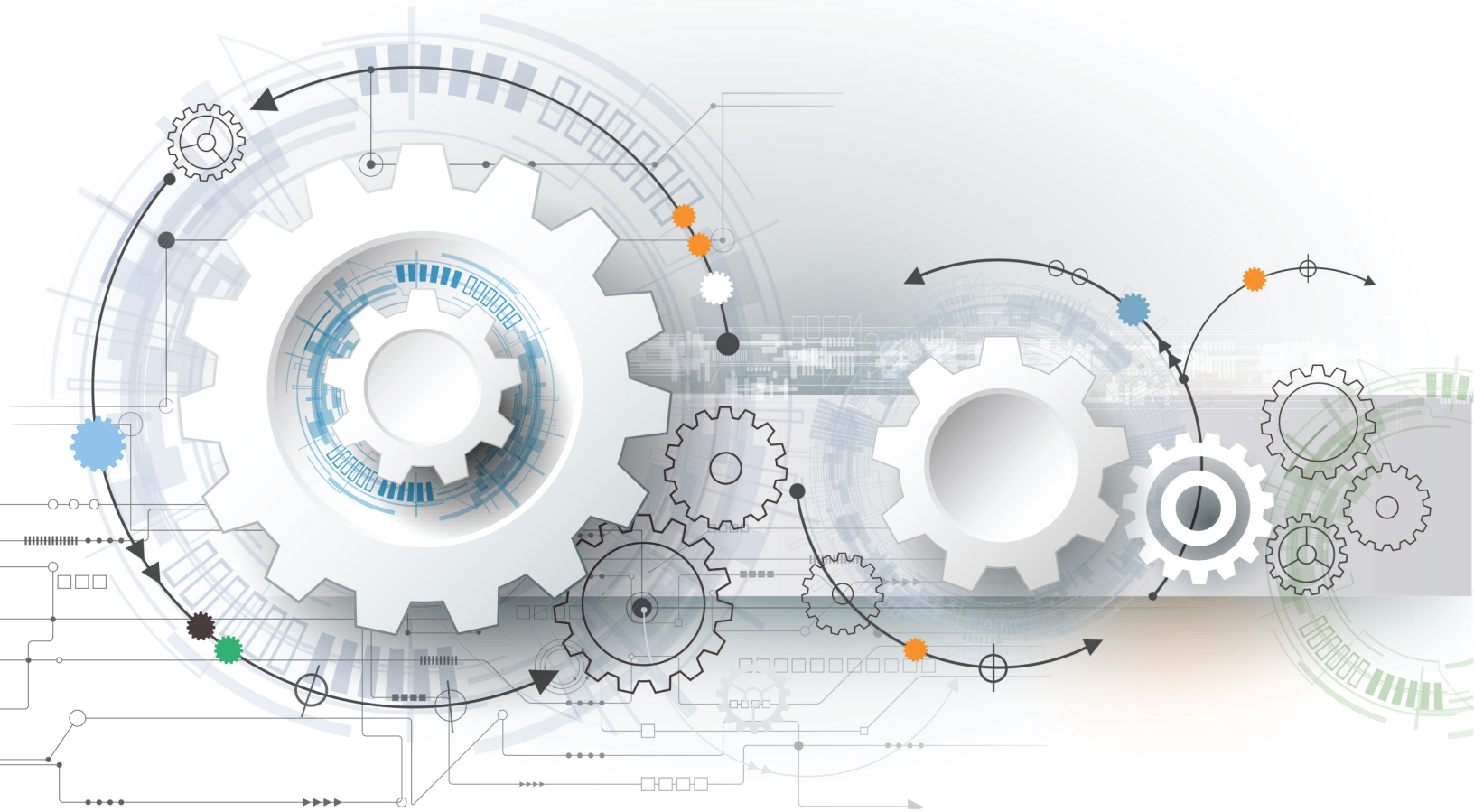
Working Paper
2018-6

Investigating Test Prep Impact on Score Gains Using Quasi-Experimental Propensity Score Matching

RAEAL MOORE, PHD
EDGAR SANCHEZ, PHD
MARIA OFELIA SAN PEDRO, PHD



ACT.org/research

ABOUT THE AUTHORS

Raeal Moore is a senior research scientist specializing in survey methodological research and research on education best practices in P-12 schools.

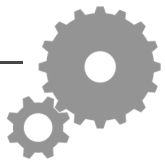
Edgar Sanchez, a senior research scientist in the Validity and Efficacy Research Department at ACT, works on predictive modeling of student educational outcomes such as enrollment, persistence, and graduation. His current research focus is on the efficacy of test preparation.

Maria Ofelia "Sweet" San Pedro is a research scientist in the Learning, Assessment and Navigation Experiences group within ACT Research. specializing on learning science and learning analytics. She researches on K-12 learning experiences, assessment, and instruction within digital learning environments.

ACKNOWLEDGEMENTS

The authors wish to thank Krista Mattern, Jeff Allen, Jeff Schiel, and Ty Cruce for their feedback on previous manuscripts.

ACT WORKING PAPER SERIES



ACT working papers document preliminary research. The papers are intended to promote discussion and feedback before formal publication. The research does not necessarily reflect the views of ACT.

Abstract

A quasi-experimental method and test-retest design were used to explore the impact of participating in test preparation, whether the impact of test preparation depends on the first ACT® score, and the impact of specific test preparation activities. Test preparation improved students' retest scores and this effect did not differ depending on students' first ACT score. Working with a private tutor improved retest scores for those who reported preparing, whereas engaging in other test preparation activities was not statistically related to score gains. Students who reported feeling inadequately prepared for the second test had ACT Composite scores that were lower than those students who felt adequately prepared.

Keywords: test preparation, propensity score matching, quasi-experimental design, ACT

Introduction

Performance in high-stakes, standardized college entrance exams (e.g., SAT, ACT) provides important information that is used by postsecondary institutions in their admission process (Clinedinst, 2014). Higher exam scores increase the chances of a student getting admitted in more selective schools and eligibility for merit-based scholarships (Carnevale & Rose, 2003; Doyle, 2006). Given this context, test preparation programs have emerged to address students' desire to improve their test scores whether by "improving the skills measured by the test or by improving the skills for taking the test, or both" (Messick, 1982). Test developers and other service providers have developed a number of test preparation solutions, including workbooks, practice tests, coaching, class curricula, as well as in-person or online tutoring packages. These solutions can be characterized by instruction aimed at developing skills and abilities, practice and exposure to practice problems that resemble the actual text, as well as the awareness and practice of test-taking strategies. Test preparation activities therefore vary by duration and intensity.

Despite the growth of the test preparation industry (Barnes Reports, 2017), the existing research on the efficacy of test preparation has shown mixed results depending on the approach to test preparation, population tested, or study design/methodology (Lane, Raymond, & Haladyna, 2015). On average, research shows test preparation has small to moderate positive effects on actual test scores, with average test score gains falling within the margin of error for standardized college entrance exams. A study by Briggs (2001) showed test preparation solutions such as commercial preparation classes and tutoring had very small effects on ACT subject test scores, in which score increases did not exceed a full score point. The National Association for College Admissions Counseling (NACAC) reviewed efficacy studies on test preparation and also

found minimal effects on test scores – an average gain of approximately 30 points on the SAT (Briggs, 2009). Meta-analyses on test preparation for the SAT or similar achievement tests show a modest test score gain of 0.25 standard deviations. Assuming similar gains for the ACT test, this equates to a gain of about 1.3 to 1.7 points for the subject tests and the Composite score (Bangert-Drowns, Kulik, & Kulik, 1983; Montgomery & Lily, 2012; Powers, 1993).

Irrespective of engaging in test preparation activities, research has found that students who retest tend to increase their score. A meta-analysis that looked at practice effects found increases in test scores that were approximately 0.25 standard deviations for students who retook a cognitive ability test when assessed between the first and second take (Hausknecht et al., 2007). For the ACT test, retesting has shown a 1 to 2 point scale score increase (Andrews & Ziomek, 1998). The exact gain associated with retesting is dependent upon a number of factors, including the number of times a student has tested, in what grade a student takes the ACT test, and how much time has elapsed between test administrations (ACT, 2017b).

Aside from efficacy studies, research on test preparation has looked into factors of practice opportunities in test preparation and motivational factors in students that may result in test score gains. Appelrouth, Zabrucky, and Moore (2015) used regression analyses to examine the relationship between SAT score gains and several factors of test preparation sessions. Their findings showed that time spent on individual tutoring was positively associated with an increase in the SAT Total score – with each additional hour spent on individual tutoring increasing the final SAT score by 2.34 points (e.g. roughly 13 hours of individual tutoring is associated with an increase of 30 points or an SAT Total score gain of 0.25 SD). In addition, the time students start test preparation was also positively associated with SAT score gains – students who start test preparation before June of their junior year would stand to gain 4.3 points for every preceding

month. As such, a student who would start preparing for the SAT in February of their junior year would score 34.4 points higher than a student who would start preparing in September of their senior year (Appelrouth, et al., 2015).

For the ACT test, internal research also shows that test preparation is only modestly associated with gains in ACT test scores, with higher increases related to length of preparation. Students who took and prepared for a second ACT test gained an average of 1.4 points on their second ACT Composite score (Schiel & Valiga, 2014a). Students who prepared for the second test for over 20 hours attained Composite scores 0.7 point higher, on average, compared to those who reported 3-6 hours of preparation (Schiel & Valiga, 2014b). Short-term preparation activities (e.g. commercial workbooks) were found to be associated with 1.2 to 1.5 point increases in ACT Composite score, while longer-term activities (e.g. high school coursework) were associated with 2.5 to 5.8 point increases in ACT Composite score (ACT, 2005).

In addition to factors related to practice opportunities (i.e., length of test preparation), studies have also looked into the relationship between motivational factors and test performance in students' test preparation and test-taking, such as awareness of their test preparedness or their self-efficacy. Mulvenon, Stegman, and Ritter (2005) evaluated the impact of anxiety, pressure, and self-efficacy on student performance on standardized tests. They found students' perceptions of negative pressure about testing negatively influenced their test performance, while students' perceptions of their ability in math and reading were positively associated with their performance in math and reading assessments. Hong, Sas, and Sas (2006) looked into test preparation awareness between high-achieving and low-achieving high school students for an algebra test. Their findings show that high-achieving students were more aware of their test preparedness in cognitive areas (competence, understanding of the material, and study behavior). Peng, Hong,

and Mason (2014) modeled the relationships among motivational variables in test-taking and test preparation strategies of 10th graders and found students with high self-efficacy applied more effort in test-taking and performed better than students with low self-efficacy. Findings in these studies are consistent with self-efficacy having a positive influence on student achievement (Bandura, 1993).

Although numerous and extensive, existing studies that have looked at effects of test preparation on SAT or ACT test performance have had some limitations. According to the previously mentioned NACAC report (Briggs, 2009), more than 30 studies have been conducted since 1953 evaluating the effect of coaching on the SAT; only two such studies had examined the effect of coaching on the ACT, and few were conducted on students taking the tests since 2000. Many previous studies also used small samples that were not necessarily representative of high school students taking the SAT or ACT (Briggs, 2009), or used research designs that raised concerns about either internal or external validity (Lane et al., 2015). As test preparation becomes widely accessible through different delivery systems, large-scale studies of test preparation efficacy that involve a variety of test preparation activities become more important to understanding the value and impact of test preparation activities on both the ACT and SAT.

In this paper, we examine the impact of participating in test preparation prior to retaking an ACT test. The present study aims to address the call by Briggs (2009) for additional and more robust research on the efficacy of test preparation programs for the ACT. This study uses a fairly large sample of ACT-tested students and a quasi-experimental design that takes advantage of propensity score matching techniques to make causal claims related to the following research questions:

1. Using a pretest-posttest design, do students who participate in test preparation have larger score gains relative to students who did not participate in test preparation? Does the test preparation effect depend on students' pretest scores?
2. Among students who participated in test preparation, is the number of hours spent participating in each of ten test preparation activities related to retest scores?
3. Among students who participated in test preparation, do their own beliefs that they might have been ill-prepared to take the test, regardless of the test preparation activities they engaged in, impact retest scores?

We focus on the effects of test preparation on ACT Composite score as it is most commonly used and referenced by students and admissions offices alike.

Method

Data. The data for this study were taken from a previously conducted study on test preparation (Schiel & Valiga, 2014a; Schiel & Valiga, 2014b). In that study, an online survey was administered to a random sample of 76,000 ACT test takers who had completed the ACT for the first time in April or June 2012 and retested in October 2012 or completed the ACT for the first time in September, October, or December 2012 and retested in April 2013. These students were invited via email to participate in a survey about their test preparation activities. Of these students, 9,654 students responded to the survey (12.7% response rate).

We were interested in students that had retested for the ACT, had not used test preparation aids prior to their first test, but had used test preparation aids prior to their second test. These selection criteria and the use of propensity score matching (PSM; detailed in the Data Analysis section) resulted in a total number of 2,660 students in the analysis sample. Of these, 1,330 students indicated that they did prepare for their second test and another 1,330 students

indicated that they did not prepare for their second test and thus were selected as a comparison group. Student self-reported demographic and background information (e.g., race, parental income) as well as academic performance indicators (e.g., high school GPA) were provided by students at the time of ACT test registration.

Participants. Of the 2,660 students used in the analysis for this study, most were female (62%), in their junior year (85%) of high school, and White (66%) (See Table 4). Approximately 47% of study participants were middle income ($36,000 - $100,000) with fewer percentages of students coming from lower- and higher-income families. The study sample closely resembled the population of 2013 ACT test takers (58% White, 13% African-American, 15% Hispanic/Latino, 4% Asian, and 4% other/multi-race) but differed somewhat on gender (58% female, 42% male; ACT, 2013). Survey respondents also had a slightly higher ACT Composite test score (M = 22.5, SD = 4.7) than the national average (M = 20.9, SD = 4.8).

Measures.

Survey of Test Preparation Activities for the ACT. The ACT test preparation survey consisted of three sections that measured test preparation, how well students thought they were prepared, and amount of time allocated to 10 preparation activities. From this survey, data on test preparation activity, lack of academic preparation for the ACT, and amount of test preparation the student participated in were collected.

Test preparation. Students were asked to indicate whether or not they prepared for the first and/or second ACT test administration. This survey question was written such that we asked students about test preparation activities "outside of normal classroom participation." This was done in an attempt to isolate activities beyond normal classroom instruction.

*Lack of Preparation***.** This measure consisted of students' answers to three survey items: "I have not yet taken the class(es) necessary for doing well on one or more areas of the test"; "Some areas of the test had not been covered at all or had not been covered adequately in my high school classes"; and "I realized I had not done anything to prepare myself for taking this type of test." For each item, students were instructed to answer yes or no. If a student said yes to any one of the three questions, they were coded as perceiving themselves to have a lack of preparation for the ACT test.

Amount of test preparation. The survey also asked students about their exposure to 10 specific test preparation activities. These activities ranged from interaction with online test materials to self-directed workbooks to one-on-one support (see Table 1). Students were asked to indicate the number of hours they spent on each of the activities in preparation for the second test. Using a six-point scale (0 = zero hours spent on activity; 1= 1-5 hours; 2 = 6-10 hours; 3 = 11 or more hours spent on activity), respondents indicated their amount of test preparation activity. Example items include "Worked with a private tutor or consultant" and "Took a commercial test-preparation course(s)."

Table 1. Test Preparation Activities Investigated in the Survey

Test Preparation Activities
Practice Test in ACT's Free *Preparing for the ACT*
ACT Online Prep
Another Web-Based Test-Preparation Program
The Real ACT Prep Guide
Another Test Prep Workbook
Test Prep Workshops/Courses Offered by the High School
Commercial Test Prep Course
Private Tutor or Consultant
One-on-One with a High School Teacher
Other Test Prep Software

ACT Composite Score. The ACT is a curriculum-based battery of four multiple-choice tests of educational achievement—English, mathematics, reading, and science—and an optional writing test. The ACT, typically taken in the 11th or 12th grade, measures students' academic readiness for college in key content areas. The ACT Composite score is the arithmetic mean of the four academic test scores rounded to the nearest whole number and reported on a scaled score from 1 to 36.

Statistical Controls. In order to better isolate the effect of test preparation, we included additional student characteristics in our analysis. Each student characteristic was included because of the theoretical relationship it has with ACT performance (Radunzel & Noble, 2012; Sanchez, 2013). In this study, we statistically controlled for the impact of whether students commonly experience test anxiety; prior ACT Composite score; whether the student took trigonometry in high school; whether the student took physics in high school; whether advanced placement courses in mathematics, science, and English were taken in high school; high school GPA; whether students felt they needed assistance in core content; whether high school attended was public or private; whether a fee waiver was used to take the ACT; highest parent education; student's grade level; student's gender and race; and parents' income. Most of these student academic and demographic characteristics were collected at the time of registration while information about test anxiety was collected from the survey.

Data Analysis.

Propensity Score Matching (PSM). In studying the effect of test preparation, the interest is in estimating the causal effects of engaging in test preparation while controlling for confounding variables, such as parents' income and prior student achievement. However, random assignment in educational settings raises ethical and logistical concerns (Walker,

Hoggart, & Hamilton, 2008). Propensity score matching presents an alternative to randomization that models the assignment of students to the treatment conditions; in this case, the assignment of students to either receive test preparation or not, using an estimated propensity score (Rosenbaum & Rubin, 1983; see Nagengast, Marsh, & Hau, 2013, for a discussion of how propensity score matching is a better alternative to covariate adjustment). This estimated propensity score measures the probability of being assigned to the test preparation group given the covariates entered into the propensity score matching analysis. Through this process, we created control and treatment groups that are comparable on the propensity score which incorporates all included covariates.

In this study, we used the SAS software macro "OneToManyMTCH" (Parsons, 2004). This macro uses a greedy, nearest-neighbor, matching algorithm to identify a matched sample for the control group, using an eight to one digit match. We used one-to-one matching between treatment and control groups. The calculation of the propensity score was based on a logistic regression model that included 13 related covariates. Of those, 10 covariates were retained using stepwise selection and forced inclusion (see Table 2).[1]

[1] Listwise deletion was used to handle missing data.

Table 2. Student Characteristics Considered and Retained in the Propensity Score Model

Retained in Propensity Score Model	Student Characteristic
✓	Prior ACT Composite score
✓	High school GPA
✓	Interaction between prior ACT Composite and high school GPA
✓	Gender (Female)
✓	Minority membership (African American, Hispanic, American Indian, or Native Hawaiian)
✓	Family Income
✓	Highest parental Education
✓	Self-reported experience of anxiety during testing
✓	Fee waiver status indicator
	Taken math coursework beyond Algebra II
	Taken Science coursework beyond Chemistry
	Taken advanced mathematics coursework
✓	Taken advanced science coursework
	Taken advanced English coursework

Note: To qualify for an ACT fee waiver, a student must meet certain criteria. For example, they must be currently enrolled in the 11th or 12th grade; a US citizen or testing in the US, US territories, or Puerto Rico; and must meet one or more indicators of economic need listed on the ACT Fee Waiver form. For the full criteria for eligibility, see https://www.act.org/content/dam/act/unsecured/documents/FeeWaiver.pdf. These students are eligible to receive a fee waiver to cover the cost of taking the ACT.

Linear regression. Two linear regression models were estimated to address the research questions. The first regression analysis sought to determine if students who participate in test preparation activities have a higher ACT Composite retest score than those who did not prepare for the test, controlling for academic and demographic characteristics. These academic and demographic characteristics were included in the porosity score model and balanced by matching. We include them here in an effort to better capture the relationship between

potentially confounding variables, activities of interest, and our outcome of ACT test score. Interactions between preparation status and demographic information, and between the students' first ACT score and test preparation status, were tested. This allowed us to test whether there is a differential impact of preparation for student subgroups (i.e., race, gender, and parent income) and across the range of prior ACT Composite score.

We sought to investigate whether the amount of time allocated to various test preparation activities and/or the student's self-perception of test preparedness might have contributed to gains among students who participated in test preparation. Therefore, the second regression model, including only those students who participated in test preparation, sought to determine whether the amount of time allocated to a series of preparation activities and/or the student's self-perception of test preparedness impacted students' ACT Composite retest score. In addition to the variables entered into the model for the first research question, we included in this second regression model the number of hours spent on 10 test preparation activities and an indicator for whether students thought they were ill prepared to take the second test, regardless of the preparation activities engaged in. The 10 test preparation activities were treated as continuous measures in the regression model since dummy coding each activity for four categories would have resulted in estimating many coefficients relative to the reduced sample size used for research question two. For those test preparation activities found to be statistically significant in the linear model, pair-wise comparisons across the bins were conducted.

Results

Propensity Score Matching. Table 3 illustrates the balancing that was achieved in the control and treatment groups by using propensity score matching. This table shows the standardized mean difference between groups on each of the covariates investigated. Prior to matching, large group differences were observed for minority group membership, parental income, highest parental education, test anxiety, using a fee waiver, and high school GPA. After matching, the standardized mean differences between the treatment and control groups for all covariates were well below 10%.

For continuous variables the standardized mean difference is $d = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$, where X_1 and X_2 denoted the sample means, and s_1^2 and s_2^2 denote the sample variances. For categorical variables, the standardized mean difference is $d = \frac{(\widehat{P_1} - \widehat{P_2})}{\sqrt{\frac{[\widehat{P_1}(1-\widehat{P_1}) + \widehat{P_2}(1-\widehat{P_2})]}{2}}}$, where $\widehat{P_1}$ and $\widehat{P_2}$ denote the proportion or mean of a binary variable in the treatment and control groups. An extension of this binary case is available which makes use of multivariate Mahalanobis distance to handle multinomial variables (Yang & Dalton, 2012).

In addition to this empirical estimation of balance, the distributions of each covariate were examined. For illustrative purposes, the distribution of the propensity score for the treatment and control groups after matching are displayed in Figure 1. The distributions for both groups are very similar. Similar trends were found for other continuous variables and similar proportions were found across groups for categorical variables.

Table 3. Propensity Score Matching Balancing Results

	Pre-Match Treatment Mean	Pre-Match Control Mean	Pre-Match Difference	Pre-Match Standardized Difference (%)[a]	Post-Match Treatment Mean	Post-Match Control Mean	Post-Match Difference	Post-Match Standardized Difference (%)
Minority Status	0.17	0.26	-0.10	-23.49	0.21	0.22	-0.01	-1.84
Low-Income	0.17	0.28	-0.11	-25.58	0.22	0.24	-0.02	-4.64
Female	0.65	0.62	0.03	6.73	0.64	0.60	0.03	6.97
Highest Parental Education	5.59	4.96	0.63	31.72	5.29	5.25	0.03	1.74
Test Anxiety	0.30	0.24	0.06	13.53	0.27	0.27	0.00	0.85
Fee Waiver	0.16	0.29	-0.13	-31.71	0.20	0.22	-0.02	-4.08
Taken Algebra 2 or Beyond	0.56	0.58	-0.02	-4.13	0.55	0.56	-0.01	-2.57
Taken Physics or Beyond	0.28	0.28	0.01	1.14	0.27	0.25	0.02	4.96
Taken Advanced Mathematics	0.49	0.49	0.01	1.59	0.50	0.49	0.01	1.80
Taken Advanced Science	0.48	0.46	0.02	3.13	0.47	0.47	0.00	0.30
Taken Advanced English	0.56	0.54	0.01	2.80	0.55	0.55	0.00	0.91
First ACT Composite Score	22.63	22.75	-0.12	-2.44	22.53	22.46	0.07	1.44
High School GPA	3.60	3.52	0.08	16.21	3.56	3.56	0.01	1.37

[a] The standardized difference percentage presented in this table in $d * 100$. Where d is the standardized mean difference.

Note: Standardized difference percentages are indexes that measure the effect size between two groups. Among propensity score literature a common criteria for concluding similarity of groups is a standardized difference percentage of less than 20%. For a more detailed treatment on standardized differences the reader is directed to Yang & Dalton (2012).

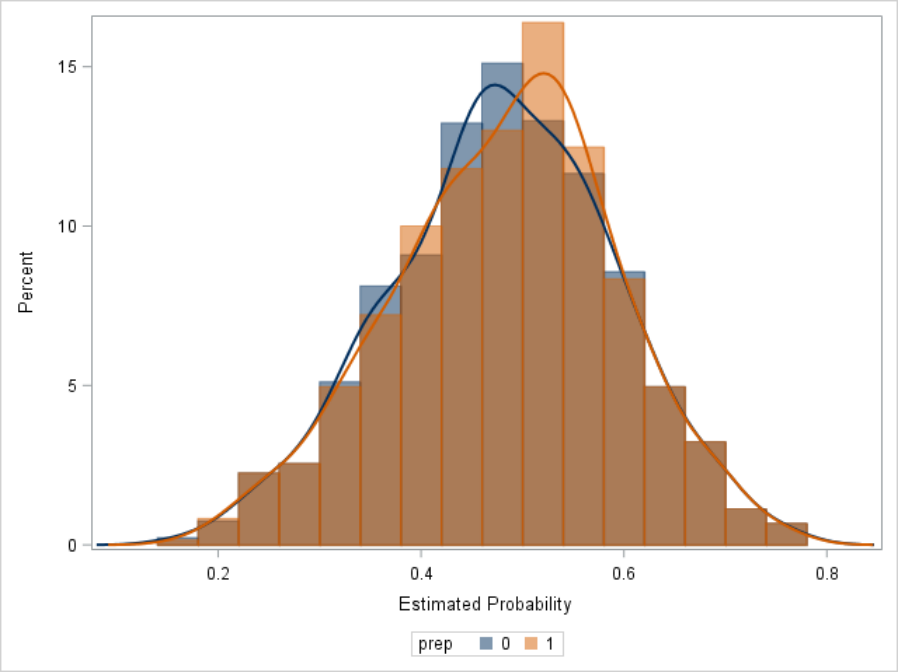


Figure 1. Distribution of propensity score for the treatment (1) and control (0) groups

Descriptive statistics. After identifying the matching control group via PSM, 2,660 students were used as the analytic sample to answer the first research question; the 1,330 students who prepared for the second test were used to answer the second research question. These 1,330 students were the treatment condition for the first research question. Table 4 shows means and standard deviations for the variables of interest. Students who participated in test preparation before the retest had a slightly higher ACT Composite score relative to students who did not. Covariates are similar across the two groups.

Table 4. Descriptive Statistics, by Test Preparation Status and Overall

	No Preparation (n=1,330)		Preparation (n=1,330)		Total (n = 2,660)		Population (N= 76,000)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Test Preparation	0.00	0.00	1.00	0.00	0.50	0.50	--	--
First ACT Composite Score	22.46	4.85	22.53	4.55	22.50	4.70	20.92	4.78
Second ACT Composite Score[a]	23.33	4.98	24.12	4.83	23.72	4.92	22.02	4.99
High School GPA	3.56	0.46	3.56	0.47	3.56	0.47	3.40	0.53
Taken Algebra 2 or Beyond	0.58	0.49	0.56	0.50	0.57	0.50	0.50	0.50
Taken Physics or Beyond	0.30	0.46	0.31	0.46	0.31	0.46	0.28	0.45
Taken Advanced Mathematics	0.49	0.50	0.50	0.50	0.49	0.50	0.35	0.47
Taken Advanced Science	0.47	0.50	0.47	0.50	0.47	0.50	0.34	0.47
Taken Advanced English	0.55	0.50	0.55	0.50	0.55	0.50	0.41	0.49
Skills Needed	0.53	0.50	0.54	0.50	0.54	0.50	0.62	0.48
Public School	0.86	0.34	0.82	0.39	0.84	0.37	0.82	0.38
Parents' Income								
Low Income (<$36K)	0.24		0.22		0.23		0.30	
Middle Income ($36K-$100K)	0.47		0.48		0.47		0.43	
High Income (>$100K)	0.29		0.30		0.30		0.27	
Parent's Highest Education Level								
High School or Less	0.16		0.15		0.15		0.19	
Business/Technical School	0.04		0.04		0.04		0.04	
Some College, No Degree	0.10		0.11		0.11		0.12	
Associates Degree	0.10		0.10		0.10		0.10	
Bachelor's Degree	0.31		0.31		0.31		0.29	
Graduate Study	0.21		0.21		0.21		0.17	
Doctorate or Professional Degree	0.08		0.08		0.08		0.08	
Race								
White	0.65		0.67		0.66		0.57	
Minority	0.22		0.21		0.21		0.30	
Asian	0.06		0.06		0.06		0.05	
Two or More Races	0.04		0.03		0.03		0.03	
No Response	0.04		0.03		0.03		0.04	
Female	0.60		0.64		0.62		0.58	
Student Grade Level								
Grade 10	0.07		0.07		0.07		0.04	
Grade 11	0.84		0.86		0.85		0.81	
Grade 12	0.08		0.06		0.07		0.14	
Test Anxiety	0.24	0.43	0.24	0.43	0.24	0.43	--	--
Test Preparation Activities								
Practice Test in ACT's Free *Preparing for the ACT*	--	--	0.74	0.96	0.37	0.77	--	--

Table 4 (continued)

	No Preparation (n=1,330)		Preparation (n=1,330)		Total (n = 2,660)		Population (N= 76,000)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
ACT Online Prep	--	--	0.40	0.75	0.20	0.57	--	--
Another Web-Based Test-Preparation Program	--	--	0.39	0.80	0.19	0.60	--	--
The Real ACT Prep Guide	--	--	0.75	1.04	0.38	0.82	--	--
Another Test Prep Workbook	--	--	0.75	1.07	0.38	0.84	--	--
Test Prep Workshops/Courses Offered by the High School	--	--	0.69	1.10	0.35	0.85	--	--
Commercial Test Prep Course	--	--	0.31	0.80	0.15	0.58	--	--
Private Tutor or Consultant	--	--	0.41	0.89	0.20	0.66	--	--
One-on-One with a High School Teacher	--	--	0.30	0.67	0.15	0.50	--	--
Other Test Prep Software	--	--	0.23	0.62	0.11	0.46	--	--
Perception of Inadequate Preparation	--	--	0.28	0.45	0.32	0.47	--	--

Note: Test preparation activity items were only asked to students who indicated they prepared for the second test.
[a]Approximately 12% of second ACT scores were missing for the population.

RQ1: The Impact of Test Preparation on Retest Scores. Table 5 presents the full and reduced regression models used to determine if there was a statistically significant difference in the ACT Composite retest score between those students who prepared for the second test and those who did not. In the full model, results showed that there was no statistically significant interaction between first ACT score and whether the student prepared for the ACT. This means there were no differential effect of test preparation on students' retest score by how well students did on the first test. There were also no significant interaction terms between preparation and race/ethnicity or gender.

After controlling for student demographic information and academic performance, a statistically significant difference in the retest ACT Composite score for the two groups was found in both the full and reduced models. Using the reduced model, the adjusted ACT

Composite mean for the test preparation and non-test preparation groups were 24.33 and 23.63, respectively. Test preparation was the second strongest predictor of ACT Composite scores. The reduced model explained 87.5% of the variation in ACT Composite scores.[2,3]

Table 5. Multiple Linear Regression Full and Reduced Models Predicting ACT Composite Scores from Test Preparation Status (n= 2,660)

	Full Model				Reduced Model			
	b	SE	β	t	b	SE	β	t
(Constant)	23.63	0.04		584.62	23.62	0.04		582.26*
Test Preparation	0.70	0.07	0.07	10.30*	0.70	0.07	0.07	10.22*
First ACT Composite Score	0.86	0.01	0.82	76.35*	0.87	0.01	0.83	80.56*
Test Preparation X First ACT Composite Score	0.00	0.02	0.00	0.27				
Taken Algebra 2 or Beyond	0.23	0.08	0.02	2.88*	0.24	0.08	0.02	3.07*
Taken Physics or Beyond	0.04	0.08	0.00	0.55	-	-	-	-
Taken Advanced Mathematics	0.19	0.10	0.02	2.022*	0.26	0.08	0.03	3.23*
Taken Advanced Science	0.12	0.09	0.01	1.25	-	-	-	-
Taken Advanced English	0.06	0.09	0.01	0.71	-	-	-	-
High School GPA	0.76	0.11	0.07	6.96*	0.77	0.11	0.07	7.08*
High School GPA X First ACT Composite Score	0.08	0.02	0.04	4.632*	0.08	0.02	0.04	4.83*
Skills Needed	-0.22	0.08	-0.02	-2.94*	-0.23	0.08	-0.02	-2.98*
Parents' Income[a]								
Low Income (<$36K)	-0.02	0.11	0.00	-0.22	-0.14	0.09	-0.01	-1.51
High Income (>$100K)	0.25	0.09	0.02	2.89*	0.34	0.08	0.03	4.05*
Fee Waiver	-0.20	0.11	-0.02	-1.85	-	-	-	-
Public School	-0.29	0.10	-0.02	-3.02*	-0.31	0.10	-0.02	-3.23*
Parents' Education[b]								
Business/Technical School	0.23	0.19	0.01	1.20	-	-	-	-
Some College, No Degree	-0.04	0.14	0.00	-0.31	-	-	-	-
Associates Degree	0.13	0.15	0.01	0.89	-	-	-	-
Bachelor's Degree	0.10	0.12	0.01	0.78	-	-	-	-
Graduate Study	0.10	0.14	0.01	0.73	-	-	-	-
Doctorate or Professional Degree	0.65	0.17	0.04	3.914*	-	-	-	-
Students' Race[c]								
Minority	-0.38	0.10	-0.03	-3.77*	-0.43	0.10	-0.04	-4.49*
Asian	-0.20	0.15	-0.01	-1.30	-0.21	0.15	-0.01	-1.41

[2] The proportion of variance explained is predominantly driven by prior test score. The proportion of variance explained without test prep is 0.866. With test prep included in the model the proportion of variance explained increases to 0.871.

[3] To check for the effect shared variance due to the matching performed, cluster robust standard errors that accounted for the paired clusters was implemented in a separate model. Accounting for this shared variance did not dramatically impact estimates of model. For example, the maximum difference in estimates for the reduced model in research question 1 was 0.01223.

Table 5 (continued)

	Full Model				Reduced Model			
	b	SE	β	t	b	SE	β	t
Two or More Races	-0.51	0.19	-0.02	-2.68*	-0.50	0.19	-0.02	-2.62*
No Response	0.08	0.20	0.00	0.41	0.01	0.19	0.00	0.03
Female	-0.20	0.07	0.02	2.72*	-0.21	0.07	0.02	2.97*
Student Grade Level	-0.31	0.10	-0.02	-3.18*	-0.30	0.09	-0.02	-3.09*
Test Anxiety	-0.05	0.08	-0.01	-0.64	-	-	-	-
Interactions								
Test Preparation X Minority	-0.29	0.19	-0.01	-1.52	-	-	-	-
Test Preparation X Asian	-0.18	0.30	0.00	-0.59	-	-	-	-
Test Preparation X Two or More Races	-0.53	0.38	-0.01	-1.40	-	-	-	-
Test Preparation X No Response	1.09	0.39	0.02	2.780*	-	-	-	-
Test Preparation X Female	-0.06	0.14	0.00	0.41	-	-	-	-
Test Preparation X Low Income	0.09	0.19	0.00	0.46	-	-	-	-
Test Preparation X High Income	0.43	0.17	0.02	2.625*	-	-	-	-

Note: Variables are grand mean centered.
- dropped in the reduced model.
* $p < .05$
[a] Moderate income is the reference category
[b] High School Diploma or less is the reference category
[c] White is the reference category

RQ2 and RQ3: Impact of Amount of Time in 10 Preparation Activities and Ill-Prepared Perceptions on Retest Scores. To facilitate a better understanding of the context for the regression analysis that investigated which test preparation activities improved students' ACT scores, we first provide descriptive information on the number of activities engaged in by students' demographic characteristics (Table 6).

More than 65% of the students reported participating in one to three test preparation activities, and an additional 27% of students reported participating in 4 to 7 activities. Very few students (7%) reported participating in eight or more activities. A trend emerged between key student demographic characteristics and the number of activities students participated in.

Specifically, the percentage of minority students, percentage of fee waiver students, and percentage of students with low family income increases as the number of test preparation activities engaged in increases.

Table 6. Student Demographic Information, by Self-Reported Number of Test Preparation Activities

Number of activities	Frequency	%Female	%Fee Waiver	Average HSGPA	Race/Ethnicity					Family Income		
					%White	%Minority	%Asian	Two or more Races	No Response	Low Income	Middle Income	High Income
Unknown	10	40.00	30.00	3.61	60.00	10.00	20.00	10.00	0.00	50.00	40.00	10.00
1-3	873	62.20	17.41	3.60	72.16	16.72	5.96	3.55	1.60	18.33	49.37	32.30
4-7	359	68.52	22.01	3.50	57.94	27.86	6.69	2.51	5.01	27.30	46.52	26.18
8+	88	62.50	35.23	3.39	54.55	32.95	4.55	3.41	4.55	35.23	40.91	23.86

A second linear regression analysis, including only those students who participated in test preparation, was conducted to determine whether the amount of time allocated to 10 test preparation activities had a positive impact on students' retest score (See Table 7). In addition, we looked to see if students' perception of inadequate preparation impacts retest score given the importance of this variable as shown in the literature.

Table 7. Multiple Linear Regression Full and Reduced Model Predicting ACT Composite Scores From Amount of Time in 10 Test Preparation Activities (n= 1,330)

	Full Model				Reduced Model			
	b	SE	β	t	b	SE	β	t
(Constant)	23.98	0.06		412.72*	23.99	0.06		413.20
First ACT Composite Score	0.88	0.02	0.83	54.40*	0.88	0.02	0.82	55.25*
Taken Algebra 2 or Beyond	0.14	0.11	0.01	1.23*	.154	.110	.016	1.398
Taken Advanced Mathematics	0.37	0.12	0.04	3.20*	0.36	0.12	0.04	3.12*
High School GPA	0.47	0.15	0.05	3.10*	0.48	0.15	0.05	3.17*
High School GPA X First ACT Composite Score	0.07	0.02	0.03	2.73*	0.07	0.03	0.03	2.69*
Skills Needed	-0.32	0.11	-0.03	-2.96*	-0.32	0.11	-0.03	-2.92*
Parents' Income[a]								
Low Income (<$36K)	-0.08	0.14	-0.01	-0.61	-0.08	0.14	-0.01	-0.58
High Income (>$100K)	0.41	0.12	0.04	3.49*	0.44	0.12	0.04	3.68*
Public School	-0.34	0.13	-0.03	-2.59*	-0.37	0.13	-0.03	-2.85*
Students' Race[b]								
Minority	-0.58	0.14	-0.05	-4.21*	-0.57	0.14	-0.05	-4.13*
Asian	-0.25	0.21	-0.01	-1.18	-0.23	0.21	-0.01	-1.09
Two or More Races	-0.80	0.28	-0.03	-2.88*	-0.78	0.28	-0.03	-2.82*
Race Category No Response	0.46	0.31	0.02	1.50	0.59	0.31	0.02	1.92
Female	-0.20	0.10	0.02	1.97*	-0.19	0.10	0.02	1.80
Student Grade Level	-0.37	0.14	-0.03	-2.63*	-0.37	0.14	-0.03	-2.62*
Test Preparation Activities								
Practice Test in ACT's Free *Preparing for the ACT*	0.01	0.06	0.00	0.21	-	-	-	-
ACT Online Prep	-0.02	0.07	0.00	-0.27	-	-	-	-
Another Web-Based Test-Preparation Program	0.04	0.07	0.01	0.49	-	-	-	-
The Real ACT Prep Guide	0.10	0.05	0.02	1.98	-	-	-	-
Another Test Prep Workbook	0.10	0.05	0.02	1.95	-	-	-	-
Test Prep Workshops/Courses Offered by the High School	0.08	0.05	0.02	1.76	-	-	-	-
Commercial Test Prep Course	0.01	0.07	0.00	0.14	-	-	-	-
Private Tutor or Consultant	0.16	0.06	0.03	2.70*	0.20	0.06	0.04	3.53*
One-on-One with a High School Teacher	-0.12	0.09	-0.02	-1.35	-	-	-	-
Other Test Prep Software	-0.04	0.09	-0.01	-0.46	-	-	-	-
Perception of Inadequate Preparation	-0.28	0.11	-0.03	-2.48*	-.322	.113	-.030	-2.85*

* $p < .05$

Note: Variables are grand mean centered. The reduced model for RQ1 was used as a starting point for answering RQ2.

- dropped in the reduced model.

[a] Moderate income is the reference category

[b] White is the reference category

Among the 10 test preparation activates investigated, only the number of hours working with a private tutor or consultant had a significant impact on retest scores. Using the reduced model, those who did not engage in this test preparation activity had an adjusted mean ACT Composite score of 24.31 while those who had 11 hours or more of this test preparation activity had an adjusted mean score of 24.91. Pair-wise comparisons of the four bins that measured the amount of time allocated to a tutor or consultant showed that students who participated in 11 hours or more with a tutor or consultant had a statistically higher ACT Composite re-test score than those who did not spend any hours on the activity. No other comparisons were found statistically different.

From the same model, perceived inadequate preparation was also a statistically significant predictor of students' retest scores.[4] Students who self-reported feeling inadequately prepared for the second test had an ACT Composite score that was 0.322 scale score point below that of students who felt adequately prepared. The model explained 86.6% of the variation in ACT Composite scores.[5]

Discussion

We examined the impact of using test preparation prior to taking the ACT a second time. Emphasis was placed on those students who did not use test preparation materials before taking the ACT the first time. This line of research is critical given the potential inequalities in using test preparation products for college entrance examinations. There is no question that the use of

[4] Among students who reported using one of the 10 activities in question, the correlation between number of activities undertaken and self-reported feeling of being underprepared to take the ACT was -0.032 (p=0.2365). Due to the low correlation, it is unlikely that the number of test preparation activities engaged in mediates the effect of under-preparation on ACT Composite score.

[5] The proportion of variance explained is predominantly driven by prior test score. The proportion of variance explained without test prep is 0.864. With test prep included in the model, the proportion of variance explained increases to 0.865.

standardized tests to aid college admissions decisions creates a very high stakes situation for graduating high school students. It is because of this fact that the test preparation industry has seen such strong growth in the United States (Barnes Reports, 2017). It is also the reason why a NACAC discussion paper (Briggs, 2009) has called for greater rigorous research to be conducted on the effects of test preparation for the ACT. The present research study contributes to this needed body of evidence.

This study uncovered a statistically significant impact of test preparation on ACT retest scores. On average, those who participated in test preparation prior to taking the ACT a second time had an ACT Composite score 0.71 scale score points above those who did not. To put this score difference in context, the standard deviation for the ACT Composite score in the 2016 tested population was 5.6 (ACT, 2017a). This is consistent with prior test preparation studies that have found a positive effect of test preparation.

Additionally, we found that the effects of test preparation did not differ by how well students had done on a prior administration of the test. Most importantly, we also found that the effects of test preparation did not differ by race/ethnicity, gender, or family income. This suggests that test preparation is equally effective for minority students as for majority students and equally effective for females as for males. The findings pertaining to family income suggest that financial means, and any associated access to test preparation activities, overall, did not impact ACT retest scores.

However, when we investigated specific types of test preparation activities, we found working with a private tutor or consultant has a statistically significant impact on ACT retest scores. This type of one-on-one, personalized activity was found to have a notable effect on

retest scores. This could have important practical significance in the admissions process for schools who utilize cutoffs when evaluating prospective students (Briggs, 2009).

At first blush, this finding may seem to contradict the finding that family income did not have a differential effect on test preparation efficacy, as families with higher incomes may be more likely to be able to afford private tutoring.[6] The two findings can be reconciled by a nuanced understanding of what the models used are testing and the sample differences. In the first model, we are looking at the effect of family income on test preparation efficacy regardless of the type of test preparation used or the duration of time spent on those activities on students who did or did not use test preparation. In the second model, we are examining the impact of hours using specific test preparation activities while controlling for the effects of family income only on students who did a test preparation activity. This finding therefore suggests that regardless of family income, private tutoring has a positive effect on ACT retest scores.

In future research, we believe it will be beneficial to further tease out the impacts of participation in specific test preparation from the intensity of participating in that activity while considering access issues (e.g. financial status). Due to the relatively small number of low and middle income students that reported working with private tutors, we were unable to fully examine the interaction between family income and the number of hours working with a private tutor or consultant. This line of inquiry is needed in future work to further understand how access to and use of a private tutor might lead to inequality issues for low income students. Further, we recommend that future research measure the amount of time allocated to different test preparation activities continuously.

[6] In the current sample, the Spearman correlation between family income and number of hour spent working with a private tutor or consultant was 0.078 ($p<0.0001$).

A final notable result of this study was that students' self-perception of preparation for the ACT test was a significant predictor of retest performance. This study highlights students' own understanding of their academic standings in relation to content mastery and potential performance on tests. It would be beneficial to utilize this type of self-knowledge to help connect students with study aids that can help them raise their self-perceived preparation by increasing their actual preparation for the ACT test.

While this study took an important step forward in providing causal evidence for the impact test preparation has on standardized tests such as the ACT, we have to make note of the fact that study participants had an average ACT Composite test score that was 1.6 scale score points above the national average. This is worth considering because when we speak of students who use test preparation materials, we can imagine different types of test preparation users including lower achieving students who need supplementary instruction to strengthen skills and higher achieving students who are using test preparation materials to review content and increase their chances of getting an exceptional score.

In sum, this study contributes to the test preparation literature by evaluating the effects of test preparation on the ACT, as Briggs (2009) has noted, is sorely needed. Moreover, unlike the typical study of test preparation, this study employed a rigorous research methodology that allowed for a causal examination of efficacy.

References

ACT. (2005). *Issues in college readiness: What kind of test preparation is best? College Readiness, 1-3.* Iowa City, IA: ACT. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/best_testprep.pdf

ACT. (2017a). *ACT profile report - National: Graduating class 2016.* Iowa City, IA: ACT. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/P_99_999999_N_S_N00_ACT-GCPR_National.pdf

ACT. (2017b). *ACT technical report.* Iowa City, IA: ACT. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf

Andrews, K. M., & Ziomek, R. L. (1998). *Score gains on retesting with the ACT assessment.* Iowa City, IA: ACT.

Appelrouth, J. I., Zabrucky, K. M., & Moore, D. (2015). Preparing students for college admissions tests. *Assessment in Education: Principles, Policy & Practice, 24*(1), 78-95.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117–148.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. C. (1983). Effects of coaching programs on achievement test performance. *Review of Educational Research, 53*(4), 571-585.

Barnes Reports. (2017). *Exam preparation & tutoring industry* (NAICS 611691). World Industry & Market Outlook Report. Bonita Springs, FL: Barnes Reports.

Briggs, D. C. (2001). The effect of admissions test preparation: evidence from NELS:88. *Chance 14*(1), 10-18.

Briggs, D. C. (2009). *Preparation for college admission exams* (2009 NACAC Discussion Paper). Alexandria, VA: National Association for College Admission Counseling.

Carnevale, A. P., & Rose, S. J. (2003). *Socioeconomic status, race/ethnicity, and selective college admissions (*A Century Foundation Paper). New York, NY: The Century Foundation.

Clinedinst, Melissa E. (2014). *2014 state of college admission.* Washington, DC: National Association for College Admission Counseling.

Doyle, W. R. (2006). Adoption of merit-based student grant programs: An event history analysis. *Educational Evaluation and Policy Analysis, 28*(3), 259-285.

Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology, 92*(2), 373-385.

Hong, E., Sas, M., & Sas, J. C. (2006). Test-taking strategies of high and low mathematics achievers. *The Journal of Educational Research, 99*(3), 144-155.

Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2015). *Handbook of test development*. New York, NY: Routledge.

Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for educational and testing practice. *Educational Psychologist, 17*(2), 67–91.

Montgomery, P., & Lilly, J. (2012). Systematic reviews of the effects of preparatory courses on university entrance examinations in high school-age students. *International Journal of Social Welfare, 21*(1), 3-12.

Mulvenon, S. W., Stegman, C. E., & Ritter, G. (2005). Test anxiety: A multifaceted study on the perceptions of teachers, principals, counselors, students, and parents. *International Journal of Testing, 5*(1), 37-61.

Nagengast, B., Marsh, H. W., & Hau, K. T. (2013). Effects of single-sex schooling in the final years of high school: A comparison of analysis of covariance and propensity score matching. *Sex roles, 69*(7-8), 404-422.

Parsons, L. (2004, May). *Performing a 1:N case-control match on propensity score.* Paper presented at the meeting of the SAS User Group International, Montreal, Canada.

Peng, Y., Hong, E., & Mason, E. (2014). Motivational and cognitive test-taking strategies and their influence on test performance in mathematics. *Educational Research and Evaluation, 20*(5), 366-385.

Powers, D. E. (1993). Coaching for the SAT: A summary of the summaries and an update. *Educational Measurement: Issues and Practice, 12*(2), 24-30.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Radunzel, J., & Noble, J. (2012). *Predicting long-term college success through degree completion using ACT [R] Composite score, ACT Benchmarks, and high school grade point average.* Iowa City, IA: ACT.

Sanchez, E. I. (2013). *Differential effects of using ACT® College Readiness Assessment scores and high school GPA to predict first-year college GPA among racial/ethnic, gender, and income groups.* Iowa City, IA: ACT.

Schiel, J., & Valiga, M. (2014a). *How ACT scores change with test preparation.* Iowa City, IA: ACT.

Schiel, J., & Valiga, M. (2014b). *How the amount of time spent on test preparation relates to ACT scores.* Iowa City, IA: ACT.

Walker, R., Hoggart, L., & Hamilton, G. (2008). Random assignment and informed consent: A case study of multiple perspectives. *American Journal of Evaluation, 29*(2), 156-174.

Yang, D., & Dalton, J. E. (2012, April). *A unified approach to measuring the effect size between two groups using SAS®*. Paper presented at the SAS Global Forum, Orlando, FL. Retrieved from http://support.sas.com/resources/papers/proceedings12/335-2012.pdf