by another lawsuit and a new act calling for reforms.

## **2.7 Scholastic Aptitude Test (SAT)**

Adapted from: The Leicester SAT Experience. Written by Gerard Mangelot and Mario Tongol. April 30, 1998.

The Scholastic Aptitude Test was created in 1926 by the College Entrance Examination Board. The exam was aimed at determining high school students' preparedness for college. It was comprised of several lengthy essays. The test was taken by a small percentage of college-bound students and was used by a few college admissions staffs primarily to identify working class students worthy of scholarship support to attend Ivy League Colleges. It was normalized using preparatory school graduates who, on the whole, did well at Harvard and its sister schools. Since its creation, the SAT has changed considerably. Standardized multiple-choice questions were added to facilitate grading. Eventually, the essays were dropped altogether leaving only the standardized portion which was influenced by the IQ testing movement of the period (Wirtz et al. 1977).

The Educational Testing Service, created in 1947 by the College Board, is currently responsible for administering the SAT. Many colleges and universities require applicants to submit an SAT or ACT score for consideration. Therefore, most high schools "require", or strongly urge, college-bound students to take the SAT. The test is comprised of a Mathematical and a Verbal part. The score range for both parts is on a



scale of 200 to 800. The Mathematical portion is geared at the material taught typically in grades one to nine or up to and including geometry and algebra. The Verbal part determines a student's word relationship and reading skills (Crouse and Trusheim 1979).

High school students as well as college admissions staffs depend on the SAT to help them judge whether an applicant will fit in and flourish at a college. It starts when prospective students compare their SAT scores with the average published SAT scores for a certain college to decide whether to bother to apply. College admission staff members use the SAT to determine who to accept and decline for admissions, but it is not their prime decision guide in most cases. High school program difficulty and grades come first, then SAT and class rank. The SAT is valued as a control on local high school grading standards and the difficulty of the courses offered. However, it is in the "benchmark" decisions that the influence of the SAT is more subtle and insidious.

Those with SAT scores that are very high do not have the rest of their file examined very closely. One reader is considered enough and there is no committee review. Those with scores below a given level will not have the rest of their file reviewed at all or are deemed too great a risk to be eligible for scholarship support. In combination, SAT scores can have a considerable impact in determining access to competitive schools where the admissions office is looking for a weakness to use as grounds for exclusion. Even where a full committee review of the whole file is undertaken the focus of discussion is often discrepancies between the high school grades and SAT score construed as evidence of "over" and "under" achievement.

However, the whole idea of trying to predict future college performance is both necessary and controversial. Students who have worked hard and done well in high school are expecting their reward in the form of a range of options and possibly

scholarship support. For them, the SAT looms as a potential threat to their investment. Other students, who have not applied themselves consistently, can make up for it in one day on one test. It could be a “last chance” to showcase their capabilities, as aptitude or promise.

The reason for the inclusion of SAT background and data in this study is to examine whether the MCAS, like the SAT before it, can be shown to have elements of aptitude as well as achievement. Previous studies, which will be mentioned later, have shown that the SAT, which is viewed as an aptitude test, also has elements of achievement affecting scores and the combination favors certain types of learners as identified by the MBTI. Therefore, core curriculum and course preparation can account for up to a 400-point difference in SAT scores. However, comparably prepared students still differ in average by learning style in ways that are predictable and meaningful given Jungian theory, the basis for the MBTI.

## **2.8 ACT Assessment™**

The ACT Assessment™, or “A-C-T”, is a national college admission examination that sometimes replaces the SAT, or is taken in combination with the SAT by college-bound students. This SAT competition is based in Iowa and the ACT is more popular in the mid-West and south than the SAT. ACT results are accepted by almost all U.S. colleges and universities in lieu of the SAT anywhere in the country, but it is the underdog primarily in the Northeast. As noted earlier, the ACT is curriculum based, designed to cover what is taught in typical high schools and text books. By contrast, the SAT, which is an aptitude based test, is designed so as to avoid testing content, or what

one has been taught (and hopefully learned) at the level of facts. It is an assessment of reasoning ability, so as not to favor students from stronger secondary schools. The ACT is “designed to assess high school students’ general educational development and their ability to complete college-level work” (<http://www.act.org/aap/index.html>). The ACT covers four skill areas: English, Mathematics, Reading, and Science Reasoning. The ACT-Plan is given as a 10<sup>th</sup> grade test and it includes 215 multiple-choice questions and takes approximately 3 hours and 20 minutes to complete, with actual testing time being 2 hours and 55 minutes. One more feature of the examination is that it contains a questionnaire which provides the student with valuable information for career and educational planning. The PSAT is normally administered early in 11<sup>th</sup> grade. The ACT or SAT is then taken primarily in late 11<sup>th</sup> or early 12<sup>th</sup> grade.

The ACT is administered on five national test dates, which occur in October, December, February, April, and June. Some states also offer the test in September. The registration fee for the ACT is \$23, and \$26 in Florida. The ACT Assessment™ is prepared according to three standards: Standards for Educational and Psychological Testing, Code of Professional Responsibilities in Educational Measurement, and Code of Fair Testing Practices in Education. The test has been administered since the fall of 1959, and it has been offered in all 50 U.S. states since 1960.

The test is scored differently from both the MCAS and the SAT/PSAT. The highest possible score on the ACT Assessment™ examination is 36, and the odds of this happening are 1 in 8,131. A conversion chart between ACT and SAT scores can be found in Appendix C. It is of interest to us here because of its greater philosophical and content area alignment with the MCAS than the SAT, and the fact that the Fitchburg Public schools encourage its administrations, and are an area testing center.

## 2.9 Overview of Statistics

Statistics, the mixture of art and science behind determining the relationships between different individuals in groups, is the backbone of research into learning styles as they apply to academic testing. A detailed synopsis of statistical analysis is beyond the scope of this report, though a brief description of the concepts employed in the analysis of data in this project would go far to assist the untrained reader as to the meaning of the numbers, and results, we cite. To that end, a brief overview of the major concepts used in this report follows.

Statistical analysis is based on the scientific application of hypothesis testing to large data sets. Hypothesis testing itself is somewhat esoteric, with the accurate description of the nature of the theory which one wishes to test being difficult to develop. Suffice it to say that with statistics one tries to disprove a hypothesis, and typically the hypothesis is that there is no connection between 2 variables in the data set we are studying. For instance, if one wished to see if the men in a certain school are taller than the women there, one would compare the heights of as many men and women as possible. The hypothesis would be the null hypothesis that there is no difference between the heights. By disproving the null hypothesis, one would demonstrate that there is a difference, and hence can go on to describe that difference. This method of disproving a hypothesis that claims there are no special features in the data refers to testing the null, or no-difference, hypothesis.

Most hypothesis testing demands that a certain amount of assuredness be achieved in reaching a conclusion. This is represented by the p-value for the hypothesis

test, a decimal fraction of the total number of times that the result of our hypothesis test would result in the null hypothesis. Typically, this value is set low, perhaps 0.05 or lower. This value means that 5%, or 1 in every 20 data items, or less behave as in the null hypothesis. If less than 5% of the data conform to the no-difference rule, it is typically accepted that the difference is real, and the differences found in the data is statistically significant. Significance is merely a measure of confidence that the finding we see bears on the hypothesis we are testing. If our p-value was 0.2, then 20% of the data would conform to the null hypothesis but we would consider it invalid even if the test resulted in that whole 20%. Fundamentally, this means that we can set the bar for rejecting the null hypothesis as high or low as we want, but we need it low enough to matter in the real world.

What is done with the data at hand is dependent on the nature of the data at hand. Sometimes one has continuously increasing or decreasing test scores, for instance. Other times a variable consists of a list of distinct, discrete values that may be possible for a trait, such as people being male or female. Statistics work to compare these continuous and discrete data values in a variety of ways. Different methods are employed to study numerical, non-numerical, ordered and non-ordered, and other types of data. Generally, speaking though all types of analysis begin with getting related data in the first place.

Data collection is often a difficult process especially when large numbers of people or large organizations are involved. The problem is that data must not only be gathered accurately, but it is necessary to gather data from large numbers of cases that were selected in a random fashion. The large numbers are necessary because the more individuals in a sample, the more certain one can be that the null hypothesis has been disproved for any given difference or pattern found in the data. If the sample size is 100,

then a p-value of 0.05 may be reached easily if there is a sampling error. Let us say that the sampling is not very random, and we end up with 6 data cases which conform to the null hypothesis. Under the 100 size sample this says that the null hypothesis is true. But let us say that we have a sample size of 200, and this time, even with our not quite random sampling we pick up those same 6 conformists and 2 more, we have 8 null hypothesis data points. For the sample size of 200 and p-value of 0.05, then up to 10 data points may conform to the null hypothesis and yet we still reject it. What this means is that for larger and larger samples, we reduce the probability that an analysis will be so mistaken that reverses the null hypothesis test. Typically, statisticians strive for samples sizes near 1000 to ensure a reliable result of their analysis, but they are all forced to work with what they can get, and gathering even 200 complete cases can often be difficult.

With this concept of p-values and hypothesis testing in mind the present statistical analysis proceeds along one of two major paths, depending on what the data look like. By "look like", statisticians mean if they conform to some pattern that is easily recognizable. For instance, for many tests in school students' scores are scaled so as to make a bell curve, or normal distribution. This type of distribution indicates behavior of the sample tending toward an average value, or put another way, there are fewer outstanding students, and fewer total failures, than there are average students. Statisticians typically employ some initial tests to describe what distribution of the data look like.

These tests are often referred to as descriptive statistics, and include some common tools such as finding the sample's mean, median, and mode. More importantly, measures are based on the actual spread, or range, of the data values in the sample. Clustering of the data according to its ranking by quarters (4 equal groups) is referred to

as interquartile range, or the location of 25% and 75% of the data. That statistic gives an idea of general spread. A better idea is provided by the variance, or variation within the range of the sample. The variance is the sum of the distance of the data values in the sample from the mean. This variation can be compared between different samples to determine if their spread is greater or lesser than one another or so great in absolute terms that it really can't be described very well by the summary statistic. The standard deviation is the square root of the variance, a value illustrating the typical deviation of individual cases from the sample mean. Tied to this is the sample's standard error, the deviation of multiple samples from the same population from the sample mean. Together, this substantial list of statistics provides an idea of the spread of the data, the shape of its distribution, and how accurate any sampling measure has been.

Other descriptive statistics are commonly used, such as the confidence interval, the coefficient of variation (comparison of variation between samples), and skew of the data toward one extreme. Together, from these statistics, one gets a picture necessary to understand basics about the behavior of different samples, if one knows enough, and can be sure of some features such as normality on given variables of the distribution, it is possible to proceed with parametric statistical tests. If one is uncertain about some traits of the data, or if the data are non-normal, it is best to err on the conservative side and employ ranked, or non-parametric, statistical tests. For this project, the important analyses will involve Pearson correlation coefficient and standard regression techniques based on them.

Pearson correlation is the process of estimating the amount of variation in one variable, or set of data, due to the size and variation in another variable within the same sample. The meaning of the data is interpreted from the resulting  $r$  value of the test, from

-1 to 0, and to 1. This corresponds to perfect negative correlation to perfect positive correlation, with no correlation at all being in the middle at the zero point. A Pearson correlation value of 0.91 results in 83% of the variation in the size of the tested variable is "explained by" (can be predicted in terms of) variation in the other variable. If the p-value is very low, for instance 0.005, then the correlation explains 83% of the variation, and the chance that there really is no difference in the larger population despite the strong difference (correlation) in the sample at hand is only 5 chances out of 1,000.

The second major analytical technique used in this study is standard regression. Standard linear regression tests to determine the strength of the prediction value of a cause, or x, variable in determining the affected variable, or y. The strength of the relationship is often reported in terms of Pearson correlation values, and the slope and intercept for the best fit line through the x-y paired data provides the predictive mechanism for using the regression technique. This test assumes at the very least complementary variances for the tested variables, a linear fit to the data, and perfect determination of the x values. If the assumptions hold, this tool becomes incredibly useful in going a step beyond just describing patterns and into determining relationships. Its major advantage over the Pearson correlation coefficients on which it is based is the capacity to use multiple independent variables to predict values in a single dependent variable. It is essentially a multiple correlation coefficient.

The statistical analyses used in this report conform to the descriptions above, with some additional tools employed in the study of variance. What the reader may understand about the statistical process employed is that the data were studied, described using basic techniques, their variation studied using a variety of comparative methods, and finally major conclusions drawn on the basis of correlation and regression values.



For a further treatment of the methods, or a deeper understanding of the meaning of the numbers, consult any introductory statistics text. We used "Applied Statistics for Engineers and Scientists", by Petrucci, et al, published by Prentice Hall, Upper Saddle River, NJ, in 1999.

## **2.10 Previous Studies and Findings**

Adapted from: The Leicester SAT Experience. Written by Gerard Mangenot and Mario Tongol. April 30, 1998.

### **2.10.1 Previous MCAS Studies**

The MCAS is still new. Hence, published studies on the MCAS have been very scarce despite the huge effort in the state to disseminate the actual results to the schools in a useful form. In fact, there has been no data to present in this report from any previous WPI study. Newspaper reports are about all that we have seen and these present many numbers and system averages, but do not go into deep analysis of the data. The project being proposed in this document is one of the first, if not the first, study to create a database containing MCAS and MBTI data. Hence, this study will probably be the first to see if learning style as determined by the MBTI can help one identify students who are at risk to under perform on the MCAS, as it did on the PSAT. Finally, this study is also the first, to our knowledge, to make a comparison between an achievement and aptitude test, of which the MCAS claims to be the former, by putting the SAT results in the same file with the MCAS results. However, there have been extensive studies done on standardized tests, such as the SAT and PSAT, in relation to learning style. These studies

will be summarized in the following sections since the current proposed project also deals with other standardized tests other than the MCAS. The MBTI manual includes correlations of the MBTI dimensions with various standardized tests (including the PSAT and SAT) as well.

### 2.10.2 Previous ACT Studies

A statistical analysis done by McElhaney (1998) showed that the correlation between MCAS scores and Plan ACT scores were reasonably strong (0.5-0.6). This makes sense due to the fact that both the ACT and MCAS are achievement tests. The data for this analysis was collected from 10<sup>th</sup> graders in the Fitchburg school system that took the MCAS in the 1998-1999 school year. Below is a table of the correlation coefficients between sections of the Plan ACT and sections of the MCAS for these students. Notice that most of the coefficients are above 0.5, which means that the correlations are strong. Therefore, a student’s performance on one of the tests (i.e. the MCAS) can be used to predict the same student’s performance on the other (i.e. the ACT) at levels considerably better than chance. A 0.5 correlation means that 25% of the variance in the dependent variable (MCAS) is explained by the independent variable (ACT). For a 0.6 correlation, the percent of variance explained is a robust 36%.

	<b>MCAS-10 English Scaled</b>	<b>MCAS-10 Math Scaled</b>	<b>MCAS-10 Science Scaled</b>
<b>ACT-Plan English</b>	0.613	0.541	0.535
<b>ACT-Plan Reading</b>	0.564	0.442	0.430
<b>ACT-Plan Rhetorical</b>	0.554	0.507	0.522
<b>ACT-Plan Algebra</b>	0.535	0.684	0.588
<b>ACT-Plan Geometry</b>	0.374	0.574	0.397
<b>ACT-Plan Math</b>	0.500	0.673	0.561

<b>ACT-Plan Science</b>	0.506	0.531	0.511
<b>ACT-Plan Usage</b>	0.611	0.512	0.500
<b>ACT-Plan Composite</b>	0.639	0.633	0.591

Further, other research has reported even higher levels of correlation between students' PSAT and ACT Plan scores, in the 0.7-0.8 range. This research was presented in an oral presentation by Eric Tapley in Gainesville, FL, in March of 2000 and is fully documented in his IQP report submitted on March 13, 2001, but was not available to us at the time our analysis was completed.

### **2.10.3 Previous SAT Studies**

There have been several studies comparing personality type, cognitive learning style, and performance on standardized testing conducted by Worcester Polytechnic Institute students in area schools during the past few years. Each of these studies resulted in strong evidence to support the theory that test scores on the PSAT or SAT vary in ways that favor certain types of learners. Though studies have been performed here since 1995, the majority of our hypotheses will be based on the study by John Pieper, which concluded in early 1997. Pieper's project generated by far the largest, most inclusive (PSAT) data set of all. When Keith McCormick completed his work analyzing these data, it had the most detailed analysis of all of the data sets. Pieper and McCormick's study involved students from four Worcester Public High Schools (WPS), Nashoba Regional High School and Massachusetts Academy of Math and Science. The following sections will summarize the major results and findings of that study.

Previous studies have shown that the "sensing" versus "intuitive" dimension of the MBTI is by far the strongest indicator of performance on the SAT found on the two

measures (MBTI and GCSI) utilized in this study. Another common indicator of SAT performance is the “perceiving” versus “judging” section of the MBTI. The “extroversion” versus “introversion” and “thinking” versus “feeling” portions of the MBTI did not result in any significant findings in the previous studies, though the literature includes reports favoring the introverts, and Isabel Myers-Briggs expected that to be the case. Indeed, she expected E/I to be the second strongest predictor. In Pieper’s study, the second strongest indicator was the Remote Associates dimension of the GCSI, but there is no MCAS data available in a data set with the GCSI results, so we will not go into those findings here.

In the Worcester Public School (WPS) System, intuitive (N) types scored, on average, 122 points higher than sensing (S) types. In each of three studies, this trend proved true with the intuitive (N) advantage ranging up to 162 points. In WPS 49 of the 122 point intuitive (N) advantage came in the math section while 73 points came in the verbal section.

The intuitive (N) advantage is probably due to the multiple choice format of the SAT. Students with a strong intuitive (N) sense would be able to pick out the right answer from a list even if they could not have figured it out without clues. The discrepancy between that advantage in the math and verbal sections can probably be attributed to the “grid- in” portion of the math section. Since the answers are not listed in this section, the intuitive advantage in recognizing answers rather than generating them is neutralized when the possible alternatives are not offered.

Previous studies have also found an advantage in SAT performance among perceiving (P) types. In WPS perceiving (P) types scored 45 points higher on the SAT

than judging (J) types. This trend was supported by data from other previous studies. In these studies the perceiving advantage ranged up to 111 points.

The perceiving advantage can be attributed to perceiving types taking in more information before coming to a decision. This allows them to more fully understand the question instead of picking the first answer that looks possible or reasonable. As with the intuitive (N) advantage, the perceiving (P) advantage is predominantly found in the verbal section. This could be in part be attributed to the influence of the “grid- in” style questions as well. Also, this can in part be attributed to the typical “perceivers’ ” superior reading comprehension on the verbal section (Pieper 1997). They tend to pick up details on the first read before they know what they are looking for, and focus their attention. J’s focus early, and look for only what is relevant. They will generally have to go back through once they know what to look for.

Pieper’s study in the Worcester Public School system included an analysis of the sixteen possible personality types in the MBTI versus performance on the PSAT. The highest scoring MBTI type was INFJ. The lowest scoring MBTI type was ESTJ. In Nashoba Regional High School, the study showed the highest scoring type was INTJ. The lowest was, again, ESTJ. These two types were about 300 points apart in both cases. Pieper’s report did not include average PSAT scores for all of the sixteen types. He reported the top three versus the bottom three for Worcester. These two groups were about 250 points apart. Tongol and Mangenot hoped to do that analysis again in Leicester, but the cohort size (110 combined) was too small to support an analysis that would divide the “sample” into sixteen categories (1998).

Previous studies also reported similar findings to Pieper’s study for the MBTI. Kibbler and McTague reported a 108 point advantage for Intuitive (N) types over Sensing

(S) types. They also reported a 102 point advantage of Perceiving (P) types over Judging (J) types. Batey, Brezniak, and Purohit (1995) reported a striking 162 point advantage for Intuitive (N) types over Sensing (S) types. They also reported a 111 point advantage of Perceiving (P) types over Judging (J) types. In addition, Batey et al. found a 68 point advantage of Feeling (F) types over Thinking (T) types. See the chart below.

***MBTI Type Advantages Found in Previous Worcester Area Studies***

	<b>Batey et al.</b>	<b>Kingsland et al.</b>	<b>Pieper</b>
<b>Intuitive Advantage</b>	+162	+108	+122
<b>Feeling Advantage</b>	+68	Unreported	Unreported
<b>Perceiving Advantage</b>	+111	+102	+45
<b>Cohort Sizes</b>	229	276	1267 (Jr + Soph)

**2.10.4 Average SAT Scores from Previous Studies**

One more piece of information that is vital to our study is the average SAT scores at various schools from former studies. Only Pieper’s PSAT findings will be noted here because Kingsland et al’s previous research was done on the original SAT before it was changed in 1995. Batey et al. used the first available version of the “New” SAT. Batey et al. used only the first SAT score, not the highest one as Kingsland et al did. Pieper reported average Verbal PSAT scores in the Worcester Public School system ranging between 399 to 443 points depending on the high school. Math scores ranged between 375 and 428 points. In Nashoba the Verbal average PSAT score was 506 points and the Math average was 501 points. At the Massachusetts Academy of Math and Science, the average Verbal PSAT score was 669 points and the Math PSAT score was 706 points.

All of these scores are expected to move up about 50 points from the PSAT to later SAT scores. The Worcester Scores were also deflated by the administration's attempt to get all students, even those not college bound, to take the PSAT that year. Unlike the WPS, at Nashoba High School, the 85%-90% of the students in the junior class who took the PSAT and who participated in the study were planning to attend college or at least considering it as an option.

***PSAT Averages From Pieper's Study***

	<b>WPS</b>	<b>Nashoba</b>	<b>Mass Academy</b>
<b>Verbal PSAT</b>	419	506	669
<b>Math PSAT</b>	406	501	706
<b>Total</b>	825	1007	1375
<b>Cohort Sizes</b>	1267 (Jr + Soph)	308 (Jr + Soph)	35 (Jr Only)

A simple examination of these numbers shows that the Mass Academy scores are by far the highest and the intuitive advantage diminishes as the scores rise. This is expected since the students are an exclusive pre-selected group that took the PSAT before as part of their admission process to the academy. Here, the E/I relationship become more important. Nashoba is the second highest overall. Since it is a somewhat suburban/rural public school, the Leicester results were expected to somewhat mimic the Lancaster students in the Nashoba district, but Leicester is more of a working class town than Bolton or Stow, the other two towns in the Nashoba Regional System. The WPS results are, on average, the lowest. Since many students at the Worcester schools are from lower socio-economic backgrounds, and many are not college bound, this is to be expected. Pieper does not report PSAT scores by social class, but the larger literature does so, documenting a strong relationship.

Finally, Tongol and Mangenot made some discoveries that will eventually coincide with this study. They found that in terms of MBTI type, the Sensing (S)/ Intuitive (N) and Judging (J)/ Perceiving (P) parameters gave not strong, but moderate findings in relation to PSAT/SAT performance (1998). Intuitive (N) and Perceiving (P) types can “outscore their counterparts consistently when they are compared only to those of their own class year and academic program” (Mangenot, et al. 1998). This was an extremely interesting finding and was one of the foundations for which the proposed project was built. Can the same Intuitive and Perceiving parameters as determined by the MBTI predict strong performance on the MCAS as they did for the PSAT and SAT? In order to answer this question, it first must be determined whether or not the MCAS shows elements of aptitude and not only achievement.



### **3 General Hypotheses**

As mentioned in section 2.10 of this report, certain parameters of the MBTI, especially the Sensing (S)/ Intuitive (N) and Judging (J)/ Perceiving (P) types, have been shown to somewhat predict how a student will perform on the SAT. It is best that the data is analyzed under a control variable, i.e. for those classmates in the same academic program. When this is not done, MBTI relationships can be exaggerated as in Pieper's WPS study, or can be obscured altogether, as in the Leicester study. This is due to the impact of differences in preparation and which learning types have the greatest access to the more challenging Advanced Placement and Honors courses in a given high school. Therefore, the correlations that will be made between learning style via the MBTI and the SAT should also be strong because the students are of the same class.

In relation to the MBTI and individual subject grades or overall GPA relationship, it is hypothesized that those with the higher GPA's will tend to be certain learning types, although it is unknown which ones at this point. Isabel Myers predicted that "judging" students would do better in class day-by-day, but the "perceivers" would outperform them on standardized tests like the PSAT. Students with a higher GPA are often encouraged to take more challenging courses, and therefore, they are more prepared for standardized examinations than other students. However, some decline, so as to keep their grades higher in a less challenging program. It is also thought that classroom performance in certain subjects, such as science and English, will be a predictor for performance on the same sections of the MCAS rather than general GPA predicting average scores.

Since both the ACT and MCAS are achievement tests, it is expected that certain learning types determined by the MBTI will be more prone to underperformance than others. This follows the logic that the ACT and MCAS are highly correlated as shown in the literature review. Also, certain MBTI types have under performed on sections of standardized tests which have elements of achievement in them, indicating that certain MBTI types should be expected to perform similarly on the MCAS. While there has been little to no research in this area, the hypothesis has a foundation. Remember that the SAT has been shown to have elements of achievement in it. Also, previous research has shown correlations between the PSAT/SAT and ACT, as mentioned in section 2.10 of this report. Therefore, if certain learning styles by MBTI standards give some students an advantage on the sections of the PSAT/SAT that are supposedly “achievement sections”, the same MBTI types will give the same students an advantage on both the ACT and MCAS assessment exams. It was also shown before in section 2.10 that many sections of the MCAS and ACT show strong correlations. Therefore, performance on sections of one test should predict performance of similar sections on the other test.

Getting back to the main focus of the project, which is the effect of MBTI type on MCAS performance, it is expected that there will be certain types that will have advantages and others that will be at risk. In section 4.5 of this report, there are questions of correlation listed. It is expected that there will be significant correlations between most, if not all, of the relationships that are stated.

## **4 Methodology**

### **4.1 Objectives**

The scope of this project goes well beyond trying to find if there is a relationship between the dimensions of the MBTI and the MCAS. By itself, that is not a complex analysis. However, if there are correlations between these two entities that finding will give rise to new hypotheses, a search for more correlations and replications of existing findings may be necessary. For instance, since there has been no other MCAS/MBTI study, this information should be compared to the findings of prior studies on the SAT/PSAT, grades and level of course difficulty, the Plan ACT (if available). Other analyses may be called for by the school systems involved. Whatever they deem necessary to develop an understanding of which students are at risk to under perform on this standardized test is worth doing, so long as they can provide the necessary data. Ethnicity and social class are likely to concern them.

Therefore, in this study, we seek to find in-depth relationships between MCAS scores, overall and by subject, using all four specific MBTI parameters, including discrete and continuous scores for each. Based on previous studies with the PSAT and SAT, the analysis will be focused on the Sensing (S) and Intuitive (N) aspects of the MBTI. This variable has been shown to be the most correlated with test scores on other standardized examinations.

The second part of this study will involve a brief guide to identifying students who are at risk to under perform on the MCAS so that the Fitchburg and Worcester public school systems can handle any plans to use this lead time as they see fit. It must

be stressed that it is not the intention of this study to instruct the school systems how to deal with these students. Suggestions for such ends will not be a part of this study. However, it is the intention of this project team to aid the school systems in looking at MCAS data help identify at-risk students and explain why the test is especially challenging for them. If time permits and the school systems are willing, we would also like to set up a meeting or presentation in which we present our findings and lend our insight as to why students may be receiving lower scores on the MCAS. We would also like to do this in order to answer any questions that the school system officials may have pertaining to the study in person. However, if asked for advice on how to handle students with risk factors, we prefer to leave that to the professionals who have been trained in that particular field, as we can offer only our opinions.

#### **4.2 Current Status of Worcester and Fitchburg MCAS**

Because of the MCAS, the Worcester and Fitchburg public school systems are in dire straights. Both schools are seriously under performing on the statewide test, and if the problem isn't resolved, a large percentage of their students will not be able to graduate from high school. The following table shows some statistics for the Science and Technology scores on the MCAS. These figures are from eighth graders who took the MCAS in 1999. They are the class of 2003, which takes the 10<sup>th</sup> grade test the first time this year, 2001.

**1999 MCAS Scores (8<sup>th</sup> grade)**

<b>School System</b>	<b>Rank</b>	<b>Advanced (280 - 260)</b>	<b>Proficient (259 - 240)</b>	<b>Needs Improv. (239 - 220)</b>	<b>Failing (219 - 200)</b>
Westborough	11	21%	44%	22%	13%
Algonquin	15	7%	39%	35%	19%
Nashoba	27	8%	30%	34%	27%
Quabbin	55	6%	33%	36%	26%
Leicester	64	3%	37%	32%	28%
Marlborough	80	6%	24%	27%	43%
Narragansett	?	5%	18%	30%	46%
Leominster	91	3%	20%	28%	48%
Fitchburg	118	1%	10%	21%	66%
Worcester	120	2%	13%	18%	66%
Boston	127	1%	8%	15%	71%
Statewide		5%	23%	27%	45%

It can be seen from the table that in Fitchburg, Worcester and especially Boston, over two thirds of the students are likely to fail the test and thus, not be eligible for graduation. Besides the fact that many whom would be normally eligible to graduate will not, there is a large percentage of students in the urban districts that cannot seem to meet the state's minimum required knowledge base. There is currently a movement afoot attacking the MCAS, supported by educators. It is obvious that something must be done to improve this situation if it really reflects educational performance and levels of learning by the school systems. If not, the test will create great hardship, and many educators comfortable with the idea of evaluating schools want to see the provision that will keep some students from graduating repealed.

### **4.3 Parameters of Study**

The study contained students who attended high school in the Worcester and Fitchburg public school systems. All of these students graduated from high school in the year 1999, 2000, or 2001<sup>1</sup>. The MBTI was administered to them when they were in 10<sup>th</sup> or 11<sup>th</sup> grade, which was in 1997, 1998, or 1999 for most of them. The MCAS exam was also administered when they were sophomores. The number of students in the Worcester Public School system that took the MBTI and (presumably the MCAS) was about 1200, and there were 182 who took the MBTI (and presumably both exams) in the Fitchburg school system. Therefore, the total number of students available for study from both school systems was 942. This number was large enough to provide a reliable study and give meaningful results. Of these 942 students, the database contained somewhere in the neighborhood of 750-850 students from both systems for whom we received MCAS and PSAT/SAT scores. A large portion of the data came from the Worcester Public School system, simply because there are four high schools there, and a significantly larger number of 10<sup>th</sup> graders for which there was MBTI data. Information for every student was not available. Hence, the database contained lapses and we took a loss of up to 200 students that took the MBTI.

### **4.4 Database Creation**

Again, the number of students that were contained in the usable Worcester database from both school systems totaled just under 1200 students. Encouraging responses to our

---

<sup>1</sup> The class of 2001 data from Worcester was added in to reach 1000 cases, as this was more statistically reliable. This cohort was missing data from Doherty High School., which serves the middle class area.

request for MCAS data from Worcester were received and data was provided. Individual transcript data (subject, grade, course number) was received from both Worcester and Fitchburg, SAT data was also obtained from both school systems for about 50% of the students who take that test, Plan ACT data from Fitchburg, any other well-known achievement tests that students in both systems have taken, and any remaining MBTI data from a few students in the Fitchburg school system were requested as well. Letters were written to contacts in both school systems to request these data<sup>2</sup>. As detailed in the letter, a meeting was scheduled with the school systems before statistical analysis in order to outline what the school system would like the project to focus on in addition to our goals.

Once the database was created, many parameters were included, such as class, year when the test was taken, MBTI type, MCAS scores in all subject areas, verbal, math, and overall SAT scores, individual subject grades, ACT scores (where applicable), are also by subject and composite. Microsoft Excel<sup>®</sup> was used to build the database, and it was organized well enough so that it could be utilized by the school systems involved to see if they could have identified students at risk to under perform on the MCAS with such a mix of variables had it been created before the students took that test at the end of the 10<sup>th</sup> grade.

#### **4.5 Status of Project/Explanation of Findings**

This study of the MCAS in terms of the MBTI was born from an IQP project began in late 1998 by Michael McElhaney under the guidance of John Wilkes. While the scope of the original project was large it has been wisely cut to more manageable pieces

<sup>2</sup> This was Mr. Lamey from Fitchburg and Dr. Mostue from Worcester- refer to Appendix C for the letters.

and attacked by a number of IQP teams in the interim, with a key aspect being examined by the team of Ken Fountain and John Stambaugh. Their goal, which is the focus of this project report, was to look at the central relationship between the MBTI preferences and performance on the MCAS, a lofty goal which links together many other pieces of related work.

Michael McElhaney was brought on board to assist in completion of the reporting phase as well as to drive a large portion of data merger and analysis through. During the first few months of 2001 Ken, John, and Mike worked together pulling the last of what the project needed together. The first run analyses and preliminary results were obtained by the last day of the term, and the majority of the written report was presented in first draft form. Unfortunately there remains infrastructure analyses, redrafting of the final report, and database refining to complete during the Spring of 2001.

As it stands therefore this project report is incomplete. It correctly expresses the design considerations, hypotheses, and methodology employed in pursuing the questions which were asked but it does not contain final conclusions. Draft conclusions, along with another 40 pages of discussion and results have been put together but are not yet ready for publication. Given that Michael McElhaney must submit the project in its present state this document represents a charting document whose counterpart, a full, refined explanation of the analyses results, will be released later in Spring. The student team agreed that given Michael's contributions in terms of report writing and content, as well as the data handling and analyses done, that his efforts may come to a close.

The final instantiation of the results sections of the report, along with discussions developed over the course of the project, will be included in the follow up project report produced by Ken Fountain and John Stambaugh. The final data will be made available



via Professor Wilkes, and a handful of analyses may be added on the part of project consultants. Please be patient in waiting for well described findings from this study, as MCAS testing now represents a fact of life for our communities. By better understanding performance on the measure, we are more prepared to help our children succeed.

## **9 Project Evaluation**

The Interdisciplinary Qualifying Project (IQP) is designed for students to do research in an area which integrates science and technology in society. It is the opinion of this research team that the IQP differs from the Major Qualifying Project in two major ways. The first is that more times than not, the IQP is done outside of a student's major. The second, and most important, is that the IQP seeks to explore a society-technology issue or social problem in society, which relates to some sort of policy issue. This issue usually has something to do with a social or political system (involves classical or social science) and the educational system, which is the focus of this project. The MQP, on the other hand, seeks to pioneer an area of scientific or engineering research which has had little to no consideration paid to it. This is not always the case, but in this day in age, where the university is trying to keep up with technological advances, it is more efficient and recognizable to be part of a development team than a problem solving (quality) team.

It is with these thoughts in mind that this evaluation of the MCAS/MBTI project that we write down what this project has done to heighten the skills that both do and do not apply to the career paths we have or will choose in the future. All of the members of the team were able to perform all of the relevant skills involved with database creation and statistical analysis. We gained experience in data handling as well as negotiating with school system officials for data that was not necessarily available for public viewing. Also, while the three of us had taken a statistics and math course before, this project was an actual application of that theory to a real-life problem in which many things were at stake for the schools, teachers, and most of all, the students. We see the MCAS and SAT as devices, or instruments (technologies), designed to assess a social

system, which involves design choices, or tradeoffs. Finally, it provided experience in writing an extensive and detailed report of our findings, so that someone who was not familiar with the situation at all would understand what we intended to do, what we did do, and what we suggest others do in the future to make this research even more meaningful.

However, this project was not always a pleasurable experience. It was very hard to obtain data in a useful form from the school systems. Time was an issue as the school officials had other responsibilities and higher priorities than catering to our needs on the spot. Hence, we did not always get the idea that our time and contributions were valued. We had brief windows in which to help out, and were not on the payroll, so waiting was not making us money, nor making progress. It was hard to be patient and then expend our time fixing dozens and dozens of database errors others had made, and not wonder if we would ever get to use our real skills. Other personal responsibilities on the part of the team members also hampered efforts to meet some deadlines. However, with any research, not everything goes according to schedule, and the manner in which we adapted to these obstacles was a large part of the success of the project. In the end, we made progress, contributed, and answered most of our questions. The next project team will hopefully start in a much better place than we did.

## BIBLIOGRAPHY

An ACT Program For Educational Planning. (2000, September). *ACT Assessment® Homepage* [Online]. Available on the World Wide Web: <<http://www.act.org/aap/index.html>>

Batey, Danielle et al. "Cognitive Bias In The Scholastic Aptitude Test." Unpublished Interactive Qualifying Project, Worcester Polytechnic Institute, 1995.

Crouse, James et al. The Case Against The SAT. Chicago: The University of Chicago Press, 1988.

Kingsland, David et al. "Learning Style Biases In The Scholastic Aptitude Test." Unpublished Interactive Qualifying Project, Worcester Polytechnic Institute, 1995.

Mangenot, Gerard C. and Mario L. Tongol. "The Leicester SAT Experience." Unpublished Interactive Qualifying Project, Worcester Polytechnic Institute, 1998.

Massachusetts Comprehensive Assessment System. (2000, September). *Massachusetts Department of Education* [Online]. Available on the World Wide Web: <<http://www.doe.edu.mass/mcas>>

McElhaney, Michael F. "MBTI Data And Its Application To MCAS." Unpublished Interactive Qualifying Project Report, Worcester Polytechnic Institute, 1998.

Pieper, John B. "Cognitive And Learning Style Biases In The SAT/PSAT." Unpublished Interactive Qualifying Project, Worcester Polytechnic Institute, 1997.

Wirtz, Willard et al. On Further Examination: Report Of The Advisory Panel On The Scholastic Aptitude Test Score Decline. New York: College Entrance Examination Board, 1977.