1-1-2016

Meta-Analysis of the Predictive Validity of Scholastic Aptitude Test (SAT) and American College Testing (ACT) Scores for College GPA

Muhammet Curabay
University of Denver

META-ANALYSIS OF THE PREDICTIVE VALIDITY OF SCHOLASTIC APTITUDE TEST (SAT) AND AMERICAN COLLEGE TESTING (ACT) SCORES FOR COLLEGE GPA

A Thesis

Presented to

the Faculty of the Morgridge College of Education

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

Muhammet Curabay

November 2016

Advisor: Dr. Antonio Olmos

Author: Muhammet Curabay
Title: META-ANALYSIS OF THE PREDICTIVE VALIDITY OF SCHOLASTIC APTITUDE TEST (SAT) AND AMERICAN COLLEGE TESTING (ACT) SCORES FOR COLLEGE GPA
Advisor: Dr. Antonio Olmos
Degree Date: November 2016

Abstract

The college admission systems of the United States require the Scholastic Aptitude Test (SAT) and American College Testing (ACT) examinations. Although, some resources suggest that SAT and ACT scores give some meaningful information about academic success, others disagree. The objective of this study was to determine whether there is significant predictive validity of SAT and ACT exams for college success. This study examined the effectiveness of SAT and ACT scores for predicting college students' first year GPA scores with a meta-analytic approach. Most of the studies were retrieved from Academic Search Complete and ERIC databases, published between 1990 and 2016. In total, 60 effect sizes were obtained from 48 studies. The average correlation between test score and college GPA was 0.36 (95% confidence interval: .32, .39) using a random effects model. There was a significant positive relationship between exam score and college success. Moderators examined were publication status and exam type with no effect found for publication status. A significant effect of exam type was found, with a slightly higher average correlation for SAT compared to ACT score and college GPA. No publication bias was found in the study.

Acknowledgements

First of all, I am grateful to The Almighty God for completing this thesis. I also want to express my sincere thanks to Ministry of National Educational in Turkey for providing me a chance to come to the USA for good education.

Additionally, I would like to thank several people for their assistance, guidance and support. The completion of this thesis could not have been possible without my thesis advisor Antonio Olmos, and my advisor Kathy Green. Thanks for your assistance, encouragement, and support helped me to understand my subject to complete my thesis.

I also thank Dr. Nicholas Cutforth, Duan Zhang, Turker Toker, Cahit Polat, Priyalatha Govindasamy, Dareen Alzahrani in Research Methods and Statistic Department. To my friends Ramazan Karatas, Fatih Gok, Fatih Kunkul, Erdal Asker,and Yunus Ozturk.

In addition, I appreciate Serkan Perkmen, the first person, who encourage me to come to the USA. I owe thanks to my former University professors Tuncay Saritas, Emin Korkusuz, Aysen Karamete, Gulcan Ozturk, and Aydin Okcu.

Finally, I want to thank the most important people in my life. My father Mehmet Curabay, my mother, Nefika Curabay and my sister Rahime Curabay. I also thank people who vouched for me before I started my education, my uncle Talat Solbay, and his family.

Table of Contents

List of Figures

Chapter One: Introduction

Standardized tests have existed since the late 19th century. As a result that, these tests have become critical for high school graduations and college admissions (Webb, 2013). Standardized tests are considered to provide realistic and objective quantitative results of students' academic success. Before standardized exams, schools and teachers were creating and assessing exams – those were not the same all around the US - for their students ("Standardized Tests - ProCon.org," 2016). Nowadays, The Scholastic Aptitude Test (SAT) and American College Testing (ACT) exam are the two most important standardized tests taken by college candidate students in the US. To assess predictive validity, researchers receive scores on a current measure and compare them with future scores on the desired outcome. A high correlation shows that the selection method works well. A low correlation indicates poor predictive validity, and the selection method is not beneficial. The purpose of this study was to examine the validity of the SAT and ACT in predicting college grade point average (GPA).

SAT and ACT

Since several American Universities decided to use SAT scores of the candidate students to make decisions -accept or reject- about those students' applications, the SAT exam has become critical. According to College Board statistics, not only around 1.7 million students took the SAT in 2015 ("2015 College-Bound Seniors Total Group Profile Report," n.d.), but also the ACT was taken by more than 1.8 million secondary school students per year ("About ACT," n.d.). The statistics show that the number of test takers of the ACT has been increasing each year gradually since its inception, and also, ACT takers' numbers were more than the SAT exam takers' number in 2011 for the first time. In 2011, the SAT was taken by 1,664,479, whereas the ACT was taken by 1,666,017 students (Pope, 2012).

The literature suggests the SAT is a good predictor of academic success. As the research of Ditchkoff, Laband, and Hanby (2003) states, high-school grade-point average (HSGPA) and Academic College Testing (ACT) or Scholastic Aptitude Test (SAT) scores are invaluable sources of making predictions of academic success. It is clear that there is a positive relationship between the ACT Composite score and high school grade-point average (GPA) and 1st-year academic performance. In addition to this, first-year academic performance is the best predictor of second- and 3rd-year retention.

For the students who attend the top colleges, the SAT exam scores are less efficient than HSGPA in predicting the success of students' first-year grade point average (FGPA). On the other hand, HSPGA is a better predictor than the SAT for the success of the students who attend less selective colleges (Kobrin & Michel, 2006). There is a positive and consistent correlation between SAT score and cumulative GPA as students' progress through their college careers ("The SAT: A Robust Predictor of College Success," n.d.). SAT scores also predict which students are likely to return for the second and third year of college ("The SAT: A Robust Predictor of College Success," n.d.). A study shows that the prediction of first-year academic success is equal with SAT and HSGPA, with correlations of 0.37 (Patterson, Mattern, & Kobrin, 2009).

Another important exam for students is the ACT. Honken and Ralston (2013) pointed out that ACT score and GPA have a positive correlation. Numerous researchers say that it is possible to predict future academic success by using ACT and SAT scores. On the other hand, Perez (2002) states that college performance cannot be ascertained by using either exam. Both exams are coachable, advantaging students who can afford to spend $800 or more on test preparation classes. While there is much open deliberation surrounding how coachable exams are, various studies have shown that scores on exams can be essentially supported through rigorous coaching (Perez). The most significant reason why students' exam scores can be increased through coaching is the exam's format and narrow range of the content chosen by the companies. (Perez). To maximize ACT scores, learning some valuable good test taking strategies is useful ("Different Tests, Same Flaws: A Comparison of the SAT, SAT II, and ACT," n.d.). Also, it is true that these exams have some disadvantages for some students, because of their gender and nationalities. SAT scores overpredicted for males and Asians, Hispanics, and blacks; i.e., the results show that these groups have lower grades than their SAT score would predict. On the other hand, SAT scores underpredicted the grades received by females and whites; i.e., these groups acquired better grades than would be anticipated from their SAT scores (Lynn & Mau, 2001). Because of typically lower SAT Scores of low SES students, they may be encouraged to attend colleges by their friends and lecturers (Walpole, 2003). Higher verbal SAT Scores may let the students attend top institutions, whereas higher quantitative scores would lead them to have a higher college GPA (Walpole, 2003).

Different correlations between test scores and college success have been found by various researchers. Coyle (2015) found a correlation between ACT score and 1st-year academic performance of .38. Boyraz (2015) found a correlation between ACT score and 1st-year academic performance of .22. Although both studies were published in the same year, there was a substantial difference found in the correlation between exam scores and college GPA. This difference in the correlations found suggest it may be fruitful to seek variables that moderate the relationship. Publication status seems associated with differences in correlations found for the same exam type with studies conducted in the same year. For example, in one of the published studies (Sinha, 2011) the correlation was .59. However, in an unpublished study (Romeo, 2011), the correlation was .21. As a result, exam type (SAT/ACT) and publication status (published/unpublished) were included in the present study as potential moderators of effect size.

Meta-analysis is defined as the application of statistical procedures used to collect findings of empirical findings of individual studies to integrate and evaluate them (Wolf, 1986). For this reason, meta-analysis was an invaluable research tool for this study.

Research Questions

The main research questions were:

1. Are SAT (Scholastic Aptitude Test) and ACT (American College Testing) scores predictive of college success? What is the mean correlation between exam score (combining SAT and ACT) and college GPA?

2. Are publication status (published or unpublished) and exam type (SAT or ACT) significant moderators of the predictive validity of exam score for college GPA?

Chapter Two: Method

The data were collected by searches of Academic Search Complete, ERIC, PyscINFO, and Google Scholar. Studies were identified which had correlations between SAT/ACT tests and academic success. Results of *t*-tests or ANOVA were converted to a Pearson's correlation coefficient. Package Metafor (R) was used to calculate effect sizes from studies. Microsoft Excel was used to convert scores from *t*-tests or ANOVA to correlation. Publication status and test type were used for moderator analyses. Converting t-tests and ANOVA to r obtained by the formulas:

$$r = \sqrt{\frac{t^2}{t^2+df}} \tag{1}$$

$$r = \sqrt{\frac{F}{F+df}} \tag{2}$$

Instruments

The SAT I was created in the racist eugenics action goes to the 1920s. As a psychometrician working on the Army Alpha Test (an "IQ" test used with World War I enlistees that was utilized to legitimize racial sorting), Carl Brigham, was given a job by ETS to develop a test to separate more intelligent students from the others for college education in 1925 (Perez, 2002). ACT started to become a popular and useful exam in 1959. At the time, it was the third major manufactured admission test. The ACT, made by E.F. Lindquist (who likewise outlined the Iowa Test of Basic Skills) and Ted McCarrel, was first planned to rival the SAT I (Perez). According to the test's manufacturer, there is a significant difference between the ACT and the SAT I. It is that ACT was very close to the classroom curriculum. The ACT Assessment was not an IQ test or relevant to IQ tests. On the contrary, the ACT questions are linked to classroom learning (Perez). ACT, Inc. states that by using the ACT scores, it is possible to measure high schoolers' general education outcomes and possibility of completion college education work. As an exam, the SAT is created by the College Board, a for-profit corporation. It is administered for the College Board by the Educational Testing Service. Although the original design of SAT I was not a good match for higher school curriculum, thanks to the adjustments done in 2016, the SAT now has a very close relationship with high school curricula.

One of these two exams' scores, SAT and ACT is a requirement of almost all college and universities' application processes. The institutions believe that these exams give them information about the students' readiness to the college education and future success (Schneider & Dorans, 1999). In general, SAT and ACT scores of most students are very close (Schneider & Dorans). The College Board's research shows that there is a strong correlation between these two exams, with the range of r = .89 to .92. On the other hand, their focus areas are not similar. While SAT focuses on verbal and mathematical skills, ACT focuses on the high school curriculum (Schneider & Dorans). The strong relationship between these exams' scores could be because of having a similar format.

They are timed, multiple-choice tests normed on national samples of students (Perez, 2002). If a student does not have the ability to use the limited time efficiently, s/he cannot demonstrate his/her real abilities in the final score. For example, research has shown that the timed exams create disadvantages for the female students and the students whose first language is other than English (Perez). The findings of such research may explain why females get higher grades in high school and college exams, whereas they get lower SAT and ACT scores (Perez).

Predictive validity refers to how accurately a test can predict some future outcome such as academic success (Green, 1991). Tests, for example, the Graduate Record Examination or the Scholastic Aptitude Test, are intended as predictors of academic success at the graduate and undergraduate levels, respectively. If the test has more support for validity, it has a higher correlation with the outcome measure. Factors affecting the value of a predictive validity coefficient include the time between measurement of the predictor (the test) and the outcome. Prediction tends to be more accurate over shorter than over longer time periods. A second factor is whether the test is used for selection purposes. If so, those persons not selected would be unavailable for assessment on the outcome measure. This preselection is likely to reduce the variance of scores and so reduce the predictive validity. This reduced variance is called restriction of range. Restriction of range occurs whenever design or circumstances abbreviate the values of one or both variables being correlated, and participants are intentionally excluded from the study because of having a lower score than a certain criterion (Weber, 2001).

The ACT is an exam which includes 215 multiple-choice questions to complete in 3 hours and 30 minutes. It also has a short break ("The ACT Test Help and Frequently Asked Questions," n.d.). The ACT is used to measure students' academic success in English, Math, Reading, and Science. Students' scores are created by counting the correct answers. The incorrect answers do not have any effect on the score. ACT Composite Score is from 1 to 36 (average of four test scores).

The SAT is 3 hours plus 50 minutes for the optional essay. The essay had a separate score in March 2016. The score is created based on correct answers. No penalty is given for incorrect answers ("The ACT Test Help and Frequently Asked Questions," n.d.). The SAT Math section has 58 items in 1 hour and 20 minutes. The Evidence-Based Reading and Writing Reading Test include 52 items with the time limit of 65 minutes; the Writing, and Language Test has 44 items, in 35 minutes. The Essay section is optional and takes 50 minutes. Scaled scores range from 200 to 800 for Evidence-Based Reading and Writing; 200 to 800 for Math; and from 2 to 8 for the Essay.

Searching Scope

The first strategy was to use Summon@DUto identify potential studies for inclusion in the meta-analysis. The second strategy was searching specific databases. Academic Search Complete and Eric (ProQuest) were relevant databases for finding studies about college admission. After identifying studies, Google Scholar and the Web of Science were useful for forward and backward citation searching. The main inclusion criterion was reporting a correlation between tests and academic success; if a t-test or ANOVA result was provided, it was converted to a correlation.

Lastly, advanced searching was used to get added relevant results. Use of Boolean operators is the best approach to advanced searching in databases. Starting from general search by using keywords (GPA, SAT, ACT e.g.), then "AND," "OR," "NOT" Boolean operators were used to get more accurate results. At the same time, the asterisk (*) was used as a truncation symbol for extending specific words. For example, predict* retrieved these words: prediction, predictive, predictor, predicting.

Inclusion Criteria

- The first inclusion criteria were articles must be either published studies or grey literature (reports, working papers, theses, dissertations, government documents) from between 1990 and 2016 written in English.
- Appropriate studies must describe the relationship between test scores and college success. Eligible studies must have quantitative results and sample sizes.
- Studies need to have the correlation between test score (ACT, SAT) and college GPA or a *t*-test, ANOVA, or chi-square statistic.

Exclusion Criteria

- Studies were excluded if they were published before 1990.
- Studies which did not have a relational statistic between SAT or ACT score and College GPA were excluded.
- Qualitative studies, which did not have descriptive statistics and correlations or *t*-test, ANOVA, or chi-square statistics, were excluded.

See Appendix A for a flowchart of the search process and Appendix C for citations of included studies.

Publication bias

In areas where selectivity exists, a synthesis based on only the published results may be biased. The use of prospective registries of studies to be included in systematic reviews has been suggested (Berlin & Ghersi, 2005) because it provides an unbiased sampling frame for the elimination of publication bias. Carrying out as comprehensive a search as possible when obtaining literature for a synthesis will help minimize the influence of publication bias. In particular, this may involve searching the gray literature for studies not formally published (Cooper, Hedges, & Valentine, 2009). The researcher searched unpublished sources such as reports, books, and dissertations.

Methods for identifying publication bias

First, the funnel plot was used to tested for publication bias. The expectation is that the plot should appear symmetric on the distribution of effect sizes and funnel-shaped if no bias is present. It is essentially a scatterplot of measure of study size against a measure of effect size (Cooper, Hedges, & Valentine, 2009). If publication bias is present, we might expect some suppression of smaller, unfavorable, and nonsignificant studies that could be identified by a gap in one corner of the funnel and hence could yield asymmetry in the plot (Cooper, Hedges, & Valentine, 2009).

Another method used to address publication bias is the Fail Safe N method. Both Rosenthal's (1979) and Orwin's (1983) approach were implemented in assessing the impact of publication bias. This method considers the question of how many new studies are required to bring the overall treatment effect to non-significance (Rosenthal). Trim and Fill is another method that was used for addressing the problem of publication bias.

In this method, smaller studies are omitted until the funnel plot is symmetrical (trimming). The trimmed funnel plot is used to estimate the true “center” of the funnel, and then the omitted studies and their missing “counterparts” around the center are replaced (filling). This provides an estimate of the number of missing studies and an adjusted treatment effect, including the “filled” studies (Sterne, Egger, & Smith, 2001). Lastly, the researcher used Egger’s linear regression. A test of the null hypothesis that $\beta_0 = 0$ (no funnel plot asymmetry) can be derived from the usual regression output produced by statistical packages (Rothstein, Sutton, & Borenstein, 2006).

Analysis

For analyzing data, the Package Metafor (R) was used. Microsoft Office Excel was utilized for converting values that were not presented in a regular format (e.g., t-, F-statistic) to a correlation. First, the data file was input in CSV format. Next, the appropriate statistic was selected in Package Metafor (R). Then, heterogeneity was checked. According to whether there was significant heterogeneity or not, the theoretical argument for how to treat effect sizes was chosen (fixed effect, random effect). After that, the effect was computed. The effect size shows the relationship between the predictor (SAT, ACT), and the outcome (GPA). Finally, moderator analysis was performed by publication status and exam type before tests were implemented to find publication bias.

Chapter Three: Results

In this chapter, the researcher provides the results of the meta-analysis exploring the predictive validity of students' SAT and ACT scores for GPA. First, a description of the studies is presented. Then, the findings of the meta-analysis are displayed. The results of the heterogeneity test, moderator analyses, and publication bias are then provided. The meta-analysis was conducted using the Package Metafor®. Also, SPSS, Excel, and Comprehensive Meta-Analysis (CMA; Borenstein et al., 2005) were employed to create figures and tables.

Search Results

To identify relevant studies, the researcher searched specific databases. Studies were mainly gathered from Academic Search Complete and ERIC (ProQuest) databases. Moreover, other ProQuest and EBSCOhost databases were checked to get relevant studies. Dissertations and Theses (ProQuest) was helpful to obtain dissertations. Also, Google Scholar and Web of Science were used for forward and back citation searching.

First, the researcher used a general search. Key search terms included (SAT, Scholastic Aptitude Test, ACT, American College Testing, GPA, Grade Point Average, college success, predict*, correlate*). After getting more than a thousand results, the researcher looked for key terms on titles and abstracts. Four hundred seventy-six studies were collected after duplicate studies were removed. Finally, 48 studies were selected for the quantitative analysis.

The researcher obtained 60 effect sizes from the 48 studies. Table 1 lists the included studies. Eight studies had more than one effect size. For example, the researcher retrieved four effect sizes from Chowdhury's (2013) study. Table 1 displays studies in alphabetical order by first author's last name. Table 1 also lists studies' sample size and the correlation used to calculate the effect size. Moreover, every study has publication status and exam type listed, which are the two variables used in the moderator analysis.

Table 1. Included Studies

Study	Year	Sample Size (N)	Effect Size (r)	Publication Status	Exam Type
Baker	2016	390	0.42	published	ACT
Bardi	2011	91	0.26	published	ACT
Berry	2009	165781	0.357	published	SAT
Boyraz	2015	484	0.22	published	ACT
Brian	2002	4871	0.265	unpublished	SAT
Carolyn	1994	386	0.36	published	ACT
Chowdhury	2013	92	0.44	published	ACT
Chowdhury-2	2013	105	0.29	published	ACT
Chowdhury-3	2013	57	0.39	published	ACT
Chowdhury-4	2013	69	0.49	published	ACT
Combs	2001	383	0.37	unpublished	SAT
Conard	2005	289	0.28	published	SAT
Coyle	2008	161	0.29	published	SAT
Coyle	2015	1174	0.38	published	SAT
Coyle-2	2008	88	0.22	published	ACT
Coyle-2	2015	1094	0.34	published	ACT
Coyle-3	2008	980	0.35	published	SAT
Coyle-4	2008	898	0.33	published	ACT
Cutrona	1994	418	0.28	published	ACT

DeBerard	2004	204	0.3	published	SAT
Gibb	2002	109	0.46	published	SAT
Guerrero	2000	1142	0.244	unpublished	SAT
Haemme-Fem	2012	237	0.39	published	ACT
Haemmer-mal	2012	1105	0.25	published	ACT
Hollomon	1995	664	0.14	unpublished	ACT
House	1997	148	0.211	published	ACT
Jackson	2003	219	0.33	published	ACT
Keiser	2015	1976	0.28	published	ACT
Keiser-2	2015	56516	0.36	published	SAT
Kirby	2005	93	0.46	published	SAT
Kirby-2	2005	154	0.26	published	SAT
Komarraju	2012	375	0.36	published	ACT
Kraft	2014	125	0.4	published	SAT
Lindley-	2002	313	0.43	published	ACT
Marsh	2008	123	0.44	published	SAT
Marsh-2	2008	100	0.43	published	ACT
Mullen	1995	23064	0.44	unpublished	ACT
Myers-Black	1992	89	0.26	unpublished	ACT
Myers-White	1992	326	0.53	unpublished	ACT
O'Malley	1996	175	0.39	unpublished	SAT
Paszczyk	1994	428	0.21	unpublished	ACT
Patton	1998	6496	0.18	unpublished	ACT
Pettijohn	1995	42	0.41	unpublished	ACT
Redding	1999	76	0.52	unpublished	ACT
Robbins	2010	299	0.154	unpublished	ACT
Romeo	2011	182	0.21	unpublished	SAT
Romeo	2013	143	0.21	published	SAT
Royalty	1994	160	0.16	unpublished	ACT
Schlenker	2013	234	0.39	published	SAT
Schmitt	2009	1155	0.539	published	SAT
Seymour	1994	104	0.197	published	ACT
Shepperd	1993	101	0.4	published	SAT
Singleton	2009	225	0.48	published	SAT
Sinha	2011	836	0.59	published	SAT
Strang	2013	162	0.467	published	SAT
Valencia	1991	99	0.51	unpublished	ACT
Wagerman	2007	131	0.25	published	SAT
Westrick	2015	169818	0.38	published	ACT
Zeng	2002	136	0.31	unpublished	ACT
Ziomek	1996	2959	0.42	unpublished	ACT

Results of Meta-Analysis

Results are presented as follows: effect size using both fixed effects and random effects models are presented followed by the forest plot to provide a summary of individual studies and a radial plot to represent aggregate data. Then moderator analyses results are provided for publication status and exam type. Lastly, funnel plots, Fail Safe N method results, and Trim and Fill method results are presented to indicate the problem of publication bias.

Effect Size

In this study, the first question was: are SAT (Scholastic Aptitude Test) and ACT (American College Testing) scores predictive of college success? What is the correlation between exam score (combining SAT and ACT) and college GPA?

The researcher computed 60 effect sizes to answer the first question. According to the fixed effect model, a statistically significant relationship between exam scores and college success was realized. Exam scores are predictive of college success. The average effect size was $ES = 0.38$, $k = 60$, $p < 0.001$, standard error = .0015. The fixed effects model makes sense if there is the reason to believe that all the studies are functionally identical, and the goal is to compute the common effect size, which would then be summed up to different cases of this same populace (Borenstein, Hedges, & Rothstein, 2007). The random effects model aims to point out the mean of a distribution of effects, not to estimate only one trues. In the random effects model, small studies cannot be discounted by giving them small weights. Although that study's estimate may be imprecise, it is still useful because of being the only study which provides an estimate (Borenstein). Also, studies have more balance under a random effects model. Under the fixed effects model, larger studies share most of the total weight. Under the random effects model, extreme studies is related to their sizes. If they are large, they lose their effects, whereas if they are small, they gain influence.

In this study, because of two main reasons, the researcher adopted a random effects model. First, due to heterogeneity of variance, $Q > \chi^2$ (k-1), the null hypothesis of homogeneity of effect sizes was rejected, Q (59) = 930.19, $p < .0001$. When the null hypothesis of homogeneity of effect size is rejected by finding a statistically significant Q statistic, adopting a random effects model is an option. However, this practice may not encourage us, because we should use the random effects model to understand whether there is a common effect size for all the studies or not, instead of focusing on the statistical test's outcome (Borenstein, 2009). Also, in this study, two studies had large sample sizes. Under the fixed effects model, they were more heavily weighted. The researcher wanted to get more balance to share relative weight based on sample size. If the published literature is used to collect useful studies, the random effect model becomes more appropriate (Borenstein, 2009). In this study, 70% of the studies were published.

With the random effects model, there was also a small average correlation between exam scores and college success, $r = .36$, $p < .001$, standard error = .0166. Homogeneity of effect sizes was violated as with the fixed effects model, Q (59) = 930.19, $p < .0001$.

Exam scores were predictive of college success. The confidence interval was (.32, .39). When comparing the fixed effects model confidence interval (.3821, .3880), because of the standard error differences the random effects model had a wider confidence interval.

Forest Plot

A forest plot indicates the point estimate of studies with their confidence intervals and also shows the overall effect size. A forest plot represents uncertainty and the summary effect and indicates the extent to which each study contributes to the overall result (Cooper, Hedges, & Valentine, 2009). The diamond in Figure 1 represents the summary effect size at the bottom of the plot. The plot provides some meaningful data on the statistics which allow us to make correct interpretations and shows anomalies like attention required outliers (Borenstein, 2005). Before starting to use the funnel plot and statistical tests to figure out biases, most researchers generate a forest plot to get a visual idea of the effect size for the study (Borenstein). In this study, the researcher obtained two forest plots to compare fixed effect and random effect models. Figure 1 exhibits the forest plot for the random effects model; Figure 2 presents the forest plot for the fixed effects model. One of the main differences between random effects and fixed effects models is the studies' relative weight. The fixed effects model contains a wide range of weights (as reflected in the size of the boxes) whereas there is a narrow range of weights under the random effects model (Borenstein, 2009). For example, to compare large and small studies under both models, one of the largest studies (Berry, 2009) has two times the weight of Baker's (2016) study under the random effects model. On the other hand, under the fixed effects model, Berry (2009) has 410 times the weight of Baker (2016). The fixed effects model had a larger effect size (ES = .38) than the random effects model (ES =.36). In this study, the confidence band for 39 studies intersected the mean effect size, while 21 studies did not intersect the mean effect size under the random effects model. Under the fixed effects model, 39 studies intersected the mean effect size as well.

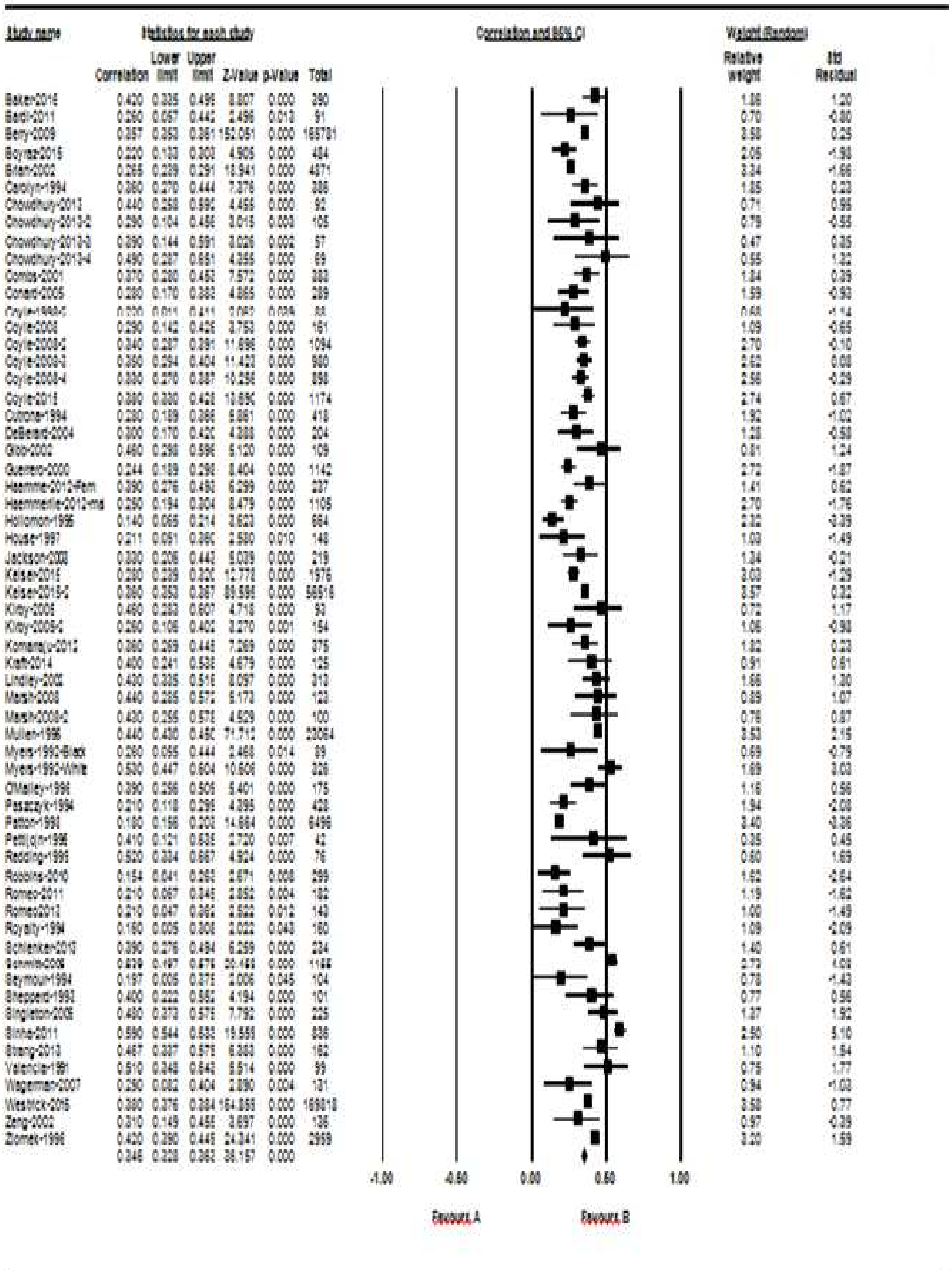


Figure 1. Forest Plot of Effect Sizes (Random Effects)

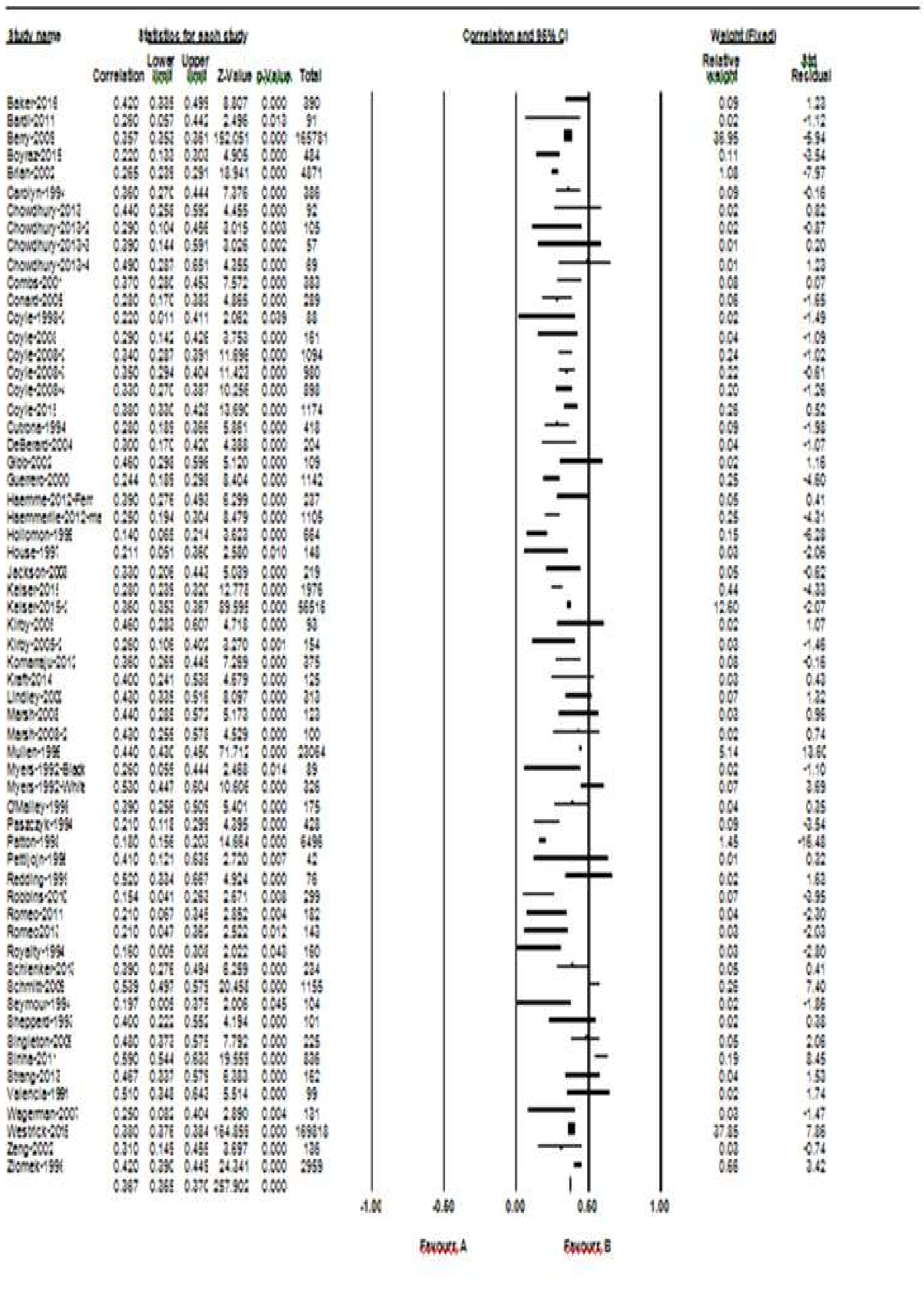


Figure 2. Forest Plot of Effect Sizes (Fixed Effects)

Radial Plot

Galbraith's radial plot is an alternative way to visualize results of the meta-analysis. Galbraith plots are used to facilitate examination of heterogeneity which also covers the detection of outliers (Cabrera & Higgins, 2010). The data are presented in the middle of the figure with the inverse of the standard error of the x-axis and the standardized estimate on the y-axis; a two-unit change in the standardized estimate is equivalent to the 95% confidence interval (Cooper, Hedges, & Valentine, 2009). In the radial plot, the larger studies are located on the y-axis. The reason for getting different radial plots under fixed effects and random effects models is their formula. The random effects model adds the estimate of the between-studies variance (τ2). As a result, under the fixed effects model large studies are separate from other studies. However, under a random effects model, studies scatter more closely to each other. Also under the random effects model, detecting outliers are easier than under the fixed effects model. Figures 3 and four present the radial plots, the inverse of the standard error of the x-axis and the standardized estimate as the y-axis in 95% confidence interval (Cooper, Hedges, & Valentine). Under the fixed effects model, because of the relative weight, the big studies suppressed standardized estimate scores. In contrast, under random effects model studies, relative weights are more stable. The big studies did not suppress the standardized estimate score. The range of standardized estimates is wider.

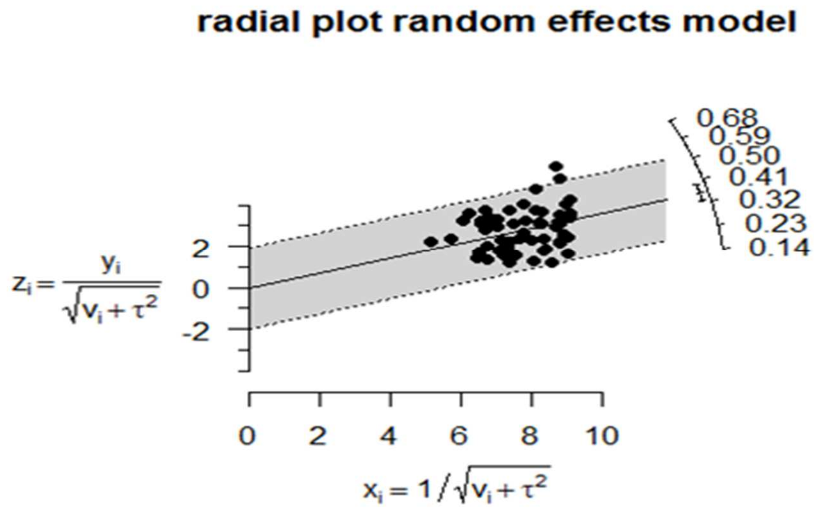


Figure 3 Radial Plot (Random Effects)

Moderator Analysis

In this study, the second question was: are exam type (SAT and ACT) or publication status (published or unpublished) moderators of the relationship between exam score and GPA?

The null hypothesis that the effect size values are homogeneous was not rejected. There was no statistically meaningful difference in effect size by publication status. See Appendix B for details of this moderator analyses.

Exam type was another moderator. Under the random effects model the Q statistic also was significant ($X^2(1) = 4.77, p < 0.05$). There was a significant difference between SAT and ACT in predicting college GPA. However, the effect sizes were quite similar: the correlation for ACT was .3271 and for SAT it was .3573 (random effects model).

Publication Bias

In order to assess the impact of publication bias, unpublished studies can be included in the meta-analysis and can be compared regarding whether these unpublished studies have smaller effect sizes than published ones (Card, 2015). If the researcher does not realize any differences between these two different study types, this means there is no evidence of publication bias (Card).

For meta-analysis, publication bias is one of the biggest problems. When researchers find null results, they are less likely to publish their studies. Publication bias shows its effect on the conclusion if the published literature is not representative of studies on the topic, in that the available results likely show a stronger overall effect size than if all studies were considered (Card, 2015). Searching grey literature (dissertations, reports) is one of the solutions to avoid publication bias. Three different approaches to examining publication bias were used in this study (Funnel Plots, Fail Safe N, and Trim and Fill).

Funnel Plot

A funnel plot is a graphic way of detecting publication bias. The funnel plot is a scatterplot of the effect sizes found in studies about their sample size (Card, 2015). If there is no publication bias, sampling error occurs randomly, because the studies are distributed symmetrically about the mean effect size (Borenstein, 2009). If publication bias is present, we might expect some suppression of smaller, unfavorable, and nonsignificant studies that could be identified by a gap in one corner of the funnel and hence could yield asymmetry in the plot (Cooper, Hedges, & Valentine, 2009). Figure 5 shows that there was no asymmetry in the plot. The researcher did not detect publication bias in this study. Although the plot looks symmetrical, some studies appear to be outside the funnel. It is common to see that large studies seem over the top of the graph and aggregate around the mean effect size, whereas the smaller studies seem at a lower level of the top of the graph and spread around the mean effect size (Borenstein, 2009). Normally the increase in the sample size has a negative correlation with standard error, and the expectation is their location should observe close to the mean effect size. Studies spread out over large areas, and some studies are located out of the funnel plot. The researcher compared four articles whose funnel plots look like this funnel plot to check any problems with the funnel plot. According to Leta, Alemayehu, Seyoum, and Bezie (2016), this type of funnel plot did not suggest the existence of publication bias. It shows the large degree of heterogeneity among studies.

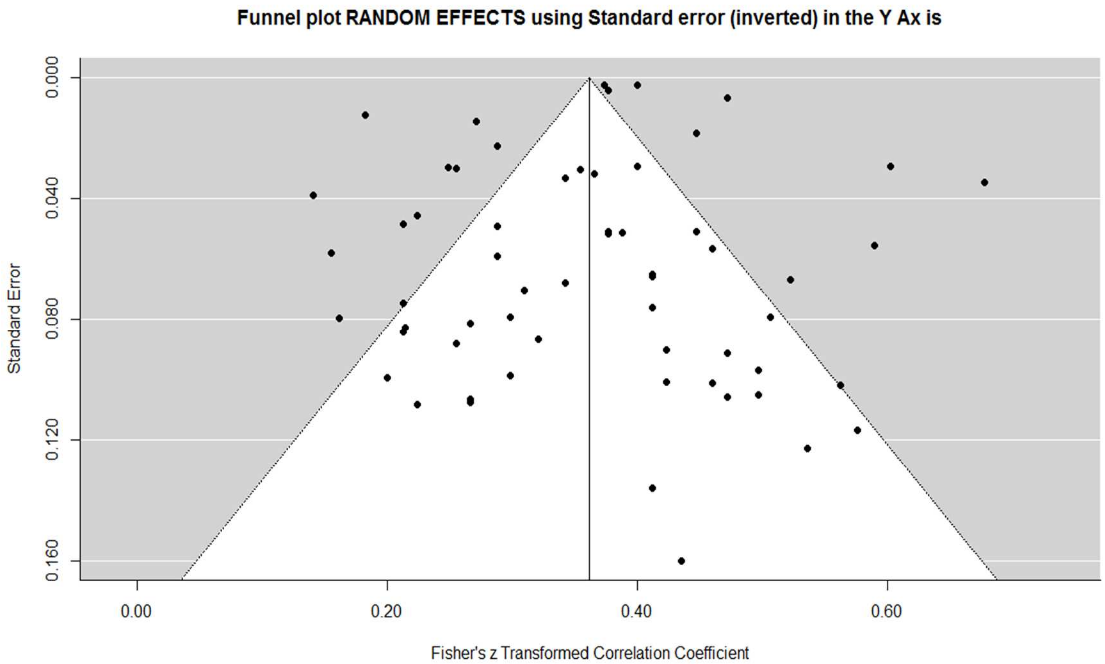


Figure 4. Funnel Plot of Random Effect Size Values by Standard Error

Fail-safe N Calculation Using the Orwin Approach

Orwin's (1983) approach is an alternative for a Fail-Safe N calculation. In this approach, the fail-safe number (N_{ES}) would tell us how many excluded studies with an average effect size of zero would be needed before the true effect size would be reduced to the smallest meaningful effect size (Card, 2015). In this research, 61 missing studies bring the overall effect to a specified level of .18. Metafor provided a target effect size by dividing the mean effect size in half. Another target effect size could have been determined, according to Cohen (1969). When the determined target effect size is lower than target effect size, the number of missing studies will increase.

Fail-safe N Calculation Using the Rosenthal Approach

The Fail Safe N method considers the question of how many new studies averaging a null result are required to bring overall treatment effect to non-significance (Rosenthal, 1979). Rosenthal's approach determines the number of unpublished studies, with an average observed effect of zero, there would need to be to reduce the overall z-score to non-significance (Cooper, Hedges, & Valentine, 2009). In this study, when 273,022 missing studies are added, the p-value would be nonsignificant. According to the formula when the *p*-value is small, the Fail-safe N is high. When the *p*-value is close to a .05 significance level, Fail-safe N is small.

Duval and Tweedie's Trim and Fill

The key idea behind the funnel plot is that in the absence of bias the plot is symmetric about the summary effect size (Borenstein, 2005). If there are more small studies on the right than on the left, our apprehension is that there might be studies missing from the left. The trim and fill procedure builds directly on this idea by imputing the effect sizes for the missing studies, adding them to the analysis, and then recomputing the effect size (Borenstein, 2005).

Trim and Fill uses an iterative methodology to remove the most extreme studies from the positive side of the funnel plot, recalculating the effect size at each iteration until the funnel plot is symmetric about the new effect size (Borenstein, 2005). In this study, as shown in Figure 6, there are no unfilled circles, and an estimated number of missing studies on the left side is 0.0 (SE = 4.55) under a random effects model. Egger's and

Kendall's tests are not significant. If one is significant, there is some evidence of potential bias.

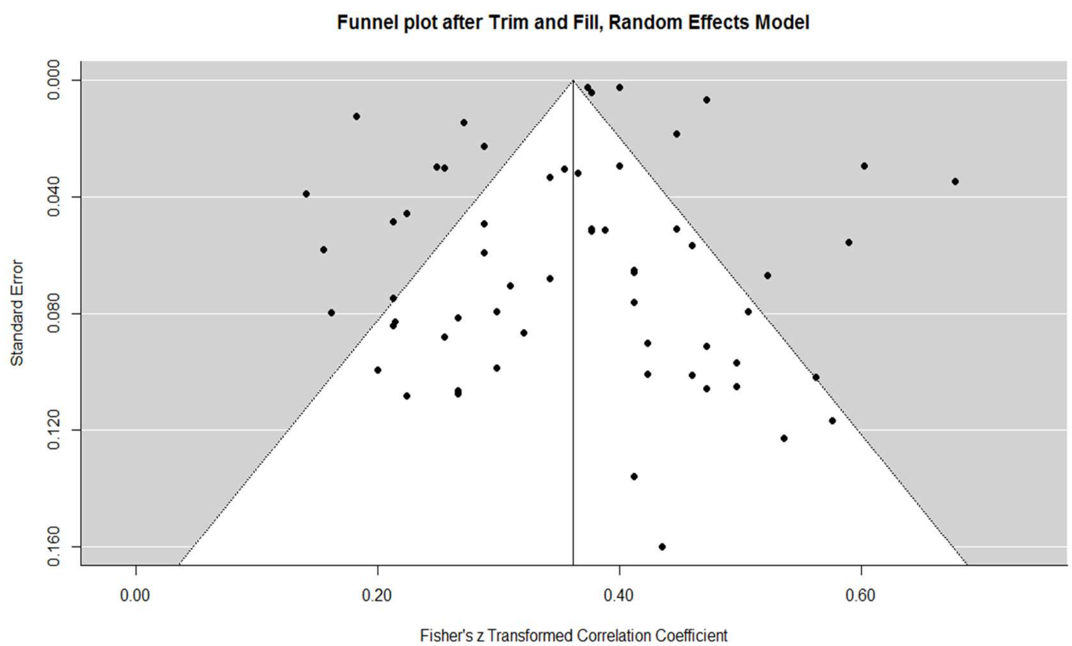


Figure 5. Funnel Plot of Random Effect Size After Trim and Fill

Chapter Four: Discussion

This study addressed two main research questions.

1. Are SAT (Scholastic Aptitude Test) and ACT (American College Testing) scores predictive of college success? What is the mean correlation between exam score (combining SAT and ACT) and college GPA?
2. Are publication status (published or unpublished) and exam type (SAT or ACT) significant moderators of the predictive validity of exam score for college GPA.

After computing 60 effect sizes, an overall effect size of .36 (95% CI = .32 - .39) was determined. The researcher found that there was a statistically significant positive relationship between exam score (SAT, ACT) and college success. Moreover, according to Cohen's (1992) guidelines, the effect size of .36 is equivalent to a medium to large effect. When predicting students' academic performance, a small amount of variance ($.36^2$) is explained by students' exam scores. Pritchard and Wilson (2003) claimed no single factor or set of factors (e.g., demographic, academic, social, emotional) predict individual student success. There is a multitude of factors that influences the way students adjust to college. However, exam score was a significant predictor of college success. But, getting low scores on the exams does not necessarily imply a low college GPA will be obtained nor high exam scores imply a high college GPA.

The average effect size was produced for the predictive validity of ACT and SAT scores and college GPA in this study. The results were very close in both random effects (ES=.36) and fixed effects models (ES=.38). In the literature, Keiser (2015) is one of the large current studies about predictive validity of SAT and college GPA. Keiser collected data from more than 50,000 students. An effect size of 0.36 was found by Keiser. Westrick (2015) is the other large current study about predictive validity of ACT and college GPA. He had a sample size of more than 100,000. His effect size was 0.38. When comparing this study and other two different studies, the results are consistent with results found in prior research were random effects and fixed effects model.

Moderator analysis was computed, and the Q statistic was not significant for publication status. Significant differences were not found in effect size between published and unpublished studies. Moderator analysis was also computed for exam type. The Q statistic was significant for exam type. There was a difference between ACT and SAT score in effect size, though the difference was small. According to the College Board's research, the correlation between SAT and ACT scores ranges from .89 to .92. There is a strong positive correlation between SAT and ACT scores. Although they have some minor differences in test content and test, both exams are good predictors of college success. Students who obtain high scores on the SAT exam, most probably get high scores on the ACT exam. In different states, a different exam is more popular, but almost every four-year college in the US accepts ACT and SAT scores.

This study had limitations. First, the data had a large degree of heterogeneity. One of the hypothesized reasons for heterogeneity was publication status. The researcher performed a moderator analysis using publication status. However, the researcher did not find a significant result for publication status. Also, included studies represented different types of samples. For example, some studies focused on high academic level students. Their mean standardized scores and mean college GPA's are higher than in the general population. When the researcher looked at the correlation between exam scores and college GPA for high-GPA samples, lower correlations were found, likely because of range restriction. These studies yielded small correlations. Also, some studies did not represent the whole population. For example, Guerrero's (2000) study is one of the studies which was located outside of the funnel plot. Guerrero's (2000) sample size was 1,142, and the *r* was 0.244. In his study, about half of the students were Asian. The correlation for Asian students alone was 0.09. Heterogeneity is affected when studies vary in representation of a general population and in sample size. Another reason for the heterogeneity is some large studies had high correlations, and they were located outside of the funnel plot. For example, Schmitt's (2009) study involved data from different colleges, and this study used corrected GPA to standardize across the different colleges.

Future Research

In this study, the researcher assessed the predictive validity of the Scholastic Aptitude Test (SAT) and American College Testing (ACT) overall scores for College GPA. Before searching relevant databases, the researcher wanted to focus on both gender and ethnicity. For example, Lynn and Mau (2001) found that SAT scores were less accurate in predicting the grades of males and Asians, Hispanics, and African Americans. In other words, these groups did not obtain as good grades as would be predicted from their scores. Conversely, SAT scores were underpredicted grades for females and whites. These groups obtained better grades than would be predicted from their SAT scores (Lynn & Mau, 2001). The researcher searched other studies, which had correlations between SAT/ACT score and college GPA by gender and ethnicity. Unfortunately, not enough studies were found to allow a meta-analysis. In the future, more studies are needed that include gender and ethnicity as variables. Also, exam coaching will be another potential moderator variable for predicting college GPA.

References

Anzures-Cabrera, J., & Higgins, J. (2010). Graphical displays for meta-analysis: An overview with suggestions for practice. *Research Synthesis Methods, 1*(1), 66-80.

Borenstein, M. (2005). Software for publication bias. *Publication bias in meta-analysis: Prevention, assessment and adjustments,* 193-220.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2005). *Comprehensive Meta-Analysis, Version 2.* Englewood, NJ: Biostat.

Borenstein, M., Hedges, L., & Rothstein, H. (2007). Meta-analysis: Fixed effect vs. random effects. *Meta-analysis. com.*

Borenstein, M. (2009). *Introduction to meta-analysis*. Chichester, U.K: John Wiley & Sons.

Card, N. A. (2015). *Applied meta-analysis for social science research.* Guilford Publications.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd edition). New York, NY: Russell Sage Foundation.

Different Tests, Same Flaws: A Comparison of the SAT, SAT II and ACT. (n.d.). Retrieved from http://fairtest.org/different-tests-same-flaws-comparison-sat-sat-ii-a

Ditchkoff, S. S., Laband, D. N., & Hanby, K. (2003). Academic performance of transfer versus" native" students in a wildlife Bachelor of Science program. *Wildlife Society Bulletin,* 1021-1026.

Dorans, Neil J., and Dianne Schneider. 1999. *Concordance between SAT I and ACT Scores for Individual Students*. Princeton, NJ: College Entrance Examination Board.

Honken, N. B., & Ralston, P. A. (2013). High-Achieving High School Students and Not So High-Achieving College Students: A Look at Lack of Self-Control, Academic Ability, and Performance in College. *Journal of Advanced Academics, 24*(2), 108-124. doi:10.1177/1932202x13482466

Green, K. E. (1991). *Educational testing: Issues and applications.* NY: Garland Pub.

Kobrin, J. L., & Michel, R. S. (2006). The SAT as a predictor of different levels of college performance (The College Board Research Report No. 2006-3). New York: The College Board.

Leta, S., Alemayehu, G., Seyoum, Z., & Bezie, M. (2016). Prevalence of bovine trypanosomosis in Ethiopia: a meta-analysis. *Parasites & Vectors, 9*(139).

Lynn, R., & Mau, W. C. (2001). Ethnic and sex differences in the predictive validity of the scholastic achievement test for college grades. *Psychological Reports, 88,* 1099-1104.

Patterson, B. F., Mattern, K. D., & Kobrin, J. L. (2009). *Validity of the SAT for predicting FYGPA: 2007 SAT validity sample*. Statistical Report No. 2009-1.College Board.

Pope, J. (24 September, 2012). SAT scores edge down; ACT now more popular exam. *Associated Press.*

Pritchard, M. E., & Wilson, G. S. (2003). Using emotional and social factors to predict student success. *Journal of College Student Development, 44*(1), 18-28.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* John Wiley & Sons.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. doi:10.1037/0033- 2909.86.3.638.

Sinha, R., Oswald, F., Imus, A., & Schmitt, N. (2011). Criterion-focused approach to reducing adverse impact in college admissions. *Applied Measurement in Education, 24*(2), 137-161. doi:10.1080/08957347.2011.554605

Standardized Tests - ProCon.org. (2016, June 2). Retrieved July 6, 2016, from http://standardizedtests.procon.org/

Snyder, T. D., & Dillow, S. A., 2013. *Digest of Education Statistics,* 2012. NCES 2014-015. National Center for Education Statistics.

Sterne, J. A., Egger, M., & Smith, G. D. (2001). Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal, 323*(7304), 101.

The ACT Test Help and Frequently Asked Questions. (n.d.). Retrieved from http://www.act.org/content/act/en/products-and-services/the-act/help.html

The SAT: A Robust Predictor of College Success. (n.d.). Retrieved from http://highered.colorado.gov/Academics/Groups/AdmissionTransferReview

Walpole, M. (2003). Socioeconomic status and college: How SES affects college experiences and outcomes. *Review of Higher Education*, 27(1), 45-73.

Webb, B. (2013). Standardized Tests. In P. L. Mason (Ed.), *Encyclopedia of Race and Racism* (2nd ed., Vol. 4, pp. 128-130). Detroit: Macmillan Reference USA. Retrieved July 7, 2016, from Gale Virtual Reference Library.

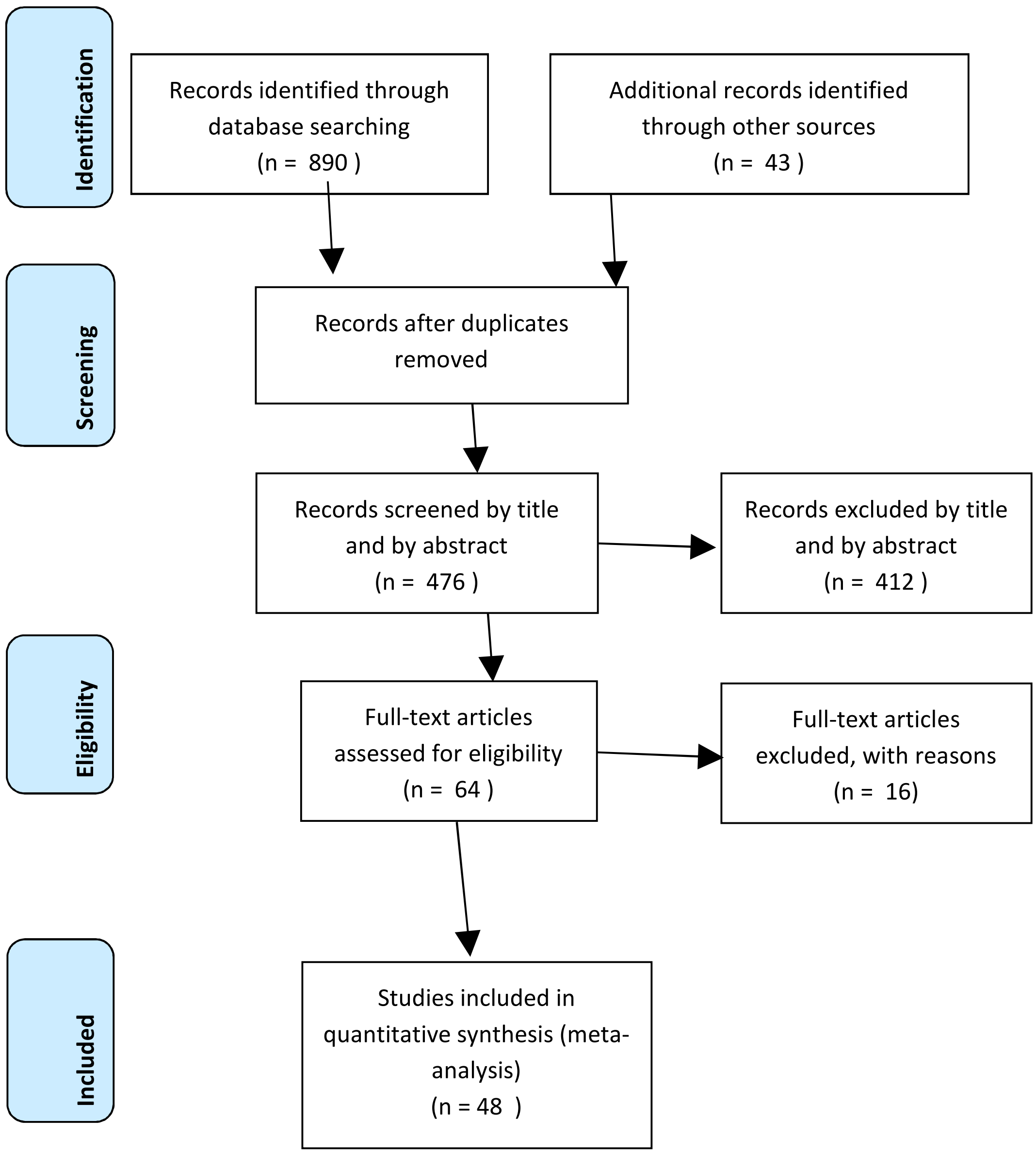
Weber, D. A. (2001). *Restriction of range*: The truth about consequences and corrections.

2015 College-Bound Seniors Total Group Profile Report. (n.d.). Retrieved from https://secure-media.collegeboard.org/digitalServices/pdf/sat/total-group-2015.p

Appendix A

FLOW CHART

Chart from Moher, Liberati, Tetzlaff, & Altman, The PRISMA Group (2009).



Appendix B

Moderator Analysis Results

Results for subgroups (fixed effect model):

	k	COR	95%-CI	Q	I^2
group = published	42	0.3673	[0.3646; 0.370]	289.66	85.8
group = unpublished	18	0.3647	[0.3564; 0.373]	640.19	97.3

Test for subgroup differences (fixed effect model):

	Q	d.f.	p-value
Between groups	0.33	1	0.5651
Within groups	929.86	58	< 0.0001

Results for subgroups (fixed effect model):

	k	COR	95%-CI	Q	I^2

group = act 35 0.3777 [0.3741; 0.3813] 631.86

94.6%

group = sat 25 0.3573 [0.3538; 0.3608] 236.15

89.8%

Test for subgroup differences (fixed effect model):

	Q	d.f.	p-value
Between groups	62.18	1	< 0.0001
Within groups	868	58	< 0.0001

Results for subgroups (random effect model):

k COR 95%-CI Q I^2

group = act 35 0.3271 [0.2939; 0.3595] 631.86

94.6%

group = sat 25 0.3573 [0.3481; 0.3948] 236.15

89.8%

Test for subgroup differences (random effects model):

	Q	d.f.	p-value
Between groups	4.77	1	0.029

Results for subgroups (random effects model):

	k	COR	95%-CI	Q	I^2
group = published	42	0.3645	[0.3490; 0.3798]	289.66	85.8
group = unpublished	18	0.3171	[0.2463; 0.3845]	640.19	97.3

Test for subgroup differences (random effects model):

	Q	d.f.	p-value
Between groups	1.77	1	0.183

Appendix C

INCLUDED STUDIES

Baker, M. R., Frazier, P. A., Greer, C., Paulsen, J. A., Howard, K., Meredith, L. N., ... Shallcross, S. L. (2016). Sexual victimization history predicts academic performance in college women. *Journal of Counseling Psychology, Advance Online Publication.* http://dx.doi. org/10.1037/cou0000146. Benotsch, E. G., Koester, S., Luck

Bardi, M., Koone, T., Mewaldt, S., & O'Connor, K. (2011). Behavioral and physiological correlates of stress related to examination performance in college chemistry students. *Stress*, *14*(5), 557-566.

Berry, C. M., & Sackett, P. R. (2009). Individual differences in course choice result in underestimation of the validity of college admissions systems. *Psychological Science*, *20*(7), 822-830.

Boyraz, G., Granda, R., Baker, C. N., Tidwell, L. L., & Waits, J. B. (2015). Posttraumatic stress, effort regulation, and academic outcomes among college students: A longitudinal study. *Journal of Counseling Psychology.* Advance online publication. doi:10.1037/cou0000102

Chowdhury, M. I., & Wheeling, B. (2013). Determinants of major field test (MFT) score for graduating seniors of a business school in a small mid-western university. *Academy of Educational Leadership Journal*, *17*(1), 59.

Combs, J. E. (2001). *Academic self-efficacy and the overprediction of African American college student performance* (Doctoral dissertation, University of California, Santa Barbara).

Conard, M. A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality*, *40*(3), 339-346.

Coyle, T. R., & Pillow, D. R. (2008). SAT and ACT predict college GPA after removing g. *Intelligence*, *36*(6), 719-729.

Coyle, T. R., Snyder, A. C., Richmond, M. C., & Little, M. (2015). SAT non-g residuals predict course specific GPAs: Support for investment theory. *Intelligence*, *51*, 57-66.

Cutrona, C. E., Cole, V., Colangelo, N., Assouline, S. G., & Russell, D. W. (1994). Perceived parental social support and academic achievement: An attachment theory perspective. *Journal of Personality and Social Psychology*, *66*(2), 369.

DeBerard, M. S., Spielmans, G. I., & Julka, D. L. (2004). Predictors of academic achievement and retention among college freshmen: A longitudinal study. *College Student Journal*, *38*(1), 66.

Fields, V. (1991). An investigation into factors affecting academic success associated with on-campus living experiences for African-American undergraduate students at Iowa State University. Unpublished doctoral dissertation, Iowa State University, Ames. IA.

Gibb, B. E., Zhu, L., Alloy, L. B., & Abramson, L. Y. (2002). Attributional styles and academic achievement in university students: A longitudinal investigation. *Cognitive Therapy and Research*, *26*(3), 309-315.

Guerrero, B. G. (2001). *An analysis of academic, demographic, and non-cognitive factors that influence academic performance during the freshman year in college* (Doctoral dissertation, ProQuest Information & Learning).

Haemmerlie, F., & Montgomery, R. L. (2012). Gender differences in the academic performance and retention of undergraduate engineering majors. *College Student Journal, 46*(1), 40–45.

House, J. D., & Keeley, E. J. (1997). Predictive validity of college admissions test scores for American Indian students. *The Journal of Psychology*, *131*(5), 572-574.

Hu, N. B. (2002). Measuring the weight of high school GPA and SAT scores with second term GPA to determine admission/financial aid index–a case study. In *annual meeting of the Association for Institutional Research, Toronto, Canada (ERIC Document reproduction Service No. ED 473 071)*.

Jackson, T., Weiss, K. E., Lundquist, J. J., & Hooper, D. (2003). The impact of hope, procrastination, and social activity on academic performance of Midwestern college students. *Education*, *124*(2), 310.

Keiser, H. N., Sackett, P. R., Kuncel, N. R., & Brothen, T. (2016). Why women perform better in college than admission scores would predict: Exploring the roles of conscientiousness and course-taking patterns. *Journal of Applied Psychology*, *101*(4), 569.

Kirby, K. N., Winston, G. C., & Santiesteban, M. (2005). Impatience and grades: Delay-discount rates correlate negatively with college GPA. *Learning and individual Differences*, *15*(3), 213-222.

Komarraju, M., Ramsey, A., & Rinella, V. (2013). Cognitive and non-cognitive predictors of college readiness and performance: Role of academic discipline. *Learning and Individual Differences*, *24*, 103-109.

Lindley, L. D., & Borgen, F. H. (2002). Generalized self-efficacy, Holland theme self-efficacy, and academic performance. *Journal of Career Assessment*, *10*(3), 301-314.

Marsh, C. M., Vandehey, M. A., & Diekhoff, G. M. (2008). A comparison of an introductory course to SAT/ACT scores in predicting student performance. *The Journal of General Education*, *57*(4), 244-255.

Mullen, R. (1995). *Indicators of grade inflation* (AIR 1995 Annual Forum Paper). Columbia: University of Missouri—Columbia.

Myers, R. S., & Plyes, M. R. (1992). Relationship among high school grades, ACT test scores, and college grades. *Mid-South Educational Research Association,* 2-15. Retrieved from http://eric.ed.gov/PDFS/ED353317.pdfhttp://eric.ed.gov/PDFS/ED353317.pdf

Parrott-Robbins, R. J. (2010). *Identifying the Value of the ACT Score as a Predictor of Student Success in Respiratory Care, Radiography, and Nursing at Southeast Kentucky Community and Technical College*. (Mississippi State University, Ph.D., 2010). *DAI*, *71* (8), 2734A. (Accession No. AAI3412668)

Paszczyk, S.L. 1994. *A comparative analysis of ACT scores and final GPAs of Chicago State University undergraduate students.* Chicago, IL: Chicago State University.

T. K. Patton, "Dierential prediction of college performance between gender," Report No. TM029407, Middle Tennessee State University. (ERIC Reproduction Service No. ED427030), 1998.

Pettijohn, T. F. (1995). Correlations among college students' grade point averages and American College Test scores. *Psychological Reports*, *76*(1), 336-338.

Robertson-Kraft, C., & Duckworth, A. L. (2014). True Grit: Trait-level Perseverance and Passion for Long-term Goals Predicts Effectiveness and Retention among Novice Teachers. *Teachers College Record,* 1-27.

Romeo, E. M. (2013). The Predictive Ability of Critical Thinking, Nursing GPA, and SAT Scores on First-Time NCLEX-RN Performance. *Nursing Education Perspectives*, *34*(4), 248-253.

Schlenker, B. R., Schlenker, P. A., & Schlenker, K. A. (2013). Antecedents of academic engagement and the implications for college grades. *Learning and Individual Differences*, *27*, 75-81.

Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology*, *94*(6), 1479.

Shepperd, J. A. (1993). Student derogation of the Scholastic Aptitude Test: Biases in perceptions and presentations of College Board scores. *Basic and Applied Social Psychology*, *14*(4), 455-473.

Singleton, R. A., & Wolfson, A. R. (2009). Alcohol consumption, sleep, and academic performance among college students. *Journal of Studies on Alcohol and Drugs*, *70*(3), 355-363.

Sinha, R., Oswald, F., Imus, A., & Schmitt, N. (2011). Criterion-focused approach to reducing adverse impact in college admissions. *Applied Measurement in Education*, *24*(2), 137-161.

Strang, K. D. (2013). University accreditation and benchmarking: Pedagogy that increases student achievement. *International Journal of Educational Research*, *62*, 210-219.

Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality*, *41*(1), 221-229.

Villella, J. W. (1996). *Freshmen involvement in a Pennsylvania state university and its effect on grade point average and satisfaction with the institution*. Retrieved from ProQuest Dissertations and Theses. (UMI No. 9717004)

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT® scores, high school grades, and SES. *Educational Assessment*, *20*(1), 23-45.

Zeng, L., Simonsson, M., & Poelzer, H. (2002, April). *Teacher certification tests: Variables that predict pass/fail status on elementary professional development examination for preservice teachers.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Ziomek, R. L. & Andrews, K.M. (1996). *Predicting the College Grade Point Averages of Special Tested Students from Their ACT Assessment Scores and High School Grades* (ACT Research Report Series, 96-7). Iowa City, IA: American College Testing