Understanding the Benefits of the *Official ACT Prep Guide* on ACT Composite Score

Edgar I. Sanchez

ACT.org/research

Author

Edgar I. Sanchez

Edgar Sanchez, a senior research scientist in the Validity and Efficacy Research department at ACT, works on predictive modeling of student educational outcomes. He is currently focusing on the efficacy of test preparation programs.

Acknowledgements

The author would like to thank Raeal Moore, Joann Moore, and Jeff Allen for their review and feedback on earlier versions of this report.

Conclusions

In this study, I examined purchase of the *Official ACT Prep Guide*, an ACT-authored workbook meant to prepare students to take the ACT®. I estimated the effects of purchase on ACT scores using a variety of quasi-experimental design (QED) methods, including propensity score matching (PSM), coarsened-exact matching (CEM), Mahalanobis distance matching (MDM), and inverse-probability of treatment weighting (IPW). I evaluated each of the four QEDs and interpreted the estimates of the effect of purchasing the *Guide*. All four methods were successful in attaining good covariate balance for both first-time and repeat tested students; however, the matching methods utilized a very low percentage of the total sample. This was most dramatically seen for the CEM method.

This study found a positive effect for purchase of the *Guide* on ACT Composite score for both first-time and repeat tested students. This effect was larger for repeat tested students than first-time tested students. This study further found that the magnitude of this effect differed by student subgroups. For first-time and repeat tested students, the *Guide* had a larger effect for African American and Hispanic students than for White students. Furthermore, for both first-time and repeat tested students, low-income students saw a larger effect of purchasing the *Guide* than higher income students.

So What?

Given the importance of college entrance exams like the ACT, it is vital that we understand the factors at play when students prepare to take the ACT. Preparing to take the test can help students demonstrate their true ability. This study demonstrates the positive effect of the *Guide*.

This study also demonstrated that some quasi-experimental methods worked better than others and that by averaging the effect across methods we can attain an overall estimate of the effect of the *Guide*.

Now What?

With these research-backed findings, we can help connect students, particularly students from underserved backgrounds, with a cost-effective way to improve their ACT scores.

Contents

Introduction

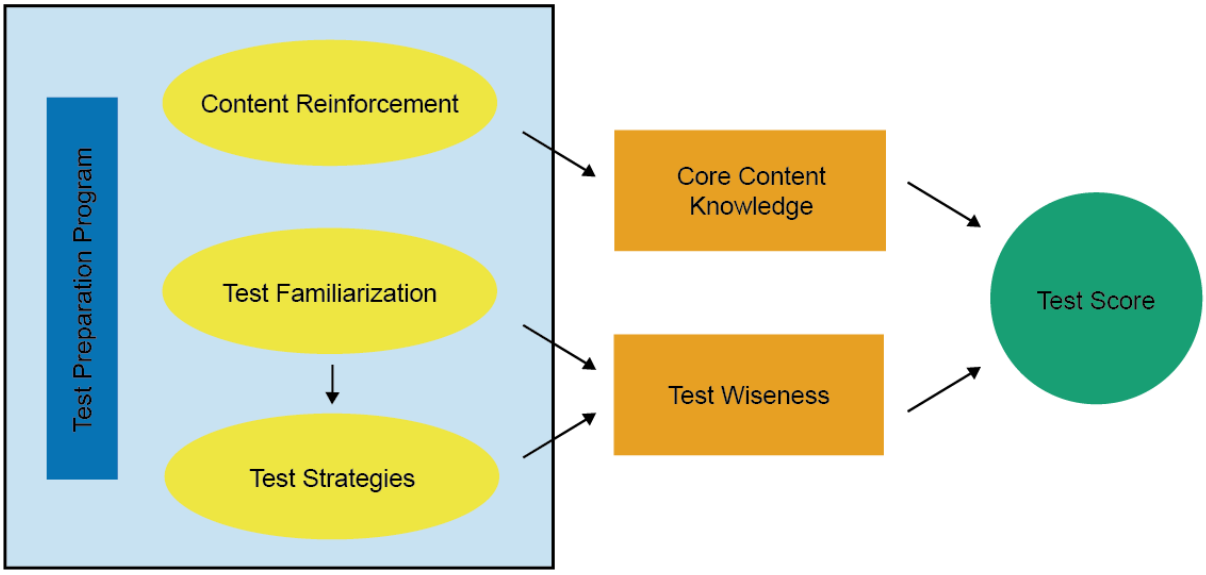
The ACT® test is oriented toward the general content areas of high school and college instructional programs and can be used to assess what students have learned in high school in their core coursework. Increased high school coursework is associated with higher ACT scores (ACT, 2019); and students who take four or more years of English and three or more years each of high school mathematics, social science, and natural science courses have notably higher ACT subject and Composite scores (ACT, 2006). Further, high school GPA and coursework taken accounts for between 64% and 77% of total variance in predicting ACT subject and Composite scores (McNeish, Radunzel, & Sanchez, 2015).[1] ACT scores are indicators of academic preparation that are important for evaluating prospective postsecondary applicants and identifying students in need of extra academic support.

Given the importance of standardized tests in the college entrance process, it is understandable why so many students invest time preparing for the tests. In fact, in the 2018-2019 academic year, almost 64% of ACT test-takers reported using some type of test preparation material.[2] The effective use of test preparation materials can have several beneficial outcomes for students.

Components of Test Preparation

Three areas of emphasis for test preparation include content reinforcement, test familiarization, and testing strategies (see Figure 1). Content reinforcement functions to refresh students on content they should be learning in school and offers opportunities to skill up in preparation for the exam. For students who are unfamiliar with the test, test familiarization can serve to prepare for the format of the test as well as an opportunity to practice taking the exam, thereby reducing construct irrelevant variance in test scores. Testing strategies include things like wrong answer elimination strategies and time management. The joint goals of test familiarization and testing strategies are to improve test wiseness, which is a “subject’s capacity to utilize the characteristics and formats of the test and/or the test taking situation to receive a high score” (Millman, Bishop, & Ebel, 1965, p. 707). The skill of test wiseness is independent of core content knowledge.

Figure 1. Test Preparation Conceptual Model



Test wiseness can then have a direct effect on the assessment score and reduce construct irrelevant variance in the score. For example, one study found that students who spent more hours preparing for the ACT reported less anxiety during testing (Steedle, 2018). Since non-achievement factors such as test wiseness impact student performance, test developers are left with a couple of options. First, they can attempt to minimize the impact of non-achievement-based factors on their tests. This can, and should, be done by test developers through thoughtful test design and a rigorous development process (Eignor, 2013). The reduction in the effect of such non-achievement-based factors can increase the validity of test scores (Messick, 1993).

Second, they can attempt to reduce the effect of non-achievement-based factors for students. Test developers can attempt to democratize non-achievement-based skills among all test-takers such that they do not unduly benefit or hinder any student's performance on the test (e.g., by providing efficacious test preparation that addresses all three goals mentioned to all students; Bishop & Davis-Becker, 2016).

Efficacy of Test Preparation

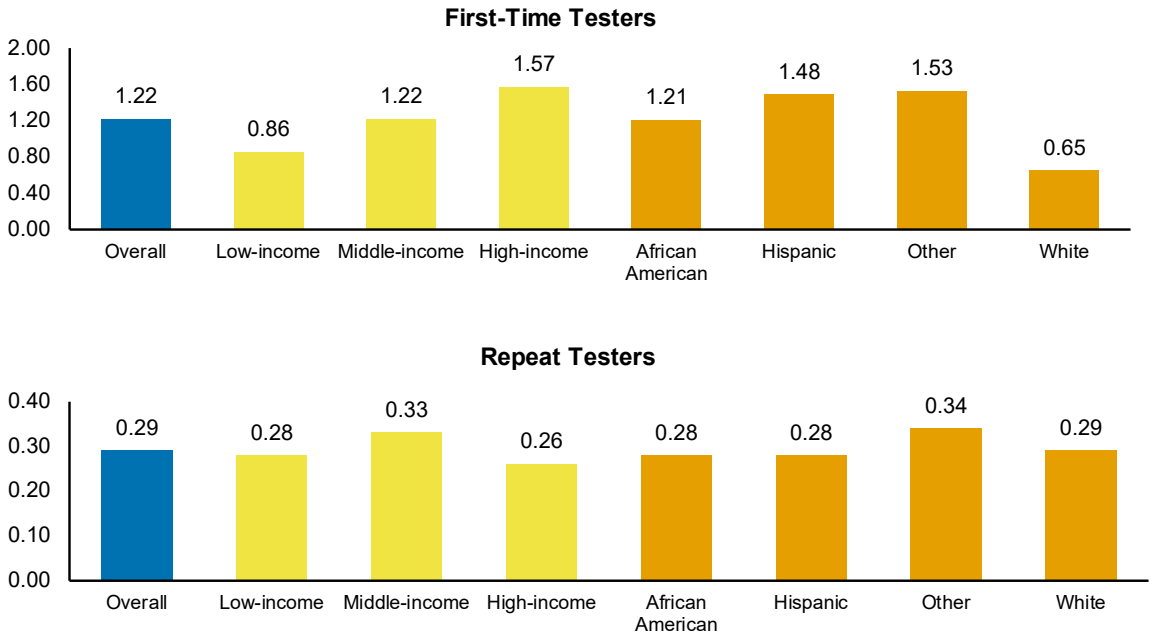
When students use test preparation, they tend to see about a 0.25 standard deviation increase on standardized test scores and about a one-point increase in their ACT Composite score. This is an estimate across many types of test preparation that can vary on the basis of duration, intensity of activities, delivery format (i.e. online vs. workbooks), implementation fidelity, and quality (Payne & Allen, 2019; Sanchez, 2018; Sanchez, 2019a; Sanchez, 2019b; Sanchez & Harnisher, 2018; Moore, Sanchez, & San Pedro, 2019). Briggs (2001) found that test preparation had small effects on ACT subject test scores, with score increases for ACT English, mathematics, or reading not exceeding a full score point. Some studies have found that underserved students, including lower-income and minority students, tend to benefit more from test preparation than other students (Sanchez, 2018; Sanchez & Harnisher, 2018). These two studies, for example, found that lower-income and minority students saw a greater benefit from participation in test preparation than higher-income and White students.

A further review of efficacy studies by the National Association for College Admissions Counseling (NACAC) found that test preparation has small and positive effects on SAT test scores—an average gain of around 40 points (0.20 SD units) on SAT total score (Briggs, 2009).[3]

Meta-analyses on test preparation for the SAT or similar achievement tests show comparable findings. Montgomery and Lily (2012) showed that students who received SAT coaching had higher scores compared to students who did not receive coaching (23.5-point difference for SAT-Verbal and 32.7 points for SAT-Math; 0.23 and 0.28 standard deviation units, respectively). Bangert-Drowns, Kulik, and Kulik (1983) showed that coaching increased scores on achievement tests (other than SAT) by 0.25 standard deviations. Powers (1993) also found that coaching programs produced a test score gain of around 0.21 standard deviation for the math and 0.14 for verbal sections of SAT (15 to 25 points each). Yet another meta-analysis of 50 studies found increases in test scores of approximately 0.25 standard deviations for students who retook a cognitive ability test when assessed between the first and second test (Hausknecht et al., 2007). If a similar 0.25 standard deviation pattern also holds for the ACT, we would expect test preparation to result in an ACT Composite score increase of about 1.5.

Prior research has demonstrated that the effect of test preparation can differ for first-time and repeat test-takers. For example, Sanchez (2019b) demonstrated that the estimated effect of ACT Online Prep on ACT scores is impacted by the inclusion of a prior ACT score in the analysis model. For example, this study estimated that purchase of AOP for first-time testers resulted in an increase in ACT score of 1.22 points (see Figure 2), or 0.21 standard deviation units. For repeat testers, this estimate was 0.29 point, or 0.05 standard deviation units. The difference could be due to differences in the analysis model, or due to test preparation having a greater impact for first-time examinees.

Figure 2. Estimated Effects of AOP Purchase on ACT Composite Score, Overall and by Subgroup



Other research on ACT retesters suggests that the score gain for using some type of test preparation is about 0.7 point (Moore, Sanchez, & San Pedro, 2019), or 0.12 standard deviation units. There are a couple of possible explanations for the differences in effect estimates. First, prior experience taking the test could function as a practice effect and familiarize students with the test. Therefore, the effect of additional test prep activities is smaller for repeat testers because all retesting students benefit from the practice effect, leaving less to be explained by additional test prep activities. Prior ACT test scores are excellent predictors of later ACT test scores. Therefore, the inclusion of a prior test score accounts for a significant portion of the variance in post-test scores thereby reducing the variance explained by other factors.

Goals of the Study

The literature on test preparation efficacy tends to focus on larger and more in-depth programs such as workshops or in-person and online offerings. While these types of programs are ubiquitous in the test preparation space, there is another arguably more accessible type of test preparation: workbooks. Workbooks are a low-cost, easily obtainable way to prepare for standardized tests. In the case of college entrance exams, students can purchase these workbooks locally and online, can check them out from school or local libraries, or get the workbooks from peers or family members who have already taken the exam. In the present study, I examined the effects of purchasing the *Official ACT Prep Guide*, a workbook published by the makers of the ACT, on ACT Composite score. Given the widespread use of this product (e.g. in the 2018-2019 academic year over 300,000 copies were sold), it is important to estimate the impact of purchasing this workbook.

Researchers often endeavor to make causal inferences about intervention efficacy. In the case of randomized experiments, doing so is straightforward. However, for potentially many reasons, including ethical or financial reasons, randomization to an intervention may not be possible. Barring the ability to conduct a randomized controlled trial, it is possible to make use of existing observational data to support casual inferences. In this case, methods must be implemented that account for the inherent self-selection bias in choosing to utilize the intervention. Much has been written about making causal inferences utilizing observational data (Austin, 2009; Austin, 2011; Rubin 1973, 1979; Heckman, Ichimura, & Todd 1998; Robins & Ritov, 1997; Robins, Hernan, & Brumback 2000; Rubin & Thomas 2000; Hirano, Imbens, & Ridder 2003; Sato & Matsuyama 2003; Kurth et al. 2006). I refer interested readers to those documents for more information on these methods and the underlying statistical theories. In addition to these references, there are also several practitioner guides focused on the implementation of these procedures (Austin & Stuart, 2015; Burgette, Griffin, & McCaffrey, 2020; Caliendo & Kopeinig, 2008; D'Agostino, 1998; McCaffrey et al., 2013; Olmos & Govindasamy, 2015).

When examining the impact of a test preparation solution, such as the *Official ACT Prep Guide*, researchers have a plethora of analytical tools at their disposal. Each has advantages and disadvantages, and researchers attempt to select the most appropriate analytical methodology to answer their research questions. The estimate of the effect varies depending upon the analytical method chosen, however. This leaves open a question about the appropriateness of the analytical method chosen. For this reason, the present study will explore four quasi-experimental designs (QEDs) for estimating the effect of purchasing the *Guide*.

Although various QEDs can be employed, one class of approaches attempts to account for non-random assignment by balancing treatment and comparison groups on baseline observable characteristics. Two popular methods are propensity score matching (PSM) and inverse probability of treatment weighting (IPW). Their popularity is evidenced by over 14,000 and 17,000 Google Scholar search results of articles from 2019 that used or mentioned these methods. Recent work by King and Nielsen (2019), however, raises important concerns about PSM, suggesting it can increase imbalance. They argue that using other matching techniques would be more appropriate.

In this study, I used propensity score matching (PSM), coarsened-exact matching (CEM), Mahalanobis distance matching (MDM), and inverse-probability of treatment weighting (IPW) to explore the effect of purchasing the *Guide*. In the matching and weighting methods investigated, propensity scores (i.e. the probability of treatment participation) are first modelled using an appropriate model such as logistic regression. For the matching methods, specific rules are used to identify control students that are "matched" on characteristics to treatment students. These matched students are then used in a final analysis to evaluate efficacy. For the weighting method, the propensity score is used to calculate a weight for each person in the sample, and weighted analysis are then used to evaluate efficacy.

In the propensity score literature, it is often advised that the method selected should be the one that provides the best balancing on baseline student characteristics. In this study, I examined the performance of each of the four QEDs and interpreted the estimates of the impact of purchasing the *Guide*.

In this study I explored the following research questions:

1. How do four QED methods compare in attaining covariate balance and model fit for both first-time and repeat ACT test-takers?

2. What is the impact of purchasing the *Official ACT Prep Guide* on ACT Composite score for first-time and repeat ACT test-takers?

3. How does the impact of purchasing the *Official ACT Prep Guide* on ACT Composite score vary by student subgroup?

Methods

Analytical Sample

This study focused on students who purchased the 2018-2019 *Official ACT Prep Guide*, which was sold from late May 2018 until mid-April 2019, and who reported not using other forms of test preparation. Specifically, this sample purchased the workbook when they registered for a national ACT test date. All students took the ACT exam at least once in June, July, September, October, or December of 2018 or in February or April of 2019. The comparison group included all students who took the ACT on the same national test dates and reported not using test preparation. Both groups were also limited to high school students with valid ACT Composite scores and no missing data.

I examined the effects of the *Guide* for both first-time and repeat test-takers. First-time test-takers were those who had only taken the ACT test once during the study window. Repeat test-takers were students who had taken the ACT at least twice by the end of the study window. These specifications resulted in a total sample of 336,070 first-time test-takers and 172,133 repeat test-takers. For first-time test-takers, there were 13,760 students who purchased the *Guide* and 322,310 control students. For repeat test-takers, there were 3,747 students who purchased the *Guide* and 168,386 control students.

In Table 1, we see that about half of the students in the first-time test-taker sample were in the 11th grade and most anticipated completing at least a bachelor's degree. There was a greater percentage of White students than African American or Hispanic students among those that purchased the *Guide*. Additionally, students who purchased the *Guide* were more likely to be from families with higher income levels and have parents who had at least a bachelor's degree. Finally, on average, students who purchased the *Guide* took the ACT one month earlier than students who did not.

Table 1. Background Characteristics of the First-Time Test-Taker Study Sample

	Purchase of *Guide*		
	No	Yes	Total Group
Expected Student Education Attainment (%)			
Bachelor's Degree	49.39	46.56	49.28
Associate/Voc-Tech	3.26	1.76	3.20
Graduate/Professional	41.40	49.09	41.71
Other - No Response	5.94	2.59	5.81
Student Grade Level (%)			
9-10	10.87	8.61	10.78
11	51.34	58.49	51.64
12	37.78	32.90	37.58
Race/Ethnicity (%)			
White	57.46	70.96	58.01
African American	10.39	4.47	10.15
Asian	5.00	5.12	5.01
Hispanic	18.31	9.80	17.96
Other	5.55	5.38	5.54
Prefer Not to Respond/Missing	3.29	4.27	3.33
Family Income (%)			
< $36K	18.51	7.52	18.06
$36K - $60K	13.96	8.50	13.73
$60K - $100K	18.30	17.79	18.28
> $100K	28.27	46.46	29.02
Missing	20.97	19.72	20.92
Gender (%)			
Female	55.96	54.20	55.88
Male	44.04	45.80	44.12
Math Coursework (%)			
All Other Math Patterns Not Missing	45.82	45.07	45.79
Beyond Algebra I, Geometry, & Algebra II	54.18	54.93	54.21
Science Coursework (%)			
All Other Science Patterns Not Missing	65.84	64.56	65.79
Biology, Chemistry, & Physics	34.16	35.44	34.21
Highest Parental Education (%)			
HS Or Less	16.79	6.53	16.37
Some College	21.26	16.92	21.08
Bachelor's	29.76	34.09	29.93
Beyond BA	25.25	39.15	25.82
Missing	6.95	3.31	6.80
ACT Composite (Mean(SD))	21.56 (5.16)	22.68 (5.18)	21.61 (5.16)
High School GPA (Mean(SD))	3.48 (0.50)	3.51 (0.49)	3.48 (0.50)
Months to Graduation (Mean(SD))	12.80 (6.68)	13.89 (6.41)	12.85 (6.68)

In Table 2, we see that the background characteristics for repeat test-takers share some similarities to those of first-time testers. Students who purchased the *Guide* were similar to students who did not purchase the *Guide* on expected educational attainment, grade level, science coursework taking, prior ACT Composite score, HSGPA, and number of prior ACT tests. There were more White students and fewer African American and Hispanic students among those that purchased the *Guide* relative to those that did not. Additionally, there were more lower-income students (i.e., students from households with a family income below $60,000) among those that did not purchase the *Guide*. There were also slightly fewer students who had taken math coursework beyond Algebra II among purchasers of the *Guide*. Repeat tested purchasers of the *Guide* also took their most recent test about one month after non-purchasers of the *Guide* and took their prior test about two months sooner than those not purchasing the *Guide*.

Table 2. Background Characteristics of the Repeat Test-Taker Study Sample

	Purchase of *Guide*		
	No	Yes	Total Group
Expected Student Education Attainment			
Associate/Voc-Tech	1.47	1.41	1.47
Bachelors	45.51	44.97	45.50
Graduate/Professional	49.61	51.37	49.65
Student Grade Level			
9-10	5.70	5.79	5.71
11	35.79	34.21	35.75
12	58.51	59.99	58.54
Race/Ethnicity			
White	65.55	71.66	65.68
African American	8.78	6.33	8.73
Asian	5.24	4.08	5.21
Hispanic	11.45	7.93	11.38
Other	5.20	5.10	5.20
Prefer Not to Respond/Missing	3.78	4.91	3.81
Family Income			
< $36K	13.96	6.99	13.81
$36K - $60K	13.61	9.69	13.53
$60K - $100K	21.00	19.24	20.96
> $100K	33.86	47.08	34.15
Missing	17.56	17.00	17.55
Gender			
Female	56.52	53.43	56.45
Male	43.48	46.57	43.55

Table 2. Background Characteristics of the Repeat Test-Taker Study Sample—continued

	Purchase of *Guide*		
	No	Yes	Total Group
Math Coursework			
All Other Math Patterns Not Missing	40.45	43.90	40.52
Beyond Alg I, Geom, Alg II	59.55	56.10	59.48
Science Coursework			
All Other Sci Patterns Not Missing	66.97	67.79	66.98
Bio, Chem, Phys	33.03	32.21	33.02
Highest Parental Education			
HS or less	11.34	6.08	11.22
Some College	19.65	17.32	19.60
Bachelor's	33.01	34.24	33.04
Beyond BA	31.36	39.15	31.53
Missing	4.64	3.20	4.61
ACT Composite (mean(SD))	23.54 (5.09)	23.36 (5.09)	23.53 (5.09)
Prior ACT Composite (mean(SD))	22.24 (4.75)	22.04 (4.74)	22.24 (4.75)
High School GPA (mean(SD))	3.62 (0.43)	3.58 (0.44)	3.62 (0.43)
Months to Graduation (mean(SD))	10.46 (5.67)	11.20 (6.25)	10.48 (5.69)
Number of Prior ACT Tests (mean(SD))	1.66 (1.06)	1.50 (0.93)	1.66 (1.06)
Number of Days since Most Recent ACT Test (mean(SD))	296.36 (294.67)	231.36 (249.56)	294.94 (293.92)

Looking across first-time and repeat test-takers, we saw some notable differences. For example, the first-time test-taker sample was about 50% 11th graders while the repeat test-taker sample was about 60% 12th graders. There were slightly more African American and Hispanic students in the first-time test-taker sample than the repeat test-taker sample. There were also fewer high-income students in the first-time vs. repeat test-taking samples. Finally, the students in the repeat test-taker sample had higher ACT Composite scores and HSGPA than students in the first-time test-taker sample.

Measures

The Official ACT Prep Guide. The *Guide* is a self-paced workbook that aims to help students prepare for the ACT. It includes five practice tests as well as explanations for all correct and incorrect answer choices. It also includes online content with additional practice questions, and it allows users to create quizzes in each content area. Additionally, the workbook includes test-taking strategies for each subject. Importantly,

it familiarizes students with the test instructions, format, and subject areas as well as helps students gauge their strengths and areas for improvement within the test subjects.

The workbook is comprised of five parts:

1. Part one helps students get acquainted with the ACT test. Logistical issues about the test are reviewed and students are given a variety of general strategies to apply when taking the test.
2. In the second part, students are walked through how to take a practice exam in similar conditions as the real test and how they can use those results to identify strengths and areas for improvement.
3. Part three walks students through each section of the ACT and reviews the content covered, the types of questions used, and strategies for doing their best on each subject.
4. Part four consists of additional practice tests with explanations for each answer and concludes with helping students understand their test scores.
5. In the concluding part of the workbook, students are walked through the process of registering for the ACT test and ensuring that they are ready for test day.

Due to the nature of the workbook as a self-paced test preparation resource, no information was available about how students used the workbook. In this study, we examine effects of purchase of the workbook, which does not necessarily imply usage of the workbook.[4]

Student Profile Section. At ACT registration, students were asked demographic and background questions. One component of this information is the student profile section. From this self-reported information, we use the following student characteristics in the study: highest parental education, student grade level, race/ethnicity, expected educational attainment, family income, and gender.

Course Grade Information Section. Students also provided information relating to their courses taken in high school and their grades in those courses. I included indicators for having taken mathematics beyond Algebra II and having taken Biology, Chemistry, and Physics. These course-taking patterns are indicative of taking more rigorous math and science coursework in high school. Self-reported grades in up to 23 courses in English, mathematics, social studies, and natural science are averaged to calculate students' high school GPA (HSGPA). It has been shown that this self-reported HSPGA correlates highly with students' transcript GPA (Sanchez & Buddin, 2015).

Months to Graduation. The number of months between a student's most recent ACT test date and May of their senior year in high school was included as an indicator of educational progress.

Months Elapsed Between Pre- and Post-ACT Tests. For repeat test-takers, I also included the number of months elapsed between ACT test administrations. This time factor accounts for the opportunity to learn and improve between the two tests, including regular instruction and schooling.

Number of Times Tested. For repeat test-takers, the total number of prior ACT tests taken was also included in models. As noted, repeated exposure to the test can serve as a form of practice and test preparation while also increasing test familiarity and reducing test anxiety.

Analysis

In the use of propensity scores for QEDs, there is typically a two-stage process. First, a propensity score model is fit that models the likelihood of selection into the intervention. In this model, often a logistic regression model for binary outcomes, we ascertain the probability of treatment given student baseline characteristics. We want to include any characteristic that may impact the likelihood of treatment but is not impacted by the treatment. The second stage involves fitting a model for the outcome of interest given the treatment and the student characteristics.

In order to arrive at a more robust understanding of the effect of the *Guide*, I make use of four QED designs: Propensity Score Matching (PSM), Coarsened Exact Matching (CEM), Mahalanobis Distance Matching (MDM), and Inverse Probability of Treatment Weighting (IPW). Matching preprocessing methods (i.e. CEM, MDM, and PSM) were done in R (3.6.0) using the MatchIt package (version 3.0.2) and the matched sample was then used in a linear regression model to estimate the effect of purchasing the *Guide*. IPW was conducted by fitting a logistic regression model with R, and average treatment effect on the treated (ATT) weights were subsequently calculated and used in a linear regression model.

In PSM, nearest neighbor matching is used to identify the single treatment and control pair with a propensity score with a similar value (Austin, 2011). The propensity score is the probability of self-selection into treatment. Although 1-to-1 matching is typically used, it is possible to use other ratios. Additionally, a caliper may be used to limit the maximum difference between treatment and control pair propensity scores to ensure that matches with large differences are not selected when close matches are not available. MDM is similar to PSM in its process; however, the distance between pairs is calculated using the Mahalanobis distance.

In CEM, each predictor is first coarsened, or binned, into groups. Then individuals with the same coarsened variables are place in the same stratum. Weighting is then used such that treated individuals are weighted to 1 and control individuals are weighted to equal the number of treated individuals divided by the number of control individuals in the same stratum (King, Nielsen, Coberley, Pope, & Wells, 2011). If 1-to-1 matching is used, the weights for both the treatment and control groups will equal 1. In IPW, the propensity score is used to weight the sample rather than pruning the sample based on propensity score matches. For both comparability purposes and the interest of this study, we use ATT weights. ATT is the average effect of the treatment among those who ultimately receive the treatment (Austin, 2011). Following this weighting method, individuals in the treatment group are weighted to 1 while individuals in the control group are weighted based upon propensity of self-selection into treatment (i.e., pj/[1-pj]; Austin & Stuart, 2015).[5] Weights were truncated at the 1st and 90th percentile of weight values.[6]

In this study, I first examined the covariate balance of each QED to determine the ability of these methods to create comparable treatment and control groups. The PSM, CEM, and MDM matching methods may result in samples with similar group-level characteristics, but each may result in different individuals being selected because of different algorithms. Because this study seeks to draw conclusions about the impact of using the *Guide*, it is prudent to examine the fit of the linear models used to estimate the effect of purchasing the *Guide*. Fit may vary across the different QED methods because they use different samples or different weights.

In examining the effect of the *Guide* on ACT test scores, I present the estimated marginal mean (EMM) ACT Composite score for different groups.[7] EMMs were calculated using the EMMEANS package (version 1.4.6) in R. These EMMs present the marginal means based on proportional representation of factor combinations in the sample.[8] As such, the EMMs for groups of research interest are adjusted for the means of other factors in the model.

Results

Research Question 1: How do the four QED methods compare in attaining covariate balance and model fit for both first-time and repeat ACT test takers?

First-Time Testers

Logistic Model. All the QED methods examined depended upon a first-stage implementation of a logistic model which used student characteristics to predict purchase of the *Guide*. In this study, all four of these methods utilized the same logistic regression model. While there are several alternative ways to predict purchase of the *Guide*, in this study, I utilized logistic regression. The student characteristics included highest parental education, gender, student educational level, race/ethnicity, high school GPA, expected educational attainment, family income, mathematics and science coursework, and the number of months to graduation to predict purchase of the *Guide*. Among these predictors, only science coursework taken was not a significant predictor of purchase behavior. Of interest, students whose highest parental education was high school or less or who did not provide parental education were less likely than students with a parent with a bachelor's degree to purchase the *Guide*. Students in the 9th or 10th grade are much less likely, and students in the 12th grade are much more likely, to purchase the *Guide* compared to 11th graders. Additionally, Hispanic and African American students were less likely than White students to purchase the *Guide*.

Table 3. Logistic Regression Model Predicting First-time Tester Purchase of the Guide

Student Characteristics	Coefficient	Std Error	Odds Ratio
Maximum Parental Education			
High School or Less	-0.56***	0.04	0.57
Some College	-0.11***	0.03	0.90
Beyond a Bachelor's Degree	0.19***	0.02	1.21
Missing Education	-0.59***	0.06	0.55
Male	-0.03*	0.02	0.97
Student Ed Level			
9th & 10th Grade	-1.41***	0.05	0.24
12th Grade	0.65***	0.03	1.92
Race/Ethnicity			
African American	-0.74***	0.04	0.48
Hispanic	-0.42***	0.03	0.66
Asian	0.04	0.04	1.04
Other	-0.09**	0.04	0.91
Prefer Not to Respond/Missing	0.14***	0.04	1.15
High School GPA	0.28	0.18	1.32
High School GPA Squared	-0.09***	0.03	0.91
Family Income			
< $36K	-0.22***	0.09	0.80
$60K–$100K	0.30***	0.04	1.35
> $100K	0.67***	0.04	1.95
Missing	0.45***	0.03	1.57
Taken Mathematics Beyond Algebra II	-0.08***	0.04	0.92
Taken Biology, Chemistry, and Physics	0.02	0.02	1.02
Number of Months to Graduation	0.08***	0.02	1.08

Note: Reference categories are bachelor's degree, female, 11th grade, White, associate degree, $36K–$60K, not taking mathematics beyond Algebra II, and not taking Biology, Chemistry, and Physics. *p<0.1; **p<0.05; ***p<0.01.

Covariate Balance. Prior to the implementation of QED postprocessing methods, we can see that there were considerable differences (i.e. an absolute standardized mean difference [SMD] of 0.1 or higher) between students who purchased the *Guide* and those who did not on student characteristics such as highest parental education, race/ethnicity, expected educational attainment, and family income (see Table 4). We can see that the implementation of the PSM, CEM, MDM, and IPW methods all succeeded in reducing the SMDs below a conservative absolute threshold of 0.1. In this sense, all methods attained acceptable covariate balance. When we look at the magnitudes of

the postprocessed absolute SMD, we see that the CEM and MDM methods provided the best absolute covariate balance.

While all methods succeeded at covariate balance, we can see that the 1-to-1 matching methods (i.e. PSM, MDM, and CEM) resulted in many cases not being used in the final postprocessed sample.[9] This was not the case for the IPW method which used the entire sample and made adjustments via weighting. In fact, both the PSM and MDM methods utilized 8% of the total sample, the CEM method utilized 6% of the total sample, and the IPW method utilized 100% of the sample.

Table 4. Covariate Balance Attained by the Four QED Methods Examined for First-Time Testers

Student Characteristics	Preprocessed Standardized Mean Difference	Postprocessed Standardized Mean Difference			
		PSM	CEM	MDM	IPW
Highest Parental Education					
High School or Less	-0.42	0.00	0.00	0.00	-0.01
Some College	-0.12	0.00	0.00	0.00	-0.01
Bachelor's Degree	0.17	-0.02	0.00	0.00	-0.01
Beyond a Bachelor's Degree	0.28	0.01	0.00	0.00	0.02
Missing Education	-0.20	0.00	0.00	0.00	-0.01
Male	0.04	0.00	0.00	0.00	0.00
Student Ed Level					
9th & 10th Grade	-0.08	0.01	0.00	0.00	-0.01
11th Grade	0.09	-0.01	0.00	0.00	0.02
12th Grade	-0.10	0.02	0.00	0.00	-0.01
Race/Ethnicity					
White	0.30	-0.04	0.00	0.00	0.01
African American	0.15	-0.03	0.00	0.00	-0.01
Hispanic	-0.29	0.00	0.00	0.00	-0.01
Asian	0.01	0.01	0.00	0.00	0.00
Other	-0.01	0.01	0.00	0.00	0.00
Prefer Not to Respond/ Missing	0.05	0.03	0.00	0.00	0.01
High School GPA	0.05	-0.05	0.00	-0.02	-0.01
Expected Student Education Attainment					
Associate Degree	-0.29	0.03	0.00	0.00	0.00
Bachelor's Degree	0.05	-0.05	0.00	-0.01	-0.01
Graduate/Professional	-0.06	0.00	0.00	0.00	0.01
Other/No Response	0.15	-0.01	0.00	0.00	0.00
Family Income					
< $36K	-0.21	0.02	0.00	0.00	-0.01
$36K - $60K	-0.11	0.01	0.00	0.00	-0.01
$60K–$100K	-0.42	0.00	0.00	0.00	-0.01
> $100K	-0.01	0.00	0.00	0.00	0.02
Missing	0.36	0.01	0.00	0.00	-0.01

Table 4. Covariate Balance Attained by the Four QED Methods Examined for First-Time Testers—continued

Student Characteristics	Preprocessed Standardized Mean Difference	Postprocessed Standardized Mean Difference			
		PSM	CEM	MDM	IPW
Taken Mathematics Beyond Algebra II	-0.03	-0.01	0.00	0.00	-0.01
Taken Biology, Chemistry, and Physics	0.02	-0.01	0.00	0.00	0.00
Number of Months to Graduation	0.03	0.00	0.00	0.00	0.02
N Treatment	13,760	13,760	10,533	13,760	13,760
N Control	322,310	13,760	10,533	13,760	322,310

While the matching was effective at creating homogenous groups for comparison (i.e., a comparison of demographics across purchase behavior but within method) it is important to also examine the impact of each methodology on the demographic makeup of the samples (i.e., compare demographics across methods). We can see in Table 5 that each methodology largely resulted in very similar groups of students who did and did not purchase the *Guide*. We do see some minor differences between the No Purchase group in the IPW method and other No Purchase groups. For example, 51% of the No Purchase group in the IPW method were 11th graders while that number ranged from 58% to 62% for other methods.

Table 5. Characteristics of the First-time test Taker Sample by QED Method and Purchase Group—continued

	Method							
	PSM		CEM		MDM		IPW	
Student Characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Expected Student Education Attainment								
Associate/Voc-Tech	1.61	1.76	0.59	0.59	1.76	1.76	3.26	1.76
Bachelors	46.58	46.56	46.62	46.62	46.61	46.56	49.39	46.56
Graduate/Professional	49.53	49.09	51.22	51.22	49.05	49.09	41.40	49.09
Other - No Response	2.27	2.59	1.57	1.57	2.59	2.59	5.94	2.59
Student Grade Level								
9-10	8.46	8.61	7.99	7.99	8.61	8.61	10.87	8.61
11	59.76	58.49	61.72	61.72	58.49	58.49	51.34	58.49
12	31.78	32.90	30.29	30.29	32.90	32.90	37.78	32.90
Race/Ethnicity								
White	72.74	70.96	80.39	80.39	70.96	70.96	57.46	70.96
African American	3.91	4.47	2.57	2.57	4.47	4.47	10.39	4.47

Table 5. Characteristics of the First-time test Taker Sample by QED Method and Purchase Group—continued

Student Characteristics	Method							
	PSM		CEM		MDM		IPW	
	No	Yes	No	Yes	No	Yes	No	Yes
Asian	4.82	5.12	3.61	3.61	5.12	5.12	5.00	5.12
Hispanic	9.70	9.80	7.78	7.78	9.80	9.80	18.31	9.80
Other	5.07	5.38	3.28	3.28	5.38	5.38	5.55	5.38
Prefer Not to Respond/ Missing	3.76	4.27	2.37	2.37	4.27	4.27	3.29	4.27
Family Income								
< $36K	7.53	7.52	5.91	5.91	7.53	7.52	18.51	7.52
$36K - $60K	8.68	8.50	6.94	6.94	8.50	8.50	13.96	8.50
$60K–$100K	17.73	17.79	17.08	17.08	17.78	17.79	18.30	17.79
> $100K	45.78	46.46	51.28	51.28	46.45	46.46	28.27	46.46
Missing	20.28	19.72	18.79	18.79	19.73	19.72	20.97	19.72
Gender								
Female	54.16	54.20	55.63	55.63	54.14	54.20	55.96	54.20
Male	45.84	45.80	44.37	44.37	45.86	45.80	44.04	45.80
Math Coursework								
All Other Math Patterns Not Missing	44.32	45.07	42.31	42.31	45.03	45.07	45.82	45.07
Beyond Alg I, Geom, Alg II	55.68	54.93	57.69	57.69	54.97	54.93	54.18	54.93
Science Coursework								
All Other Sci Patterns Not Missing	64.53	64.56	65.29	65.29	64.73	64.56	65.84	64.56
Bio, Chem, Phys	35.47	35.44	34.71	34.71	35.27	35.44	34.16	35.44
Highest Parental Education								
HS or less	6.46	6.53	4.90	4.90	6.53	6.53	16.79	6.53
Bachelor's	34.64	34.09	35.65	35.65	34.10	34.09	29.76	34.09
Beyond BA	38.60	39.15	41.35	41.35	39.15	39.15	25.25	39.15
Missing	3.23	3.31	2.84	2.84	3.31	3.31	6.95	3.31
Some College	17.06	16.92	15.27	15.27	16.90	16.92	21.26	16.92
ACT Composite (mean(SD))	22.62 (5.12)	22.68 (5.18)	23.14 (5.15)	23.28 (5.11)	22.52 (5.13)	22.68 (5.18)	21.56 (5.16)	22.68 (5.18)
High School GPA (mean(SD))	3.53 (0.47)	3.51 (0.49)	3.60 (0.42)	3.60 (0.42)	3.51 (0.48)	3.51 (0.49)	3.48 (0.50)	3.51 (0.49)
Months to Graduation (mean(SD))	14.01 (6.30)	13.89 (6.41)	14.24 (6.05)	14.22 (6.05)	13.87 (6.34)	13.89 (6.41)	12.80 (6.68)	13.89 (6.41)

Linear Model Fit. To evaluate the effect of purchasing the *Guide*, a linear model was fit to predict ACT Composite score with 12 predictors and no interactions (Table 6). These predictors were included because their relationship with ACT Composite score has been documented in prior research. There is also evidence supporting the relationship between coursework taken and ACT Composite score (ACT, 2013a; ACT, 2013b; ACT, 2013c; ACT, 2013d; Allen, 2015; McNeish, Radunzel, & Sanchez, 2015; ACT, 2006). High school GPA is included as a measure of prior achievement. Research has shown that ACT Composite scores vary by demographics such as race/ethnicity, family income, and gender (McNeish, Radunzel, & Sanchez, 2015). Expected student education attainment is used as a proxy for student motivation. Student grade level and months to graduation reflect opportunity for additional school learning and retesting prior to graduation and college admission.

Both the logistic model, described previously, and this linear model provide unbiased estimates of regression coefficients if each is specified correctly. The current analysis uses a doubly robust methodology which includes the same predictors in both the propensity score (logistic) and linear regressions. By doing so, the estimate is robust to misspecification in either the logistic or linear regression model. In this main effects model, there was a significant positive effect for purchase of the *Guide* in all four methods used (Table 6). The effect estimate across the four methods ranged from 0.14 point for the CEM model to 0.22 point for the IPW model.

Table 6. Estimated Coefficients from Linear Model Predicting ACT Composite Score by QED Method for First-Time Testers

Predictor	PSM	CEM	MDM	IPW
Purchase of the *Guide*	0.18***	0.14**	0.18***	0.22***
Maximum Parental Education				
Some College	0.37***	0.14	0.30***	0.52***
Bachelor's Degree	1.03***	0.84***	0.99***	1.24***
Beyond a Bachelor's Degree	1.75***	1.59***	1.69***	1.90***
Missing Education	1.35***	1.16***	1.26***	1.01***
Student Grade Level				
11th Grade	0.25*	-0.12	0.29**	-0.11***
12th Grade	-1.02***	-1.76***	-0.96***	-1.40***
Race/Ethnicity				
African American	-2.05***	-2.07***	-2.06***	-2.31***
Hispanic	-0.74***	-0.48***	-0.57***	-1.00***
Asian	1.47***	1.90***	1.41***	1.10***
Other	0.08	0.28*	0.06	-0.19***
Prefer Not to Respond/Missing	0.14	0.96***	0.35***	0.06*

Table 6. Estimated Coefficients from Linear Model Predicting ACT Composite Score by QED Method for First-Time Testers—continued

Predictor	PSM	CEM	MDM	IPW
Expected Student Education Attainment				
Associate Degree/Voc-Tech	-1.18***	-0.59	-1.11***	-1.26***
Graduate/Professional	1.32***	1.28***	1.29***	1.28***
Other/No Response	0.38**	0.18	0.37**	-0.17***
Family Income				
< $36K	-0.19	-0.32**	-0.22*	-0.42***
$60K - $100K	0.29***	0.27**	0.27***	0.25***
> $100K	0.74***	0.81***	0.71***	0.69***
Missing	0.90***	0.90***	0.81***	0.62***
Male	0.91***	1.02***	0.95***	0.85***
Number of Months to Graduation	-0.06***	-0.11***	-0.06***	-0.10***
High School GPA	-9.31***	-12.01***	-8.36***	-8.34***
High School GPA (Quadratic)	2.04***	2.42***	1.89***	1.89***
Taken Mathematics Beyond Algebra II	1.91***	2.07***	1.98***	1.94***
Taken Biology, Chemistry, and Physics	1.02***	0.99***	0.99***	1.06***
N Counts	27,520	21,066	27,520	336,070
R^2	0.42	0.41	0.42	0.42
Adjusted R^2	0.42	0.41	0.42	0.42
F Statistic	790.36***	590.15***	792.65***	9,660.14***

Note: *p<0.1; **p<0.05; ***p<0.01

The Normal Q-Q plots in Figure 3 show clear linearity of residuals for the PSM, CEM, and MDM methods. There is some slight departure from linearity with the IPW method. The Residual vs. Fitted plots in Figure 4 are very similar across all four methods, with slight deviations from a linear relationship between predictor variables and ACT Composite score at the upper and lower extremes of the distribution. The Residuals vs. Leverage Plots (Figure 5) did not identify any influential cases worth examining. Finally, the Scale Location Plots (Figure 6) suggest we have homoscedasticity, residuals being spread equally across predictors, across methods. That said, there appears to be an issue with residuals at the lower end of the ACT Composite scale, and this issue is slightly more pronounced for the IPW method. Taken together, these series of diagnostic plots suggest that all linear models of ACT Composite score fit the data well.

Figure 3. Normal Q-Q Plots for each QED Method for First-Time Testers

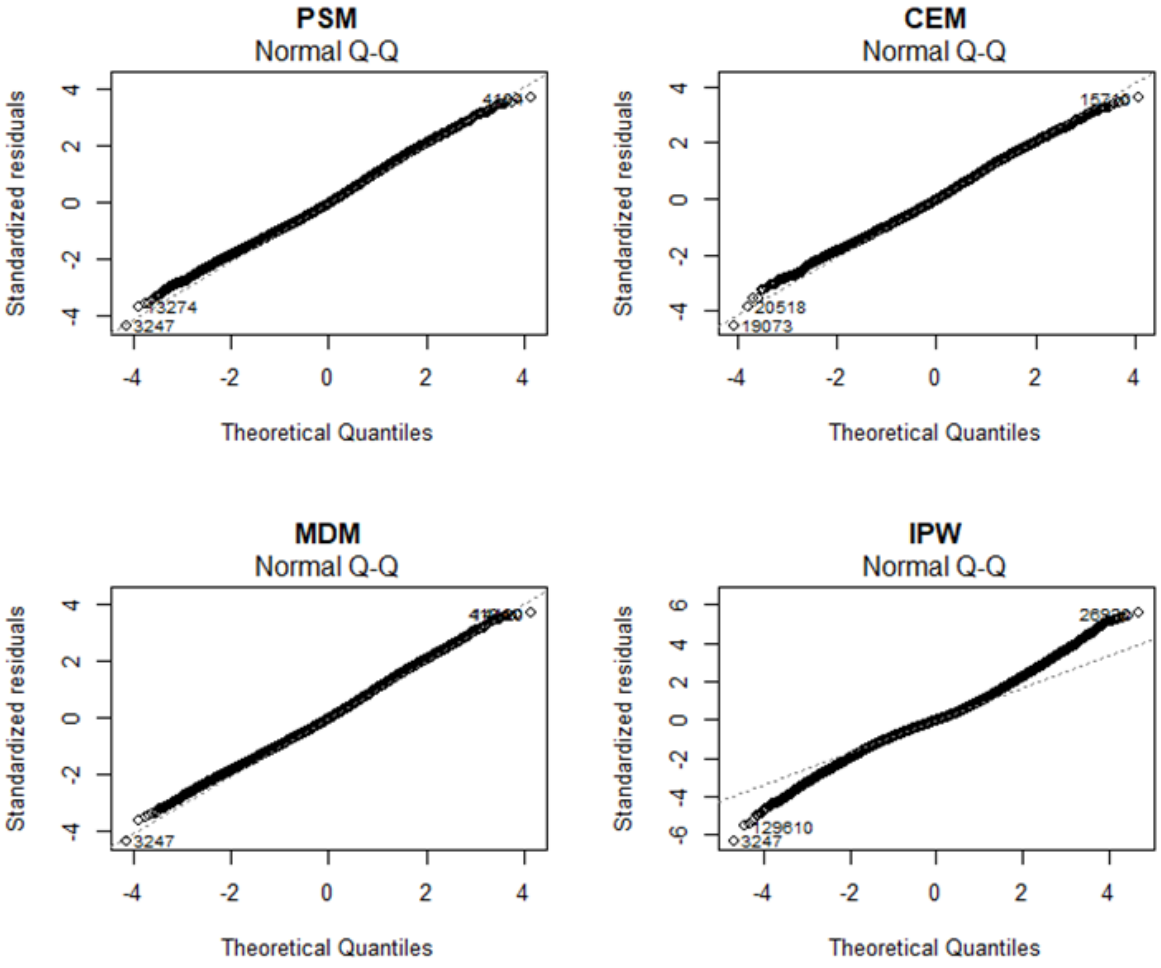


Figure 4. Residual vs. Fitted Plots for each QED Method for First-Time Testers

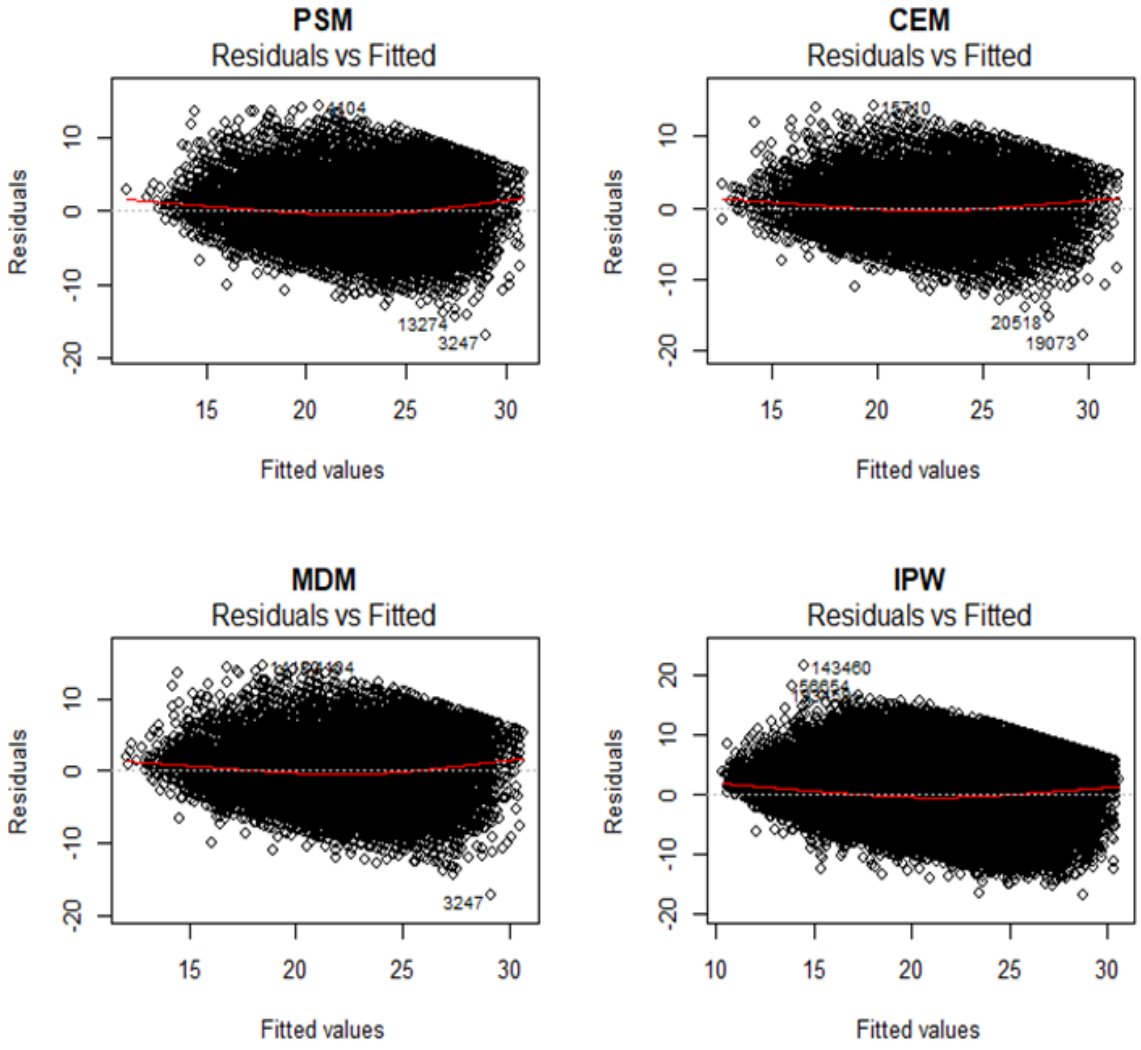


Figure 5. Residuals vs Leverage Plot for each QED Method for First-Time Testers

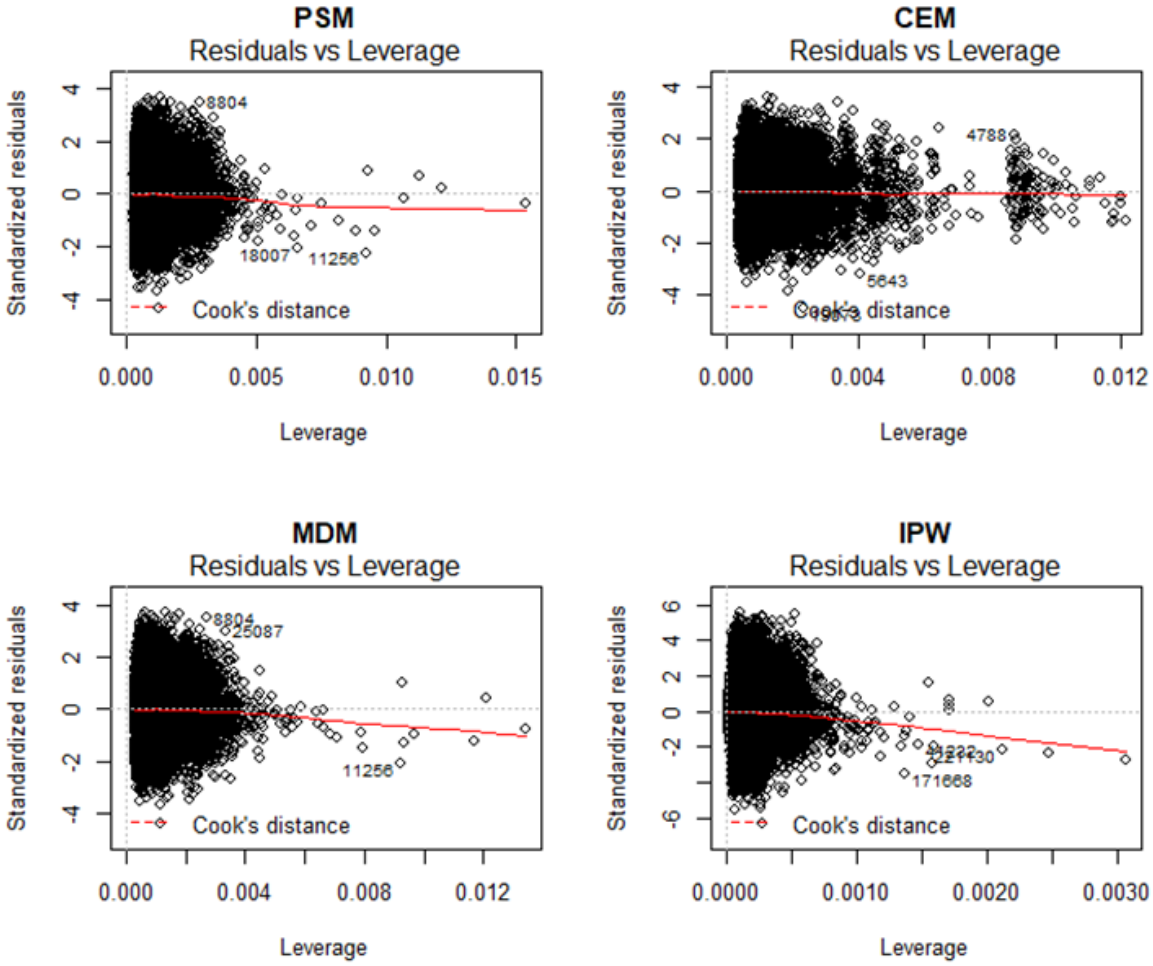
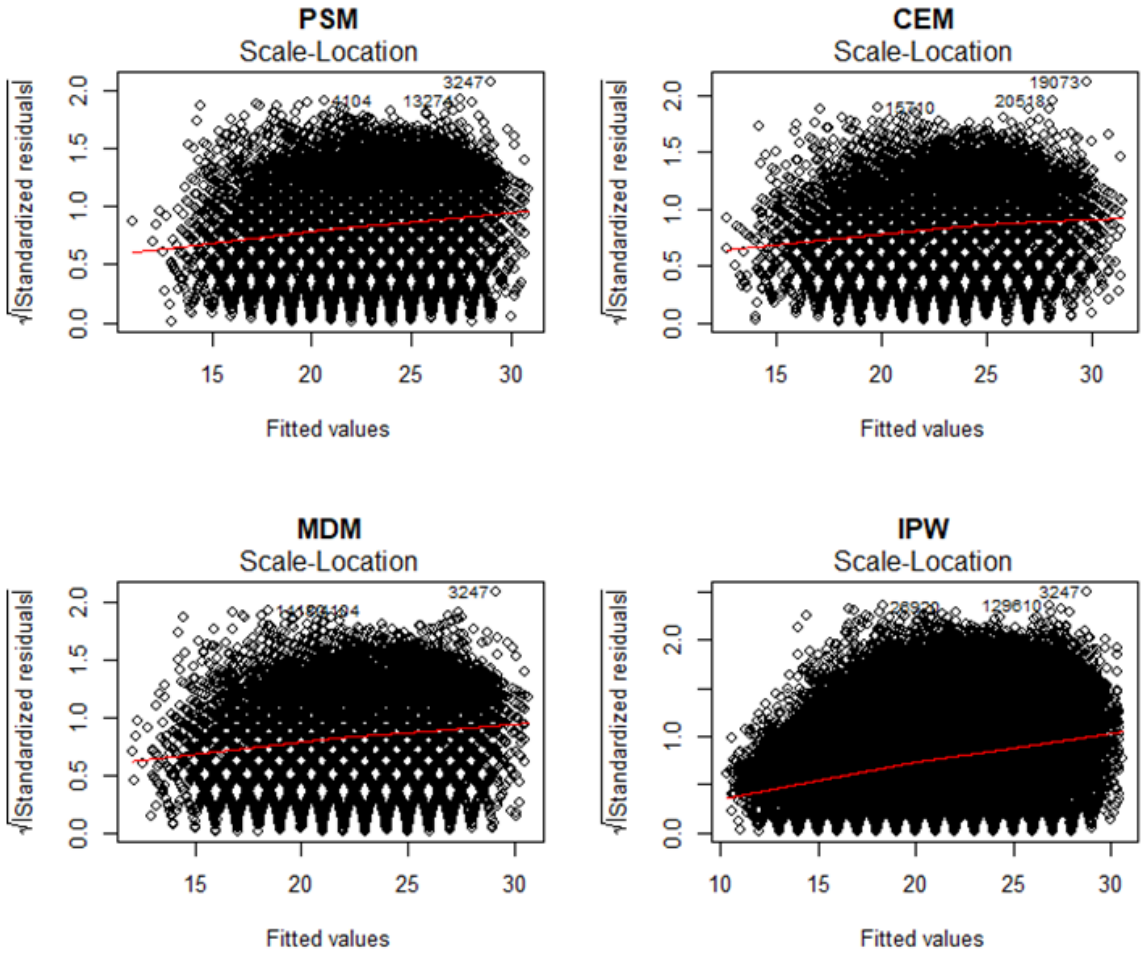


Figure 6. Scale Location Plots for each QED Method for First-Time Testers



Linear Model with Interactions Fit. A linear model was fit to predict ACT Composite score from the 12 predictors included in the main effects model and seven interactions (see Table 7). Interactions between the purchase of the *Guide* and grade level, race/ethnicity, expected educational attainment, family income level, and gender were investigated. Interactions between race/ethnicity and family income as well as between HSGPA and grade level were included as control predictors.

Table 7. Estimated Coefficients from Linear Model with Interactions Predicting ACT Composite Score for First-Time Testers, by QED Method

Predictor	PSM	CEM	MDM	IPW
Purchase of the *Guide*	0.16	0.60**	0.04	0.03
Maximum Parental Education				
Some College	0.31***	0.08	0.25**	0.44***
Bachelor's Degree	0.98***	0.79***	0.94***	1.15***
Beyond a Bachelor's Degree	1.68***	1.52***	1.63***	1.80***
Missing Education	1.31***	1.13***	1.22***	0.98***
Student Grade Level				
11th Grade	-0.74	0.73	-1.11	-1.44***
12th Grade	-0.37	1.06	-0.16	-2.01***
Race/Ethnicity				
African American	-2.72***	-2.39***	-2.61***	-2.51***
Hispanic	-1.06***	-0.53*	-1.12***	-1.24***
Asian	0.68*	0.53	-0.04	0.1
Other	-0.63*	-1.01*	-0.89**	-0.84***
Prefer Not to Respond/Missing	-1.51***	-3.05**	-1.49***	-1.07***
Expected Student Education Attainment				
Associate Degree/Voc-Tech	-1.10***	-0.46	-0.96***	-1.29***
Graduate/Professional	1.34***	1.24***	1.28***	1.27***
Other/No Response	-0.27	-0.32	-0.22	-0.22***
Family Income				
< $36K	-0.11	-0.22	-0.43**	-0.31***
$60K–$100K	0.18	0.38**	0.01	0.05
> $100K	0.48***	0.96***	0.29**	0.41***
Missing	0.50***	0.78***	0.2	0.50***
Male	0.79***	0.90***	0.85***	0.83***
Number of Months to Graduation	-0.07***	-0.11***	-0.06***	-0.10***
High School GPA	-8.98***	-10.49***	-7.54***	-8.49***
Taken Mathematics Beyond Algebra II	1.89***	2.06***	1.96***	1.92***
Taken Biology, Chemistry, and Physics	1.04***	1.00***	0.99***	1.06***
Purchase X Grade Level Interaction				
Purchase X Grade 11	-0.17	-0.25	-0.1	-0.11
Purchase X Grade 12	-0.72***	-0.69***	-0.56***	-0.69***
Purchase X Race/Ethnicity Interaction				
Purchase X African American	0.45*	0.74**	0.44*	0.50***
Purchase X Asian	-0.004	0.42	0.16	0.36***
Purchase X Hispanic	0.52***	0.17	0.18	0.49***

Table 7. Estimated Coefficients from Linear Model with Interactions Predicting ACT Composite Score for First-Time Testers, by QED Method—continued

Predictor	PSM	CEM	MDM	IPW
Purchase X Other	0.37*	0.42	0.40*	0.47***
Purchase X Missing/Prefer Not to Respond	0.12	-0.39	-0.31	0.14
Purchase X Expected Education Attainment Interaction				
Purchase X Associate/ Voc-Tech	-0.26	-0.41	-0.36	-0.03
Purchase X Graduate/Professional	-0.04	0.05	0.01	0.03
Purchase X Other/No Response	1.15***	1.01**	1.13***	1.30***
Purchase X Family Income Interaction				
Purchase X < $36K	0.41*	0.32	0.56**	0.48***
Purchase X $60K - $100K	-0.08	-0.38	-0.04	0.001
Purchase X > $100K	0.07	-0.53**	0.12	0.08
Purchase X Missing Income	0.41**	0.07	0.61***	0.59***
Purchase by Gender (Male)	0.25***	0.24**	0.19**	0.21***
Race/Ethnicity X Family Income				
African American X < $36K	0.25	-0.46	0.24	0.11
African American X $60K–$100K	0.39	-0.13	0.78*	0.21*
African American X > $100K	0.58	0.36	0.56	0.37***
African American X Missing Income	0.15	-0.85	-0.79*	-0.58***
Hispanic X < $36K	-0.83**	-1.05**	-0.33	-0.34***
Hispanic X $60K–$100K	-0.02	0.16	0.45	0.41***
Hispanic X > $100K	0.26	0.12	0.72***	0.77***
Hispanic X Missing Income	1.18***	1.47**	2.03***	0.96***
Asian X < $36K	-0.74	-0.19	-0.66	-0.39***
Asian X $60K–$100K	0.47	0.6	0.75*	0.78***
Asian X > $100K	1.15***	1.43**	1.84***	1.60***
Asian X Missing Income	1.18***	1.47**	2.03***	0.96***
Other X < $36K	-0.90*	-0.31	0.06	-0.03
Other X $60K–$100K	0.81**	1.28*	0.85**	0.63***
Other X > $100K	0.77**	1.25**	0.97***	0.95***
Other X Missing Income	0.68	1.33*	0.78*	0.40***
Prefer Not to Respond/Missing X < $36K	-0.02	3.30*	-0.14	-0.45**
Prefer Not to Respond/Missing X $60K–$100K	1.29**	4.48***	1.68***	1.00***
Prefer Not to Respond/Missing X > $100K	2.01***	4.45***	2.62***	1.54***
Prefer Not to Respond/Missing X Missing Income	1.79***	4.12***	2.20***	1.18***

Table 7. Estimated Coefficients from Linear Model with Interactions Predicting ACT Composite Score for First-Time Testers, by QED Method—continued

Predictor	PSM	CEM	MDM	IPW
HSGPA X Student Grade Level Interaction				
HSGPA X 11th Grade	0.29	-0.18	0.39	0.36***
HSGPA X 12th Grade	-0.11	-0.68*	-0.18	0.17***
N	27,520	21,066	27,520	336,070
R^2	0.42	0.42	0.42	0.42
Adjusted R^2	0.42	0.41	0.42	0.42
F Statistic	323.46***	240.99***	326.36***	3,946.94***

Note: *p<0.1; **p<0.05; ***p<0.01

The significance of the interaction terms differed by method used (Table 8). In the PSM and IPW methods, the interaction of purchase behavior with grade level, race/ethnicity, expected educational attainment, family income, and gender were significant at the 0.05 or 0.01 level. For the CEM method, the interactions of purchase behavior and grade level, family income, and gender were significant at the 0.05 or 0.01 level. In the MDM method, the interactions of purchase behavior and grade level, expected educational attainment, family income, and gender were significant at the 0.05 or 0.01 level.

Table 8. Significance of Purchase Behavior Interactions for First-time Testers

Interaction	PSM	CEM	MDM	IPW
Purchase X Grade Level	**	***	***	***
Purchase X Race/Ethnicity	***	*	*	***
Purchase X Expected Educational Attainment	***		***	***
Purchase X Family Income	***	***	***	***
Purchase X Gender	***	**	**	***

Note: *p<0.1; **p<0.05; ***p<0.01. Blanks indicate non-significance.

The fit plots for the interaction models did not differ dramatically from those of the main effect models (see Figure 7, Figure 8, Figure 9, and Figure 10). Model fit was slightly better after adding the interaction effects. Once again, these diagnostic plots suggest that all linear models of ACT Composite score fit the data well.

Figure 7. Normal Q-Q Plots for each QED Method for First-Time Tester Interaction Models

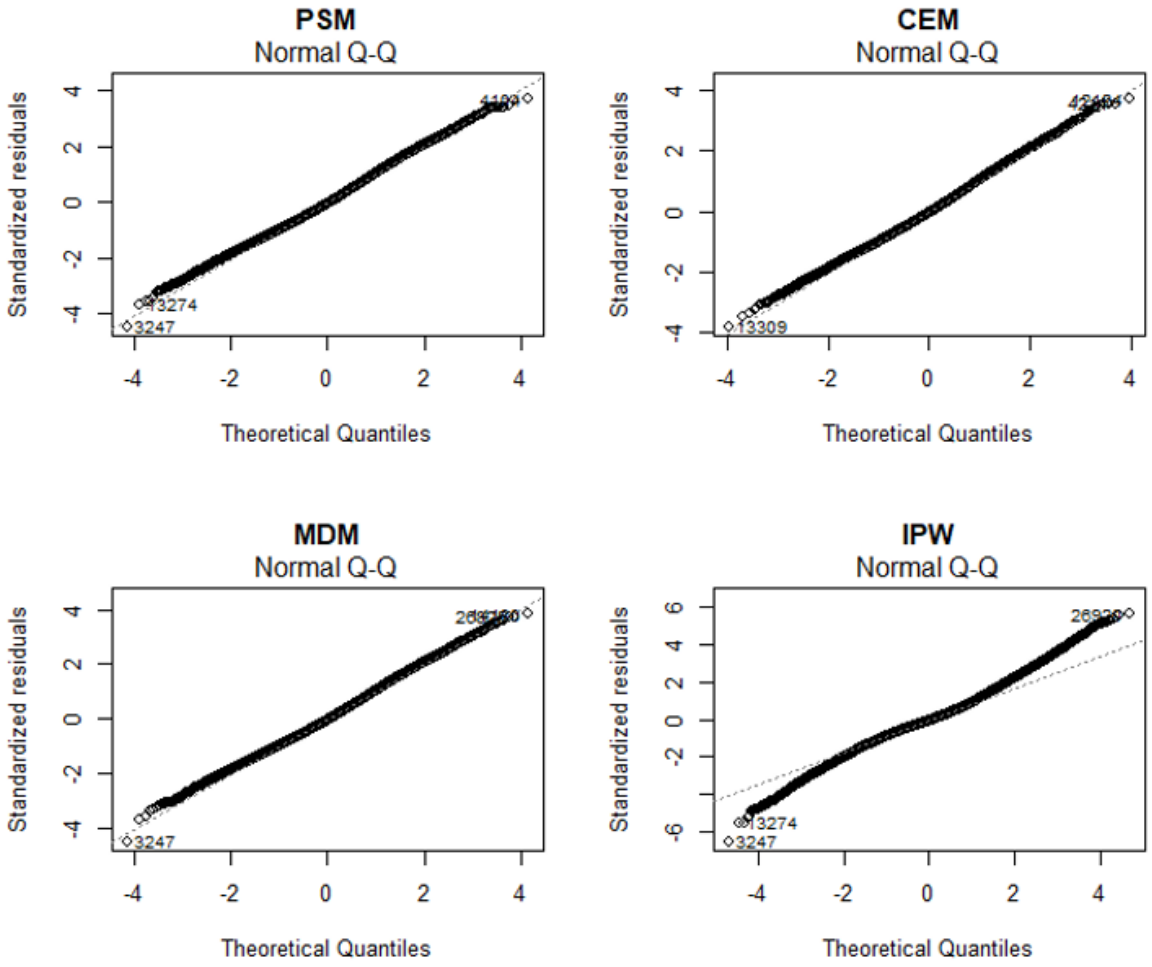


Figure 8. Residual vs Fitted Plots for each QED Method for First-Time Tester Interaction Models

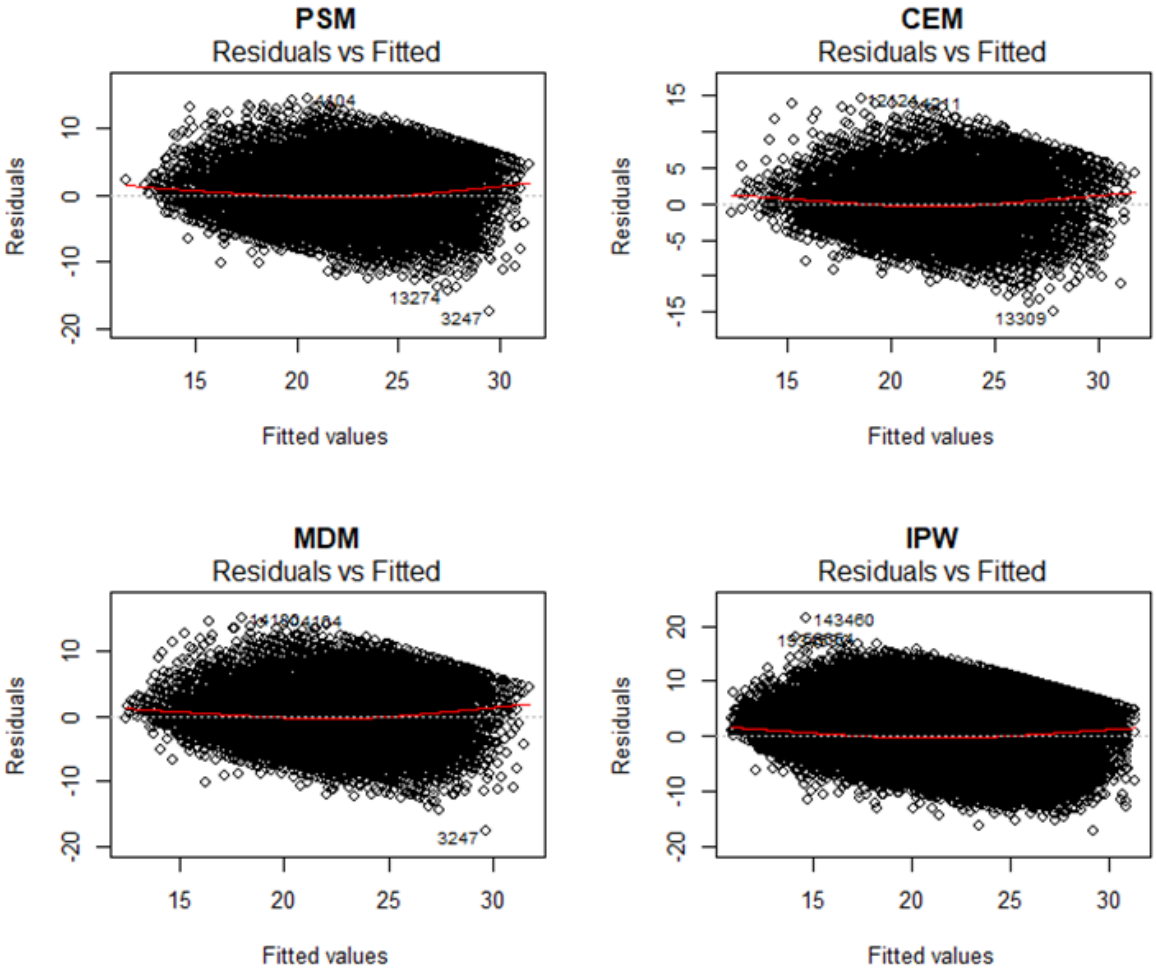


Figure 9. Residuals vs Leverage Plot for each QED Method for First-Time Tester Interaction Models

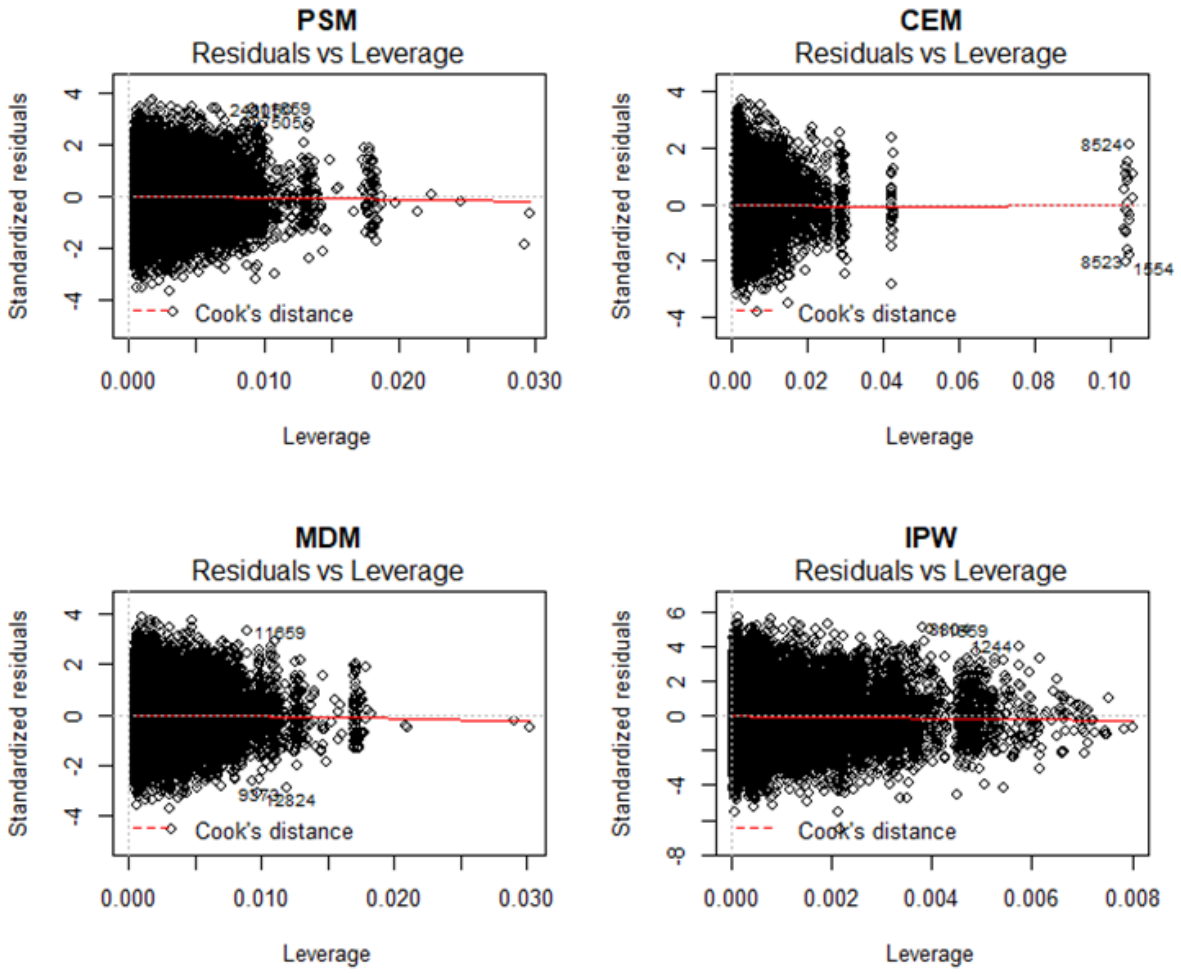
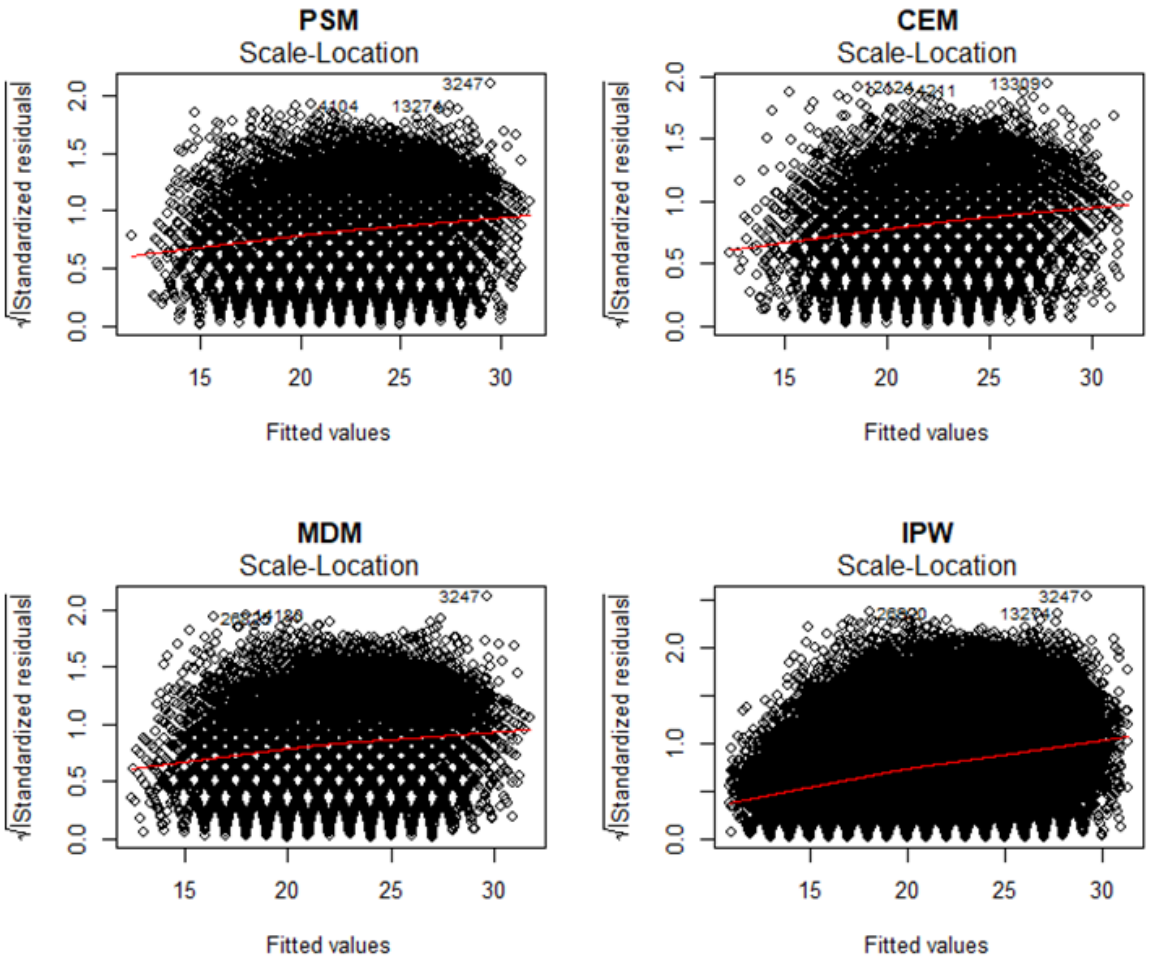


Figure 10. Scale Location Plots for each QED Method for First-Time Tester Interaction Models



Repeat Testers

Logistic Model. In addition to the predictors included in the first-time tester logistic model, this model also included the number of prior tests taken, most recent prior ACT Composite score (linear and quadratic effects), and the number of days since the most recent ACT test. Among these predictors, high school GPA (linear and quadratic effects), expected educational attainment, and mathematics and science coursework taken were not significant predictors of purchase behavior. Students whose highest parental education was high school or less or who did not provide parental education were less likely than students with a parent with a bachelor's degree to purchase the *Guide*. Students in the 9th or 10th grade were much less likely and students in the 12th grade were much more likely to purchase the *Guide* compared to 11th graders. Additionally, Hispanic and African American students were less likely than White students to purchase the *Guide*. Purchase rates increased with family income level.

Table 9. Logistic Regression Model Predicting Retester Purchase of the *Guide*

Student Characteristics	Coefficient	Std Error	Odd Ratio
Maximum Parental Education			
High School or Less	-0.40***	0.08	0.67
Some College	-0.06	0.05	0.94
Beyond a Bachelor's Degree	0.15***	0.04	1.16
Missing Education	-0.30***	0.11	0.74
Male	0.09**	0.03	1.09
Student Ed Level			
9th & 10th Grade	-1.07***	0.10	0.34
12th Grade	0.84***	0.05	2.32
Race/Ethnicity			
African American	-0.38***	0.07	0.68
Hispanic	-0.35***	0.07	0.70
Asian	-0.20**	0.09	0.82
Other	-0.06	0.08	0.94
Prefer Not to Respond/Missing	0.23***	0.08	1.26
High School GPA	-0.02	0.42	0.98
High School GPA Squared	-0.01	0.06	0.99
Expected Student Education Attainment			
Bachelor's Degree	-0.04	0.14	0.96
Graduate/Professional	0.09	0.15	1.09
Other/No Response	-0.3	0.19	0.74
Family Income			
< $36K	-0.31***	0.08	0.73
$60K–$100K	0.21***	0.07	1.23
> $100K	0.60***	0.06	1.82
Missing	0.34***	0.07	1.40
Taken Mathematics Beyond Algebra II	-0.02	0.04	0.98
Taken Biology, Chemistry, and Physics	0.01	0.04	1.00
Number of Months to Graduation	0.10***	0.01	1.11
Number of Prior ACT Tests	-0.15***	0.02	0.86
Most Recent Prior ACT Composite Score	-0.12***	0.03	0.89
ACT Composite Score Squared	0.00**	0.00	1.00
Number of Days since Most Recent ACT Test	-0.00***	0.00	1.00

Note: Reference categories are bachelor's degree, female, 11th grade, White, Associate degree, $36K–$60K, not taking mathematics beyond Algebra II, and not taking Biology, Chemistry, and Physics. *p<0.1; **p<0.05; ***p<0.01.

Covariate Balance. For the retester sample, prior to the implementation of QED postprocessing methods, there were considerable differences between students who purchased the *Guide* and those who did not on student characteristics such as highest parental education, race/ethnicity, family income, number of months to graduation, number of prior ACT tests, and number of days since most recent ACT test (see Table 10). All four methods succeeded in reducing the absolute standardized mean differences (SMD) below a conservative threshold of 0.1. In this sense, all methods attained acceptable covariate balance. Based on the postprocessed SMDs, the CEM and MDM methods once again provided the best absolute covariate balance. In this sample, the postprocessed absolute SMD was larger than the preprocessed absolute SMD in a few of the PSM and IPW cases (see Table 8). In none of these cases, however, was the absolute SMD near 0.1.

Table 10. Covariate Balance Qttained by the Four QED Methods Examined for Repeat Testers

Covariate	Preprocessed Standardized Mean Difference	Postprocessed Standardized Mean Difference			
		PSM	CEM	MDM	IPW
Maximum Parental Education					
High School or Less	-0.22	-0.01	0.00	0.00	0.00
Some College	-0.06	0.03	0.00	0.00	0.00
Bachelor's Degree	0.03	-0.02	0.00	0.00	-0.01
Beyond a Bachelor's Degree	0.16	0.00	0.00	0.00	-0.02
Missing Education	-0.08	-0.02	0.00	0.00	0.00
Male	0.06	-0.02	0.00	0.00	-0.01
Student Ed Level					
9th & 10th Grade	0.00	0.02	0.00	0.00	-0.01
11th Grade	-0.03	0.00	0.02	0.01	0.03
12th Grade	0.03	-0.01	0.00	-0.01	-0.01
Race/Ethnicity					
White	0.14	-0.01	0.00	0.00	0.03
African American	-0.10	0.00	0.00	0.00	-0.02
Hispanic	-0.13	-0.01	0.00	0.00	-0.01
Asian	-0.06	-0.01	0.00	0.00	-0.01
Other	0.00	0.01	0.00	0.00	0.01
Prefer Not to Respond/ Missing	0.05	0.02	0.00	0.00	-0.01
High School GPA	-0.08	0.01	0.00	-0.04	0.02
Expected Student Education Attainment					
Associate Degree	0.00	-0.01	0.00	0.00	-0.03
Bachelor's Degree	-0.01	0.03	0.00	0.00	-0.01
Graduate/Professional	0.04	-0.02	0.00	0.00	-0.01

Table 10. Covariate Balance Qttained by the Four QED Methods Examined for Repeat Testers—continued

Covariate	Preprocessed Standardized Mean Difference	Postprocessed Standardized Mean Difference			
		PSM	CEM	MDM	IPW
Other/No Response	-0.08	-0.01	0.02	0.00	0.03
Family Income					
< $36K	-0.27	0.01	0.00	0.00	-0.01
$36K–$60K	-0.13	0.00	0.01	0.00	-0.02
$60K–$100K	-0.04	-0.01	0.00	0.00	-0.01
> $100K	0.26	0.00	0.00	0.00	-0.01
Missing	-0.02	0.00	0.00	0.00	0.01
Taken Mathematics Beyond Algebra II	-0.07	-0.03	0.00	0.00	0.00
Taken Biology, Chemistry, and Physics	-0.02	-0.03	0.00	0.01	0.01
Number of Months to Graduation	0.12	0.01	0.03	0.02	0.04
Number of Prior ACT Tests	-0.17	0.01	0.00	0.03	0.00
Most Recent Prior ACT Composite Score	-0.04	-0.03	0.00	-0.01	-0.01
Number of Days since Most Recent ACT Test	-0.26	0.00	0.00	-0.05	-0.02
N Treatment	3,747	3,747	604	3,747	3,747
N Control	168,386	3,747	604	3,747	168,386

Once again, the 1-to-1 matching methods resulted in many cases not being used in the final postprocessed sample. Both the PSM and MDM methods utilized 4% of the total sample while the CEM method only utilized 1% of the total sample. The IPW method utilized 100% of the sample. The CEM method once again proved to be very restrictive and resulted in a very small matched sample.

In Table 11, it is apparent that PSM, MDM, and IPW largely resulted in very similar groups of students who did and did not purchase the *Guide* across subgroups. There are, however, important differences between the CEM sample and samples from the other methods. For example, the CEM sample had fewer students expecting to complete a bachelor's degree and more students planning to complete a graduate or professional degree than the other methods. There were also considerably more White students and students with a family income above $100,000 than in the other methods. Additionally, there were more students who had taken advanced mathematics coursework and had parents with an education beyond a bachelor's degree. Finally, the CEM method resulted in a sample with a shorter period between tests (by about one month) relative to the other methods. All of these differences taken together result in a concerningly different sample than other methods. As a result, the CEM results will be presented but not discussed further.

Table 11. Characteristics of the Repeat Test-Taker Sample by QED Method—continued

	Method/Purchase Status							
	PSM		CEM		MDM		IPW	
Characteristic	No	Yes	No	Yes	No	Yes	No	Yes
Expected Student Education Attainment								
Associate/Voc-Tech	1.55	1.41	0.00	0.00	1.41	1.41	1.42	1.41
Bachelors	43.53	44.97	39.74	39.74	44.73	44.97	45.09	44.97
Graduate/ Professional	52.55	51.37	59.77	59.77	51.61	51.37	51.16	51.37
Other - No Response	2.38	2.24	0.50	0.50	2.24	2.24	2.32	2.24
Student Grade Level								
9-10	5.34	5.79	3.48	3.48	5.79	5.79	5.65	5.79
11	34.19	34.21	23.68	23.68	33.65	34.21	34.04	34.21
12	60.48	59.99	72.85	72.85	60.56	59.99	60.31	59.99
Race/Ethnicity								
White	72.00	71.66	93.38	93.38	71.68	71.66	71.27	71.66
African American	6.27	6.33	1.49	1.49	6.33	6.33	6.50	6.33
Asian	4.35	4.08	0.99	0.99	4.08	4.08	4.22	4.08
Hispanic	8.11	7.93	3.15	3.15	7.90	7.93	8.22	7.93
Other	4.88	5.10	0.50	0.50	5.10	5.10	5.13	5.10
Prefer Not to Respond/Missing	4.38	4.91	0.50	0.50	4.91	4.91	5.13	5.10
Family Income								
< $36K	6.70	6.99	2.98	2.98	7.02	6.99	7.31	6.99
$36K–$60K	9.77	9.69	4.14	4.14	9.66	9.69	10.09	9.69
$60K–$100K	19.46	19.24	16.06	16.06	19.19	19.24	19.82	19.24
> $100K	47.08	47.08	67.55	67.55	47.13	47.08	45.44	47.08
Missing	17.00	17.00	9.27	9.27	17.00	17.00	17.33	17.00
Gender								
Female	52.47	53.43	56.62	56.62	53.56	53.43	53.92	53.43
Male	47.53	46.57	43.38	43.38	46.44	46.57	46.08	46.57
Math Coursework								
All Other Math Patterns Not Missing	42.62	43.90	30.63	30.63	43.82	43.90	42.97	43.90
Beyond Alg I, Geom, Alg II	57.38	56.10	69.37	69.37	56.18	56.10	57.03	56.10

Table 11. Characteristics of the Repeat Test-Taker Sample by QED Method—continued

	Method/Purchase Status							
	PSM		CEM		MDM		IPW	
Characteristic	No	Yes	No	Yes	No	Yes	No	Yes
Science Coursework								
All Other Sci Patterns Not Missing	66.56	67.79	66.23	66.23	68.13	67.79	67.35	67.79
Bio, Chem, Phys	33.44	32.21	33.77	33.77	31.87	32.21	32.65	32.21
Highest Parental Education								
HS or less	6.27	6.08	1.66	1.66	6.03	6.08	6.35	6.08
Some College	16.09	17.32	11.26	11.26	17.29	17.32	17.66	17.32
Bachelor's	35.12	34.24	38.08	38.08	34.32	34.24	34.54	34.24
Beyond BA	39.02	39.15	48.01	48.01	39.15	39.15	34.54	34.24
Missing	3.50	3.20	0.99	0.99	3.20	3.20	3.33	3.20
ACT Composite (mean(SD))	23.16 (5.00)	23.36 (5.09)	24.52 (4.36)	24.67 (4.41)	23.16 (4.89)	23.36 (5.09)	23.16 (4.92)	23.36 (5.09)
High School GPA (mean(SD))	3.58 (0.44)	3.58 (0.44)	3.75 (0.33)	3.75 (0.34)	3.60 (0.42)	3.58 (0.44)	3.59 (0.44)	3.58 (0.44)
Months to Graduation (mean(SD))	11.17 (6.06)	11.20 (6.25)	10.34 (4.79)	10.42 (4.73)	11.10 (6.02)	11.20 (6.25)	11.00 (5.92)	11.20 (6.25)
Number of Prior ACT Tests (mean(SD))	1.50 (0.88)	1.50 (0.93)	1.33 (0.68)	1.33 (0.68)	1.47 (0.86)	1.50 (0.93)	1.52 (0.89)	1.50 (0.93)
Number of Days since Most Recent ACT Test (mean(SD))	232.15 (206.46)	231.36 (249.56)	198.57 (216.60)	200.80 (213.51)	243.57 (239.74)	231.36 (249.56)	235.48 (199.84)	231.36 (249.53)

Linear Model Fit. To evaluate the main effect of purchasing the *Guide* for retesters, a linear model was fit to predict ACT Composite score with the previously discussed 12 predictors and no interactions (Table 12). In these models, the PSM, MDM, and IPW methods found a significant positive effect of purchasing the *Guide* for retested students. The average difference between students who purchased and did not purchase the *Guide* ranged from 0.28 for the MDM method to 0.32 for the IPW method.

Table 12. Estimated Coefficients from Linear Model Predicting ACT Composite Score by QED Method for Repeat Testers

Predictor	PSM	CEM	MDM	IPW
Purchase of the *Guide*	0.31***	0.14	0.28***	0.32***
Maximum Parental Education				
Some College	0.12	-1.00*	0.13	0.02
Bachelor's Degree	0.19*	-0.40	0.12	0.20***
Beyond a Bachelor's Degree	0.23**	-0.59	0.20*	0.26***
Missing Education	0.32*	0.32	0.12	0.10***
Student Grade Level				
11th Grade	-0.47***	-1.28***	-0.34**	-0.21***
12th Grade	-1.19***	-2.05***	-0.86***	-0.75***
Race/Ethnicity				
African American	-0.28***	0.39	-0.18*	-0.35***
Hispanic	0.01	0.46	0.20**	-0.09***
Asian	0.30**	0.40	0.26**	0.24***
Other	-0.07	-0.23	-0.03	-0.07***
Prefer Not to Respond/Missing	-0.06	-0.27	0.03	0.03
Expected Student Education Attainment				
Associate Degree/Voc-Tech	-0.39**	-	-0.38*	-0.34***
Graduate/Professional	0.13**	-0.08	0.08	0.19***
Other/No Response	-0.03	-0.82	-0.22	0.01
Family Income				
< $36K	0.14	-0.07	0.06	-0.02
$60K–$100K	0.13	0.61**	0.12	0.04**
> $100K	0.27***	0.81***	0.25***	0.15***
Missing	0.26***	1.03***	0.36***	0.19***
Male	0.21***	0.27**	0.27***	0.14***
Number of Months to Graduation	-0.01	-0.03	0.004	-0.002
High School GPA	-1.35**	-4.92*	-2.51***	-1.42***
High School GPA (Quadratic)	0.29***	0.80*	0.47***	0.31***
Taken Mathematics Beyond Algebra II	0.30***	0.34**	0.19***	0.28***
Taken Biology, Chemistry, and Physics	0.24***	0.03	0.19***	0.22***
Number of Prior Tests	0.09***	0.08	0.06**	0.16***
Prior ACT Composite	0.81***	0.77***	0.81***	0.79***
Prior ACT Composite (Quadratic)	0.00**	0.00	0.00**	0.00***
Number of Days since Most Recent ACT Test	0.01***	0.01***	0.01***	0.00***
N	7,494	1,208	7,494	172,133
R^2	0.85	0.82	0.84	0.85
Adjusted R^2	0.85	0.82	0.84	0.85
F Statistic	1,415.01***	195.34***	1,365.68**	32,583.03***

Note: Reference categories are Associate degree, female, 9th-10th grade, White, Bachelor's degree, $36K –$60K, not taking mathematics beyond Algebra II, and not taking Biology, Chemistry, and Physics. *p<0.1; **p<0.05; ***p<0.01. "-" indicates there were not students with an expected educational attainment of Associate Degree for the CEM model matched sample.

The Normal Q-Q plots in Figure 11 mostly show linearity of residuals across methods with some issues at the upper and lower extremes of the distribution. The departure at the upper and lower extremes was more pronounced for the IPW method. The Residual vs Fitted plots in Figure 12 are very similar across all methods investigated. The Residuals vs Leverage Plots (Figure 13), did not identify any influential cases worth examining. Finally, the Scale Location Plots (Figure 14) largely suggest homoscedasticity in each model with a slight issue with errors at the upper and lower extreme. Taken together, these series of diagnostic plots suggest that all linear models of ACT Composite score fit the data well.

Figure 11. Normal Q-Q Plots for each QED Method for Retesters

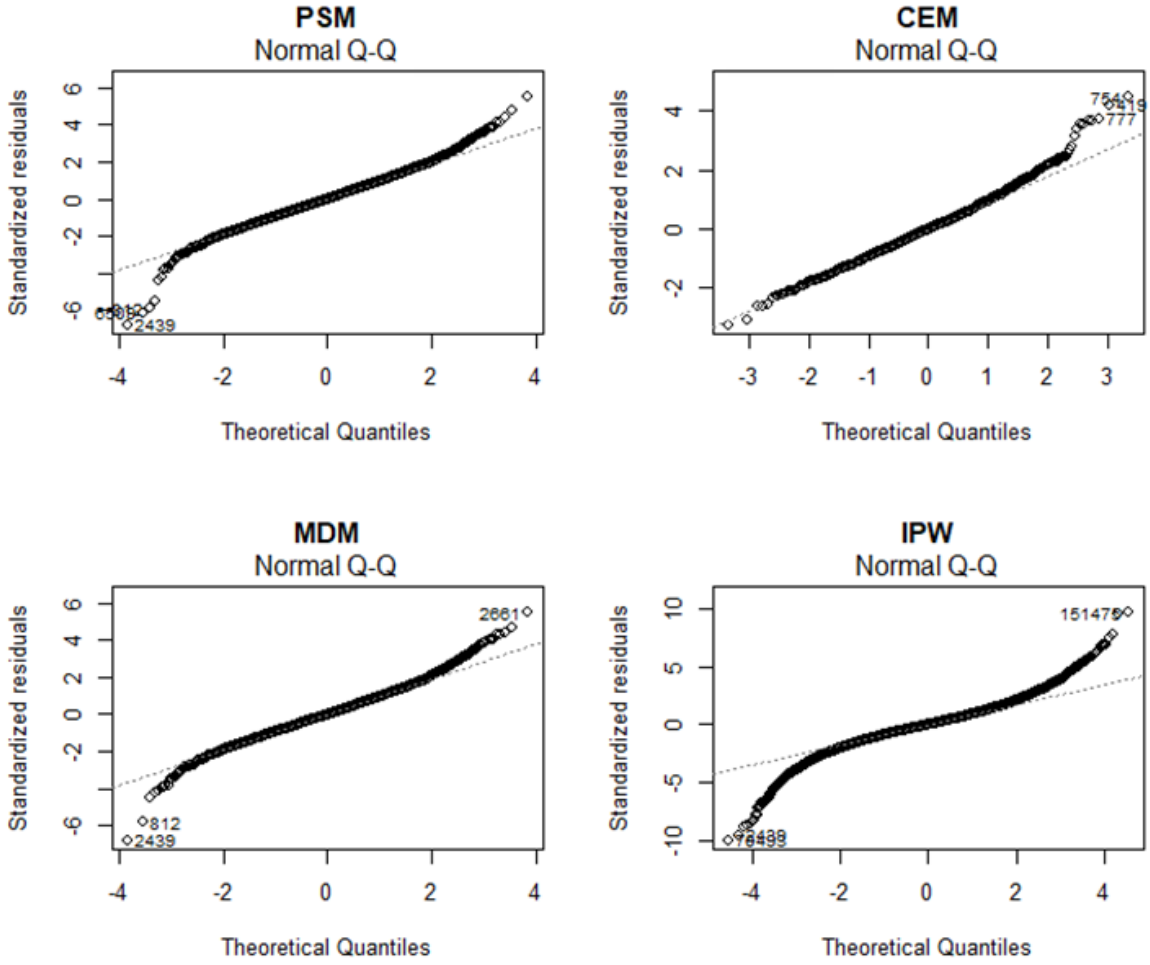


Figure 12. Residual vs. Fitted Plots for each QED Method for Retesters

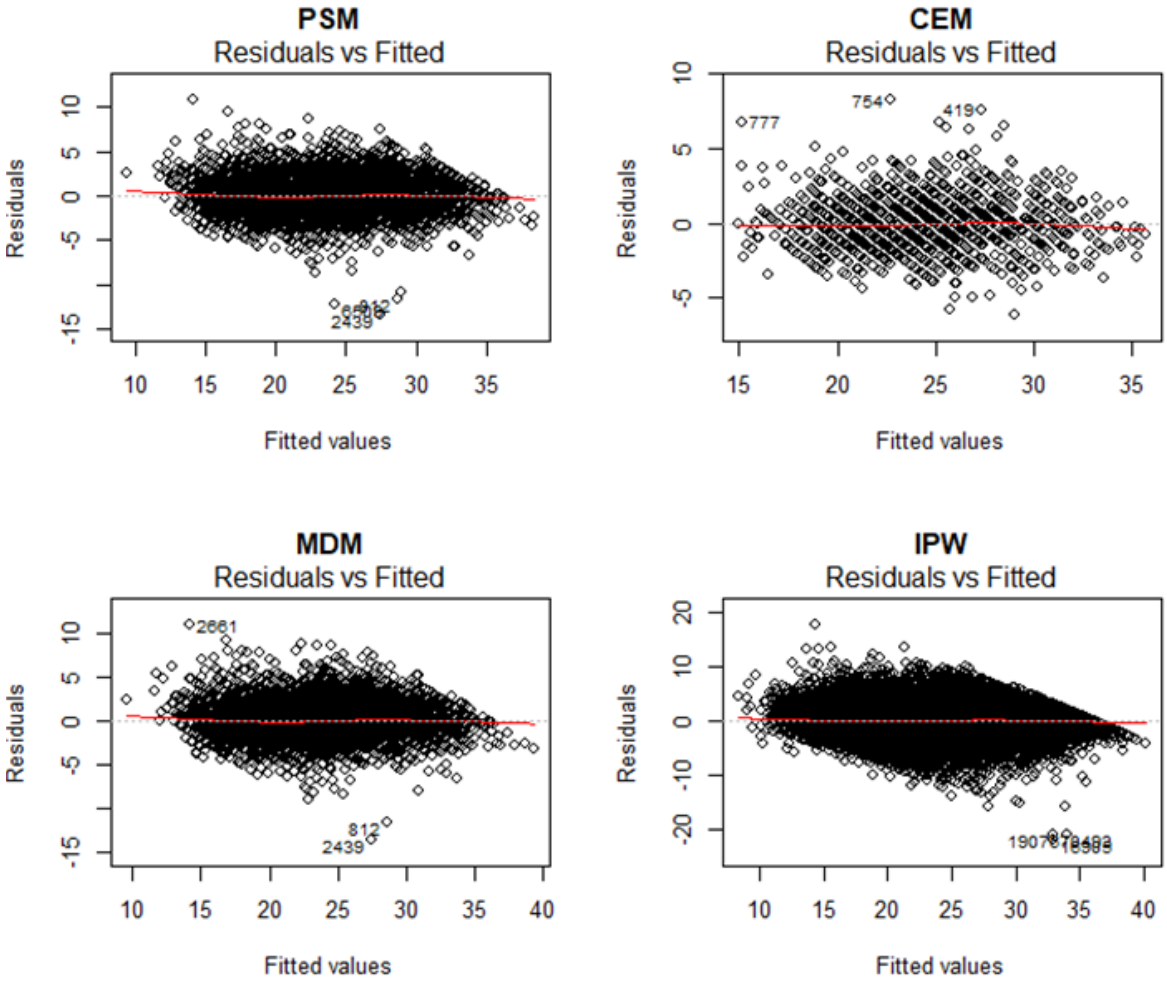


Figure 13. Residuals vs. Leverage Plot for each QED Method for Retester

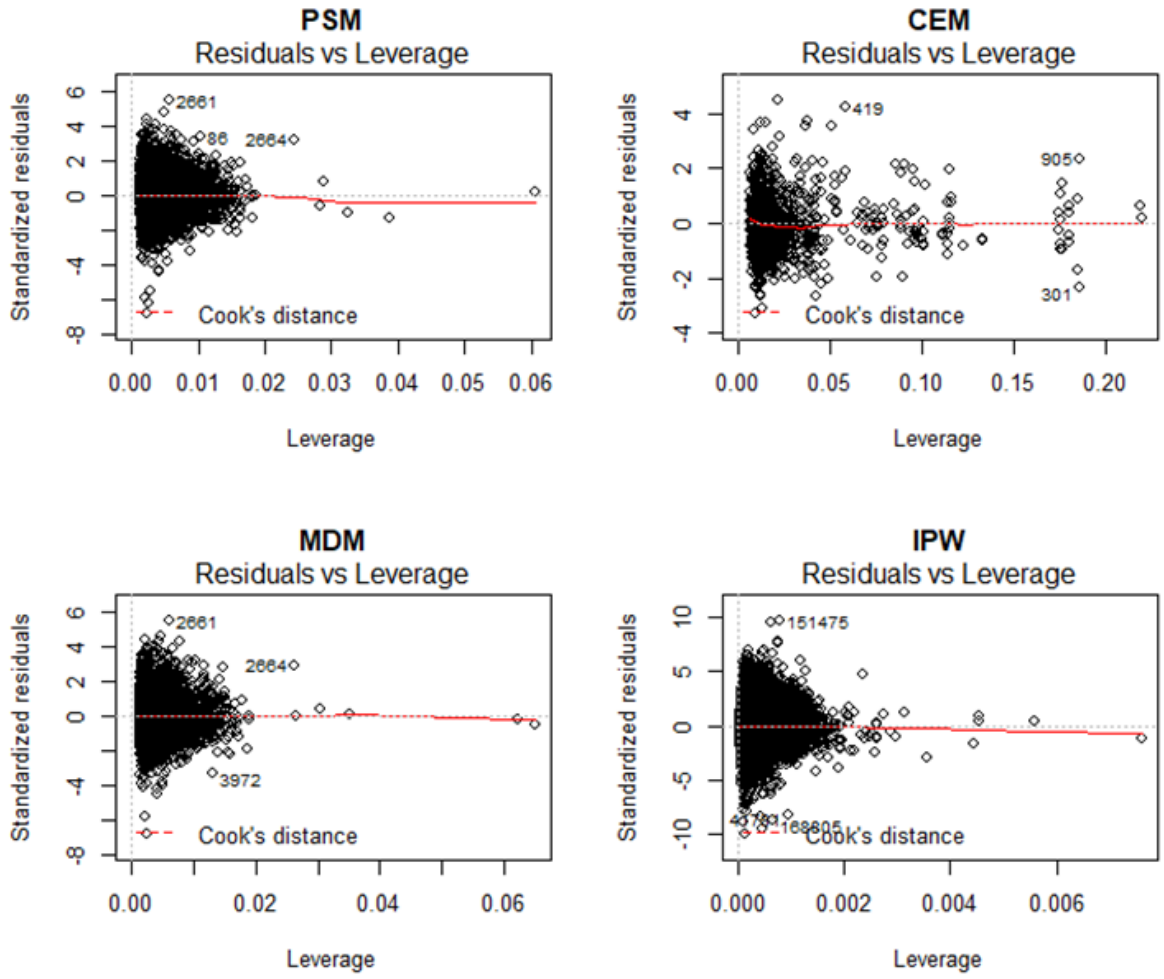
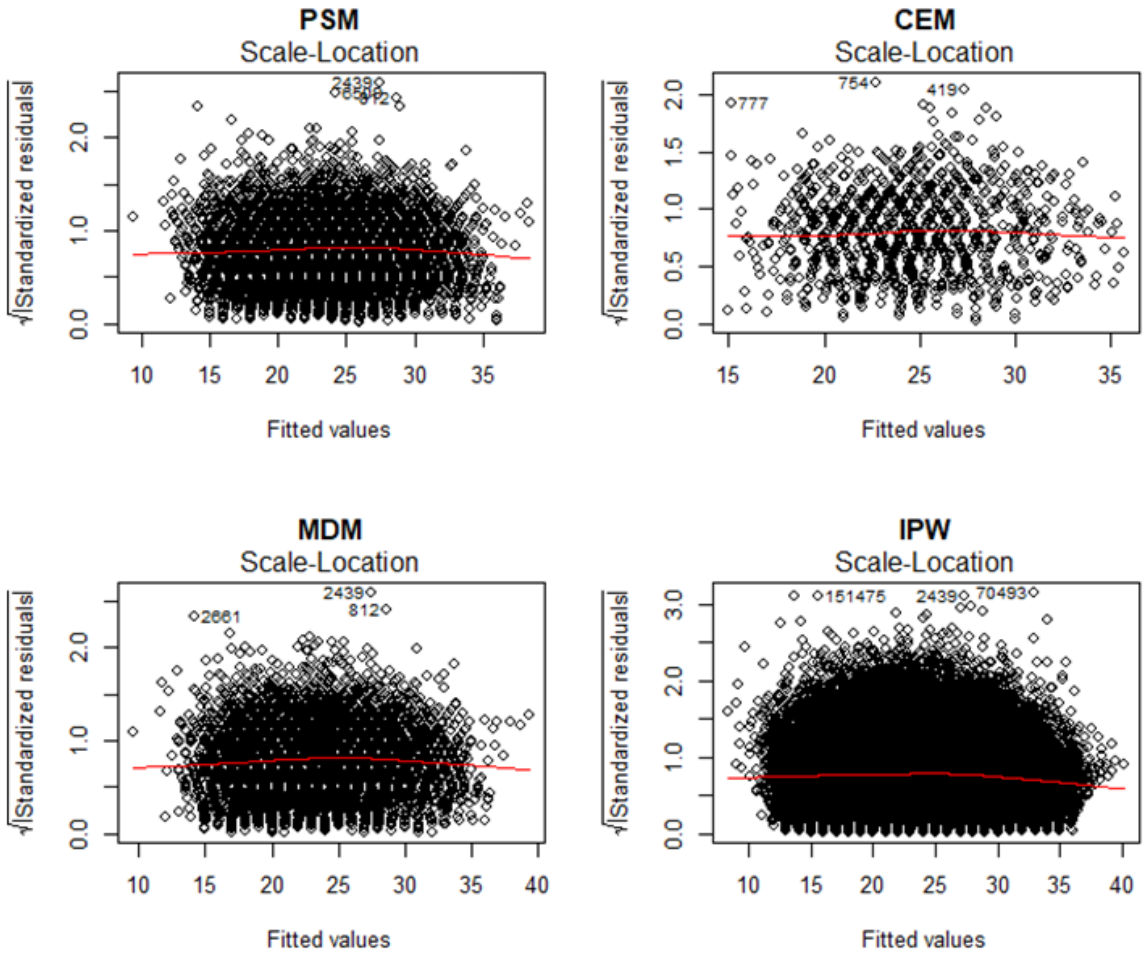


Figure 14. Scale Location Plots for each QED Method for First-Time Tester



Linear Model with Interactions Fit. In these models, I extend the main effects model for repeat testers to examine the interaction of purchase of the *Guide* with grade level, race/ethnicity, expected educational attainment, family income, and gender (Table 13). Interactions between race/ethnicity and family income as well as high school GPA and student's grade level were included as control predictors.[10] The significance of the interaction terms differed by method used (Table 14). In the PSM method, the interaction of purchase behavior with grade level and gender was significant. For the MDM method, only the interaction of purchase behavior and grade level was significant. For the IPW method, the interactions of purchase behavior and grade level, race/ethnicity, expected educational attainment, and gender were significant at the 0.05 or 0.01 level.

Table 13. Estimated Coefficients from Linear Model with Interactions Predicting ACT Composite Score by QED Method for Repeat Testers

Predictor	PSM	CEM	MDM	IPW
Purchase of the *Guide*	0.64**	1.26	0.41	0.65***
Maximum Parental Education				
Some College	0.13	-0.99*	0.15	0.02
Bachelor's Degree	0.20*	-0.39	0.13	0.20***
Beyond a Bachelor's Degree	0.23**	-0.57	0.20*	0.26***
Missing Education	0.35**	0.33	0.13	0.11***
Student Grade Level				
11th Grade	2.27*	12.26	3.10**	1.11***
12th Grade	2.17	11.28	2.83*	0.74***
Race/Ethnicity				
African American	-0.44*	0.26	-0.14	-0.38***
Hispanic	-0.35	-0.01	0.13	-0.05
Asian	0.51	-0.15	0.21	0.42***
Other	0.33	0.19	0.22	0.06
Prefer Not to Respond/Missing	0.10	0.24	0.13	-0.07
Expected Student Education Attainment				
Associate Degree/Voc-Tech	-0.51*	-	-0.41	-0.34***
Graduate/Professional	0.17**	0.02	0.07	0.20***
Other/No Response	0.08	-3.00**	-0.34	0.03
Family Income				
< $36K	0.10	0.6	-0.15	-0.01
$60K–$100K	0.06	1.00**	-0.02	0.06**
> $100K	0.18	1.06***	0.13	0.15***
Missing	0.10	1.30***	0.29*	0.23***
Male	0.10	0.27*	0.21***	0.13***
Number of Months to Graduation	-0.01	-0.03	0.002	-0.002

Table 13. Estimated Coefficients from Linear Model with Interactions Predicting ACT Composite Score by QED Method for Repeat Testers—continued

Predictor	PSM	CEM	MDM	IPW
High School GPA	-0.37	-1.63	-1.42*	-0.96***
High School GPA Squared	0.26***	0.80*	0.43***	0.30***
Taken Mathematics Beyond Algebra II	0.30***	0.35**	0.19***	0.28***
Taken Biology, Chemistry, and Physics	0.25***	0.03	0.19***	0.22***
Number of Prior ACT Tests	0.09***	0.08	0.06**	0.16***
Most Recent Prior ACT Composite Score	0.80***	0.76***	0.81***	0.79***
Prior ACT Composite Score Squared	0.002**	0.004	0.002**	0.002***
Number of Days since Most Recent ACT Test	0.005***	0.01***	0.005***	0.004***
Purchase X Grade Level Interaction				
Purchase X Grade 11	-0.59***	-0.21	-0.32	-0.48***
Purchase X Grade 12	-0.71***	-0.62	-0.52**	-0.66***
Purchase X Race/Ethnicity Interaction				
Purchase X African American	0.41**	0.26	0.2	0.26***
Purchase X Asian	-0.15	1.09	-0.03	-0.01
Purchase X Hispanic	0.30*	0.93	-0.07	0.26***
Purchase X Other	-0.18	-0.83	-0.22	-0.10
Purchase X Missing/Prefer Not to Respond	0.15	-1.00	-0.06	-0.03
Purchase X Expected Education Attainment Interaction				
Purchase X Associate/ Voc-Tech	0.27	0.00	0.10	0.12
Purchase X Graduate/Professional	-0.10	-0.20	0.01	-0.14***
Purchase X Other/No Response	-0.23	4.37***	0.24	-0.08
Purchase X Family Income Interaction				
Purchase X < $36K	0.24	-1.34	0.37	0.28**
Purchase X $60K–$100K	0.15	-0.78	0.2	0.16*
Purchase X > $100K	0.18	-0.5	0.27	0.18**
Purchase X Missing Income	0.39**	-0.54	0.27	0.25***
Purchase by Gender (Male)	0.23**	0.01	0.13	0.20***
Race/Ethnicity X Family Income				
African American X < $36K	-0.40	-	-0.16	0.02
African American X $60K–$100K	-0.09	-	0.14	0.03

Table 13. Estimated Coefficients from Linear Model with Interactions Predicting ACT Composite Score by QED Method for Repeat Testers—continued

Predictor	PSM	CEM	MDM	IPW
African American X > $100K	0.05	-	-0.19	0.08
African American X Missing Income	0.15	-	-0.50	-0.11
Hispanic X < $36K	0.40	-	0.28	-0.02
Hispanic X $60K–$100K	0.40	-	0.41	0.01
Hispanic X > $100K	0.18	-	0.09	-0.01
Hispanic X Missing Income	-0.03	-	-0.36	-0.25***
Asian X < $36K	-0.39	-	-0.21	-0.22**
Asian X $60K–$100K	-0.47	-	-0.02	-0.19**
Asian X > $100K	0.01	-	0.21	-0.14*
Asian X Missing Income	-0.12	-	0.02	-0.32***
Other X < $36K	-0.60	-	-0.08	-0.16*
Other X $60K–$100K	-0.21	-	-0.1	-0.27***
Other X > $100K	-0.25	-	-0.16	-0.07
Other X Missing Income	-0.66*	-	-0.28	-0.19**
Prefer Not to Respond/Missing X < $36K	-0.42	-	0.8	-0.07
Prefer Not to Respond/Missing X $60K–$100K	-0.21	-	0.24	0.14
Prefer Not to Respond/Missing X > $100K	-0.26	-	-0.25	0.13
Prefer Not to Respond/Missing X Missing Income	-0.34	-	-0.08	0.08
HSGPA X Student Grade Level Interaction				
HSGPA X 11th Grade	-0.63*	-3.38	-0.85**	-0.34***
HSGPA X 12th Grade	-0.79**	-3.28	-0.90**	-0.39***
N	7,494	1,208	7,494	172,133
R^2	0.85	0.83	0.84	0.85
Adjusted R^2	0.85	0.82	0.84	0.85
F Statistic	624.05***	124.94***	600.99***	14,334.40***

Note: *p<0.1; **p<0.05; ***p<0.01. Blanks indicate nonsignificance.

Table 14. Significance of Purchase Behavior Interactions for Repreat Testers

Interaction	PSM	CEM	MDM	IPW
Purchase X Grade Level	***		**	***
Purchase X Race/Ethnicity				***
Purchase X Expected Educational Attainment		**		**
Purchase X Family Income				*
Purchase X Gender	**			***

Note: *p<0.1; **p<0.05; ***p<0.01. Blanks indicate nonsignificance.

The fit plots for the retester interaction model were similar to those of the main effect models (see Figure 15, Figure 16, Figure 17, and Figure 18). These interaction fit plots indicated slightly better fit than the main effects model. Collectively, these diagnostic plots suggest that all linear retester models of ACT Composite score fit the data well.

Figure 15. Normal Q-Q Plots for each QED Method for Retester Interaction Models

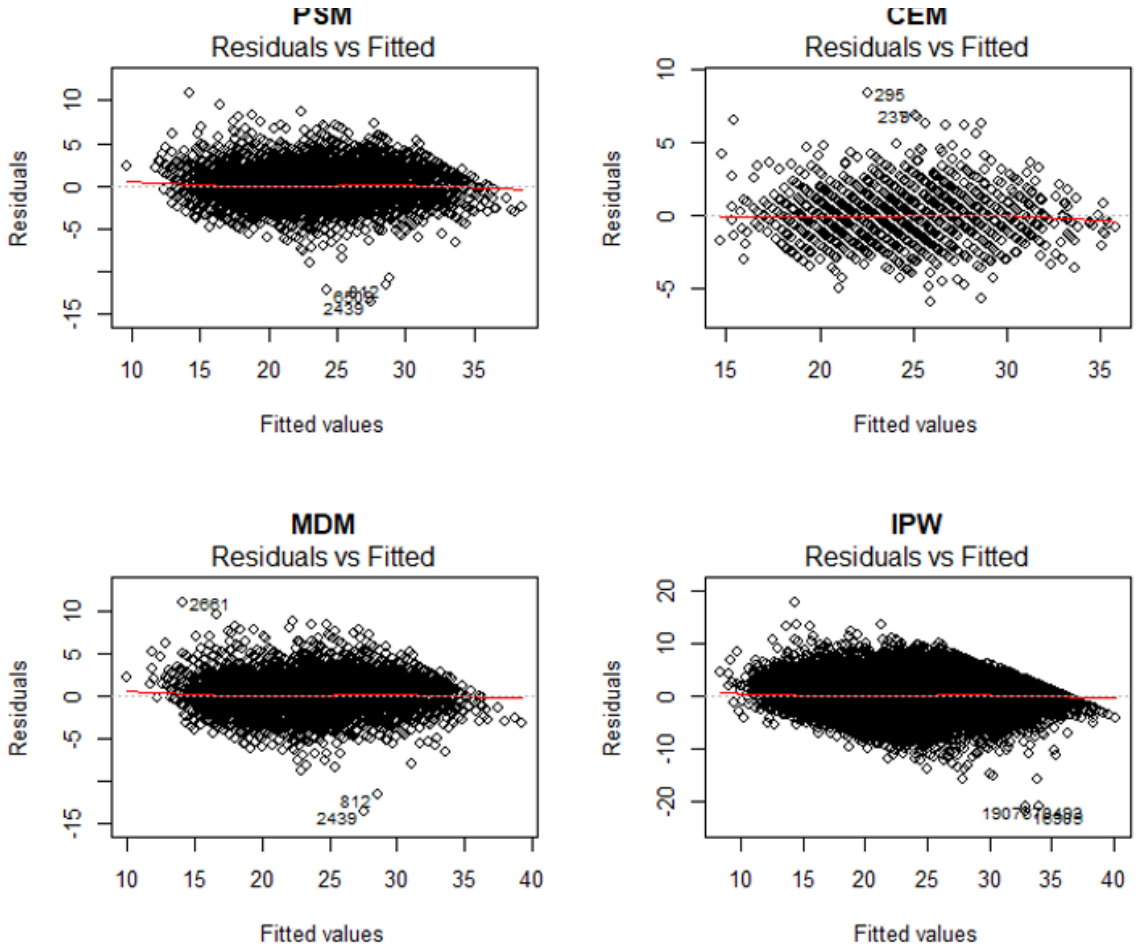


Figure 16. Residual vs Fitted Plots for each QED Method for Retester Interaction Models

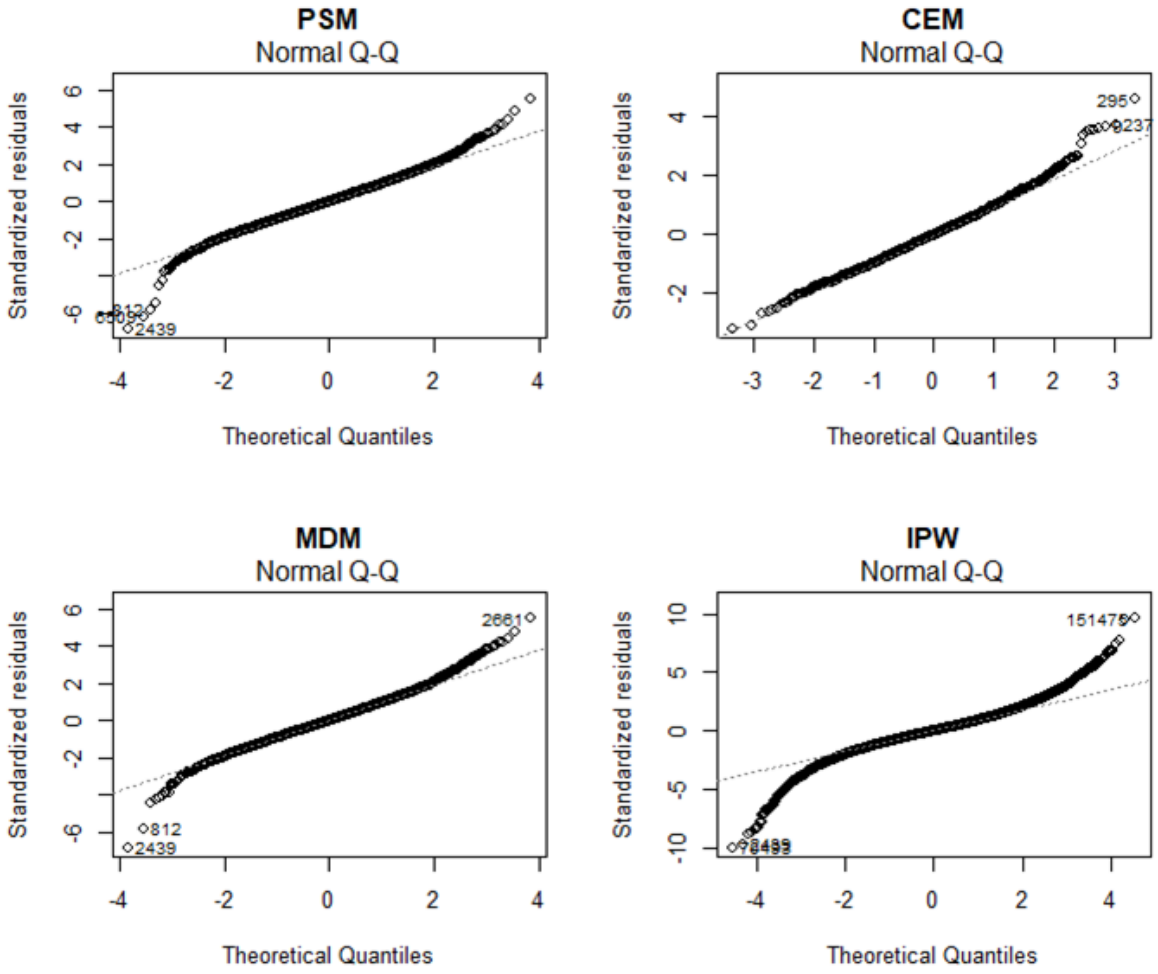


Figure 17. Residuals vs Leverage Plot for each QED Method for Retester Interaction Models

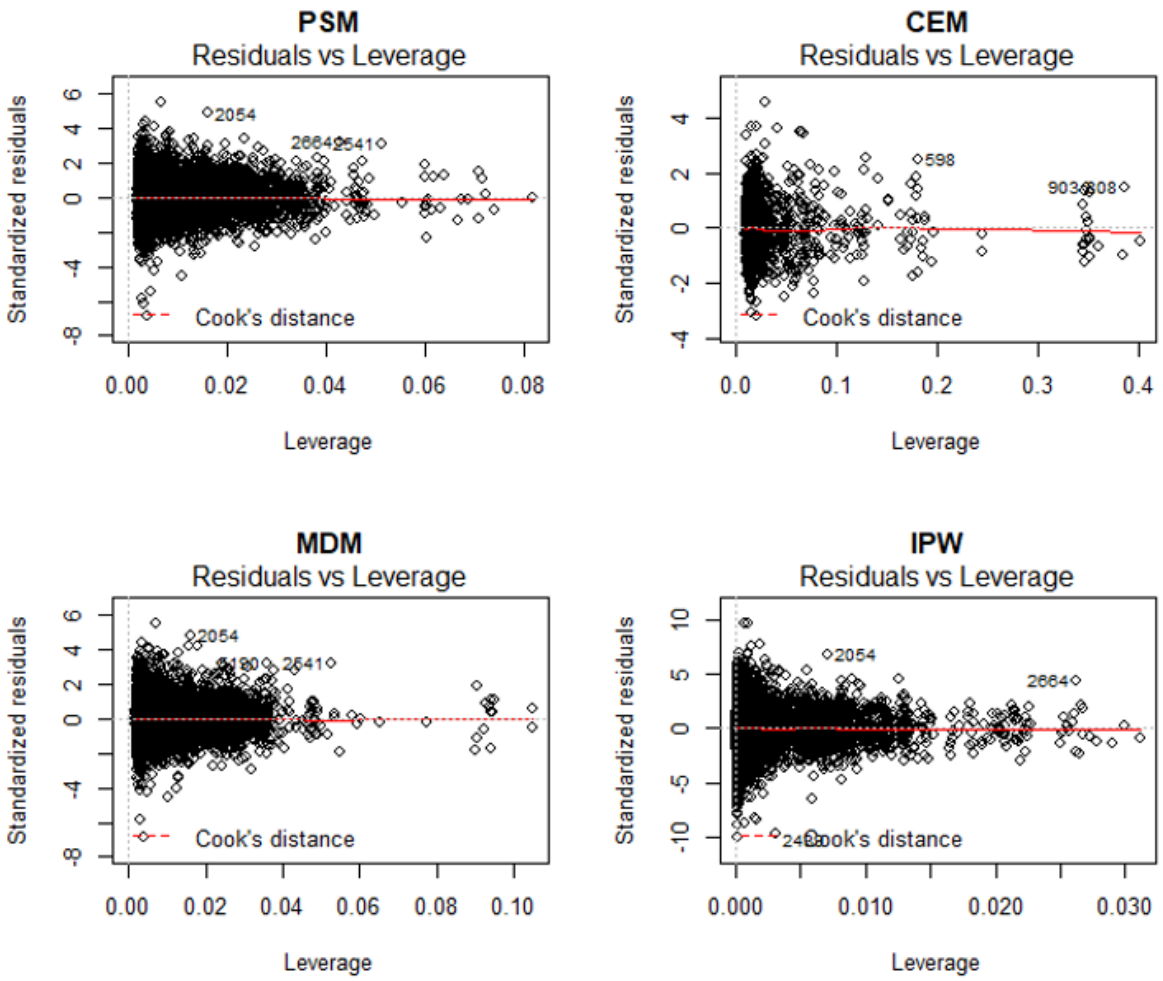
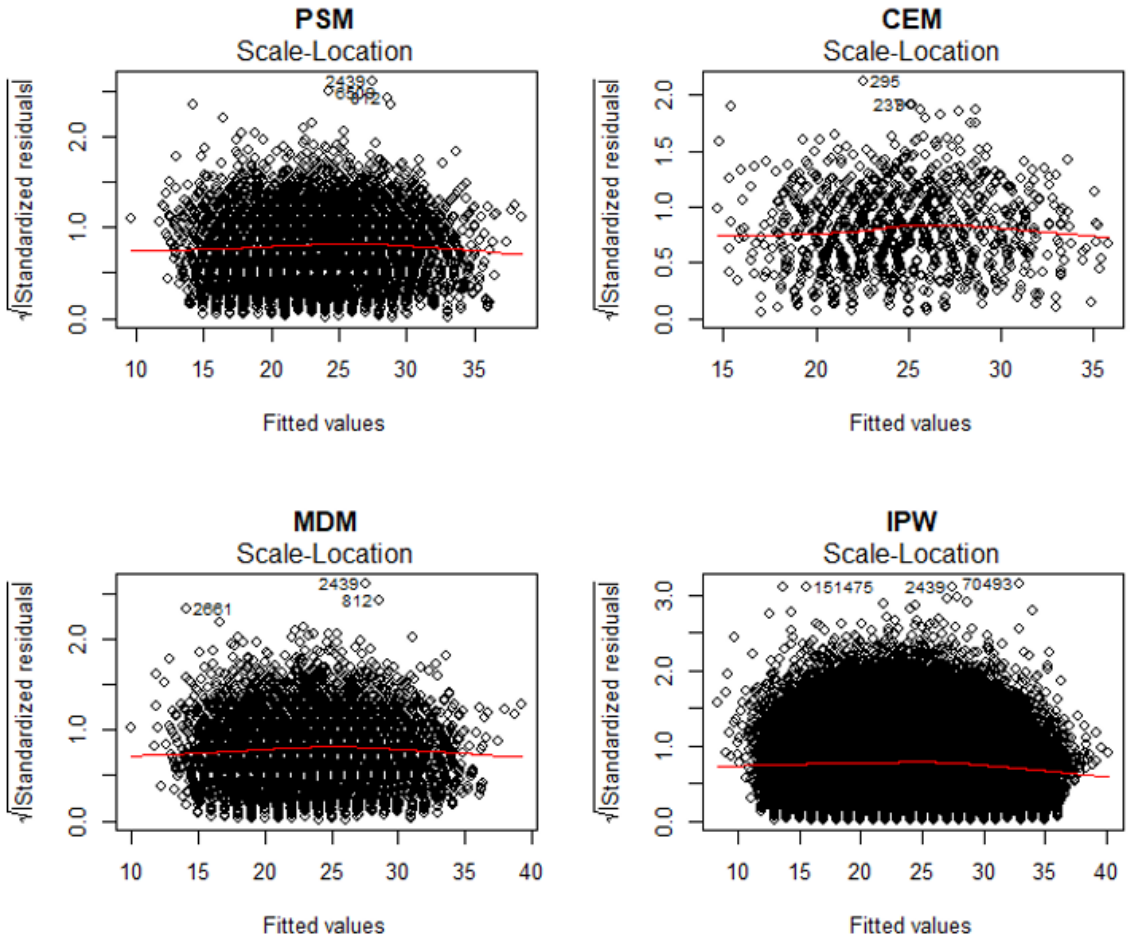


Figure 18. Scale Location Plots for each QED Method for Retester Interaction Models



Research Question 2: What is the impact of purchasing the Official ACT Prep Guide on ACT Composite score for first-time and repeat ACT test-takers?

Table 15 shows the average ACT Composite score difference between students who purchased the *Guide* and those who did not as well as its corresponding effect size for both first-time and retested students. For both first-time and retested students, the PSM, MDM, and IPW methods found significant positive effects for purchasing the *Guide*. The effect sizes, however, were all small, indicating that in these main effects models, we would estimate a conservative overall impact of purchasing the *Guide*.

Table 15. Average ACT Composite Score Increase from Purchasing the *Guide* for First-Time and Retesting Students

Sample and Method	Purchase Group Difference	Effect Size
First Time testers		
PSM	0.18***	0.03
CEM	0.14**	0.02
MDM	0.18***	0.03
IPW	0.22***	0.04
Average	0.19	0.03
Retesters		
PSM	0.32***	0.06
CEM	0.14	0.02
MDM	0.28***	0.05
IPW	0.32***	0.06
Average	0.31	0.05

Note: *p<0.1; **p<0.05; ***p<0.01. The effect size is the model specific difference between the treatment group mean and the control group mean divided by the population standard deviation. Table Averages include the PSM, MDM, and IPW methods.

Research Question 3: How does the impact of purchasing the Official ACT Prep Guide on ACT Composite score vary by student subgroup?

First-Time Testers

Table 16 presents the estimated ACT Composite score increase and effect sizes for each of the four QED methods investigated as well as the average score increase and average effect size across the PSM, MDM, and IPW methods for first-time tester subgroups. When examining student grade level, we can see that purchasing the *Guide* resulted in an average increase in ACT Composite score of almost 0.5 point for students in the 9th or 10th grade as well as for 11th graders. The average impact of purchasing the *Guide* was much smaller for 12th graders.

Table 16. Average ACT Composite Score Increase from Purchasing the *Guide* and Effect Size for First-Time Tested Students

	Estimated ACT Composite Score Increase					Effect Size				
Subgroup	PSM	CEM	MDM	IPW	Average	PSM	CEM	MDM	IPW	Average
Student Grade Level										
9th & 10th Grade	0.46***	0.44**	0.42***	0.46***	0.45	0.08	0.08	0.07	0.08	0.08
11th Grade	0.39***	0.22***	0.32***	0.56***	0.42	0.07	0.04	0.06	0.10	0.07
12th Grade	-0.34***	-0.13	-0.14*	-0.39***	-0.23	-0.05	-0.02	-0.02	-0.05	-0.04
Race/Ethnicity										
White	-0.02	0.08	0.12**	-0.01	0.04	0.00	0.01	0.02	0.00	0.01
African American	0.70***	0.91***	0.52**	0.85***	0.69	0.12	0.16	0.09	0.15	0.12
Asian	0.39*	0.58**	0.38*	1.00***	0.59	0.07	0.10	0.07	0.17	0.10
Hispanic	0.69***	0.22	0.26*	0.93***	0.63	0.12	0.04	0.04	0.16	0.11
Other	0.39*	0.45	0.49**	0.7***	0.53	0.07	0.08	0.08	0.12	0.09
Prefer Not to Respond/ Missing	0.45*	-0.12	-0.05	0.59***	0.35	0.08	-0.02	0.00	0.10	0.06
Expected Student Education Attainment										
Associate/Voc-Tech	-0.01	-0.20	-0.22	0.02	-0.06	0.00	-0.03	-0.03	0.00	-0.01
Bachelor's	0.05	0.07	0.11	0.01	0.06	0.01	0.01	0.02	0.00	0.01
Graduate/Professional	0.16**	0.16**	0.19***	0.34***	0.23	0.03	0.03	0.03	0.06	0.04
Other - No Response	1.74***	1.46***	1.56***	2.42***	1.91	0.30	0.25	0.27	0.42	0.33
Family Income										
< $36K	0.44**	0.66***	0.44**	0.56***	0.48	0.08	0.11	0.08	0.10	0.08
$36K–$60K	-0.11	0.34*	-0.01	-0.18**	-0.10	-0.02	0.06	-0.02	-0.02	-0.02
$60K–$100K	-0.23**	-0.01	-0.12	-0.25***	-0.17	-0.03	0.00	-0.02	-0.03	-0.03
> $100K	0.03	-0.08	0.09	0.07*	0.06	0.01	0.00	0.02	0.01	0.01
Missing	0.67***	0.60***	0.69***	1.09***	0.82	0.12	0.10	0.12	0.19	0.14
Gender										
Female	0.03	0.05	0.10	0.13***	0.09	0.01	0.01	0.02	0.02	0.01
Male	0.26***	0.24***	0.27***	0.38***	0.30	0.04	0.04	0.05	0.07	0.05

Note: *p<0.1; **p<0.05; ***p<0.01. The effect size is the model specific estimated ACTC score increase divided by the population standard deviation. Table Averages include the PSM, MDM, and IPW methods.

While White students did not see a notable increase across methods in ACT Composite score for purchasing the *Guide*, African American and Hispanic students saw an average increase of about 0.7 and 0.6 point, respectively. Other racial/ethnic minority groups also had an average increase with purchase of the *Guide* larger than that of White students. Across expected educational attainment levels, on average, students who expected to complete a Graduate or Professional degree or did not respond to the question saw the largest increase in ACT Composite score from purchasing the *Guide* with students who did not report their expected educational attainment seeing the largest average increase from purchasing the *Guide* (1.91 points). Most family income levels had an increase from purchasing the *Guide*; however, the largest average increase was seen for low-income students and those who did not report their family income (0.48 and 0.82, respectively). Finally, males saw a larger average increase in ACT Composite score from purchasing the *Guide* than females did (0.30 vs 0.09).

In addition to the estimated effects on ACT Composite score, Table 16 expresses the estimates as effect sizes (e.g., estimated effect on ACT Composite score, in standard deviation units). This makes it easier to compare effects across different studies that use different outcome variables. For example, on average, the effect sizes range from 0.01 to 0.33.

Repeat Testers

Table 17 presents the estimated ACT Composite score increase and the effect sizes for each of the four QED methods investigated as well as the average effect size across the four methods for repeat testers. For repeat testers, purchasing the *Guide* resulted in an increase in ACT Composite score of almost 1 point for students in the 9th or 10th grade, almost half a point for 11th graders, and a little less than a quarter of a point for 12th graders.

Table 17. Average ACT Composite Score Increase from Purchasing the *Guide* for Repeat Tested Students

	Estimated ACT Composite Score Increase					Effect Size				
Subgroup	PSM	CEM	MDM	IPW	Average	PSM	CEM	MDM	IPW	Average
Student Grade Level										
9th & 10th Grade	0.91***	0.54	0.72***	0.89***	0.84	0.16	0.09	0.12	0.15	0.14
11th Grade	0.38***	0.36	0.39***	0.42***	0.40	0.07	0.06	0.07	0.07	0.07
12th Grade	0.20***	0.05	0.18***	0.20***	0.19	0.03	0.01	0.03	0.03	0.03
Race/Ethnicity										
White	0.25***	0.13	0.30***	0.27***	0.27	0.04	0.02	0.05	0.05	0.05
African American	0.65***	-0.23	0.48***	0.58***	0.57	0.11	-0.03	0.08	0.10	0.10
Asian	0.20	0.99	0.28	0.27**	0.25	0.03	0.17	0.05	0.05	0.04
Hispanic	0.61***	0.78	0.19	0.59***	0.46	0.11	0.13	0.03	0.10	0.08
Other	-0.02	-0.71	0.05	0.15	0.07	0.00	-0.12	0.01	0.03	0.01
Prefer Not to Respond/ Missing	0.64***	-0.8	0.31	0.41***	0.45	0.11	-0.14	0.05	0.07	0.08
Expected Student Education Attainment										
Associate/Voc-Tech	0.57	-	0.35	0.44**	0.45	0.10	-	0.06	0.08	0.08
Bachelor's	0.30***	0.19	0.26***	0.34***	0.30	0.05	0.03	0.04	0.06	0.05
Graduate/Professional	0.29***	0.07	0.30***	0.29***	0.29	0.05	0.01	0.05	0.05	0.05
Other - No Response	0.55*	4.46***	0.60*	0.52***	0.56	0.09	0.77	0.10	0.09	0.10
Family Income										
< $36K	0.36**	-0.42	0.39**	0.44***	0.40	0.06	-0.07	0.07	0.08	0.07
$36K–$60K	0.03	0.65	0.00	0.08	0.04	0.01	0.11	0.00	0.01	0.01
$60K–$100K	0.19*	-0.11	0.23**	0.26***	0.23	0.03	-0.02	0.04	0.04	0.04
> $100K	0.30***	0.17	0.33***	0.31***	0.31	0.05	0.03	0.06	0.05	0.05
Missing	0.59***	0.37	0.37***	0.48***	0.48	0.10	0.06	0.06	0.08	0.08
Gender										
Female	0.20***	0.15	0.23***	0.22***	0.22	0.03	0.03	0.04	0.04	0.04
Male	0.42***	0.13	0.35***	0.43***	0.40	0.07	0.02	0.06	0.07	0.07

Note: *p<0.1; **p<0.05; ***p<0.01. The effect size is the model specific difference between the treatment group mean and the control group mean divided by the population standard deviation.

While White students only gained 0.27 point from purchasing the *Guide*, African American and Hispanic students saw an average increase of 0.57 and 0.46 respectively. Across expected educational attainment levels, all students saw an increase in ACT Composite score from purchasing the *Guide*; however, students who expected to complete an Associate/Voc-Tech degree or did not respond to the question saw the largest average increase (0.45 and 0.56, respectively). Most family income levels had a positive average increase from purchasing the *Guide* with low-income students seeing a 0.40 point gain in ACT Composite score. Finally, males saw a larger increase in ACT Composite score from purchasing the *Guide* than females did (0.40 vs 0.22). Small average effect sizes were found for 9th and 10th graders, African American and Hispanic students, students expecting to complete an Associate/Voc-Tech degree, low and missing family income levels, and male students. These small effect sizes indicate that purchase of the *Guide* has a positive effect on ACT Composite score for retesters in these subgroups.

Discussion

Given the importance of college entrance exams like the ACT, it is vital that we understand the factors at play when students prepare to take the ACT. Preparing to take the test can help students demonstrate their true ability. Test preparation can impact scores through three mechanisms: content reinforcement, test familiarization, and testing strategies. As Messick (1993) and Millman, Bishop, and Ebel (1965) pointed out, issues not related to content knowledge can play an important role in student's test performance. Therefore, it is incumbent on test developers to make every attempt to minimize the impact of factors not related to the targeted knowledge and skill constructs. It's also important to try to democratize preparation such that everyone is starting with a similar basic level of test wiseness. Test preparation is an important part of this democratization process that can simultaneously increase core content knowledge.

The use of workbooks is an extremely popular test preparation choice. In this study, I examined purchase of the *Official ACT Prep Guide*, a workbook developed by the makers of the ACT that is meant to prepare students to take the test. I evaluated a variety of QED methods for estimating the effect of purchasing the *Guide*. Specifically, I examined the use of propensity score matching (PSM), coarsened-exact matching (CEM), Mahalanobis distance matching (MDM), and inverse-probability of treatment weighting (IPW). I evaluated each of the four QEDs and interpreted the estimates of the impact of purchasing the *Guide*.

To answer this question, it is necessary to first consider the performance of each QED method in processing the sample. In the case of both first-time and repeat test-takers, all four QEDs examined were able to attain good covariate balance using a conservative threshold of an absolute standardized mean difference (SMD) of 0.1. In fact, all absolute SMDs were well below this threshold. Consistently, the CEM and MDM method achieved the lowest mean difference, often at about 0.0. As noted by King and Neilson (2019), there are theoretical concerns with the PSM methodology that could result in that method attaining worse covariate balance on observed covariates after matching. PSM did have cases where the postprocessed SMD was greater than

the preprocessed SMD. This was also the case for the IPW method in the repeat tested sample. While this was technically observed in this study for some covariates with the PSM and IPW methods, these methods still resulted in good covariate balance.

In both the first-time and repeat tested samples, the matching methods substantially reduced the size of the sample. CEM was the worst at this issue, even resulting in a sample that was so restrictive that an interaction could not be examined in the repeat tested model. When we looked at the repeat tested sample, the CEM matched sample differed from other QED methods in important ways. This means that conclusions based on this method may be different simply due to sample differences. For these reasons, I did not consider the CEM method moving forward. The MDM method also restricted the samples, but it did not do so in a manner that restricted the types of inquiries possible due to sample restrictions. The IPW method appears to have two key advantages over the matching methods. First, it is easier to implement and is conceptually easier to understand and explain to stakeholders. Second, because it utilizes weighting and not matching, it makes use of the full sample. This difference in sample utilization was dramatic, and no matching method came close to utilizing the full sample.

Based on the results for Research Question 2, we see that average ACT Composite score for students who purchased the *Guide* was higher than for students who did not purchase the *Guide*. The average mean difference between students who purchased the *Guide* and those who did not was 0.19 for first-time tested students and 0.31 for repeat tested students. It was somewhat surprising that repeat tested students had an effect that was larger in magnitude than that for first-time tested students. However, there are some key demographic differences between the first-time and repeat tested samples that should be considered. The repeat tested sample contained greater numbers of White students, students from higher income levels, and students with higher ACT Composite and HSGPA than students in the first-time tested sample. Additionally, repeat tested students had, on average, taken the ACT 1.5 times previously. Collectively, these sample differences may contribute to motivational and effort differences between first-time and repeat test-takers that may have resulted in different effects.

This analysis further shows that purchasing the *Guide* resulted in almost a one-half point increase for first-time testers not in the 12th grade (0.08 SD for 9th and 10th graders and 0.07 SD for 11th graders). Similar gains were seen for low-income students testing for the first time. African American and Hispanic students testing for the first time also saw a much larger effect of purchasing the *Guide* than White students. The average effect size of purchasing the *Guide* for repeat tested students tended to be larger in magnitude than for first-time tested students.

The estimates presented in Table 16 and Table 17 are provided for each level of the predictor, holding all other predictors constant. This is a useful way to get a picture of the independent effects for each predictor. To add some contextualization to these findings, we could ask how specific types of people with certain characteristics would benefit from purchasing the *Guide*. For example, what score increase could a low-income White, African American, or Hispanic female junior expect from purchasing the *Guide*? For a first-time tester, the White student could expect an increase in the ACT

Composite score of 0.52 with purchasing the *Guide*, while the African American and Hispanic students could expect an increase of 0.91 and 0.95 point, respectively. For retesters, the White student could expect an increase in ACT Composite score of 0.39 point while the African American and Hispanic students could expect an increase of 0.78 and 0.64 point, respectively.

Previous research has tended to find that test preparation has an approximately 0.25 SD effect on college entrance exams and cognitive achievement exams. The effects identified in this study tended to be lower than that. One possible reason for this is that the interventions studied in prior work tended to be more intensive test preparation programs while the current study examined the purchase—not use—of a self-paced workbook.

Study Limitations

Inherent in any quasi-experimental design is the limitation that we are attempting, through statistical means, to replicate random assignment to treatment. This means that while we can attempt to approximate the benefits of randomized trials it is always possible that some bias in assignment remains. This is particularly the case if there is an issue with an omitted variable that impacts our outcome.

In this study, I excluded anyone who reported using another form of test preparation. As I relied on self-reports, there is always the possibility that a student failed to report the usage of some form of test preparation. Where this may be particularly troublesome is if a student is engaged in a type of activity which they do not consider test preparation. For example, some students may not consider 1-on-1 tutoring to be test preparation but rather supplemental learning. In the scope of the definition of test preparation in this paper, if that tutor covered ACT core content areas, I may consider that test preparation. As such, it is possible that unreported test preparation could influence outcomes. This would be doubly worrisome if this systematically happened in one group and not the other. This study relies on students accurately reporting the use of this and other test preparation.

This study also is limited in its understanding of the use of the *Guide*. I was limited to purchase records for the workbook which meant there was no way to know in what ways or how often students used the workbook. In this sense, the estimates are an average across different levels of usage—from no usage at all to high, intensive usage. If we could tease out usage information, it would be possible to gain a better understanding of the effects of using the *Guide*.

Finally, I used purchase records for the *Guide* from the ACT purchase system. This may be problematic because this is not the only source to purchase the *Guide*. In fact, students can purchase the workbook from major retailers such as Amazon and Barnes & Noble, to name only two. In addition to not being able to identify if a workbook was purchased through these means, I also did not have the timing of that purchase which could impact efficacy.

Future Research

Future research that focuses on the *Official ACT Prep Guide* can make use of survey methods to clarify some assumptions made in this study and to examine effects for different levels of usage. Through survey or interview methods, future research can clarify if students have the same definition of test preparation being used in this study and if there are any additional unreported forms of test preparation being used. By doing so, future research can account for that usage and provide a more refined estimate of the efficacy of the *Guide*. This methodological choice could also open an interesting and rich learning opportunity to document and explore the usage of the *Guide*. For example, the workbook contains five full-length practice tests. It would be interesting to learn if students are spending more time on practice tests, if they are focusing on the testing strategies presented, or if they work through the book from cover to cover. These different usage strategies likely contribute to the variability in score increases seen from using the workbook. Future research should attempt to tease out these nuances.

An additional avenue of research that should be explored is the efficacy of the *Guide* for each section of the ACT. I explored the impact on ACT Composite score which is the average of the subject test scores. It is possible that this form of test preparation is equally beneficial for all sections, but it is also possible that greater increases are seen in one or more subject tests. A recent study by Payne and Allen (2019) found that another test preparation resource found the largest effect for the ACT English test. Future research should see if this finding generalizes to other forms of test preparation, indicating that English content would benefit more from test preparation, or if this is a unique finding to ACT Academy™, indicating that resource is particularly good for English content.

A final point of consideration for future research is the slight worsening of covariate balance in the present study for repeat tested students in the IPW method. While this issue is explored and documented by King and Nielsen (2019) for the PSM method, I am not aware of this property being previously documented for the IPW method. It would be worthwhile to make use of simulation studies to identify under what conditions it would be possible to exacerbate covariate balance with the IPW method. In the present study, this issue did not result in unacceptable covariate balance for the IPW method; however, it would be important to explore this phenomenon further.

References

ACT. (2013a). *English courses taken and ACT College Readiness Benchmark performance in English*. Iowa City, IA: ACT. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/Info-Brief-2013-19.pdf

ACT. (2013b). *Social science courses taken and ACT College Readiness Benchmark performance in social science.* Iowa City, IA: ACT. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/Info-Brief-2013-20.pdf

ACT. (2013c). *Science courses taken and ACT College Readiness Benchmark performance in Science*. Iowa City, IA: ACT. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/Info-Brief-2013-22.pdf

ACT. (2013d). *Mathematics courses taken and ACT College Readiness Benchmark performance in Mathematics*. Iowa City, IA: ACT. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/Info-Brief-2013-21.pdf

ACT. (2006). *Benefits of a high school core curriculum*. Iowa City, IA: ACT.

ACT. (2019). *ACT® technical manual.* Iowa City, IA: ACT.

Allen, J. (2015). *Influence of achievement in core high school courses on ACT scores.* Iowa City, IA: ACT.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in medicine, *28*(25), 3083–3107.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate behavioral research, *46*(3), 399–424.

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in medicine*, *34*(28), 3661–3679.

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.

Bishop, N. S., & Davis-Becker, S. (2016). Preparing examinees for test taking: *Guide*lines for test developers. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.). (2016). *Handbook of test development*. New York, NY: Routledge.

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

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

Burgette, L., Griffin, B. A., & McCaffrey, D. (2020). *Propensity scores for multiple treatments: A tutorial for the mnps function in the twang package*. Santa Monica, CA: Rand Corporation.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys, 22*(1), 31–72.

D'Agostino, R. B., Jr. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. S*tatistics in medicine, 17*(19), 2265–2281.

Eignor, D. R. (2013). The standards for educational and psychological testing. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbooks in psychology®. APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (p. 245–250). Washington, DC: American Psychological Association.

Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Gerrard, M. O. M. (2007). Retesting in Selection. *Journal of Applied Psychology, 92*(2), 373–385.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies, 65*(2), 261–294.

Hirano, K., Imbens, G., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica, 71*(4), 1161–1189.

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis, 27*(04), 435-454.

King, G., Nielsen, R., Coberley, C., Pope, J. E., & Wells, A. (2011). Comparative effectiveness of matching methods for causal inference. [Unpublished manuscript]. Retrieved from https://gking.harvard.edu/files/psparadox.pdf.

Kurth, T., Walker, A., Glynn, R., Chan, K., Gaziano, J., Berger, K., & Robins, J. (2006). Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of non-uniform effect. *American Journal of Epidemiology, 163*(3), 262–270.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine, 32*(19), 3388–3414.

McNeish, D. M., Radunzel, J., & Sanchez, E. (2015). *A multidimensional perspective of college readiness: Relating student and school characteristics to performance on the ACT®*. Iowa City, IA: ACT.

Messick, S. (1993). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 429–444). Phoenix, AZ: Oryx.

Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement, 25*(3), 707–726.

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.

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.

Moore, R., Sanchez, E. I. & San Pedro, S. (2019) *College entrance exams: How does test preparation affect retest scores?* Iowa City, IA: ACT.

Olmos, A., & Govindasamy, P. (2015). A practical *Guide* for using propensity score weighting in R. *Practical Assessment, Research, and Evaluation, 20*(13), 13.

Payne, J. S., & Allen, J. (2019). *An early look at ACT Academy usage and effectiveness*. Iowa City, IA: ACT.

Robins, J., Hernan, M., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11*(5), 550–560.

Robins, J., & Ritov, Y. (1997). Towards a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine, 16*(3), 285–319.

Rubin, D. (1973). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics, 29*, 185–203.

Rubin, D. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association, 74*(366), 318–328.

Rubin, D., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association, 95*(450), 573–585.

Sanchez, E. I. (2018). *ACT Online Prep helps students on the ACT*. Iowa City, IA: ACT.

Sanchez, E. I. (2019a). *Can using ACT Online Prep improve score gains?* Iowa City, IA: ACT.

Sanchez, E. I. (2019b, March). *Causal evidence for efficacy of ACT Online Prep*. Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.

Sanchez, E., & Buddin, R. (2015). *How accurate are self-reported high school courses, course grades, and grade point average?* Iowa City, IA: ACT.

Sanchez, E., I., & Harnisher, J. (2018). *The impact of ACT Kaplan Online Prep Live on ACT score gains.* Iowa City, IA: ACT.

Sato, T., & Matsuyama, Y. (2003). Marginal structural models as a tool for standardization. *Epidemiology, 14*(6), 680–686.

Steedle, J. T. (2018). *Keeping your cool: Does test anxiety bias performance on the ACT?* Iowa City, IA: ACT.

Notes

1. After accounting for HSGPA, coursework taken, high school characteristics, and noncognitive characteristics, SES-related demographics (i.e. English being spoken at home, family income and highest parental education) only explained about 1% of variance in ACT scores.

2. On national 2018-2019 ACT test dates students were asked “Did you prepare for the ACT® test using any test preparation materials (for example, The Official ACT Prep Guide, other study Guides, online materials, practice tests, tutors, or test prep courses)?” This percentage is based on students who responded to the question.

3. Effect size was calculated as the gain score divided by the 2019 SAT Total Score standard deviation.

4. Ideally a research study would examine the usage of a test preparation program, however this was not available for this study and is therefore not the focus of this paper.

5. An alternate weighting method to attain the average treatment effect is detailed in Austin & Stuart (2015)

6. These percentiles were chosen based on the distribution of weights. Weights that exceeded these percentile thresholds were set to the weight value for the threshold.

7. Estimated marginal means are also known as Least Squares Means.

8. I elected to use a proportional calculation of EMMEANS because of the unbalanced sample design. Using equal weighted EMMEANS would have provided estimates for a balanced sample design which resulted in notably different estimates and does not represent the population.

9. The CEM implementation in R does not by default implement 1 to 1 matching. This default was overridden to allow an equitable comparison to the PSM and MDM methods.

10. In the CEM model the restrictive matching algorithm resulted in low N counts for race/ethnicity and income and it was therefore not possible to test the interaction between the two characteristics as a control interaction.

About ACT

ACT is an independent, nonprofit organization that provides assessment, research, information, and program management services in the broad areas of education and workforce development. Each year, we serve millions of people in high schools, colleges, professional associations, businesses, and government agencies, nationally and internationally. Though designed to meet a wide array of needs, all ACT programs and services have one guiding purpose—helping people achieve education and workplace success.

About ACT Research

ACT Research leads the field with authority and high-quality scientific evidence in support of education and workforce practices, solutions, and services. Our mission-driven team comprises a variety of backgrounds and disciplines, and offers a wide spectrum of knowledge and skills, enabling us to deliver quality, high-impact products and services aligned to ACT's strategy and mission. Together, our research teams provide policymakers, educators, parents, and learners with research-based insights to inform their decision-making, and deliver educators and workforce development professionals with tools and services needed for education and career navigation.

ACT®

ACT.org/research