Assessment of Rationality of High School Students in the New York City Area

Jin Li

Stuyvesant High School

I have adhered to the school's policy regarding academic honesty in completing this assignment.

Submitted to Stuyvesant mathematics teacher, Dr. Bernard Feigenbaum.

2017

Abstract

Measuring one's rationality, though very a complex and difficult feat, provides a strong framework for understanding the human mind. Though it has been a relatively new area of study, the Comprehensive Assessment of Rational Thinking (CART) exam has been developed by Keith E. Stanovich to further advance the measurement of rationality. Using his research and rationality questions, a survey was created to analyze the rationality of high school students in the New York City region. The results of the test shows that age and being in Stuyvesant, among many others, are significant variables in determining a person's rationality. Recommendations include improving the rationality of students to better prepare for decision making in the real world. Applications of the CART and the results of this paper include raising awareness of the value of improved rationality for decision making.

Keywords: Rationality, CART, survey, high school students

Introduction

The study of rationality and decision making has always been a field that intrigued people. Dating back to the Classical Greek Period, philosophers like Plato and Aristotle have debated and established the foundations of the study of human reasoning. Yet only recently has psychologists vigorously studied the thought processes behind rationality. Even more recently has a test been developed to measure the degree of rationality of human beings. As a matter of fact, the first prototype of the CART (Comprehensive Assessment of Rational Thinking) exam has been completed, administered, and analyzed by Keith E. Stanovich and his fellow psychologists in the last year. Because this field of study is relatively new, the measurement of rationality still has much to bring to the understanding of human-beings.

Most people have a vague sense of what rationality is because the definition itself is ill-defined. Listed in the Oxford Dictionary, rationality is described simply as, "the quality of being based on or in accordance with reason or logic."[1] However, behaving "in accordance with reason or logic" is not specific enough to warrant a clear definition; there are many interpretations of what is reasonable and logical with respect to different fields. Yet the vague Oxford definition provides a very strong intuitive description that can very well replace the more technical definitions.

However, because many would be unsatisfied with an intuitive definition, rationality will be defined with the listed definition of Keith E Stanovich, whose work heavily influenced this paper. Stanovich uses Amitai Etzioni's normative model, stating that rationality is the "distance of thought or behavior from the optimum defined by a normative model" (Stanovich, K. E., 2016). Then an irrational behavior is one in which the behavior departs from the optimum prescribed by a normative model.[2] In addition, there are two types of rationality used by many cognitive science theorists: instrumental and epistemic. Instrumental rationality may be defined by the optimization of the individual's goal fulfillment. Then an irrational behavior based on the normative model is one in which he fails to maximize the expected value based on the available information. On the other hand, epistemic rationality is concerned with how well beliefs maps onto the actual structure of the world. Then based on the normative model, an irrational person fails to correctly apply their beliefs to the real world.

People deviate from the normative model of rationality based on various biases, which is defined in Stanovich's book as "the systematic errors that people make in choosing actions and in estimating probabilities." Heuristics refer to why people make these errors. One major example of the heuristic that causes irrationality is the interdependence of mindware[3], conflict detection, and ability to override. If the midware is available, the detection of conflict can be assessed and the person may override the initial belief. But if the subject fails to detect any reason to override their intuitive or initial response, his action may result in irrational behavior. For example, consider this question: "A bat and a ball costs $1.10 in total. The bat costs 1$ more than the ball. How much does the ball cost?" Many people give the answer that immediately comes to mind: 10 cents. They fail to detect any

[1] The Merriam-Webster and American Heritage Dictionary provides a very similar definition.

[2] For more information see *Normative-affective Factors Toward a New Decision-making Model* by Amitai Etzioni

[3] Mindware is the mental knowledge and procedures that a person uses to solve problems or make decisions.

conflict (that a 10 cent ball would result in $1.20 in total) and fail to override the initial belief.

Review of Literature

Because the measurement of rationality is relatively new, there is not a diverse literature relating to the study. Though there has been decision-making tests, they do not come close to a comprehensive study of rationality. However, Stanovich's life dedication to the topic has brought ground-breaking data and analysis to rationality. He and his colleagues have developed the CART exam and administered it three times: one to their own laboratory, another to James Madison University and the Amazon Mechanical Turk, and another to an unspecified group (Stanovich, K. E., 2016).

The testing at Stanovich's laboratory was short-form (a shorter and less comprehensive CART exam) and was administered to 372 university student volunteers with a mean age of 18.7 and standard deviation of .9. His two-hour test contained topics including disjunctive reasoning, probabilistic and statistical reasoning, belief bias syllogism, reflection versus intuition, scientific reasoning, ration bias, probabilistic numeracy, antiscience attitudes, dysfunctional personal beliefs, and conspiracy beliefs.[4]

He compared the results to SAT and AOT (Actively Open-Minded Thinking scale) scores, finding that scores of most of the subsets of the exam are correlated with SAT and AOT scores[5]. The only subset that did not correlate with the SAT and had a low correlation with the AOT is the dysfunctional personal beliefs, which involves measuring one's contaminated mindset towards science, evidence, and justifications (Stanovich K. E., 2016). His analysis also indicated that four of the subsets, namely the probabilistic and statistical reasoning, scientific reasoning, reflection versus intuition, and the probabilistic numeracy test had relatively strong intercorrelations with each other. He also found that those four categories were the strongest predictor of the total score. In addition, Stanovich found that men performed better on probabilistic and statistical reasoning, probabilistic numeracy, reflection vs intuition, and the dysfunctional personal beliefs subtest. But it is important to note that the males in his sample were also higher in cognitive ability, as they had higher SAT scores. Rationality is an subset of cognitive ability, so the fact that males are more rational may be caused by a higher cognitive ability rather than being male. The means scores for older students were also higher.

The full-form CART exam was administered to James Madison University (JMU) and the Amazon Mechanical Turk (Turk). The testing was more in-depth for these subjects and there were more topics involved. The notable differences between the two groups was the fact that the Turk sample age varied from 19 to 68, whereas the JMU sample age varied from 18 to 32, and less than 1 percent of those at JMU were 30 or older.

The Turk sample was confirmed to have a higher cognitive ability using the Cognitive Ability Composite (CAC).[6] Based on the CART exam, the Turk sample performed significantly better overall (mean

[4] These topics will not be fully explained because it is not essential to the paper.

[5] For more information, see *Assessment of Actively Openminded Thinking* by the University of Pennsylvania

[6] For more information on the test, see Heather Pillman's presentation on *Understanding the Cognitive Abilities Test™ (CogAT®).*

of 81.8 compared to 68.1) and better on 16 of the 20 subtests. There was a high correlation between the CAC with the CART exam grade and a lower correlation between the SAT score and the CART exam grade. However, the highest correlation was between the AOT and the CART exam. The CAC and the AOT predicted 57 percent of the variance in the CART total score. The three subtests that had the strongest intercorrelation was the probabilistic and statistical reasoning, scientific reasoning, and the reflection versus intuition. Comparing males to females, males performed better than females in the CART exam in both samples. But it is also important to note that the men performed better on the CAC and SAT, indicating a higher cognitive ability. Men performed better than female in 38 of the 40 comparisons of subtest, and performed notably better on probabilistic and statistical reasoning, reflection versus intuition, practical numeracy, and financial literacy and economic knowledge subsets.

The last test administered was called 38-point 2-subset CART because it was very short compared to the first two exams (total of 38 points) and had only two subsets: probabilistic and statistical reasoning and scientific reasoning. The details to this test was not comprehensive.

Testing correlations among the test, Stanovich found that there was a .96 correlation between the Full-form CART and short-form CART, .88 correlation between the Full-form CART and 38-point 2-subset CART, and a .89 correlation between short-form CART and 38-point 2-subset CART.

Though the CART exam is unique to Stanovich, there were other tests that measured processing ability that included areas of rational thinking. For example, the A-DMC (Adult Decision Making Competence) and Y-DMC (Youth Decision Making Competence) assessed consistency in risk perception using tests on probability numeracy, resistance to framing (tests whether choices are affected by irrelevant differences in problem description), resistance to sunk costs, and applying decision rules (assess how well individuals are able to use different decisions rules) (Parker, A. M., & Fischhoff, B., 2005). However, this exam contains other aspects that were not analogous to the CART exam and to rationality, making the A-DMC a weak indication of rationality.

The Halpern Critical thinking Assessment (HCTA) is also an instrument used to measure critical thinking skills. It tests hypothesis testing, likelihood and uncertainty, argument analysis, verbal reasoning, and problem solving skills. A huge aspect of the test contains areas associated with rationality (Halpern, D. F.). For example, the HCTA's hypothesis testing contains many problems related to scientific reasoning in the CART and its likelihood and uncertainty is analogous to the CART's probabilistic and statistical reasoning subset. But because the HCTA is notably less comprehensive than the CART, the HCTA is not a strong indicator of rationality.

Entering more loosely related territories, the Criteria Cognitive Aptitude Test (CCAT) measures one's ability to solve problems, digest and apply information, learn new skills, and think critically. Though there are problems in the test that relate to rationality, the test is more geared to cognitive ability, which includes aspects other than rationality. As a result the CCAT can not be used to accurately predict rationality. More well-known is the Intelligence Quotient (IQ) Exam. However, this measures one's reasoning ability and intelligence (the ability to acquire and use knowledge), which is loosely related to rationality. However, the IQ exams and the

rationality exams are well-correlated, even if IQ exams can not adequately represent rationality.

The Theory

Based on Stanovich's work, males are expected to perform better than women. Based on the analysis of the brain, male brains utilize seven times more gray matter for activities than female brains. Gray matter of the brain are responsible for information and action-processing that is a critical component of the CART exam (Jantz, G. L., 2014). The blood flow differences in the two brains may also account for the differences in the CART scores. The female brain has more natural blood flow throughout the brain at any given moment; as a result, women tend to revisit emotional memories more than men. Males, on the other hand, tend to reflect more briefly on an emotive memory, and are quick to move onto the next task. This indicates that males suffer less from bias, as bias may arise from events that had occurred earlier in the one's life and the better ability to recall these events may impact his or her decision making. Thus men are expected to do better on the probability and statistical reasoning, scientific reasoning, and belief bias subset, while doing worse on reflection versus intuition (as they are quickly to move onto the next task). However, it is important to note that men tend to have a higher cognitive ability, which is a confounded variable as higher cognitive ability may influence a person's rationality (Stanovich, K. E., 2016).

People who spend more time doing math problems are expected to score higher on the CART exam. A person who spends more time doing math problems develop their processing and reasoning abilities, thus improving their rationality. Thus they are expected to do better on the probability and statistical reasoning, scientific reasoning, and belief bias subsets. However, it is important to note the flaws of the correlation between rationality and amount of time spent doing math problems. For example, more time spent on doing math problems may simply reflect that the person is bad at math, indicating that they have a lower cognitive ability. On the other hand, a person who is good at math may spend very little time doing their math homework. In addition, different teachers assign varying amounts of homework and some students may be taking more than one math class. A very good thinker may have a teacher who assigns less homework, thus skewing the results. Another good thinker may be already good at math, making them more inclined to take more math classes, causing their time spent doing math problems to increase. That person may have scored well even if he or she had not taken more math classes. And lastly, cognitive ability may be a confounding variable. Because there are so many problems with this variable, the T Stat is expected to be small and the P-value to be insignificant. Despite these problems, it can still be reasonable to theorize a possible positive correlation.

Older people should be more rational. As people grow older, they tend to learn more about the world and practice more with decision making. A greater understanding of decision making may improve their rational thinking. In addition, experience and time improves a person's ability to think about problems. Thus older people are expected to do better on all the subsets. In addition, the brain strengthens the connections between synapses and kills off synapses that aren't used. As a result, old and brief memories that may affect biases are lost, decreasing the potential for bias. However, it is important to acknowledge that an older person may also

become more biased because of experience. Their experiences may lead to skewed thought-processes, thus making them less rational. Despite this counter argument, there is more empirical evidence that suggest older people are more rational, as older people performed better on the short-form and full-form CART exam (Stanovich, K. E., 2016). However, because the counter argument should be considered, the T Stat and P-value are expected to be insignificant.

Stuyvesant students are expected to be more rational. Stuyvesant students are required to score exceptionally well on the SHSAT, as the acceptance rate is about three percent. Because the SHSAT contains logical reasoning problems, a higher score on the SHSAT may correlate with a greater thinking skills, which improves one's rationality. In addition, Stuyvesant is heavily oriented towards math and science, which are subjects that develop a person's thinking skills. Furthermore, Stuyvesant students are more inclined to do well and try harder in school, leading to more time spent thinking. And lastly, Stuyvesant students are required to take a logic section in the beginning of geometry, which covers syllogistic reasoning and other important rational thinking skills. Thus Stuyvesant students are expected to do better on all the subsets except rational temporal discounting, as that has less to do with a better cognitive ability. As before, it is important to consider that a higher cognitive ability can be a confounded variable.

A person who sleeps more during school days is expected to be more rational. When a person sleeps, his or her brain rests neurons and forms new pathways that are crucial to memory and thinking capabilities, thus improving rationality (Pietrangelo, A., 2014). In addition, people who are sleep deprived have a harder time concentrating, decreasing their decision-making abilities. Therefore, a person who has more sleep is expected to perform better on all subsets than a person who is sleep-deprived. However, it is important to note that different people have different levels of tolerance: while some people can be fully functional with only three hours of sleep, others may still feel sluggish with nine or more hours of sleep. For this reason, despite the fact that there is an expected positive correlation, the expected T Stat and P-value are expected to be insignificant.

Lastly, a person with a higher socioeconomic status is expected to perform better on the CART exam. A person with higher status has higher access to educational opportunities and resources that may help develop their rationality. Thus that person is expected to perform better in all the subsets of the test. Yet there are problems with socioeconomic status as a variable. About 33 percent of Stuyvesant students have free lunch and another 12 percent has reduced-price lunch, indicating a huge population of students who are in a lower socioeconomic status. However, Stuyvesant students have huge access to available resources, thus contradicting the fact that a lower socioeconomic status student has less resources. In addition, more access does not necessitate efficient utilization. Many students who have a higher socioeconomic status may care less about their education while poorer students may have a greater incentive to achieve in school. For this reason, despite the fact that students in a higher socioeconomic status are expected to be moore rational, the expected T Stat and P-value are expected to be insignificant.

There are other variables that one might consider to be appropriate but are not included in the test. For example, a person with a higher GPA may indicate a higher rationality. However, it is not known whether a high GPA causes rationality, rationality causes a higher GPA, or an external variable

causes both. In addition, genetics may play a part in rationality. However, it is near impossible to know how rational predecessors were. For example, a student may have a farmer as a parent who may fail at a rationality test, not because he was born to be irrational, but because he was never given the opportunities to think critically. Therefore, no conclusions can be drawn about the rationality of the student regarding genetics.

Table 1

Name of Variable	Expected Slope Sign	Expected T-value	Expected P-value
Gender (1:M 0:F)	Positive	Large	<.05
Time Spent Doing Math Homework	Positive	Small	>.05
Age	Positive	Small	>.05
Stuyvesant Student (1:Y 0:N)	Positive	Large	<.05
Average Amount of Sleep on School Days	Positive	Small	>.05
Free Lunch (1:Y 0:N)	Negative	Small	>.05

It should be also be acknowledged that measuring rationality is a very abstract and complex task. As a result, measuring rationality in a different way can bring in different explanatory variables. Therefore, the variables described above are possible explanations of the results of this particular way of measuring rationality. Maybe in the near future, a more accurate and concrete measurement may be developed, and the variables above does not affect or explain that measurement of rationality.

Data Collection

Because this paper is based on Stanovich's book, the questions used to assess rationality comes verbatim from his Comprehensive Assessment of Rational Thinking (CART) exam. However, because his test contains many questions that may require up to three hour of testing, it would be impractical to use all the questions for the project. Instead, five questions were picked from the CART, each question representing a subset of the exam.[7] Because there were many subsets in the exam, five were chosen based on its weight in the CART exam. The questions were chosen based mostly on its readability and its concurrence with available resources and capabilities of high school students. For example, questions that required over a paragraph of reading, required excessive computations, or required certain background knowledge that high schools are not familiar with would not be chosen. The next factor in deciding the question is how well the question matches the subset that it intends to test.

This paper would numericize rationality using point values derived from each subset, which have equal weights. The subsets chosen are probabilistic and statistical

[7] Though it may be unreasonable to calculate an entire subset based on one problem, it is the only practical way given the resource and time constraints.

reasoning, scientific reasoning, reflection versus intuition, belief bias in syllogistic reasoning, and rational temporal discounting. The probabilistic and statistical reasoning subtest measures one's ability to understand and apply probability and statistics rationally to make appropriate decisions. The scientific reasoning subtest measures one's ability to apply knowledge to detect conflict, override initial beliefs, and think hypothetically. The reflection versus intuition subset also tests one's ability to override conflict, but to a greater degree. The belief bias in syllogistic reasoning subtest tests one's ability to override established world knowledge using logical conclusions. The rational temporal discounting subset measures one's instrumental rationality and their ability to maximize expected values regardless of time biases.

To represent probabilistic and statistical reasoning, the gambler's fallacy is tested in question seven of the rationality test. The gambler's fallacy is when players act as if previous events affect the next when in fact the events are independent. For example, a player who won the slot machine a few times in a row may believe that he his bound to lose the next round, even when the probability of winning the next round remains the same.

To represent scientific reasoning, people were tested on hypothesis testing and falsifiability. People have a difficult time thinking about evidence and tests that could falsify their initial hypothesis, while tending to seek to confirm their theories rather than falsifying them. For example, consider question eight: "There are four cards: 9, J, U, 2. You need to determine which card to turn over in order to determine whether the following rule holds for the deck (assuming these four cards represent the rest of the deck): If a vowel is printed on one side of the card, then an even number is printed on the other side. Which two cards do you need to turn over in order to test this rule?" Most people tend to pick U and 2 (because the two choices confirm their theory) when in fact the answer is U and 9 (flipping over 9 is essential to prove that the rule is not false).[8]

To test reflection versus intuition, a problem was chosen to detect a failure of the person's ability to override conflict detection. Consider question nine: "If it takes one minute to make each cut, how long will it take to cut a 25-foot wooden plank into 25 equal pieces?" The initial response would be 25 minutes (an irrational person's thought immediate correlates 25 cuts with 25 pieces) when in fact it is 24 minutes.[9]

To test belief bias in syllogistic reasoning, a belief bias was incorporated into a syllogistic reasoning problem by using conclusions that contradicted world knowledge. Consider this question ten: "Premise 1: All flowers are carbitops. Premise 2: All tulips are carbitops. Are all tulips flowers?" People may have a tendency to say yes because tulips are flowers in real life, when in fact the conclusion isn't necessarily true.[10]

To test rational temporal discounting, a question was made to test the participant's desire to forgo immediate gains for a greater future gain. Consider problem eleven: "If given a choice, would you prefer $340 now or $400 in 4 months?" A person who chooses $400 in four months maximizes his expected value and concurs with instrumental rationality.

The survey was done in-person, as subjects were given a sheet of paper to

[8] For more information, read Christopher Badcock's *Making Sense of Wason* article, which explains the a variation of the four card problem.

[9] To understand why, note that it takes one cut to split a wood into two pieces.

[10] To understand the solution, review the laws of syllogism.

complete. The subject would go through the problems without consulting any help, unless they were unsure of their interpretation of the questions. The sheet is then collected with respect to the person's privacy.

The survey can not be considered as a pure random sample. The surveys were handed out only to known high school students because decreasing the group range increases the accuracy of the data, especially for low numbers of observations. Because of the resource and time constraints, the sample was handed out to those who were easiest to reach. For example, surveys were handed out in the cafeteria, some classrooms, the park, and locations of various organizations. However, this should not entire skew the result because there is no intuitive relationship between the location of the participants. There is no reason to believe that people in the cafeteria is more or less rational than people who go out to lunch. Nor were surveys handed in classrooms or organizations where people tend to be more rational (i.e. math and science classrooms). In addition, surveys were not given to those who were known to have a lower or higher cognitive ability because that would skew the results. Furthermore, both the purpose and the topic of the survey was not announced to those who were willing to take the survey. This decreases non-response bias because a very unintelligible student may refuse to take a rationality survey, but will take the survey if he or she does not know what was on the survey. In addition, all the students who accepted the survey did not abandon the survey but rather completed it to the end, which further reduces response bias.

There are many reasons why this survey would fare worse if it was handed online. Because this survey required thinking and thus more time, people see that taking the survey would contain opportunity costs that aren't worth their gain from taking the survey, which is essentially nothing. On the other hand, people tend to do the survey when asked personally. In addition, answers to the problems can be searched up because some of the problems are easy to search up. Furthermore, if the participant does not understand the problem for whatever reason, he or she may guess on the problem or abandon the survey. If done by hand, the participant is entitled to seek clarification, decreasing the response bias. Lastly, there is a lower chance that data is contaminated if done in person because it would be quite obvious if the person is not taking the survey seriously.

There are many problems with the test itself. First, the test is very incomprehensive compared the CART exam. Compared to the full CART exam, which has twenty subsets and requires up to three hours to take, the survey used for this project has only five questions. When Stanovich compared the correlations among the tests that he administered, he found that a less comprehensive test has a lower correlation with the more comprehensive tests (Stanovich K. E., 2016). Then a less comprehensive exam decreases the accuracy of the exam. In addition, this test will have the same problem as the CART exam because those who have a higher cognitive ability over represents their associated group. As a result, data may be skewed if cognitive ability is not controlled. There is also the possibility that people have seen the problems before and had done it. Obviously, those who had seen the problem before have an upper hand, skewing the results of the test. In addition, there were some typos in some of the questionnaires. Though it is important to acknowledge, it should have a minuscule impact on the test because the typos were uninfluential and clarification was accessible.

Running The Test

Running a multivariable regression with the available data provides insight to the test. Based on Table 2, the Multiple R is approximately .83, indicating that the correlation between the actual rationality score and the expected rationality score is approximately .83. The adjusted R Square indicates 62 percent of variation is explained by only the independent variables that actually affect the dependent variable. The standard error is the sample estimate of the standard deviation of the error of the regression line. A standard error of .73 is relatively high in this test. The number of observations indicates the number of subjects who took the test, which is 35.

Table 2

Regression Statistics	
Multiple R	0.830051
R Square	0.688984
Adjusted R Square	0.622338
Standard Error	0.733481
Observations	35

Appendix F describes the analysis of variance (ANOVA). The total sum of squares (SS) expresses the total variation that can be attributed to various factors. The F value and Significance F statistics test the overall significance of the regression model. Because the Significance F is 4.79E-06, which is much lower than .05, the test indicates that at least some of the regression parameters are nonzero and that the regression equation does have some validity in fitting the data.

Based on Table 3, the equation of the regression is:

Rationality score *= -4.58811 - 0.57563** ***Gender*** *- 0.00759** ***Time Spent on Math Problems*** *+ 0.001087** ***Days Living as of 5/17/17*** *+ 1.95802** ***Stuyvesant*** *+ 0.000118** ***Avg Amount of Sleep on Days With School*** *- 0.23425** ***Free Lunch.***

Table 3

	Coefficients	*t Stat*	*P-value*
Intercept	-4.58811	-1.55357	.065759
Gender (1:m, 0:f)	-0.57563	-1.67401	.052633
Time Spent on Math Problems (minutes)	-0.00759	-1.15623	.1286795
Days Living as of 5/17/17	0.001087	2.305639	.0143755
Stuyvesant High School? (1:Y, 0:N)	1.95802	5.060735	1.175E-5
Avg Amount of Sleep on Days With School (minutes)	0.000118	0.08871	.464972
Free Lunch (1:Y, 0:N)	-0.23425	-0.86076	.1983415

Note that gender, Stuyvesant, and free lunch are dummy variables. For gender, 1 is assigned to males while 0 is assigned to females. For Stuyvesant, 1 represents that the subject is from Stuyvesant while 0 represents that the subject is not from Stuyvesant. Lastly, 1 indicates that the person has free

lunch, while 0 indicates that the person does not have free lunch.

The T Stat and one-tailed P-value indicates the statistical significance of the hypothesis testing. A P-value less than .05 indicates that the test result is significant, whereas a P-value greater than .05 indicates that the test result is not significant. Significance in regression hypothesis testing indicates that there is a positive linear correlation between the x-variable and the y-variable. The only x-variables that are significant based on the test are Days Living as of 5/17/17 and Stuyvesant Student or Not. For more detail, Appendix E contains the lower and upper 95% indicates the confidence interval of the coefficients. For example, for gender there is a 95 percent confidence that the true coefficient lies between -1.28 and 0.128743.

Analyzing the signs of the coefficients, one may notice that the gender and time spent on math problems have negative correlations despite the theory that they should be positive. However, because neither of the P-values of these variables are not significant, there is no reason to stress over the signs of the variables.

Appendix G describes how strongly the x-variables are correlated with one another. A correlation of .7 of higher (unless one variable is correlated to itself) is significant, and the variables must be revised to deal with the high correlation. In this test, none of the variables have correlations greater than .7. Appendix H is also a correlation matrix but it indicates the correlations between the questions. Again, none of the questions have a correlation greater than .7. In fact, the low correlations indicates that the questions are close to independent of one another, making it a better representation of the subset that it is intended to represent.

Figures 1 through 6 plots the data points of the various x variables and the rationality score. The line fit plot indicates the actual scores that corresponding to the x-variables. Appendix D shows the residual plots, which indicates the difference between the actual rationality score and the expected rationality score. All of the plots were homoskedasticity residual plots, which is one in which the residuals do not display a pattern, indicating that there is no bias and that a linear regression fits the model.

Figure 1

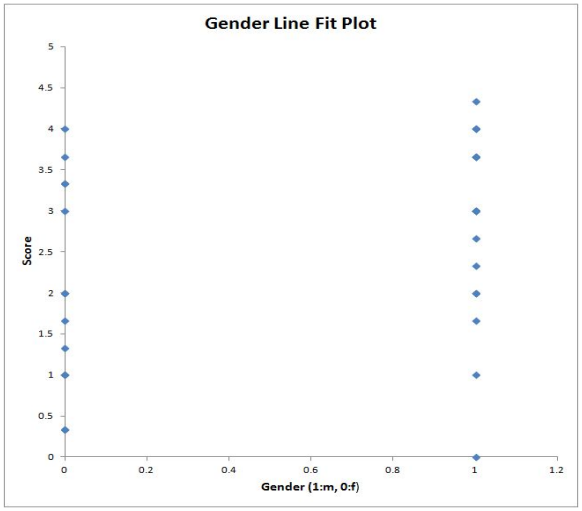


Figure 2

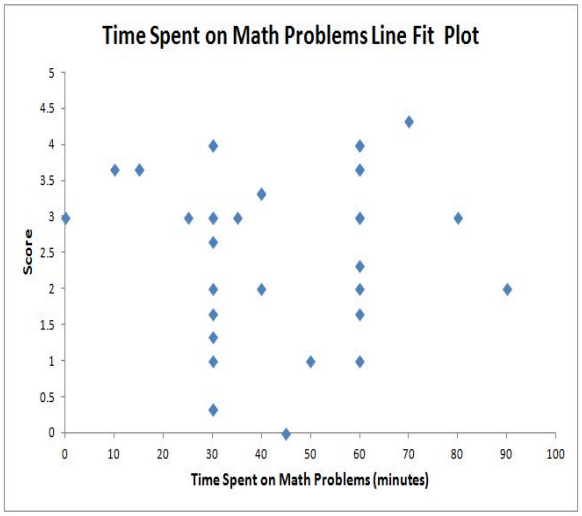


Figure 3

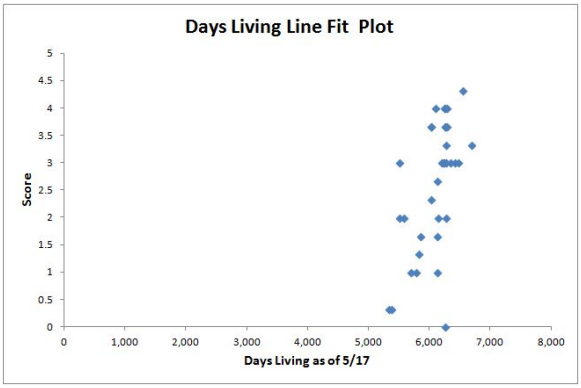


Figure 4

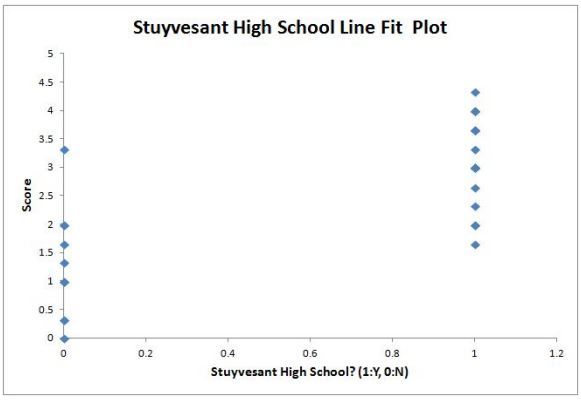


Figure 5

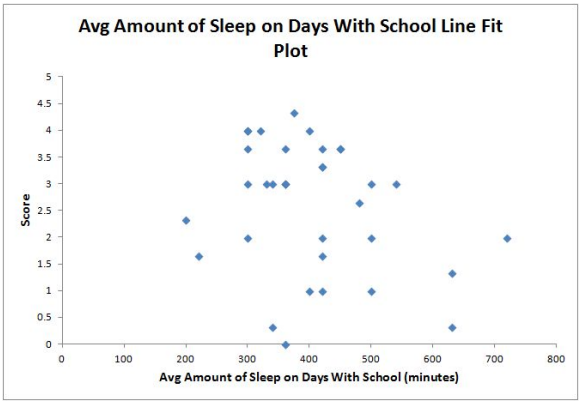


Figure 6

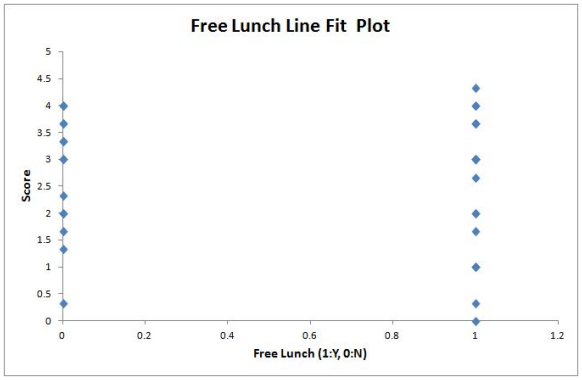
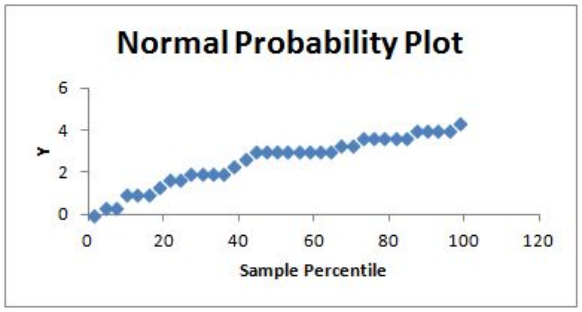


Figure 7 is a normal probability plot, which measures how normal the data is (or how close the data is to a normal distribution). Because a more linear line indicates a more normal distribution, this graph shows that the results of the data is approximately normal.

Figure 7



Conclusions and Recommended Policies

The results of the test indicates that being in Stuyvesant and amount of days living are the strongest indicators of rationality, as the p-value for both of these regression coefficients are less than .05. As a result, it can be concluded that being in Stuyvesant or living longer makes the person more rational. Because the other variables do not have a significant p-value, there is not enough statistical evidence to suggest that those variables affect rationality. However,

note that just because this test does not show that the variables do not reflect rationality, it does not mean that it does not affect rationality. A more comprehensive and extensive test may produce a statistically significant p-value for the given variables.

Based on the result, Stuyvesant students are more rational than other students, as they score almost two points higher than students from other schools. Though this is a good indication of Stuyvesant's education, it is important to note that of the 24 Stuyvesant students tested, no one scored a perfect score on the five question rationality test, which should be quite feasible. This indicates that there is still some work required by the school to improve Stuyvesant students' rationality to better prepare them for decision making in the real world. For students in other schools, even more work is required to prepare the students to think rationally.

Improving rationality in students is not a difficult task. In fact, most schools, by providing critical thinking skills and decision-making opportunities, are already improving the students' rationality. However, more should be done, as rationality is critical to the real world. The best action to take is to make students more aware of the importance of rationality. In addition, schools may implement more analytical courses and courses that force students to make rational decisions.

Applications

Rationality testing has a wide array of important applications because a notable amount of research indicates that there are links between rational thinking tendencies and real-life decision makings. For example, CART exams may be administered to appropriate fields that require high levels of rationality and low levels of bias. The rationality test may be a supplementary entrance exam to law or managerial positions, as both require a high degree of rational decision making. Those who may not be rational may take practice exams to enter their fields, which would make the prospective employee more conscious and aware about their rationality and more rational. In addition, the CART exam may be administered to students to access their rationality. Because the purpose of school is to prepare students for the real world, which involves decision making, the CART exam may be used as a benchmark for the school's progress in preparing students.

Studying rationality in humans may also help progress artificial intelligence. Analyzing humans may produce patterns and threads that can be utilized by robots (Graft, K., 2015). With a sense of rationality in robots, robots may be better suited for human interactions and decision makings. For example, rational robots may one day predict the stock market with increasing accuracy, as analyzing the stock market requires the understanding of human choices and their rationality. As noted, stock traders may not entirely make decisions based on rationality and are very permeable to biases such as emotions and past experiences (Williams, R., 2013). In addition, robots may eventually learn to defeat humans in any finite game. This has huge applications to video games, as AI can drastically improve their decision makings using results from the CART test.

Making people more aware of their rationality by administering a test may improve safety. For example, understanding temporal discounting rationality will improve the actions for those who are unable to resist temptations or lack prudence. The target group includes unmotivated students, smokers, imprudent financial decision makers, excessive drinkers, gamblers, and

physically unfit adults. In addition, those who are in active life-threatening roles may benefit from taking rationality exams to improve their rationality. For example, military leaders and generals require a strong sense of rationality, as their choices will have tremendous consequences (Jyung, C.Y., 2013).

Lastly, empirical data may be used in conjunction with theoretical analysis. Data collected from fromt his paper and from the CART exam may be used to supplement theories regarding the brain, as the brain is one of the most unresolved aspects of the human body. For example, using this data, one may draw a stronger connection between gray matter and rationality, even though there is no concrete proof that gray matter impacts rationality.

References

2017 New York City High School Directory. (n.d.). Retrieved May 22, 2017, from http://schools.nyc.gov/NR/rdonlyres/00F2DEB3-4F50-4747-A14E-E53295E078DC/0/2017NYCHSDirectoryCitywideENGLISH.pdf+

Criteria Cognitive Aptitude Test (CCAT). (n.d.). Retrieved May 22, 2017, from https://www.criteriacorp.com/solution/ccat.php

Definition of Rationality. (n.d.). Retrieved May 22, 2017, from https://en.oxforddictionaries.com/definition/rationality

Dietrich, C. (2010, February 01). Decision Making: Factors that Influence Decision Making, Heuristics Used, and Decision Outcomes. Retrieved May 22, 2017, from http://www.inquiriesjournal.com/articles/180/decision-making-factors-that-influence-decision-making-heuristics-used-and-decision-outcomes

Graft, K. (2015, September 15). When artificial intelligence in video games becomes...artificially intelligent. Retrieved May 22, 2017, from http://www.gamasutra.com/view/news/253974/When_artificial_intelligence_in_video_games_becomesartificially_intelligent.php

Halpern, D. F. (n.d.). Halpern Critical Thinking Assessment - Diane Halpern. Retrieved May 22, 2017, from https://sites.google.com/site/dianehalperncmc//home/research/halpern-critical-thinking-assessment

How Your Brain Changes With Age. (2017, March 29). Retrieved May 22, 2017, from https://www.canyonranch.com/blog/health/how-your-brain-changes-with-age/

Interpretation of an IQ score. (n.d.). Retrieved May 22, 2017, from https://www.123test.com/interpretation-of-an-iq-score/

Jantz, G. L. (2014, February 27). Brain Differences Between Genders. Retrieved May 22, 2017, from https://www.psychologytoday.com/blog/hope-relationships/201402/brain-differences-between-genders

Jyung, C.Y., Department of Vocational Education and Workforce Development, Seoul, Republic of korea, Joo, H.S., Seoul National University, Seoul, Republic of korea, & Kim, T.W., Seoul National University, Seoul, Republic of korea. (n.d.). Strategies for the Improvement of Military Service Value of Soldiers in the Army. Retrieved May 22, 2017, from http://agris.fao.org/agris-search/search.do?recordID=KR2015002254

A New Twist on a Classic Puzzle. (n.d.). Retrieved May 22, 2017, from https://www.psychologicalscience.org/publications/observer/obsonline/a-new-twist-on-a-classic-puzzle.html#.WSOepOvyvcs

Parker, A. M., & Fischhoff, B. (n.d.). Youth - Decision Making Competence (YDMC). Retrieved May 22, 2017, from http://www.sjdm.org/dmidi/Youth_-_Decision_Making_Competence.html

Pietrangelo, A. (2014, August 19). Effects of Sleep Deprivation on the Body. Retrieved May 22, 2017, from http://www.healthline.com/health/sleep-deprivation/effects-on-body

Stanovich, K. E., & Toplak, M. E. (2016). *The rationality quotient toward a test of rational thinking*. Cambridge, MA: MIT Press.

Stanovich, K. E. (2016). The Comprehensive Assessment of Rational Thinking. *Educational Psychologist, 51*(1), 23-34. doi:10.1080/00461520.2015.1125787

Stuyvesant High School. (n.d.). Retrieved May 22, 2017, from https://www.usnews.com/education/best-high-schools/new-york/districts/new-york-city-public-schools/stuyvesant-high-school-13092/student-body

Williams, R. (2013, September 22). Emotion, Not Rational Logic, Determines the Stock Market. Retrieved May 22, 2017, from https://www.psychologytoday.com/blog/wired-success/201309/emotion-not-rational-logic-determines-the-stock-market

Appendices

Appendix A: Sample Survey

1) What is your gender?

Male/Female

2) How much time do you spend doing math problems?

_____ minutes per day

3) When is your birthday? (month/day/year)

4) Are you in Stuyvesant High School?

Yes/No

5) What is your average amount of sleep on days when you have school?

_____ minutes

6) Do you have free lunch?

Yes/No

7) When playing slot machines, people win something about 1 in every 10 times. Nancy, however, won on her first three plays. What are her chances of winning the next time she plays?

Greater than 0.10 / 0.10 / less than 0.10

8) I deal you four cards from the deck. Here is what you see laid out before you on the four cards:

9 J U 2

I ask you which cards you will need to turn over in order to determine whether the following rule holds for the deck (assuming these four cards represent the rest of the deck):

If a vowel is printed on one side of the card, then an even number is printed on the other side.

Which two cards do you turn over in order to test this rule?

9) If it takes one minute to make each cut, how long will it take to cut a 25-foot wooden plank into 25 equal pieces?

_____minutes

10) Premise 1: All flowers are carbitops
Premise 2: All tulips are carbitops
Conclusion: All tulips are flowers.

Is the conclusion necessarily true?

11) If you had a choice, would you prefer $340 now or $400 in 4 months?

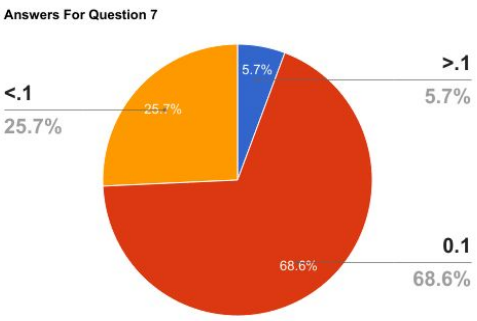
Very strongly prefer 340/ strongly prefer 340/prefer 340/ prefer 400/strongly prefer 400/ very strongly prefer 400

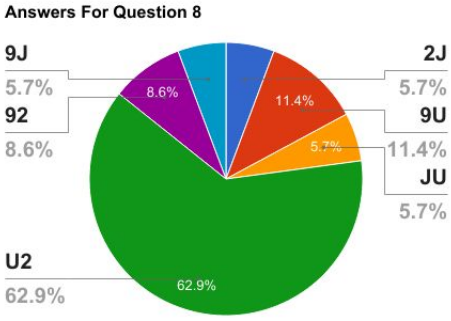
Answers:
7) .1 8) 9 and U 9)24 10) No 11) vS400

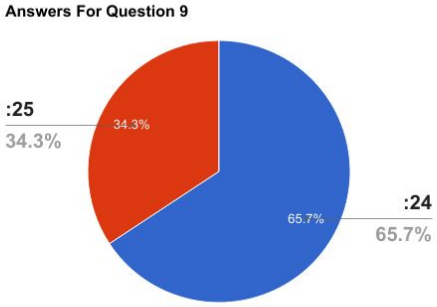
Appendix B: Data Points

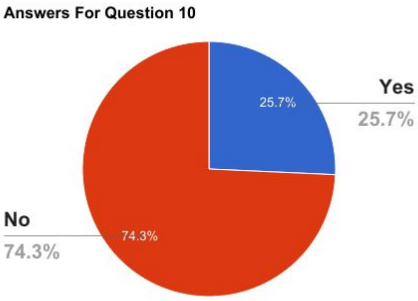
Gender (1:m, 0:f)	Time Spent on Math Problems (minutes)	Days Living as of 5/17	Stuyvesant High School? (1:Y, 0:N)	Avg Amount of Sleep on Days With School (minutes)	Free Lunch (1:Y, 0:N)	Q7	Q8	Q9	Q10	Q11	Total Score (Q7-Q11)
1	60	6,261	1	420	1	0	0	0	1	1	2
0	30	5,322	0	630	0	0	0	0	0	0.33	0.33
1	60	5,504	1	360	1	1	0	1	1	0	3
1	60	6,238	1	300	0	1	1	1	1	0	4
1	0	6,411	1	540	1	1	0	1	1	0	3
0	30	5,582	0	720	0	1	0	1	0	0	2
1	35	6,246	1	340	0	1	0	1	1	0	3
1	30	6,203	1	300	1	1	0	0	1	1	3
0	40	6,267	1	420	0	1	0	1	1	0.33	3.33
1	30	6,117	1	480	1	1	0	0	1	0.66	2.66
0	40	5,508	0	500	0	1	0	0	1	0	2
1	30	6,247	1	400	1	1	0	1	1	1	4
1	60	6,021	1	200	0	0	0	1	1	0.33	2.33
0	30	5,816	0	630	0	0	0	1	0	0.33	1.33
1	10	6,256	1	360	1	0	1	1	1	0.66	3.66
1	60	6,029	1	420	0	1	0	1	1	0.66	3.66
0	40	6,680	0	420	0	1	0	1	1	0.33	3.33
0	60	6,255	1	300	1	1	0	1	1	0.66	3.66
0	30	6,232	1	330	1	0	0	1	1	1	3
0	30	5,851	0	220	1	1	0	0	0	0.66	1.66
1	25	6,341	1	360	1	1	0	1	1	0	3
1	60	6,469	1	360	0	1	0	0	1	1	3
0	60	6,278	1	300	0	1	0	1	1	1	4
1	15	6,018	1	450	1	1	0	1	1	0.66	3.66
1	70	6,544	1	375	1	1	1	1	1	0.33	4.33
0	50	5,779	0	500	1	0	0	0	1	0	1
1	15	6,276	1	450	0	1	0	1	1	0.66	3.66
0	30	5,371	0	340	1	0	0	0	0	0.33	0.33
1	90	6,131	1	300	1	1	0	0	1	0	2
1	45	6,252	0	360	1	0	0	0	0	0	0
1	60	6,123	0	420	1	0	0	1	0	0	1
1	30	6,099	1	320	1	1	0	1	1	1	4
1	60	6,124	1	420	0	0	0	0	1	0.66	1.66
1	80	6,271	1	500	1	1	0	1	1	0	3
0	30	5,687	0	400	1	0	0	1	0	0	1

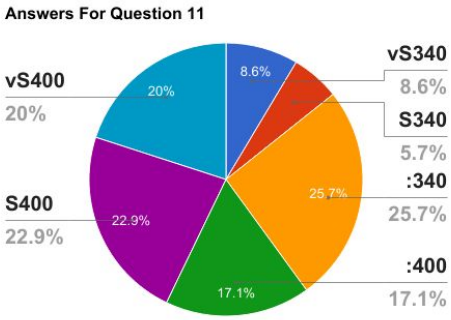
Appendix C: Answers to Rationality Quiz



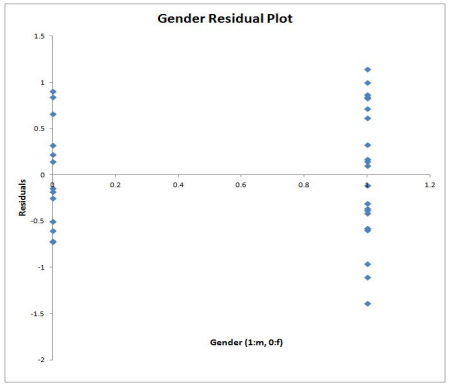


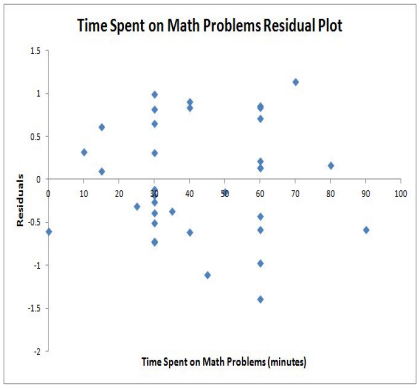


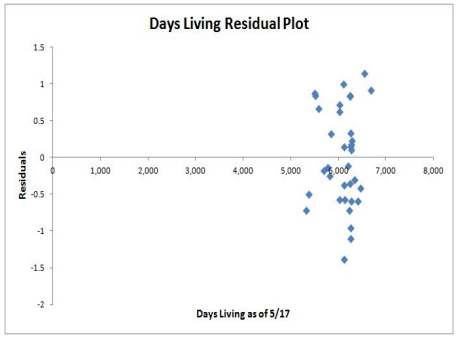


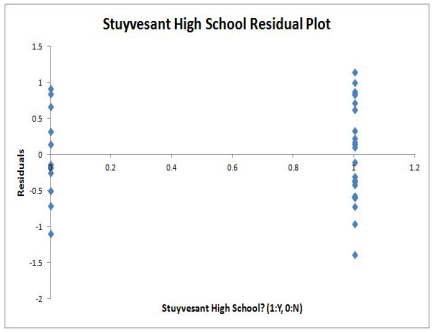


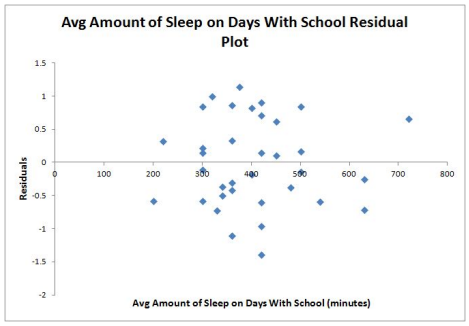
Appendix D: Residual Plots

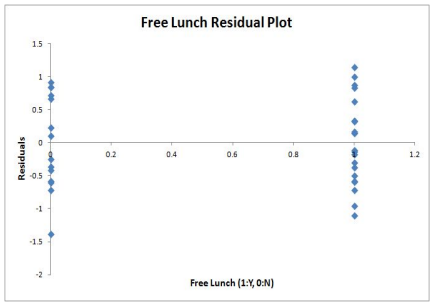












Appendix E: Regression Coefficient Results

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-4.58811	2.953263	-1.55357	.065759	-10.6376	1.461375	-10.6376	1.461375
Gender (1:m, 0:f)	-0.57563	0.343863	-1.67401	.052633	-1.28	0.128743	-1.28	0.128743
Time Spent on Math Problems (minutes)	-0.00759	0.006565	-1.15623	.1286795	-0.02104	0.005857	-0.02104	0.005857
Days Living as of 5/17/17	0.001087	0.000472	2.305639	.0143755	0.000121	0.002053	0.000121	0.002053
Stuyvesant High School? (1:Y, 0:N)	1.95802	0.386904	5.060735	1.175E-5	1.165483	2.750558	1.165483	2.750558
Avg Amount of Sleep on Days With School (minutes)	0.000118	0.001328	0.08871	.464972	-0.0026	0.002839	-0.0026	0.002839
Free Lunch (1:Y, 0:N)	-0.23425	0.272145	-0.86076	.1983415	-0.79171	0.323213	-0.79171	0.323213

Appendix F: ANOVA Results

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	33.37047	5.561744	10.33793	4.79E-06
Residual	28	15.06383	0.537994		
Total	34	48.4343			

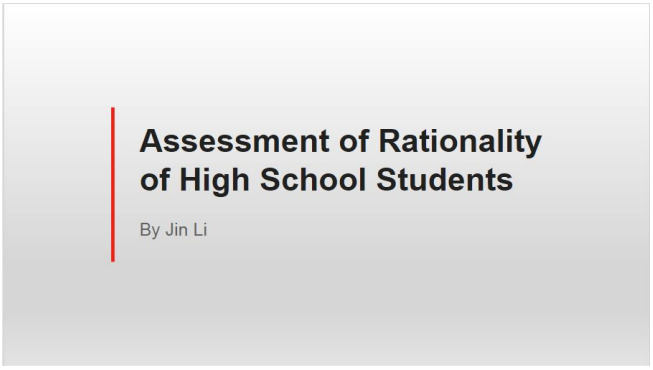
Appendix G: Correlation Matrix For Variables

	Gender	Time Spent on Math Problems	Days Living	Stuyvesant High School?	Avg Amount of Sleep on Days With School	Free Lunch
Gender	1					
Time Spent on Math Problems	0.152373	1				
Days Living	0.441336	0.127463	1			
Stuyvesant High School?	0.625956	0.159034	0.55378	1		
Avg Amount of Sleep on Days With School	-0.24607	-0.27354	-0.33194	-0.38991	1	
Free Lunch	0.217262	-0.07577	0.048227	0.075378	-0.24049	1

Appendix H: Correlation Matrix For Questions on Rationality Quiz

	Q7	*Q8*	*Q9*	*Q10*	*Q11*
Q7	1				
Q8	0.006143	1			
Q9	0.23913	0.221163	1		
Q10	0.4669	0.166667	0.180207	1	
Q11	0.056766	-0.06923	-0.09992	0.29842	1

Appendix I: Powerpoint Slides



Definition of Rationality

"The quality of being based on or in accordance with reason or logic." - Oxford Dictionary.

"Distance of thought or behavior from the optimum defined by a normative model." - Keith E. Stanovich.

Biases causes one to stray away from the normative model.

Instrumental rationality: the optimization of the individual's goal fulfillment.

Epistemic rationality: how well beliefs maps onto the actual structure of the world.

Rationality Exams

- Comprehensive Assessment of Rational Thinking (CART)
  - Created by Keith Stanovich
  - Very long and comprehensive test
- Adult Decision Making Competence (A-DMC)
  - Involves decision making, but contains questions loosely related to rationality
- Halpern Critical thinking Assessment (HCTA)
  - Contains rationality, but focuses on critical thinking, which contains other aspects other rthan rationality
- Intelligence Quotient (IQ) Exam
  - Measures intelligence, not rationality.

My Theory

- Gender
  - Males > Females.
  - Men uses seven times the amount of gray matter than women. Gray matter if responsible for information and action processing.
- Time Spent on Math Problems
  - More time spent > less time spent.
  - More practice with cognitive ability, which is related to rationality.
- Stuyvesant or Not
  - Stuyvesant > other schools.
  - Required an entrance exam that involves cognitive ability.
  - More intellectually stimulating environment.
  - First year of geometry, everyone is required to take a logic section.

- Age
  - Older > younger.
  - More time and experience to make decisions.
- Amount of Sleep
  - More sleep > less sleep.
  - When asleep, the brain forms more neurons that are crucial to thinking capabilities.
  - Difficult time concentrating when sleep deprived, which affects one's decision making.
- Socioeconomic status
  - More money > less money
  - More access to educational materials and opportunities

Survey Questions

- 5 Question Test on:
  - Probabilistic and statistical reasoning
  - Scientific reasoning
  - Reflection versus intuition
  - Belief bias in syllogistic reasoning
  - Rational temporal discounting.
- Data Collection:
  - Surveys by hand
  - Not a pure random sample

Results of Running the Test

- Significant F value
- Adjusted R-Squared of 62%
- Only Stuyvesant or not and age were significant
- All homoscedastic residual plots
- None of the variables have correlation of .7 or higher.
- None of the questions have correlation of .7 or higher.

Policies and Applications

- High school students needs to be more rational
  - No one scored perfect on a five question rationality quiz
  - Raise awareness of the value of rationality
- Applications:
  - Artificial intelligence
  - Police and military
  - Health