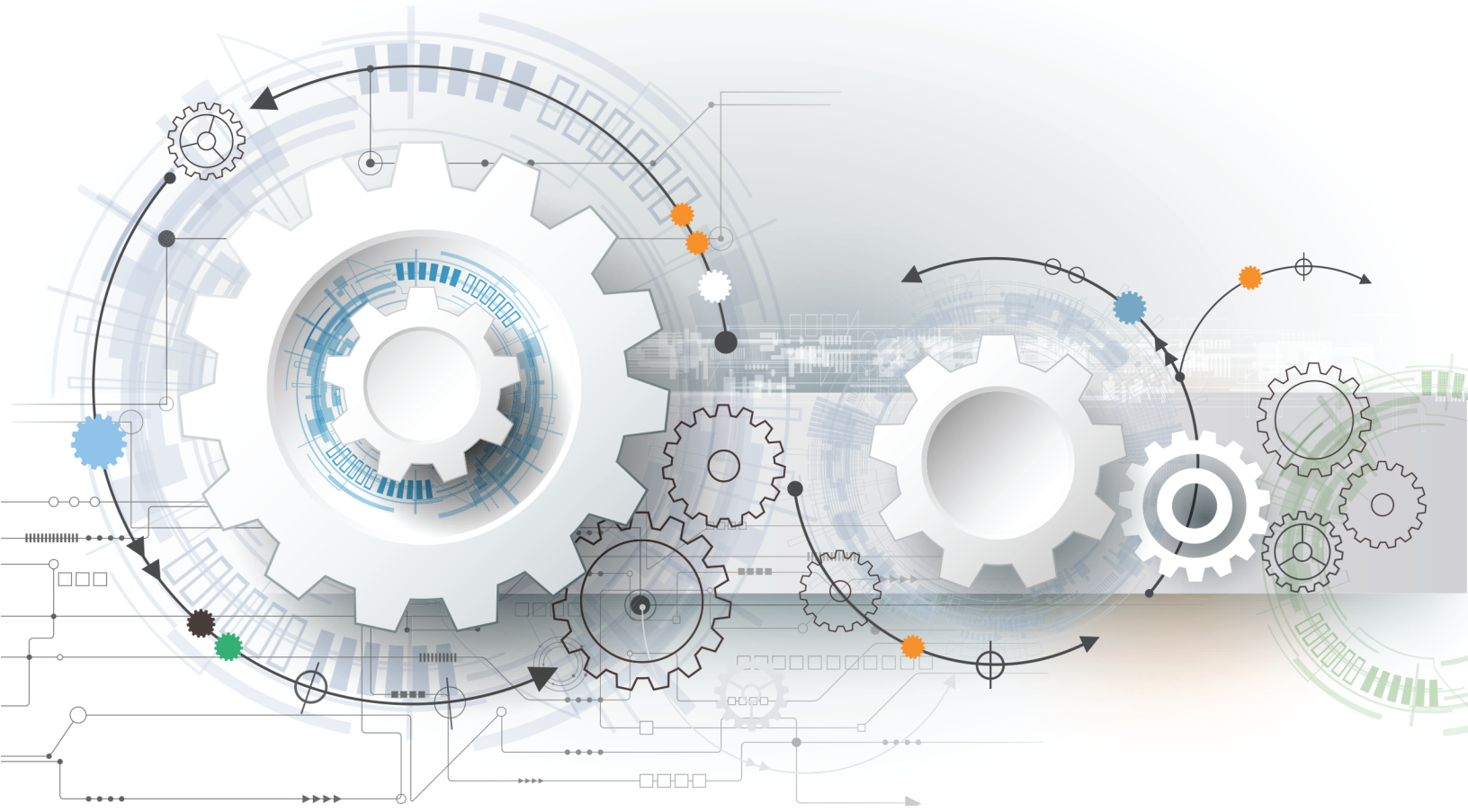
Working Paper
2020-01

Predicting Students' Chances of Completing a Degree:

How Does Superscoring Compare to Other Scoring Methods When Applicants Retest?

JUSTINE RADUNZEL, PHD, AND KRISTA MATTERN, PHD



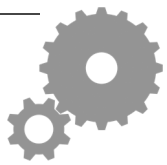
ACT.org

ABOUT THE AUTHORS

Justine Radunzel is a principal research scientist in Validity and Efficacy Research specializing in postsecondary outcomes research and validity evidence for the ACT test.

Krista Mattern is a senior director in Validity and Efficacy Research whose research focuses on predicting education and workplace success through evaluating the validity and fairness of cognitive and non-cognitive measures. Krista is also known for her work in evaluating the efficacy of learning products to help improve intended learner outcomes.

ACT WORKING PAPER SERIES



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

R1825

Abstract

Based on ACT and National Student Clearinghouse matched data for nearly 807,000 students from nearly 1,500 four-year postsecondary institutions and nearly 236,000 students from more than 900 two-year postsecondary institutions, this study evaluated the validity of various ACT® Composite scoring methods (average, highest, most recent, and superscoring) for identifying applicants who are likely to complete a degree. The study found that ACT Superscores were as predictive of completing a college degree in a timely manner as compared to the other scoring methods. The study also found that the likelihood of completing a degree for students who tested more often was underpredicted, but when examining the prediction accuracy by the number of times a student tested, superscoring resulted in the least amount of prediction error across the four scoring methods. These findings held for both the four- and two-year samples. Combined with previous research in this area on first-year college grades, this study provides evidence supporting ACT's new features of superscoring and section retesting as it suggests that selecting students' best subject scores from any test attempt (superscoring) results in a valid measure of a student's preparedness for future success.

Predicting Students' Chances of Completing a Degree:
How Does Superscoring Compare to Other Scoring Methods When Applicants Retest?

Colleges and universities evaluate applicants using a variety of factors that help inform the college admission process. For many institutions, standardized measures of academic achievement, such as the ACT® and SAT, are a critical component of that process. According to the *2019 State of College Admissions Report* of the National Association for College Admission Counseling (NACAC), 82.8% of colleges rated admission test scores as of "considerable" or "moderate" importance in the admission process (Clinedinst, 2019). Along with other academic factors such as high school grades and strength of curriculum, ACT and SAT scores are used to evaluate students' levels of readiness for college-level work and thus their likelihood of being successful in college if admitted. Validity evidence is essential to justify these uses and interpretations of admission test scores (AERA, APA, & NCME, 2014).

A substantial body of empirical research evidence exists illustrating the validity of test scores for predicting college outcomes (Allen, 2013; Allen & Sconing, 2005; Camara, Mattern, Croft, Vispoel, & Nichols, 2019; Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008; Mattern & Patterson, 2014; Radunzel & Noble, 2013; Sanchez, 2013; University of California Academic Senate, 2020; Westrick, Le, Robbins, Radunzel, & Schmidt, 2015). In particular, the findings from these studies indicate that students who earn higher ACT and/or SAT scores tend to earn higher grades in college in their first year as well as throughout their college career and return and graduate at higher rates as compared to students with lower test scores. These patterns of results hold nationally as well as by institutional characteristics (e.g., type, control, size, selectivity) and for student subgroups (e.g., gender, race/ethnicity, socioeconomic factors).

When evaluating the validity of test scores for use in college admissions, studies have typically used students' scores from their most recent, or latest, test event. However, this does

not align with how college and universities typically use ACT or SAT scores in practice. In a survey of score use policies, the College Board found that the majority of institutions use students' highest score from a single test attempt or take the highest subtest scores across all testing occasions and create a composite, which is often referred to as a "superscore" (The College Board, 2015). Other score use policies exist such as reviewing all scores (e.g., holistic review), test-optional, and test-flexible policies.

Trends in Retesting

How best to evaluate applicants who have multiple standardized test scores has become a more pressing issue as students' retesting patterns have shifted over time, and we now see more students applying to college having taken the ACT multiple times (Harmston & Crouse, 2016; Mattern & Radunzel, 2019). In particular, only roughly 1 out of 10 ACT-tested students took the ACT more than once in the early 1980s (Lanier, 1994). The percentage of students retesting jumped threefold over the course of a decade with around 30% of ACT-tested students retesting in 1992. Retesting rates continued to rise through the 90s and the early part of the 21st century with a retest rate of 44% for the 2018 ACT-tested graduate class (Mattern & Radunzel, 2019).

This recent growth in retesting has persisted despite coinciding with a substantial increase in the number of students taking the ACT as part of State and District testing, which includes a number of students who have no intentions of attending a postsecondary institution and/or use State and District testing as their sole testing event (Allen, 2015). Among a sample of over 275,000 students who enrolled in a four-year institution following high school graduation and where college grade data was available, retesting rates were higher; nearly 70% of those students took the ACT more than once (Mattern, Radunzel, Bertling, & Ho, 2018). In the fall of 2019, ACT announced that it will allow students to retest in a single subject as opposed to having to sit

for the full ACT test starting in September 2020, which may also impact retesting behavior going forward (ACT, 2019).

Validity Evidence of Admission Test Scores by Composite Scoring Method

Given that different score-use policies and practices are in place coupled with increasing retesting rates over time, understanding whether the validity of ACT scores for predicting college success varies by these different score-use policies is an important question that higher education professionals should consider when determining a score-use policy for their own institutions. A handful of studies have compared the predictive strength and/or predictive accuracy of different scoring methods of standardized admissions test scores and shed insight to this question. In general, research in this area suggests that:

1. The predictive strength of various scoring methods (first, last, highest, superscores) is fairly similar with correlations only differing by .01 to .02 (Boldt, Centra, & Courtney, 1986; Boldt, 1977; Linn, 1977; Mattern et al., 2018; Patterson, Mattern, & Swerdzewski, 2012; Roszkowski & Spreat, 2016). In general, the evidence suggests a slight advantage for average scores. When the predictive strength was estimated and evaluated by the number of test attempts, Mattern et al. (2018) also found a slight advantage for average scores but found that superscores were slightly superior when estimated over all students, irrespective of the number of test attempts.

2. As for predictive accuracy, average scores resulted in the most prediction error whereas superscores minimized the amount of prediction error among scoring methods examined (Boldt et al., 1986; Mattern et al., 2018). In particular, all scoring methods underpredicted college grades among retesters; however, average scores resulted in the greatest amount of underprediction, and superscores resulted in the

least amount of underprediction. Unfortunately, Roszkowski and Spreat (2016) examined predictive accuracy of the SAT Verbal and Math sections, separately, preventing the detection of differential prediction for various composite scores (including superscores), which are more commonly used in admissions practices as compared to individual subject scores.

In general, the research evidence supports the use of superscores in the college admission process which has led to ACT changing its position on superscoring. When evaluated from a predictive accuracy standpoint, the current evidence suggests that superscoring may be a slightly superior score-use policy as compared to the other methods examined. These findings run contrary to prior-held beliefs among many psychometricians and ACT's previous position on the issue that discouraged the use of superscoring due to concerns that superscoring would capitalize on positive measurement error and result in an overestimation of student's preparedness for future success were firmly ingrained (ACT, 2009). Findings across the studies suggest that the superiority of superscoring over other scoring methods may not be large, but the finding is consistent, and it would be hard to argue that this method is inferior to other scoring methods. That said, previous research has focused exclusively on one outcome of interest—course grades. Therefore, whether these results hold for other college outcomes is unknown and should be explored, which is the focus of the current study.

Current Study

This study extends previous research in this area by examining the predictive strength and predictive accuracy of four composite scoring methods (last, highest, average, superscoring) for another measure of student success—completing a college degree. Many scholars have argued that the true measure of student success is college graduation. Therefore, when

documenting the validity evidence of various college admission criteria, studies should evaluate not only their relationship with near-term, or proximal, indicators of college success such as first-year grade point average (GPA), but also more distal outcomes such as cumulative GPA and graduation. The current study addresses this research gap. Additionally, this is the first study to address this research question for both four- and two-year institutions. Even though two-year institutions tend to be open access and thus do not require a minimum test score for admissions, understanding whether test scores are related to future success at these types of institutions is still beneficial as test scores may be used for other purposes (e.g., course placement, identifying students in need of institutional supports and services, etc.).

Following the methodology for evaluating prediction accuracy in Mattern et al. (2018), this study explicitly tests for differential prediction by the number of retests (0, 1, 2, and 3 or more) and by different composite scoring methods. This is in contrast to some earlier studies (e.g., Boldt et al., 1986; Roszkowski & Spreat, 2016) that developed regression models for single testers and applied those models to retesters (under an assumption that outcomes for single testers are accurately predicted). Additionally, we evaluate the impact of retesting and scoring methods on validity and prediction accuracy based on ACT scores, alone and in combination with high school GPA (HSGPA), given that admission decisions are rarely, if ever, based solely on test scores. Finally, given that research has shown that students with low socioeconomic status are less likely to retest (Harmston & Crouse, 2016), we investigate the degree to which the differential prediction results can be explained by differential retesting rates by family income.

Data Source

The initial data consisted of all ACT-tested students who graduated from high school in 2010 and enrolled in a postsecondary institution in fall 2010. Initial college enrollment, subsequent enrollment, and degree completion information through summer 2017 was obtained

from National Student Clearinghouse. The sample included students beginning at both four- and two-year postsecondary institutions. However, the sample was limited to students who had ACT scores and HSGPA available, which was 96% of the initial sample. Analyses were conducted separately by institution type, where type was determined at the time of initial enrollment. Even though ACT test scores are generally not required for admissions to two-year institutions, two-year institutions often accept ACT test scores to help inform student academic readiness and course placement decisions. Moreover, it is not uncommon for students to begin college at a two-year institution and then transfer to a four-year institution. Therefore, we examined results for students attending both two- and four-year institutions. More than three-fourths (or 77.4%) of the analysis sample began at a four-year postsecondary institution: 806,795 students initially from 1,544 four-year institutions (referred to as the four-year sample) and 235,873 students initially from 925 two-year institutions (referred to as the two-year sample).

Table 1 provides a description of the institutions included in the four-year and two-year samples. The four-year institutions were diverse in terms of institutional control (37.9% public; 62.1% private), selectivity (32.8% highly selective/selective; 44.4% traditional; 22.1% liberal/open admissions policies), undergraduate enrollment size (58.7% had less than 5,000 undergraduates; 30.0% had 5,000 to under 20,000 undergraduates; 9.9% had 20,000 or more undergraduates) and location (25.1% from Northeast region; 26.2% from Midwest region; 33.6% South region; 14.6% West region). In comparison, a majority of the two-year institutions were public (90.8%) and had open/liberal admissions policies (99.2%). The two-year sample also had representation from across the four US census regions and by institution size (Table 1).

Table 1. Description of Institutions by Sample

Institution Characteristics	Categories	Four-Year Sample (k = 1,544) %	Two-Year Sample (k = 925) %
Institution control	Public	37.9	90.8
	Private	62.1	9.2
	Unknown	0.0	0.0
U.S. census region	Northeast	25.1	14.0
	Midwest	26.2	25.4
	South	33.6	36.1
	West	14.6	24.5
	Unknown	0.5	0.0
Selectivity	Highly selective/selective	32.8	0.0
	Traditional	44.4	0.3
	Open/liberal	22.1	99.2
	Unknown	0.7	0.4
Institution size	Under 1,000	10.6	5.3
	1,000 - 4,999	48.1	41.2
	5,000 - 9,999	17.4	26.6
	10,000 - 19,999	12.6	16.7
	20,000 and above	9.9	7.2
	Missing	1.4	3.0

Note. Characteristics for the postsecondary institutions were obtained from IPEDS, except for admissions selectivity. Admission selectivity was self-reported by institutions on the ACT Institutional Data Questionnaire (IDQ) as defined by the typical high school class ranks of their accepted freshmen: The majority of freshmen at highly selective schools are in the top 10%, selective in the top 25%, traditional in the top 50%, liberal in the top 75% of their high school class. Institutions with open admissions policies accept all high school graduates to limit of capacity.

The characteristics of the students in the two samples are summarized in Table 2. Female students made up more than one-half of each sample and White students nearly two-thirds of each sample. As compared to the four-year sample, the two-year sample had a higher percentage of students from families with an annual family income less than $36,000—though we note a significant percentage of students did not provide their annual family income at the time they registered to take the ACT (Table 2). The regional distributions did not differ drastically between the two samples. In terms of academic preparation, students in the four-year sample earned

higher ACT scores and HSGPAs on average than those in the two-year sample. They also tended to be more likely to test more than once with the ACT (54.7% vs. 36.7%).

Table 2. Description of ACT-Tested Students by Sample

Student Characteristics	Categories	Four-Year Sample (N=806,795) %	Two-Year Sample (N=235,873) %
Gender	Male	43.1	45.2
	Female	56.8	54.7
	Missing	0.1	0.1
Ethnicity	African American	11.5	13.9
	American Indian	0.8	1.1
	Hispanic	8.0	10.1
	Asian	4.9	3.0
	Multiracial/Other	2.6	2.6
	White	67.6	64.0
	Missing	4.6	5.3
Income	< $36,000	19.6	32.0
	$36,000 to $80,000	28.3	33.7
	> $80,000	29.8	16.7
	Missing	22.4	17.6
U.S. census region	Northeast	10.9	3.6
	Midwest	35.7	44.9
	South	37.1	36.5
	West	16.3	15.0
	Missing	0.0	0.0
Times tested	1 time	45.3	63.3
	2 times	32.7	25.7
	3 times	13.9	7.4
	4 or more times	8.1	3.6
	Mean	1.9	1.5
		Mean	Mean
Academic performance	Last ACT Composite score	23.1	18.8
	Average ACT Composite score	22.9	18.8
	Highest ACT Composite score	23.3	19.0
	Superscore ACT Composite score	23.6	19.2
	HSGPA	3.41	2.96

Note. HSGPA = high school grade point average.

Measures

ACT Composite Scores. ACT subject test scores—English, mathematics, reading, and science—from all testing administrations from a student's sophomore, junior, and senior year in

high school were obtained from the student's official ACT records. For each student in the sample, four different composite measures were calculated:

1. **Last ACT Composite score.** This score reflects the Composite score that the student earned on the last, or most recent, time they took the ACT.
2. **Average ACT Composite score.** This score is the average of all ACT Composite scores earned across test administrations/attempts, rounded to the nearest whole number.
3. **Highest ACT Composite score.** This score represents the highest ACT Composite score earned during a single administration.
4. **ACT Superscore.** This score takes the highest ACT section test scores (English, reading, mathematics, and science) across administrations and then computes the ACT Composite score based on those highest section test scores.

Number of ACT Administrations. This variable is the number of times a student took the ACT during their sophomore through senior year of high school, classified into four levels: one time, two times, three times, and four or more times. Students took the ACT 1.9 times on average for the four-year sample and 1.5 times for the two-year sample. We grouped four and more times together given the rapidly diminishing number of test-takers who took the ACT five or more times (3.0% and 1.2%, respectively).

HSGPA. HSGPA was obtained from responses to the ACT registration form, which asks students to self-report the coursework they have taken in English, mathematics, social studies, and science, and the grades earned in those courses (M = 3.41 and SD = 0.50 for four-year sample; M = 2.96, SD = 0.61 for two-year sample). Research has shown that students tend to

reliably report their coursework grades (Kuncel, Credé, & Thomas, 2005; Sanchez & Buddin, 2015).

Degree Completion. For the four-year sample, we examined earning a bachelor's degree from any institution within four years (100% of normal time) and within six years (150% of normal time) of initial enrollment. The corresponding rates for these two outcomes were 41.6% and 67.0%, respectively.

For the two-year sample, we examined earning an associate degree from any institution within three years (150% of normal time) of initial enrollment. Because some students beginning at a two-year institution transfer to a four-year institution before earning a degree, we also evaluated earning an associate or bachelor's degree within six years of initial enrollment for the two-year sample. The corresponding rates for these two outcomes were 18.5% and 40.2%, respectively.

Methods

A series of analyses were conducted to evaluate the prediction accuracy of scoring methods across retesting conditions. We began with estimating the bivariate point biserial correlations between the Composite measures and degree completion for each scoring method. Next, we fit a model that regressed the outcome (degree completion) on test scores. This model is referred to as the total group ACT score model. Then, as in our earlier paper (Mattern et al., 2018), to test for differential prediction by the number of times tested, we evaluated whether the intercepts and/or slopes of test score differed by number of testing occasions (Cleary, 1968). That is, a second model was fit that included the Composite score, indicators for the number of times tested, and interactions between the number of times tested and the Composite score. Students who took the ACT only one time served as the reference group for the number of times tested. In supplemental analyses, we added annual family income group categorized as shown in

Table 2 and the interaction between income group and the Composite score to the second model to examine the degree to which any differential prediction by number of times tested was explained by family income.

Hierarchical logistic regression models were developed to predict degree completion. Random intercept models were employed to account for students being nested within postsecondary institutions and to allow for the rates to vary across institutions. Separate models were developed for each scoring method. R^2 estimates were calculated for the generalized linear mixed effects models according to the formulas provided by Nakagawa and Schielzeth (2013; see Table 2, p. 139). To facilitate the interpretation of the results, the ACT Composite measures were centered at a value of 23 for the four-year sample and at a value of 19 for the two-year sample. For each ACT Composite measure value across the score range, prediction error was calculated by subtracting the expected probability of degree completion based on the overall total group model from the expected value based on the model that included retesting subgroup indicators and the interaction between the ACT Composite measure and the retesting indicators.[1]

Because most institutions use high school grades in addition to test scores for admissions and other purposes, additional regression models were estimated that included HSGPA as a predictor of degree completion. The extent to which differential prediction was mitigated by the inclusion of HSGPA was evaluated.

Results

Descriptive Statistics and Predictive Validity

In Table 3, the means, standard deviations, and intercorrelations of study variables are provided. As expected, ACT Superscores were the highest Composite measure (M = 23.6 for the four-year sample, and M=19.2 for the two-year sample) on average followed by highest ACT Composite scores (M = 23.3 for the four-year sample, and M = 19.0 for the two-year sample).

Average ACT Composite scores tended to be the lowest Composite measure for the four-year sample (M = 22.9). Average ACT Composite scores and last ACT Composite scores were the lowest Composite measures for the two-year sample (M = 18.8). The four scoring methods were highly correlated (*r*s ranging from .98 to .99 for both samples).

Table 3. Means, Standard Deviations, and Intercorrelations of Study Variables

#	Variable	M	SD	1	2	3	4	5
	Four-year sample (N = 806,795)							
1	Last	23.1	4.8					
2	Average	22.9	4.7	.98				
3	Highest	23.3	4.7	.99	.99			
4	Superscore	23.6	4.8	.98	.98	.99		
5	HSGPA	3.41	0.50	.53	.53	.54	.54	
6	Bachelor's degree within 4 years	41.6%		.36	.36	.36	.36	.34
7	Bachelor's degree within 6 years	67.0%		.35	.35	.35	.35	.37
	Two-year sample (N = 235,873)							
1	Last	18.8	3.9					
2	Average	18.8	3.8	.99				
3	Highest	19.0	3.9	.99	.99			
4	Superscore	19.2	3.9	.98	.98	.99		
5	HSGPA	2.96	0.61	.42	.42	.43	.44	
6	Associate degree within 3 years	18.5%		.19	.19	.20	.20	.24
7	Associate/bachelor's degree within six years	40.2%		.26	.25	.27	.27	.31

Note. All correlations are significant at $p < .0001$. M = mean or rate. SD = standard deviation. HSGPA = high school grade point average. Correlations with the binary degree completion outcomes are point biserial correlations.

For the four-year sample, the predictive strength of the four scoring methods were the same for both outcomes (*r* =.36 for four-year degree completion and .35 for six-year degree completion). For the two-year sample, the predictive strength of the four scoring methods were similar for both outcomes (*r*s ranging from .19 to .20 for three-year degree completion and .25 to .27 for six-year degree completion) with the ACT Superscore and the highest Composite score

showing the strongest relationship with degree completion. In comparison, the predictive strength of HSGPA on degree completion was estimated to be similar to that for the ACT Composite measures for the four-year sample (r =.34 and .37 for HSGPA, respectively) and slightly higher for the two-year sample (r =.24 and .31 for HSGPA, respectively).

Even when we examined the test-criterion correlations by number of testing occasions (Table 4), we found that the predictive strength of the four scoring methods were similar for both samples and outcomes. The predictive strength varied across scoring methods by at most .01 for the four-year sample and .02 for the two-year sample.

Table 4. Point Biserial Test-Criterion Correlations by Scoring Method and Times Tested

Number of Times Tested	N	Last	Average	Highest	Superscore
Four-year sample – Bach degree within 4 years / Bach degree within 6 years					
1	365,320	.36 / .35	.36 / .35	.36 / .35	.36 / .35
2	264,053	.36 / .35	.37 / .36	.37 / .35	.37 / .35
3	112,223	.34 / .33	.35 / .34	.35 / .33	.35 / .33
4 or more	65,199	.32 / .30	.33 / .31	.33 / .30	.33 / .30
Overall	806,795	.36 / .35	.36 / .35	.36 / .35	.36 / .35
Two-year sample – Assoc degree within 3 years / Assoc or bach degree within 6 years					
1	149,334	.18 / .24	.18 / .24	.18 / .24	.18 / .24
2	60,530	.20 / .27	.20 / .28	.20 / .27	.20 / .27
3	17,539	.20 / .27	.20 / .28	.20 / .27	.19 / .27
4 or more	8,470	.22 / .29	.22 / .28	.21 / .28	.22 / .28
Overall	235,873	.19 / .26	.19 / .25	.20 / .27	.20 / .27

Note. All point biserial correlations are significant at $p < .0001$.

The predictive validity of the scoring methods can also be compared by examining the ACT Composite slopes. Table 5 provides the estimated intercepts and slopes from the hierarchical logistic regression models relating the ACT Composite measures to degree completion, after accounting for the institution attended.

Table 5. Parameter Estimates (and Standard Errors) for Degree Completion Accounting for Institution Attended

Degree Completion Outcome	Parameter Estimate	Last	Average	Highest	Superscore
		Four-year sample			
4-year bachelor's	Intercept	-0.3439	-0.3245	-0.3669	-0.3969
		(0.0286)	(0.0285)	(0.0285)	(0.0285)
	Slope	0.1193	0.1205	0.1235	0.1252
		(0.0007)	(0.0007)	(0.0007)	(0.0007)
6-year bachelor's	Intercept	0.8334	0.8485	0.8148	0.7884
		(0.0252)	(0.0252)	(0.0250)	(0.0250)
	Slope	0.1175	0.1171	0.1232	0.1267
		(0.0007)	(0.0007)	(0.0007)	(0.0007)
		Two-year sample			
3-year associate	Intercept	-1.5720	-1.5634	-1.5911	-1.6127
		(0.0286)	(0.0286)	(0.0286)	(0.0286)
	Slope	0.1227	0.1230	0.1263	0.1284
		(0.0015)	(0.0015)	(0.0015)	(0.0015)
6-year associate or bachelor's	Intercept	-0.3535	-0.3451	-0.3725	-0.3947
		(0.0171)	(0.0170)	(0.0172)	(0.0173)
	Slope	0.1362	0.1361	0.1415	0.1450
		(0.0012)	(0.0012)	(0.0012)	(0.0012)

Note. All parameter estimates are significant at $p < .0001$. Each ACT Composite measure was centered at 23 for the four-year sample and at 19 for the two-year sample. The variance estimate (and standard error) associated with the random intercepts by outcome and scoring method are provided in Table A1 in Appendix A.

Even though the differences in the slopes across the scoring methods were small ($\leq .01$), the largest slope in magnitude was consistently associated with the superscoring method for each outcome and sample. The smallest slope was associated with either the last or the average scoring method. Taken together, these findings are consistent with previous research (e.g., Mattern, et al., 2018) indicating that the variation in the predictive strength of composite scoring methods is minimal.

Differential Prediction

For the topic of differential prediction, we focus within the body of this report on the results for the two outcomes associated with degree completion within six years—one for the four-year sample and one for the two-year sample. Results for the other two outcomes are provided in the Appendices.

Four-Year Sample. As shown in Table 6, the main effects and the interaction between the ACT Composite measure and the number of times tested were significantly related to six-year bachelor's degree completion for each scoring model (each overall $p < .0001$), revealing differential prediction of degree completion by the number of times a student tested. Overall, differential prediction for retesters based on the ACT Composite measures accounted for between 1.6% (for superscoring) to 3.4% (for average) of the variance in the log odds of degree completion (Table 6). By exponentiating the parameter estimates for the number of times tested, we can compare the odds of completing a degree for retesters compared to those testing only once. For example, for students with an ACT Composite score of 23, the odds of completing a bachelor's degree within six years was 1.3 (superscoring) to 1.5 (last and average) times greater for those who tested two times compared to those who tested only once, 1.8 (superscoring) to 2.3 (average) times greater for those who tested three times compared to those who tested only once, and 2.4 (superscoring) to 3.2 (average) times greater for those who tested four or more times compared to those who tested only once.[2]

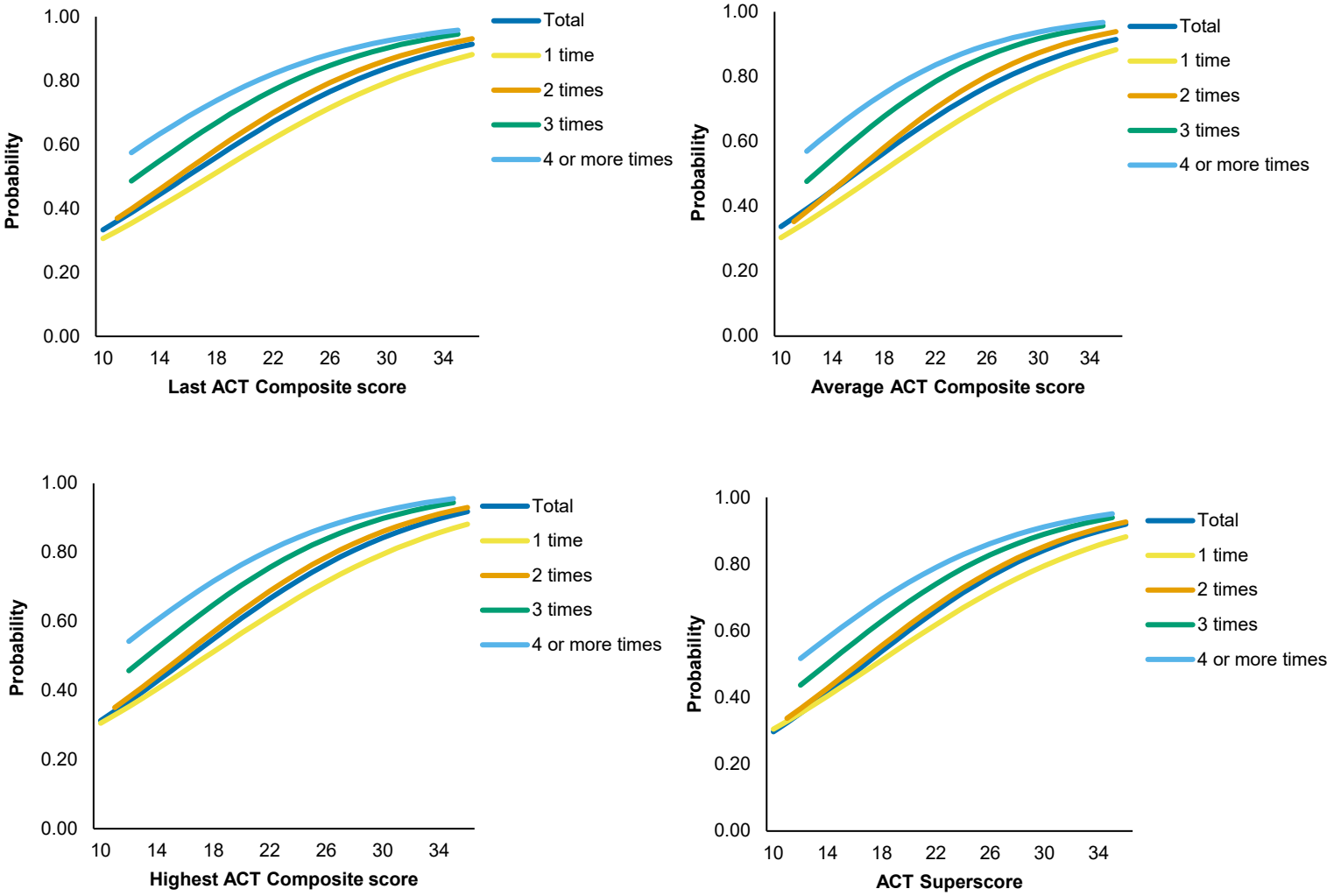
Table 6. Parameter Estimates (and Standard Errors) for the Six-Year Bachelor's Degree Completion Models for the Four-Year Sample

	Scoring Methods for ACT Composite Measure			
	Estimate (Standard Error)			
Model	Last	Average	Highest	Superscore
Intercept	0.5938 (0.0264)	0.5948 (0.0262)	0.5933 (0.0263)	0.5933 (0.0263)
ACT Composite	0.1087 (0.0010)	0.1096 (0.0010)	0.1088 (0.0010)	0.1089 (0.0010)
Times tested (2)	0.3761 (0.0066)	0.3999 (0.0067)	0.3275 (0.0066)	0.2703 (0.0065)
Times tested (3)	0.7422 (0.0093)	0.8385 (0.0097)	0.6707 (0.0091)	0.5854 (0.0091)
Times tested (4 or more)	1.0535 (0.0121)	1.1697 (0.0127)	0.9640 (0.0119)	0.8576 (0.0121)
ACT Composite * Times tested (2)	0.0168 (0.0015)	0.0238 (0.0015)	0.0192 (0.0015)	0.0192 (0.0015)
ACT Composite * Times tested (3)	0.0179 (0.0020)	0.0291 (0.0021)	0.0216 (0.0021)	0.0212 (0.0021)
ACT Composite * Times tested (4 or more)	0.0137 (0.0026)	0.0252 (0.0028)	0.0173 (0.0027)	0.0169 (0.0027)
ΔR^2 due to differential prediction	0.0271	0.0340	0.0217	0.0161

Note. Overall p values for the number of times tested and interaction between the number of times tested and ACT Composite score were both < .0001. p values for individual parameter estimates are ≤ 0.0001 unless noted otherwise: * p value ≤ 0.05; nonsignificant p values are **bolded**. Each ACT Composite measure was centered at 23. The variance estimate (and standard error) associated with the random intercepts was 0.9841 (0.0393) for Last, 0.9690 (0.0388) for Average, 0.9794 (0.0392) for Highest, and 0.9777 (0.0391) for Superscore.

Despite the fact that differential prediction of the Composite measures by the number of times tested accounted for a small fraction of the variance, the odds ratios for retesters and the probabilities presented in Figure 1 indicate that among students with the same ACT Composite score, those who retest more often are more likely to complete a bachelor's degree within six years than students who retest fewer times, even for the superscoring method.

Figure 1. Probability of Completing a Bachelor's Degree Within Six Years of Initial Enrollment by ACT Composite Scoring Method and Number of Times Tested for the Four-Year Sample

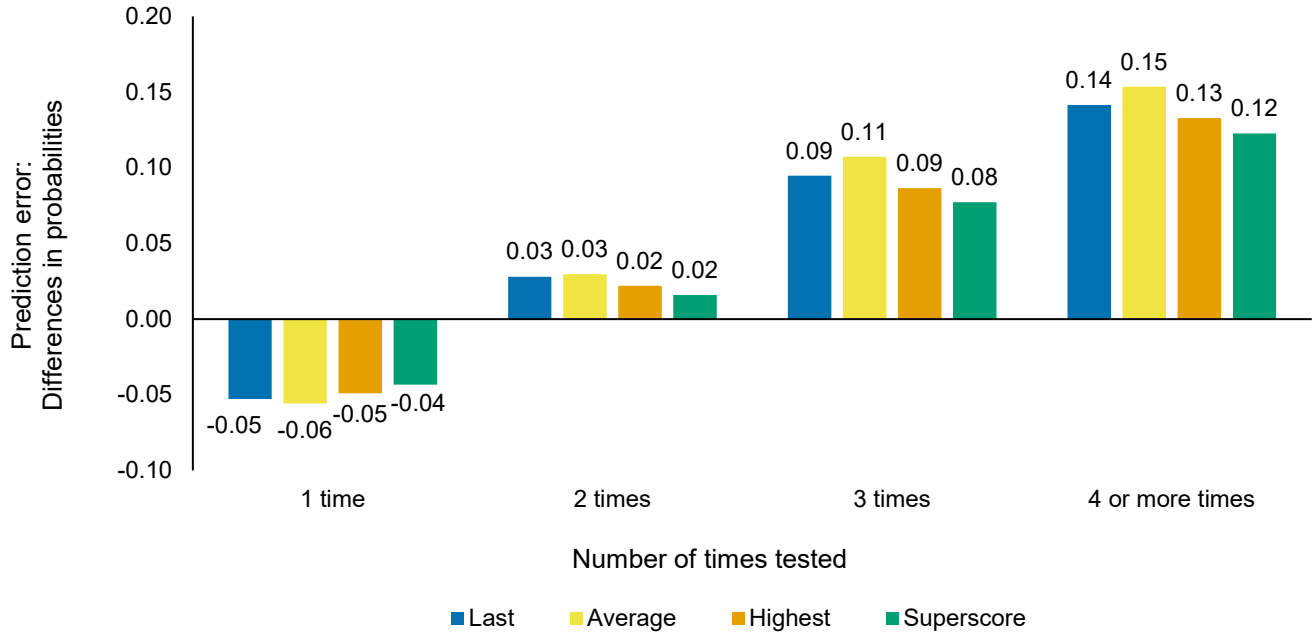


For each plot in Figure 1, the regression line for the total group (dark blue line) was provided to illustrate under- and overprediction of the probability of completing a degree by the number of times a student tested. For students who tested once, students' chances of completing a degree were overpredicted across the score range. That is, the regression line for students who tested once fell below the total group line. Therefore, predictions based on the total group model would overpredict students' chances of completing a college degree relative to a subgroup-

specific regression line. The magnitude of the overprediction varied somewhat across the ACT score scale range. These findings held for each of the scoring methods including for the superscoring method. On the other hand, the regression lines for the retesters were above the total group regression line, suggesting that students' chances of completing a degree for retesters were being underpredicted and that the magnitude of underprediction increased the more times the student tested.

Consistent with results from our earlier study (Mattern et al., 2018), the regression lines by number of times tested tend to be closest together for the superscoring and highest methods and furthest apart for the average scoring method (Figure 1). This is further illustrated in Figure 2 where the differences between a student's probability of completing a degree that takes into consideration the number of times the student tested and their predicted probability based only on their ACT Composite score are shown by ACT scoring method and the number of times tested. The values are based on an ACT Composite score of 23. As shown in Figure 2, the extent of over- and underprediction is minimized when predictions are based on the superscoring method. For example, among students who take the ACT four or more times, the magnitude of underprediction in students' predicted probabilities is 0.14, 0.15. 0.13, and 0.12 when using last, average, highest, and superscoring methods, respectively. Moreover, among retesters testing three times or testing four or more times with scores of 19 or above, the prediction error generally decreases as the ACT Composite score increases. However, across this score range, the magnitude of the prediction error for the superscoring method is consistently1 to 2 percentage points lower than when using the last method, 1 percentage point lower than when using the highest method, and 2 to 3 percentage points lower than when using the average method.

Figure 2. Magnitude of Differential Prediction on the Likelihood of Completing a Bachelor's Degree Within Six Years by Number of Times Tested and Four Composite Scoring Methods for the Four-Year Sample[3]



For the four-year sample, conclusions based on the results for the four-year bachelor's degree completion outcome were generally consistent with those reported for the six-year bachelor's degree completion outcome (see Table A2 and Figures A1 and A2 for corresponding table and figures). It was generally the case that the benefit in the reduction in the prediction error associated with the superscoring method as compared to the other scoring methods was slightly greater for this outcome. For instance, for retesters testing three times or four or more times across the score range of 19 or higher, the magnitude of the prediction error for the superscoring method was consistently 2 to 3 percentage points lower than when using the last method, 1 to 2 percentage points lower than when using the highest method, and 3 to 6 percentage points lower than when using the average method.

Two-Year Sample. Next, we examine the results from these same types of analyses for the two-year sample for completing an associate or bachelor's degree within six years of initial

enrollment. Corresponding tables and figures for the three-year associate degree completion outcome are provided in Appendix B.

As suggested in Table 7, differential prediction of degree completion existed by the number of times a student tested for the two-year sample as well; the main effects and the interaction between the ACT Composite measures and the number of times tested were significantly related to completing an associate or bachelor's degree within six years for each scoring model (each overall $p < .0001$). Overall, differential prediction for retesters based on ACT Composite score accounted for between 2.0% (for superscoring) to 3.5% (for average) of the variance in the log odds of degree completion (Table 7). For students with an ACT Composite score of 19, the odds of completing an associate or bachelor's degree within six years was 1.4 (superscoring) to 1.6 (last and average) times greater for those who tested two times compared to those who tested only once, 2.0 (superscoring) to 2.7 (average) times greater for those who tested three times compared to those who tested only once, and 2.9 (superscoring) to 4.3 (average) times greater for those who tested four or more times compared to those who tested only once.[4]

Table 7. Parameter Estimates (and Standard Errors) for the Six-Year Associate or Bachelor's Degree Completion Models for the Two-Year Sample

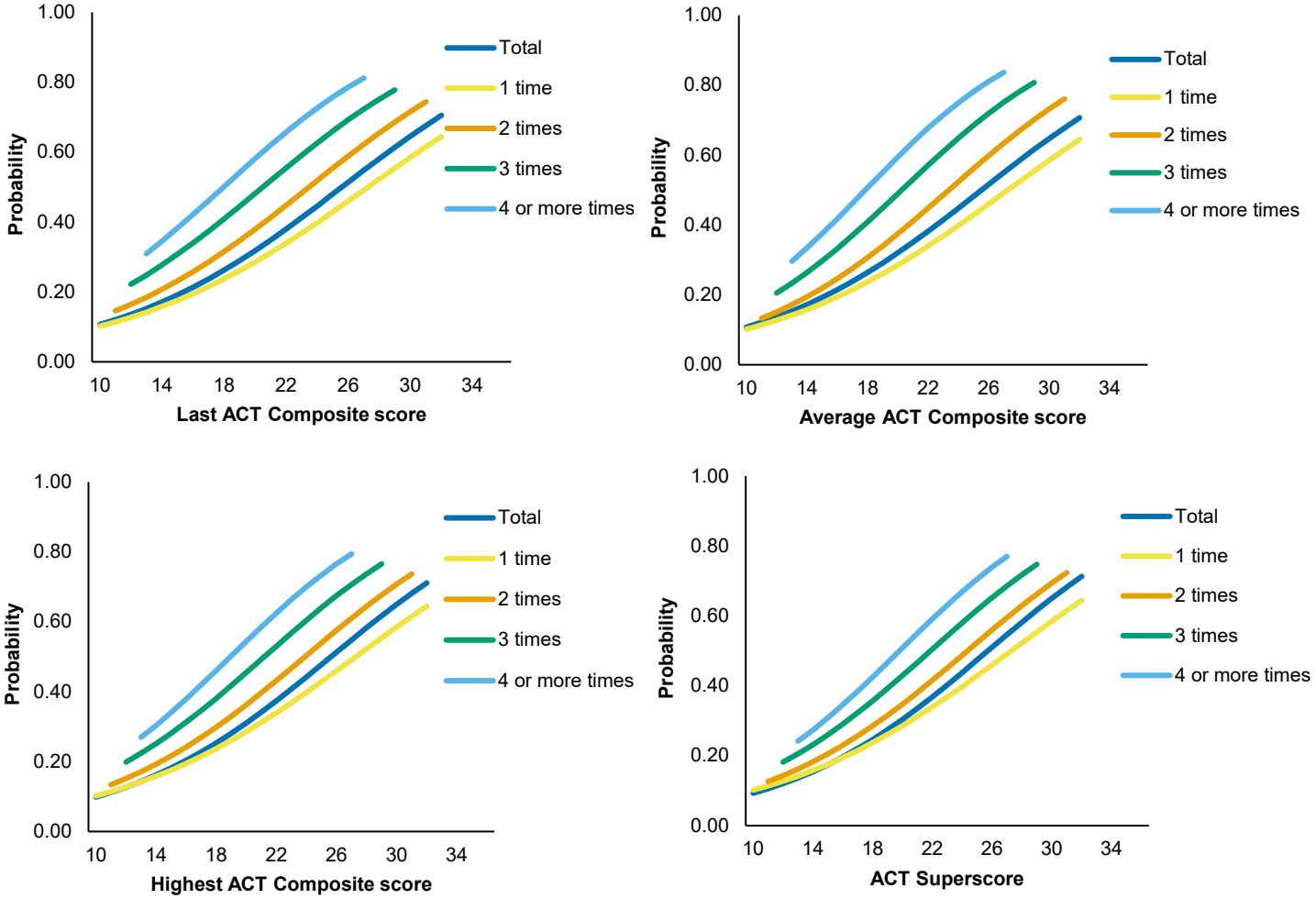
	Scoring Methods for ACT Composite Measure			
	Estimate (Standard Error)			
Model	Last	Average	Highest	Superscore
Intercept	-0.5395 (0.0194)	-0.5387 (0.0194)	-0.5393 (0.0194)	-0.5394 (0.0194)
ACT Composite	0.1260 (0.0015)	0.1261 (0.0015)	0.1260 (0.0015)	0.1260 (0.0015)
Times tested (2)	0.4718 (0.0108)	0.4817 (0.0108)	0.4096 (0.0109)	0.3447 (0.0111)
Times tested (3)	0.9083 (0.0180)	0.9886 (0.0182)	0.8130 (0.0183)	0.7120 (0.0191)
Times tested (4 or more)	1.3536 (0.0261)	1.4573 (0.0265)	1.2193 (0.0265)	1.0771 (0.0284)
ACT Composite * Times tested (2)	0.0157 (0.0028)	0.0256 (0.0029)	0.0189 (0.0029)	0.0189 (0.0029)
ACT Composite * Times tested (3)	0.0216 (0.0047)	0.0378 (0.0052)	0.0258 (0.0050)	0.0264 (0.0050)
ACT Composite * Times tested (4 or more)	0.0358 (0.0069)	0.0524 (0.0079)	0.0417 (0.0074)	0.0422 (0.0075)
ΔR^2 due to differential prediction	0.0305	0.0346	0.0254	0.0202

Note: Overall *p* values for the number of times tested and interaction between the number of times tested and ACT Composite score were both < .0001. *p* values for individual parameter estimates are ≤ 0.0001 unless noted otherwise: * *p* value ≤ 0.05; nonsignificant *p* values are **bolded**. Each ACT Composite measure was centered at 19. The variance estimate (and standard error) associated with the random intercepts was 0.2582 (0.0158) for Last, 0.2568 (0.0157) for Average, 0.2579 (0.0157) for Highest, and 0.2576 (0.0157) for Superscore.

Despite the fact that differential prediction by the number of times tested accounted for a small fraction of the variance, the odds ratios for retesters and the probabilities presented in Figure 3 indicate that among students beginning at a two-year institution with the same ACT Composite score, those who retested more often were more likely to complete an associate or

bachelor's degree within six years than students who retested fewer times. This finding was seen across the four scoring methods, even for the superscoring method.

Figure 3. Probability of Completing an Associate or Bachelor's Degree Within Six Years of Initial Enrollment by ACT Composite Scoring Method and Number of Times Tested for the Two-Year Sample

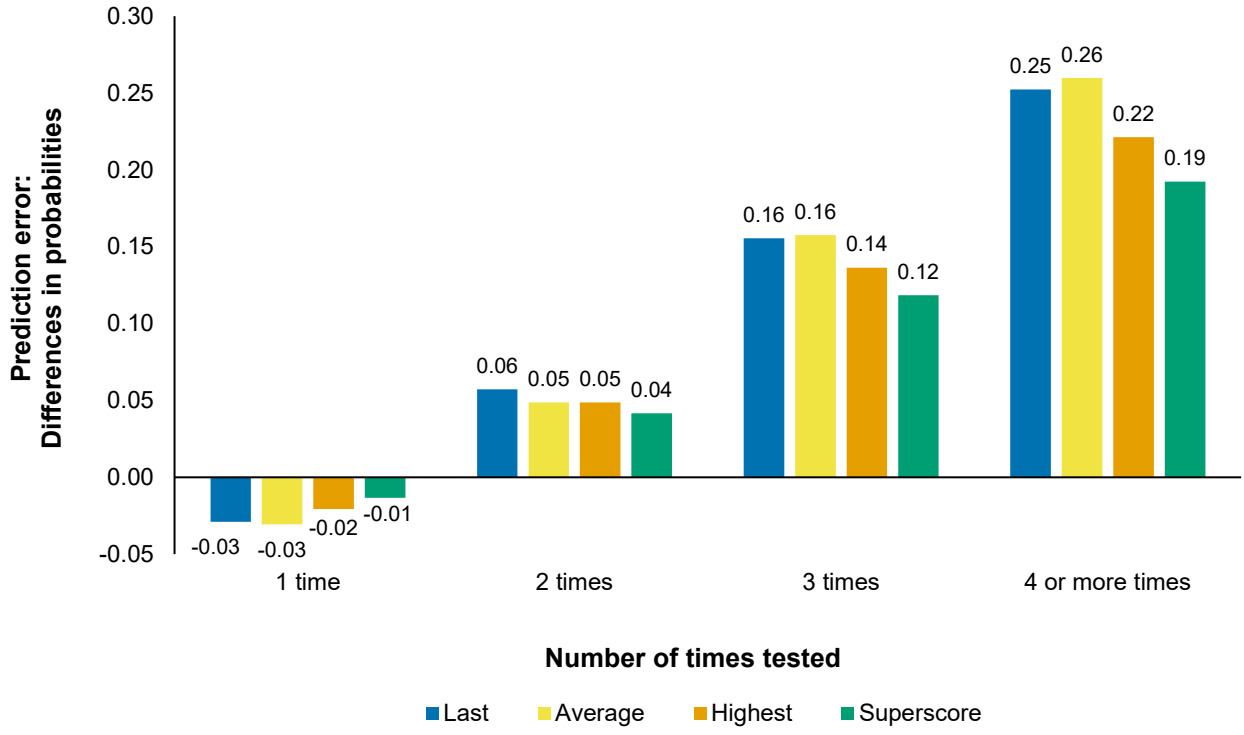


For students who tested once from the two-year sample, students' chances of completing an associate or bachelor's degree was generally overpredicted by the total group ACT Composite score line (dark blue line) with the magnitude of the overprediction increasing as ACT Composite score increased (Figure 3). These findings held for each of the scoring methods including for the superscoring method. On the other hand, the regression lines for the retesters were above the total group regression line, suggesting that students' chances of completing an associate or bachelor's degree for retesters were being underpredicted by the total group line and that the magnitude of underprediction increased the more times the students tested. This finding means that retesters were more likely to complete an associate or bachelor's degree than what is predicted based on their ACT Composite score.

Consistent with results for the four-year sample, the regression lines by number of times tested tend to be closest together for the superscoring method and furthest apart for the average scoring method (Figure 3). This is further illustrated in Figure 4 where the differences in probabilities between accounting for and not accounting for the number of times tested are shown by ACT scoring method and the number of times tested. The values are based on an ACT Composite score of 19. As shown in Figure 4, the extent of over- and underprediction is minimized when predictions are based on the superscoring method. For example, among students who take the ACT four or more times, the magnitude of underprediction in students' predicted probabilities is 0.25, 0.26. 0.22, and 0.19 when using last, average, highest, and superscoring methods, respectively. Moreover, among retesters testing four or more times with scores between 13 and 27, the prediction error generally increases as the ACT Composite score increases. Moreover, across this score range for those testing four or more times, the magnitude of the prediction error for the superscoring method is consistently 4 to 6 percentage points lower than

when using the last method, 2 to 3 percentage points lower than when using the highest method, and 4 to 7 percentage points lower than when using the average method.

Figure 4. Magnitude of Differential Prediction on the Likelihood of Completing an Associate or Bachelor's Degree Within Six Years by Number of Times Tested and Four Composite Scoring Methods for the Two-Year Sample[5]



For the two-year sample, conclusions based on the results for the three-year associate degree completion outcome were generally consistent with those reported for the six-year associate or bachelor's degree completion outcome (see Table B1 and Figures B1 and B2 for corresponding table and figures). It was generally the case that the reduction in the prediction error associated with the superscoring method as compared to the other scoring methods was slightly smaller for this outcome than for the six-year degree completion outcome. For instance, for retesters testing four or more times with scores ranging between 13 and 27, the magnitude of the prediction error for the superscoring method tended to be 2 to 4 percentage points lower than

when using the last method, 1 to 2 percentage points lower than when using the highest method, and 2 to 6 percentage points lower than when using the average method.

Statistically Controlling for Annual Family Income. Previous research on low-income students has found that their chances of long-term college success including completing a degree tend to be overpredicted based on ACT scores (Radunzel & Noble, 2013) and that they are less likely to retake the ACT (Harmston & Crouse, 2016; Mattern & Radunzel, 2019) and less likely to complete a college degree (Radunzel & Noble, 2012) than their peers.[6] Even in the current study, students from families with an annual income of less than $36,000 were less likely to retake the ACT than students from families with an annual income of more than $80,000 by 9 percentage points for both samples (51.6% vs. 60.2% and 34.7% vs. 43.9%, for four- and two-year samples, respectively).

Supplemental analyses were conducted to determine whether the differential prediction by number of times tested was simply an artifact of students' annual family income levels. We note that demographic characteristics such as annual family income are typically not included in admission models; but, to examine this topic, we added income group to the models. More specifically, the models included indicators for the number of times tested and for the income groups as well as the interactions between these indicators and the ACT Composite measures. The results from these analyses suggested that the number of times tested explained additional variance in the outcomes beyond income. First, the main effects and the interaction between the ACT Composite measure and the number of times tested were significantly related to the six-year degree completion outcomes for each scoring model after statistically controlling for differences in the intercepts and ACT Composite slopes by income group. Moreover, differential prediction by number of times tested accounted for a similar percentage of the variance in the log odds of degree completion as those shown in Tables 6 and 7.[7] These findings indicate that the

number of times tested provides meaningful explained variance in the degree completion outcomes above and beyond income. Moreover, when we examined the magnitude of the differential prediction by the number of times tested within an income group, the superscoring method resulted in the least amount of prediction error across the scoring methods.

Differential Prediction Accounting for HSGPA

Given that HSGPA also plays a prominent role in the college admission decision process, we examined the differential prediction analyses accounting for HSGPA. More specifically, we added HSGPA to the four ACT Composite score regression models and evaluated the extent to which differential prediction by number of times tested was reduced with the inclusion of this more motivationally laden construct in the model (e.g., Camara, O'Connor, Mattern, & Hanson, 2015; Mattern, Allen, & Camara, 2016).

Four-Year Sample. The percentage of variance in the log odds of six-year bachelor's degree completion explained by the joint model that included ACT Composite score and HSGPA ranged from 12.7% (average) to 13.6% (superscore) across the scoring methods; adding HSGPA increased the percentage of variance explained by 5.5 (superscore) to 6.0 (average) percentage points. As shown in Table 8, the main effects and the interaction between the ACT Composite measure and the number of times tested were significantly related to six-year bachelor's degree completion for each scoring model (each overall $p < .0001$), even after statistically controlling for HSGPA. This finding suggests that differential prediction of degree completion by the number of times a student tested existed, even after accounting for HSGPA. Adding the number of times tested and the interaction between number of times tested and ACT Composite score to the joint ACT Composite-HSGPA model explained an additional 0.9% (superscore) to 2.0% (average) of the variance in the log odds of degree completion (Table 8); these percentages were reduced by 41.7% (last) to 44.1% (superscore) over those based on the ACT Composite score

models (Table 6). Moreover, the adjusted odds ratios of completing a bachelor's degree within six years were generally reduced when HSGPA was added to the model, especially for students who tested four or more times. For example, for students with an ACT Composite score of 23, the adjusted odds of completing a bachelor's degree was 1.3 (superscoring and highest) to 1.4 (last and average) times greater for those who tested two times compared to those who tested only once, 1.7 (superscoring) to 2.0 (average) times greater for those who tested three times compared to those who tested only once, and 2.1 (superscoring) to 2.5 (average) times greater for those who tested four or more times compared to those who tested only once.[8] These findings indicate that the superscoring ACT Composite method remained the scoring method that exhibited the least amount of differential prediction, even after statistically controlling for HSGPA.

Table 8. Parameter Estimates (and Standard Errors) for the Six-Year Bachelor's Degree Completion Models for the Four-Year Sample Accounting for HSGPA

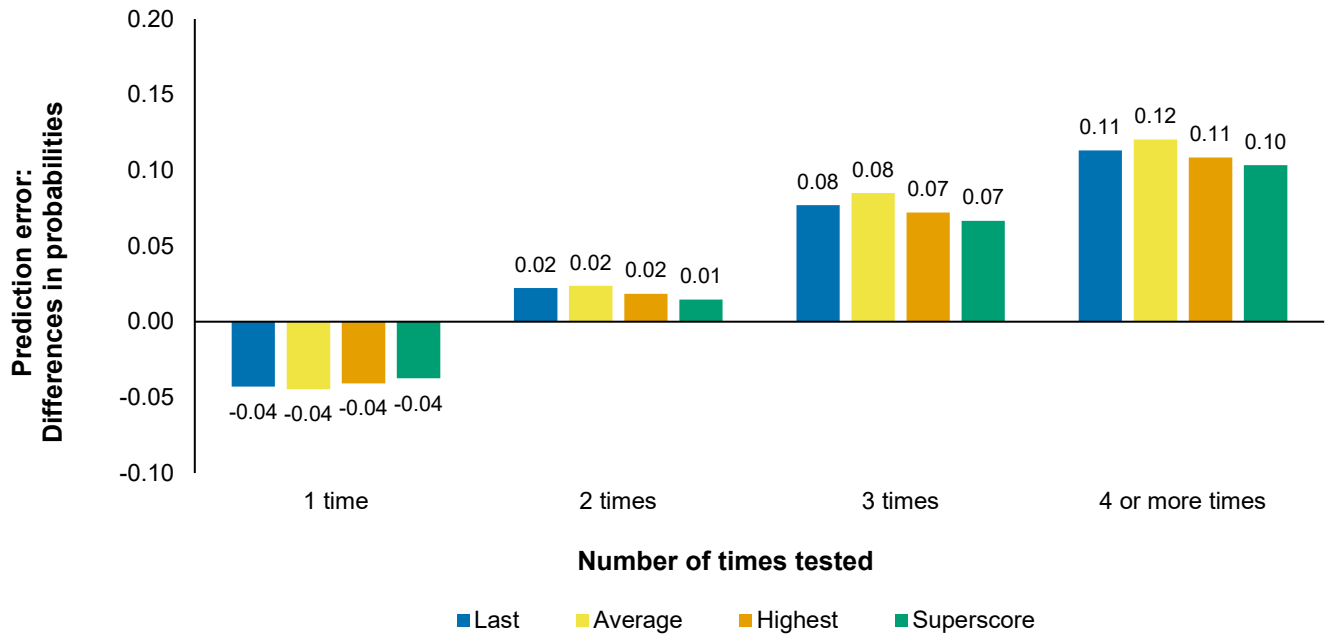
	Scoring Methods for ACT Composite Measure			
	Estimate (Standard Error)			
Model	Last	Average	Highest	Superscore
Intercept	0.6965 (0.0255)	0.6962 (0.0254)	0.6963 (0.0254)	0.6963 (0.0254)
ACT Composite	0.0656 (0.0010)	0.0664 (0.0010)	0.0655 (0.0010)	0.0655 (0.0010)
HSGPA	1.0583 (0.0067)	1.0512 (0.0067)	1.0613 (0.0067)	1.0609 (0.0067)
Times tested (2)	0.3117 (0.0068)	0.3267 (0.0068)	0.2811 (0.0067)	0.2457 (0.0067)
Times tested (3)	0.6132 (0.0094)	0.6705 (0.0098)	0.5704 (0.0093)	0.5206 (0.0093)
Times tested (4 or more)	0.8445 (0.0123)	0.9054 (0.0129)	0.7966 (0.0121)	0.7420 (0.0123)
ACT Composite * Times tested (2)	0.0125 (0.0015)	0.0168 (0.0015)	0.0138 (0.0015)	0.0137 (0.0015)
ACT Composite * Times tested (3)	0.0091 (0.0045)	0.0155 (0.0021)	0.0107 (0.0021)	0.0103 (0.0021)
ACT Composite * Times tested (4 or more)	**-0.0009** (0.0026)	**0.0040** (0.0028)	**-0.0006** (0.0027)	**-0.0010** (0.0027)
ΔR^2 due to differential prediction	0.0158	0.0197	0.0124	0.0090

Note: Overall p value for the number of times tested and for the interaction term between the number of times tested and ACT Composite score was < .0001. p values for parameter estimates are ≤ 0.0001 unless noted otherwise: * p value ≤ 0.05; nonsignificant p values are **bolded**. Each ACT Composite measure was centered at 23 and HSGPA was centered at 3.41. The variance estimate (and standard error) associated with the random intercepts was 0.9070 (0.0365) for Last, 0.8990 (0.0362) for Average, 0.9044 (0.0364) for Highest, and 0.9040 (0.0364) for Superscore.

This result is further illustrated in Figure 5 for students with an ACT Composite measure of 23 and HSGPA of 3.41. For this scenario, among students who took the ACT four or more times, the magnitude of underprediction in students' predicted probabilities was 0.11, 0.12. 0.11, and 0.10 when using last, average, highest, and superscoring methods, respectively. Compared to

the results shown in Figure 2 for students with an ACT Composite score of 23, the inclusion of HSGPA reduced the magnitude of underprediction by 1 to 3 percentage points for those who tested three times and by 2 to 3 percentage points for those who tested four or more times. Additionally, among retesters testing three times or testing four or more times with scores of 19 or above with an HSGPA of 3.41, the magnitude of the prediction error for the superscoring method was consistently1 to 2 percentage points lower than when using the last method, 1 percentage point lower than when using the highest method, and 1 to 3 percentage points lower than when using the average method. These percentage point differences were the largest among students with higher ACT Composite scores.

Figure 5. Magnitude of Differential Prediction on the Likelihood of Completing a Bachelor's Degree Within Six Years After Statistically Controlling for HSGPA by Number of Times Tested and Four Composite Scoring Methods for the Four-Year Sample[9]



For the four-year sample, conclusions based on the results for the four-year bachelor's degree completion outcome statistically controlling for HSGPA were generally consistent with those reported for the six-year bachelor's degree completion outcome (see Table A3 and Figure

A3 for corresponding table and figure). It was generally the case that the reduction in the prediction error associated with the superscoring method as compared to the other scoring methods was slightly greater for this outcome, especially among students who tested four or more times. For instance, for retesters testing four or more times across the score range of 19 or higher with a HSGPA of 3.41, the magnitude of the prediction error for the superscoring method was consistently 2 to 3 percentage points lower than when using the last method, 1 percentage points lower than when using the highest method, and 2 to 5 percentage points lower than when using the average method. These percentage point differences were the largest among students with higher ACT Composite scores.

Two-Year Sample. Next, we examine the results from these same types of analyses for the two-year sample for completing an associate or bachelor's degree within six years of initial enrollment. Corresponding tables and figures for the three-year associate degree completion outcome are provided in Appendix B.

The percentage of variance in the log odds of six-year associate or bachelor's degree completion explained by the joint model that included ACT Composite score and HSGPA ranged from 15.4% (average) to 16.0% (superscore) across the scoring methods; adding HSGPA increased the percentage of variance explained by 7.4 (superscore) to 8.1 (average) percentage points. As shown in Table 9, the main effects and the interaction between ACT Composite score and the number of times tested were significantly related to six-year associate or bachelor's degree completion for each scoring model (each overall $p < .0001$), even after statistically controlling for HSGPA. This finding suggests that differential prediction on degree completion existed by the number of times a student tested even after accounting for HSGPA for the two-year sample.

Table 9. Parameter Estimates (and Standard Errors) for the Six-Year Associate or Bachelor's Degree Completion Models for the Two-Year Sample Accounting for HSGPA

	Scoring Methods for ACT Composite Measure			
	Estimate (Standard Error)			
Model	Last	Average	Highest	Superscore
Intercept	-0.5751 (0.0198)	-0.5747 (0.0198)	-0.5754 (0.0198)	-0.5754 (0.0198)
ACT Composite	0.0720 (0.0016)	0.0722 (0.0016)	0.0719 (0.0016)	0.0718 (0.0016)
HSGPA	0.9813 (0.0094)	0.9791 (0.0094)	0.9844 (0.0094)	0.9846 (0.0094)
Times tested (2)	0.3793 (0.0111)	0.3851 (0.0111)	0.3435 (0.0112)	0.3065 (0.0114)
Times tested (3)	0.7498 (0.0185)	0.7954 (0.0186)	0.6966 (0.0187)	0.6402 (0.0196)
Times tested (4 or more)	1.1305 (0.0267)	1.1865 (0.0271)	1.0572 (0.0272)	0.9819 (0.0291)
ACT Composite * Times tested (2)	0.0103* (0.0029)	0.0150 (0.0030)	0.0113 (0.0029)	0.0110* (0.0029)
ACT Composite * Times tested (3)	0.0127* (0.0048)	0.0202 (0.0052)	0.0129* (0.0050)	0.0131* (0.0051)
ACT Composite * Times tested (4 or more)	0.0198* (0.0070)	0.0231* (0.0079)	0.0194* (0.0075)	0.0184* (0.0075)
ΔR^2 due to differential prediction	0.0180	0.0193	0.0158	0.0136

Note: Overall *p* value for the number of times tested and for the interaction term between the number of times tested and ACT Composite score was < .0001. *p* values for parameter estimates are ≤ 0.0001 unless noted otherwise: * *p* value ≤ 0.05; nonsignificant *p* values are **bolded**. Each ACT Composite score was centered at 19 and HSGPA was centered at 2.96. The variance estimate (and standard error) associated with the random intercepts was 0.2681 (0.0163) for Last, 0.2671 (0.0162) for Average, 0.2681 (0.0163) for Highest, and 0.2679 (0.0163) for Superscore.

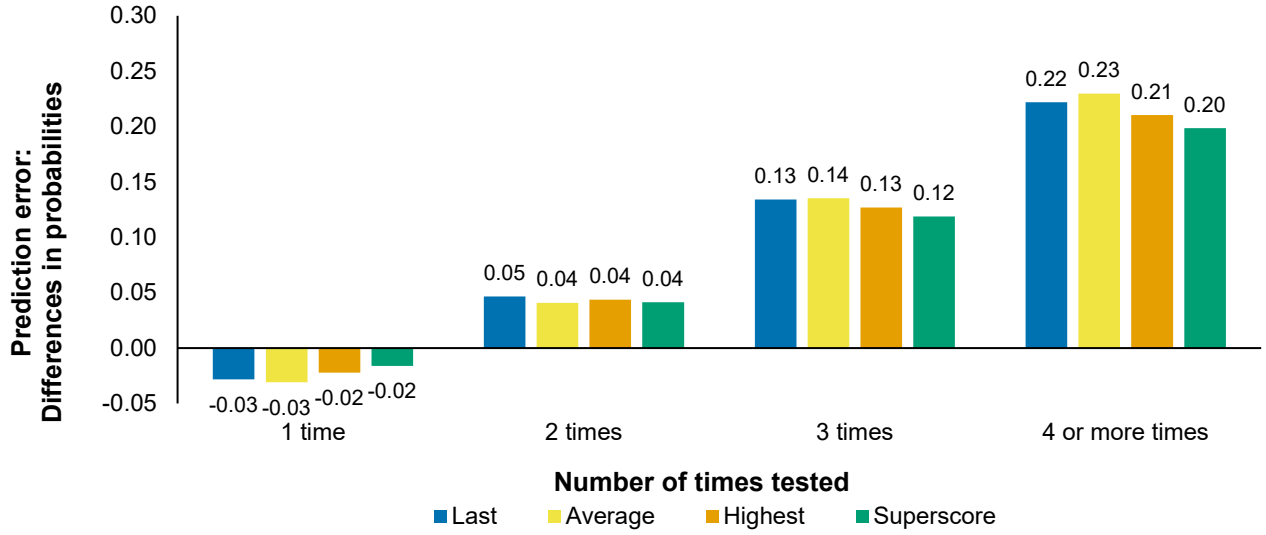
Adding the number of times tested and the interaction between number of times tested and ACT Composite score to the joint ACT Composite-HSGPA model explained an additional 1.4% (superscore) to 1.9% (average) of the variance in the log odds of degree completion (Table 9); these percentages were reduced by 32.7% (superscore) to 44.2% (average) over those based

on the ACT Composite score models (Table 7). Moreover, the adjusted odds ratios of completing an associate or bachelor's degree within six years of initial enrollment were generally reduced when HSGPA was added to the model, especially for students who tested four or more times. For example, for students with an ACT Composite score of 19, the adjusted odds of completing an associate or bachelor's degree was 1.4 (superscoring and highest) to 1.5 (last and average) times greater for those who tested two times compared to those who tested only once, 1.9 (superscoring) to 2.2 (average) times greater for those who tested three times compared to those who tested only once, and 2.7 (superscoring) to 3.3 (average) times greater for those who tested four or more times compared to those who tested only once.[10] These findings indicate that the superscoring method remained the scoring method that exhibited the least amount of differential prediction, even after statistically controlling for HSGPA.

This result is further illustrated in Figure 6 for students with ACT Composite scores of 19 and HSGPA of 2.96. For this scenario, among students who took the ACT four or more times, the magnitude of underprediction in students' predicted probabilities was 0.22, 0.23. 0.21, and 0.20 when using last, average, highest, and superscoring methods, respectively. Compared to the results shown in Figure 4 for students with an ACT Composite score of 19, the inclusion of HSGPA reduced the magnitude of underprediction for the last and average method by 2 to 3 percentage points for those who tested three times or tested four or more times. In comparison, the corresponding magnitude of underprediction was more similar between the two for the superscore and highest methods. Moreover, across the score range between 13 and 27 among retesters testing three times or testing four or more times with a HSGPA of 2.96, the magnitude of the prediction error for the superscoring method was consistently 1 to 4 percentage points lower than when using the last method, 1 to 2 percentage points lower than when using the highest method, and 1 to 5 percentage points lower than when using the average method. These

percentage point differences were the largest among students with higher ACT Composite scores.

Figure 6. Magnitude of Differential Prediction on the Likelihood of Completing an Associate or Bachelor's Degree Within Six Years After Statistically Controlling for HSGPA by Number of Times Tested and Four Composite Scoring Methods for the Two-Year Sample[11]



For the two-year sample, conclusions based on the results for the three-year associate degree completion outcome statistically controlling for HSGPA were generally consistent with those reported for the six-year associate or bachelor's degree completion outcome (see Table B2 and Figure B3 for the corresponding table and figure). It was generally the case that the reduction in the prediction error associated with the superscoring method as compared to the other scoring methods was slightly smaller for this outcome than for the six-year degree completion outcome, especially among students who tested four or more times. For instance, for retesters testing four or more times with scores ranging between 13 and 27, the magnitude of the prediction error for the superscoring method tended to be 1 to 2 percentage points lower than when using the last method, 1 percentage points lower than when using the highest method, and 1 to 2 percentage points lower than when using the average method. These percentage point differences were the largest among students with higher ACT Composite scores.

Conclusion

In conclusion, this large multi-institutional study provides support for the use of ACT Superscores to help inform college admissions, scholarship, placement, and intervention decisions. First, the current study found that ACT Superscores were as predictive of students' chances of completing a degree as other ACT Composite scoring methods that included computing the average Composite score across test administrations or using students' most recent Composite score or their highest Composite score. Second, the study found that the likelihood of completing a college degree for students who tested more often was underpredicted. That is, retesters performed better in college than what was expected based on their test scores. However, the prediction error when examined by the number of times a student tested was lower when Superscores were used instead of the other scoring methods. For students who tested only once, the superscoring method also resulted in the least amount of overprediction of a student's likelihood of completing a college degree. These findings held for multiple degree completion outcomes and for both the four- and two-year samples, as well as when statistically controlling for HSGPA. We also found that the differential prediction result by the number of times tested was not fully explained by students' income levels. Inclusion of the number of times tested and its interaction with ACT score explained variance in the log odds of the degree completion outcomes above and beyond students' income levels. Moreover, within each of the annual family income groups, the superscoring method resulted in the least amount of prediction error by number of times tested across the scoring methods.

Results from the current study were found to be consistent with those from our earlier study (Mattern et al., 2018; Mattern & Radunzel, 2019) that involved nearly 278,000 students attending 221 four-year institutions but utilized first-year grades as the outcome instead of degree completion. Taken together, findings from these studies as well as those from another

recent predictive validity study on section retesting (Radunzel & Mattern, 2020) suggest that ACT subject scores do not have to come from a single test attempt to be a valid indicator of students' college readiness, supporting both superscoring and section retesting. Moreover, the option of superscoring is in alignment with current admissions practices and policies at many postsecondary institutions. Superscoring allows students to demonstrate their academic achievement most favorably for college applications and scholarships (e.g., Cruce & Mattern, 2020). For example, in the current study, the average ACT Superscore was 0.4 to 0.5 points higher than the average latest ACT Composite measure and 0.2 to 0.3 points higher than the average highest ACT Composite measure (Table 2).

Beginning in September 2020, students who have tested more than once will have the option to send an ACT Superscore report to postsecondary institutions of their choice. The ACT Superscore report will include all test data for every test event included in the ACT Superscore, as well as the highest ACT Composite score from a single administration. Inclusion of students' highest full test scores is supported by the current study and the Mattern et al. (2018) study given that the highest method generally resulted in smaller magnitudes of differential prediction by number of times tested than the last and average scoring methods (see Figure 2 from Mattern et al. (2018) and Figures 2, 4, 5, 6, A2, A3, B2, and B3 from the current study). Additionally, the superscoring and highest methods align with how college and universities typically use ACT or SAT scores in practice (College Board, 2015).

While the inclusion of HSGPA helped reduce the magnitude of differential prediction on the likelihood of degree completion by the number of times tested, a considerable amount of prediction error remained for retesters testing three times or four or more times for both samples. These students tended to be substantially more likely to complete a degree than what was predicted based on the regression models. As suggested in the Mattern et al. (2018) study, their

underprediction may be attributable to academic behaviors that reflect behavioral manifestations of the latent construct of academic motivation (Camara, et al., 2015). Unfortunately, in the current study we did not have a pure measure of academic motivation available for a large enough sample to examine this topic. Future research should examine what factors, such as academic motivation, are related to retesting because results from such a study could potentially shed light on the development of new noncognitive admission measures.

Once the section retesting and superscoring options become operational for the ACT in September 2020, ACT is committed to monitoring how students' retesting behaviors and strategies change and whether these changes have any impact on students' ACT performance or on equity and fairness. Moreover, ACT will continue its efforts in conducting national validity studies and working with individual institutions and state systems to provide evidence supporting the use of ACT section retest scores and ACT Superscores, in combination with other measures, for college admission and course placement decisions and for identification of students who may benefit from additional academic services and supports once they matriculate to college.

References

ACT. (2009). *ACT: The first fifty years, 1959-2009*. Iowa City, IA: ACT.

ACT. (2019, October 8). *ACT test to provide new options designed to offer students more choices, faster results* [Press release]. Retrieved from http://leadershipblog.act.org/2019/10/act-test-to-provide-new-options.html

AERA, APA, and NCME. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association, American Psychological Association, National Council on Measurement in Education.

Allen, J. (2013). *Updating the ACT College Readiness Benchmarks.* Iowa City, IA: ACT.

Allen, J. (2015). *How many students use ACT State and District testing to take their sole ACT test*? Iowa City, IA: ACT.

Allen, J., & Sconing, J. (2005). *Using ACT assessment scores to set benchmarks for college readiness*. Iowa City, IA: ACT.

Boldt, R. F. (1977). *Evaluation of three methods for treating repeaters' scores*. Princeton, NJ: Law School Admission Council.

Boldt , R. F., Centra, J. A., & Courtney, R. G. (1986). *The validity of various methods of treating multiple SAT® scores.* New York, NY: College Board.

Camara, W. J., Mattern, K., Croft, M., Vispoel, S., & Nichols, P. (2019). A validity argument in support of the use of college admissions test scores for federal accountability. *Educational Measurement: Issues and Practice*, *38*(4), 12−26.

Camara, W., O'Connor, R., Mattern, K., & Hanson, M. A. (2015). *Beyond academics: A holistic framework for enhancing education and workplace success.* Iowa City, IA: ACT.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement*, *5*(2), 115−124.

Clinedinst, M. E. (2019). *2019 state of college admission.* Arlington, VA: National Association for College Admission Counseling.

The College Board. (2015). *SAT® score-use practices by participating institution.* New York, NY: The College Board. Retrieved from http://professionals.collegeboard.com/profdownload/sat-score-use-practices-list.pdf.

Cruce, T., & Mattern, K. (2020). *The impact of superscoring on the distribution of ACT scores.* Iowa City, IA: ACT.

Harmston, M. & Crouse, J. (2016). *Multiple testers: What do we know about them?* Iowa City, IA: ACT.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., and Barbuti, S. M. (2008). *Validity of the SAT® for predicting first-year college grade point average.* New York, NY: The College Board.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, *75*(1), 63−82.

Lanier, C. W. (1994). *ACT Composite scores of retested students*. Iowa City, IA: ACT.

Linn, R. L. (1977). On the treatment of multiple scores for law school admission test repeaters (Report #LSAC–77–4). In Law School Admission Council, *Reports of LSAC sponsored research: Volume III, 1975−1977* (pp. 777−784). Princeton, NJ: Law School Admission Council.

Mattern, K., Allen, J., & Camara, W. (2016). Thoughts on a multidimensional middle school index of college readiness. *Educational Measurement: Issues and Practice, 35*(3), 30−34.

Mattern, K., & Patterson, B. (2014). *Synthesis of recent SAT validity findings: Trend data over time and cohorts.* New York, NY: The College Board.

Mattern, K., & Radunzel, J. (2019). *Does superscoring increase subgroup differences?* Iowa City, IA: ACT.

Mattern, K., Radunzel, J., Bertling, M., & Ho, A. D. (2018). How should colleges treat multiple admissions test scores? *Educational Measurement: Issues and Practice, 37*(3), 11−23.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142. doi: 10.1111/j.2041-210x.2012.00261.x

Patterson, B., Mattern, K., & Swerdzewski, P. (2012). Are the best scores the best scores for predicting college success? *Journal of College Admission, 217*, 34−45.

Radunzel, J., & Mattern, K. (2020). *A case study: ACT section retest scores and superscores are predictive of first-term grades.* Iowa City, IA: ACT.

Radunzel, J., & Noble. J. (2012). *Tracking 2003 ACT-tested high school graduates: College readiness, enrollment, and long-term success*. Iowa City, IA: ACT.

Radunzel, J., & Noble. J. (2013). *Differential effects on student demographic groups of using ACT® College Readiness Assessment Composite score, ACT Benchmarks, and high school grade point average for predicting long-term college success through degree completion*. Iowa City, IA: ACT.

Roszkowski, M., & Spreat, S. (2016). Retaking the SAT may boost scores but this doesn't hurt validity. *Journal of the National College Testing Association, 2*(1), 1−16.

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

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

University of California Academic Senate. (2020). *Report of the UC Academic Council Standardized Testing Task Force (STTF).* Retrieved from https://senate.universityofcalifornia.edu/_files/underreview/sttf-report.pdf.

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

Appendix A

Table A1. Variance Estimates (and Standard Errors) of Random Intercepts from ACT Composite Models

Degree Completion Outcome	Parameter Estimate	Last	Average	Highest	Superscore
		Four-year sample			
4-year bachelor's	Intercept	1.1653	1.1566	1.1561	1.1560
		(0.0464)	(0.0461)	(0.0460)	(0.0460)
6-year bachelor's	Intercept	0.9008	0.8975	0.8893	0.8836
		(0.0363)	(0.0362)	(0.0359)	(0.0356)
		Two-year sample			
3-year associate	Intercept	0.5959	0.5971	0.5972	0.5978
		(0.0361)	(0.0361)	(0.0362)	(0.0362)
6-year associate or bachelor's	Intercept	0.1966	0.1948	0.1993	0.2032
		(0.0125)	(0.0124)	(0.0127)	(0.0129)

Note. All variance estimates are significant at $p < .0001$. The parameter estimates for the ACT Composite score models are shown in Table 5.

Table A2. Parameter Estimates (and Standard Errors) for the Four-Year Bachelor's Degree Completion Models for the Four-Year Sample

	Scoring Methods for ACT Composite Measure			
	Estimate (Standard Error)			
Model	Last	Average	Highest	Superscore
Intercept	-0.4978 (0.0292)	-0.4941 (0.0290)	-0.4972 (0.0291)	-0.4968 (0.0291)
ACT Composite	0.1170 (0.0010)	0.1179 (0.0010)	0.1172 (0.0010)	0.1173 (0.0010)
Times tested (2)	0.2180 (0.0064)	0.2325 (0.0064)	0.1686 (0.0065)	0.1123 (0.0066)
Times tested (3)	0.4663 (0.0086)	0.5467 (0.0086)	0.3963 (0.0088)	0.3134 (0.0091)
Times tested (4 or more)	0.6810 (0.0110)	0.7827 (0.0108)	0.5955 (0.0113)	0.4935 (0.0120)
ACT Composite * Times tested (2)	0.0054 (0.0014)	0.0120 (0.0014)	0.0079 (0.0014)	0.0076 (0.0014)
ACT Composite * Times tested (3)	0.0038* (0.0019)	0.0142 (0.0020)	0.0073* (0.0019)	0.0073* (0.0019)
ACT Composite * Times tested (4 or more)	**-0.0014** (0.0023)	0.0100 (0.0025)	**0.0028** (0.0024)	**0.0023** (0.0024)
ΔR^2 due to differential prediction	0.0063	0.0157	0.0080	0.0048

Note. Overall *p* values for the number of times tested and interaction between the number of times tested and ACT Composite score were both < .0001. *p* values for individual parameter estimates are ≤ .0001 unless noted otherwise: * *p* value ≤ .05; nonsignificant *p* values are **bolded**. Each ACT Composite measure was centered at 23. The variance estimate (and standard error) associated with the random intercepts was 1.2084 (0.0479) for Last, 1.1912 (0.0472) for Average, 1.2018 (0.0476) for Highest, and 1.2004 (0.0476) for Superscore.

Figure A1. Differential Prediction by ACT Composite Scoring Method and Number of Times Tested for Four-Year Bachelor's Degree Completion for the Four-Year Sample

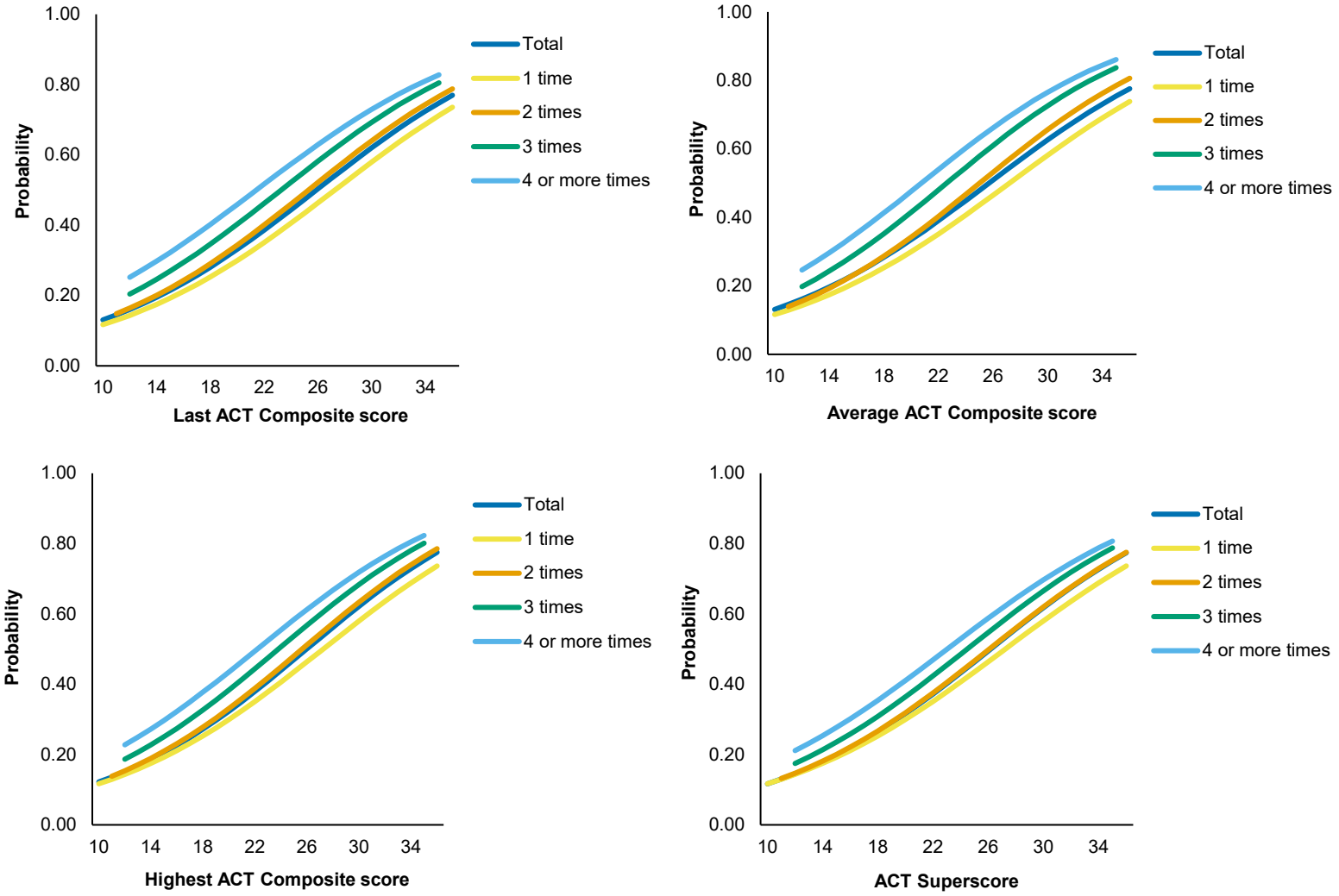


Figure A2. Magnitude of Differential Prediction on the Likelihood of Completing a Bachelor's Degree Within Four Years by Number of Times Tested and Four Composite Scoring Methods for the Four-Year Sample[12]

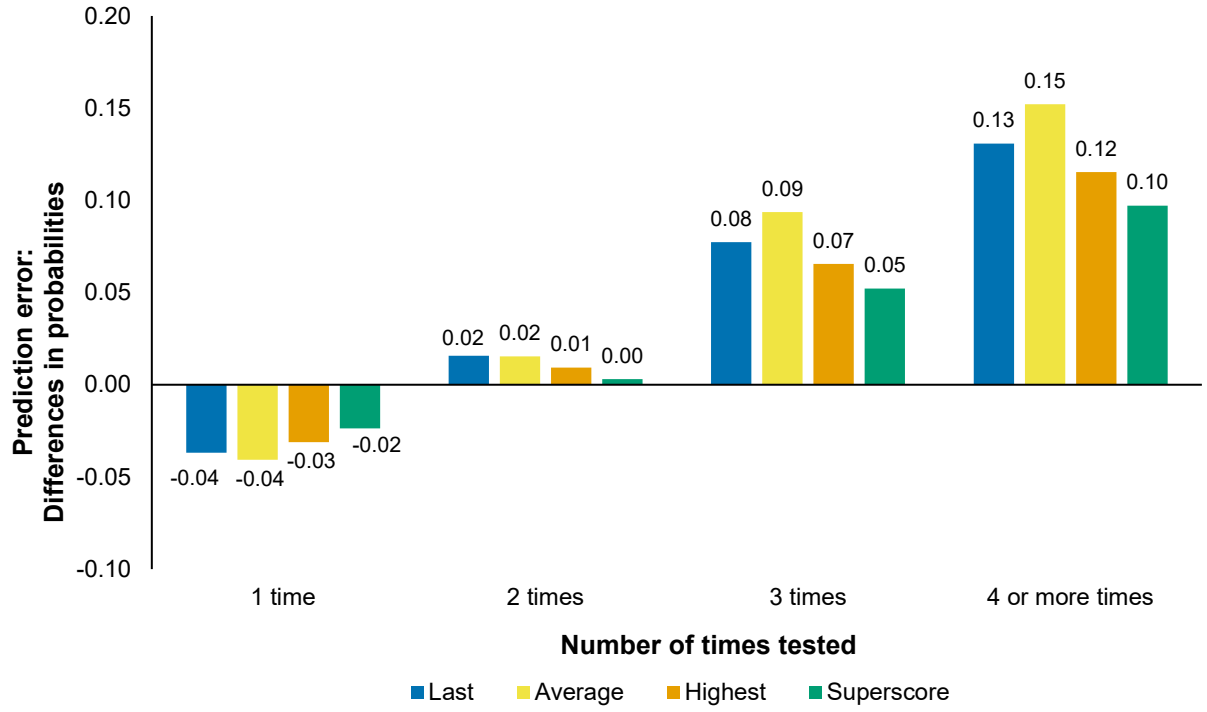
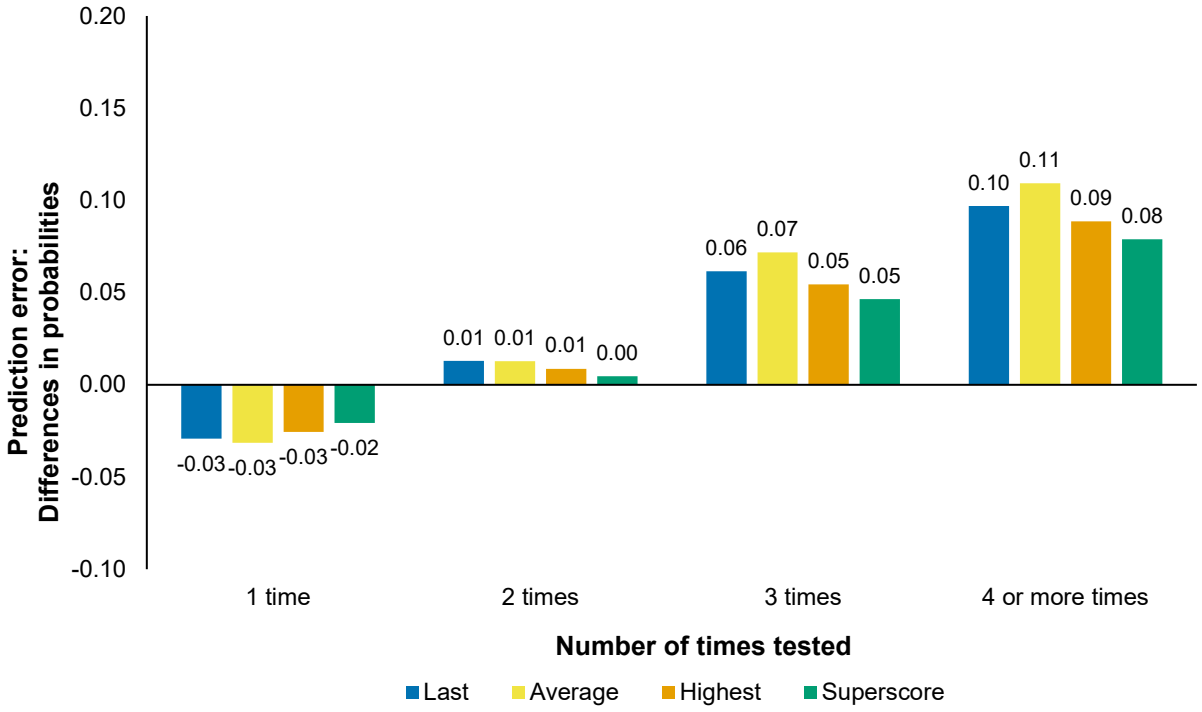


Table A3. Parameter Estimates (and Standard Errors) for the Four-Year Bachelor's Degree Completion Models for the Four-Year Sample Accounting for HSGPA

	Scoring Methods for ACT Composite Measure			
	Estimate (Standard Error)			
Model	Last	Average	Highest	Superscore
Intercept	-0.4487 (0.0288)	-0.4463 (0.0287)	-0.4483 (0.0288)	-0.4480 (0.0288)
ACT Composite	0.0765 (0.0010)	0.0775 (0.0010)	0.0766 (0.0010)	0.0766 (0.0010)
HSGPA	1.0821 (0.0072)	1.0724 (0.0072)	1.0835 (0.0072)	1.0828 (0.0072)
Times tested (2)	0.1737 (0.0065)	0.1829 (0.0065)	0.1417 (0.0066)	0.1053 (0.0067)
Times tested (3)	0.3703 (0.0088)	0.4201 (0.0087)	0.3269 (0.0090)	0.2755 (0.0093)
Times tested (4 or more)	0.5119 (0.0112)	0.5708 (0.0110)	0.4636 (0.0115)	0.4058 (0.0122)
ACT Composite * Times tested (2)	**0.0026** (0.0014)	0.0072 (0.0015)	0.0043* (0.0014)	0.0040* (0.0014)
ACT Composite * Times tested (3)	**-0.0019** (0.0019)	0.0050 (0.0020)	**0.0002** (0.0020)	**0.0003** (0.0020)
ACT Composite * Times tested (4 or more)	-0.0107 (0.0023)	**-0.0037** (0.0025)	-0.0088* (0.0024)	-0.0091* (0.0024)
ΔR^2 due to differential prediction	0.0056	0.0076	0.0038	0.0021

Note: Overall p value for the number of times tested was < .0001. Overall p value for the interaction term between the number of times tested and ACT Composite score was < .05. p values for parameter estimates are ≤ 0.0001 unless noted otherwise: * p value ≤ 0.05; nonsignificant p values are **bolded**. Each ACT Composite measure was centered at 23 and HSGPA was centered at 3.41. The variance estimate (and standard error) associated with the random intercepts was 1.1709 (0.0464) for Last, 1.1602 (0.0460) for Average, 1.1667 (0.0463) for Highest, and 1.1659 (0.0463) for Superscore.

Figure A3. Magnitude of Differential Prediction on the Likelihood of Completing a Bachelor's Degree Within Four Years After Statistically Controlling for HSGPA by Number of Times Tested and Four Composite Scoring Methods for the Four-Year Sample[13]



Appendix B

Table B1. Parameter Estimates (and Standard Errors) for the Three-Year Associate Degree Completion Models for the Two-Year Sample

Model	Scoring Methods for ACT Composite Measure Estimate (Standard Error)			
	Last	Average	Highest	Superscore
Intercept	-1.7134 (0.0292)	-1.7126 (0.0292)	-1.7134 (0.0292)	-1.7135 (0.0292)
ACT Composite	0.1187 (0.0019)	0.1189 (0.0019)	0.1187 (0.0019)	0.1188 (0.0019)
Times tested (2)	0.3378 (0.0138)	0.3442 (0.0138)	0.2860 (0.0141)	0.2305 (0.0145)
Times tested (3)	0.6647 (0.0215)	0.7241 (0.0210)	0.5928 (0.0223)	0.5174 (0.0237)
Times tested (4 or more)	0.9794 (0.0288)	1.0504 (0.0280)	0.8929 (0.0305)	0.7931 (0.0333)
ACT Composite * Times tested (2)	**0.0027** (0.0033)	0.0101* (0.0034)	**0.0049** (0.0034)	**0.0050** (0.0034)
ACT Composite * Times tested (3)	**-0.0036** (0.0052)	**0.0071** (0.0055)	**-0.0007** (0.0053)	**-0.0017** (0.0054)
ACT Composite * Times tested (4 or more)	**-0.0083** (0.0067)	**0.0019** (0.0074)	**-0.0054** (0.0071)	**-0.0037** (0.0071)
ΔR^2 due to differential prediction	0.0146	0.0169	0.0118	0.0088

Note: Overall p value for the number of times tested was < .0001. Overall p value for the interaction between the number of times tested and ACT Composite score was > 0.05. p values for individual parameter estimates are ≤ 0.0001 unless noted otherwise: * p value ≤ 0.05; nonsignificant p values are **bolded**. Each ACT Composite measure was centered at 19. The variance estimate (and standard error) associated with the random intercepts was 0.6123 (0.0370) for Last, 0.6126 (0.0370) for Average, 0.6132 (0.0370) for Highest, and 0.6131 (0.0370) for Superscore.

Figure B1. Differential Prediction by ACT Composite Scoring Method and Number of Times Tested for Three-Year Associate Degree Completion for the Two-Year Sample

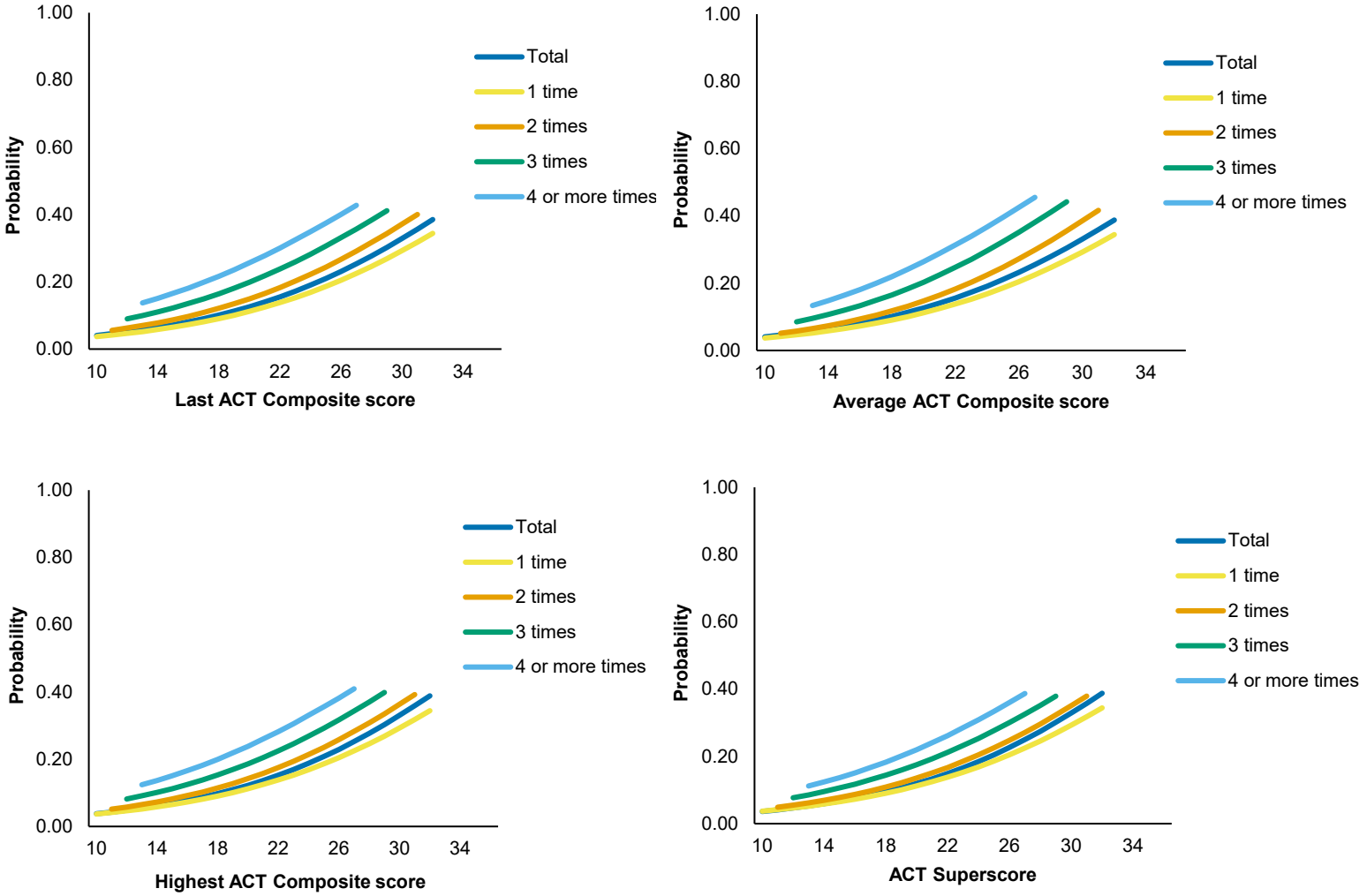


Figure B2. Magnitude of Differential Prediction on the Likelihood of Completing an Associate Degree Within Three Years by Number of Times Tested and Four Composite Scoring Methods for the Two-Year Sample[14]

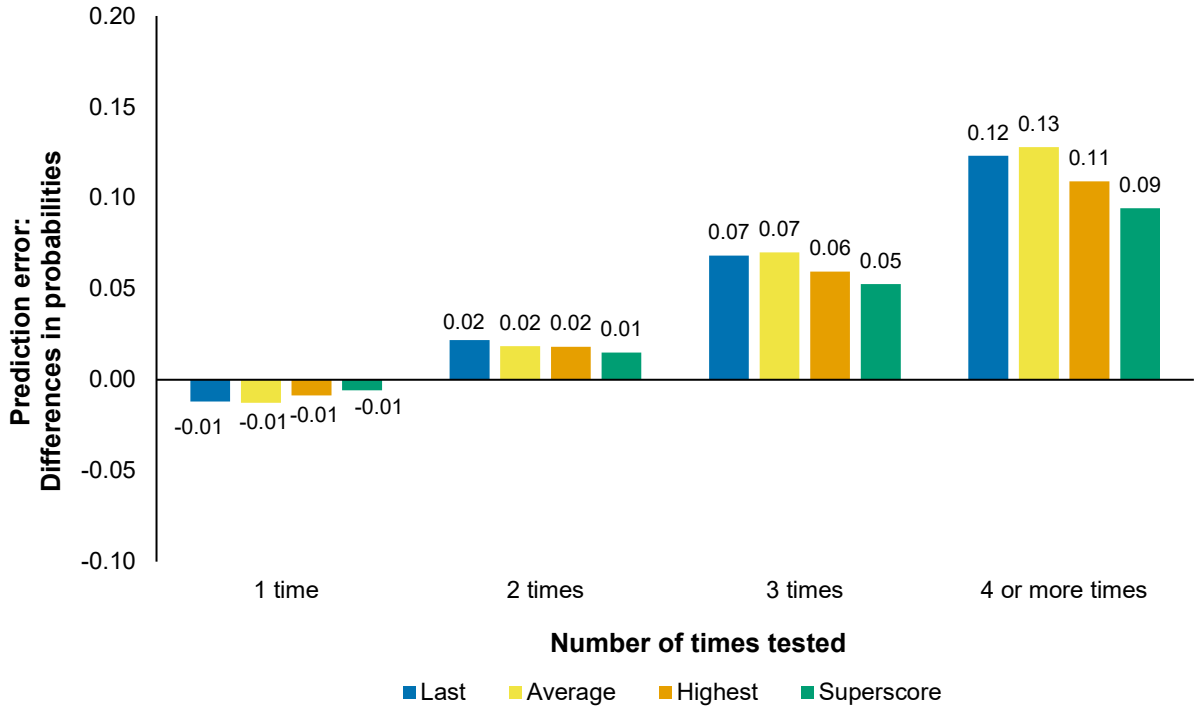
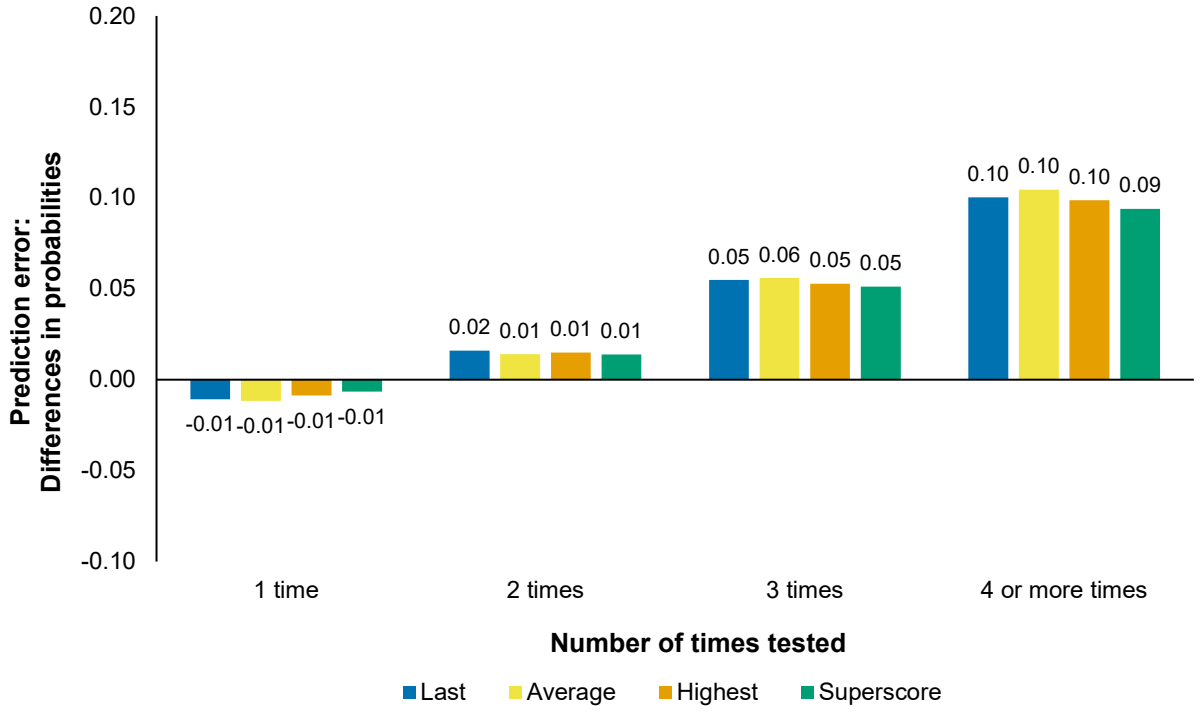


Table B2. Parameter Estimates (and Standard Errors) for the Three-Year Associate Degree Completion Models for the Two-Year Sample Accounting for HSGPA

Model	Scoring Methods for ACT Composite Measure Estimate (Standard Error)			
	Last	Average	Highest	Superscore
Intercept	-1.7946 (0.0296)	-1.7941 (0.0296)	-1.7952 (0.0296)	-1.7954 (0.0296)
ACT Composite	0.0622 (0.0021)	0.0623 (0.0021)	0.0620 (0.0021)	0.0620 (0.0021)
HSGPA	0.9947 (0.0122)	0.9935 (0.0123)	0.9979 (0.0122)	0.9983 (0.0122)
Times tested (2)	0.2349 (0.0142)	0.2382 (0.0141)	0.2099 (0.0144)	0.1827 (0.0148)
Times tested (3)	0.4897 (0.0220)	0.5169 (0.0216)	0.4595 (0.0229)	0.4284 (0.0242)
Times tested (4 or more)	0.7358 (0.0296)	0.7649 (0.0288)	0.7071 (0.0313)	0.6700 (0.0342)
ACT Composite * Times tested (2)	**-0.0014** (0.0034)	**0.0016** (0.0035)	**-0.0010** (0.0035)	**-0.0012** (0.0035)
ACT Composite * Times tested (3)	**-0.0098** (0.0053)	**-0.0063** (0.0056)	**-0.0103** (0.0055)	-0.0116* (0.0055)
ACT Composite * Times tested (4 or more)	-0.0188* (0.0069)	-0.0187* (0.0076)	-0.0207* (0.0073)	-0.0199* (0.0073)
ΔR^2 due to differential prediction	0.0063	0.0068	0.0057	0.0049

Note: Overall p value for the number of times tested was < .0001. Overall p value for the interaction term between the number of times tested and ACT Composite score was < .05. p values for parameter estimates are ≤ 0.0001 unless noted otherwise: * p value ≤ 0.05; nonsignificant p values are **bolded**. Each ACT Composite measure was centered at 19 and HSGPA was centered at 2.96. The variance estimate (and standard error) associated with the random intercepts was 0.6251 (0.0377) for Last, 0.6250 (0.0377) for Average, 0.6256 (0.0377) for Highest, and 0.6255 (0.0377) for Superscore.

Figure B3. Magnitude of Differential Prediction on the Likelihood of Completing an Associate Degree Within Three Years After Statistically Controlling for HSGPA by Number of Times Tested and Four Composite Scoring Methods for the Two-Year Sample[15]



[1] The prediction error was also evaluated on the logit scale instead of the probability scale. Results from these analyses pointed to the same general conclusion that the superscoring method resulted in the least amount of differential prediction by the number of times tested.

[2] Because of the inclusion of the interaction between the number of times tested and the ACT Composite measure, the odds ratio of degree completion for retesters will vary by ACT Composite score. To compute the odds ratios for alternative ACT Composite score values, one would first need to multiply the relevant interaction term times the alternative ACT score value minus 23, add this quantity to the parameter estimate for the number of times tested indicator, and then exponentiate the total quantity.

[3] The prediction error shown corresponds to when the ACT Composite measure is held constant at a value of 23 (parameter estimates are provided in Table 6).

[4] Because of the inclusion of the interaction between the number of times tested and the ACT Composite measure, the odds ratio of degree completion for retesters will vary by ACT Composite score. To compute the odds ratios for alternative ACT Composite score values, one would first need to multiply the relevant interaction term times the alternative ACT score value minus 19, add this quantity to the parameter estimate for the number of times tested indicator, and then exponentiate the total quantity.

[5] The prediction error shown corresponds to when the ACT Composite measure is held constant at a value of 19 (parameter estimates are provided in Table 7).

[6] The likelihood of longer-term college success including completing a degree has also been found to be overpredicted for low-income students based on HSGPA.

[7] Based on this alternative model that included income group, differential prediction by the number of times tested accounted for between 1.4% (for superscoring) to 3.1% (for average) of the variance in the log odds of six-year bachelor's degree completion for the four-year sample (values were only slightly smaller than those shown in Table 6) and between 1.8% (for superscoring) to 3.1% (for average) of the variance in the log odds of six-year

associate or bachelor's degree completion for the two-year sample (values were only slightly smaller than those shown in Table 7).
[8]See endnote 2.
[9] The prediction error shown corresponds to when the ACT Composite measure and HSGPA are held constant at values of 23 and 3.41 (parameter estimates are provided in Table 8).
[10]See endnote 4.
[11] The prediction error shown corresponds to when the ACT Composite measure and HSGPA are held constant at values of 19 and 2.96 (parameter estimates are provided in Table 9).
[12] The prediction error shown corresponds to when the ACT Composite measure is held constant at a value of 23 (parameter estimates are provided in Table A2).
[13] The prediction error shown corresponds to when the ACT Composite measure and HSGPA are held constant at values of 23 and 3.41 (parameter estimates are provided in Table A3).
[14] The prediction error shown corresponds to when the ACT Composite measure is held constant at a value of 19 (parameter estimates are provided in Table B1).
[15] The prediction error shown corresponds to when the ACT Composite measure and HSGPA are held constant at values of 19 and 2.96 (parameter estimates are provided in Table B2).