Does Superscoring Increase Subgroup Differences?

Krista Mattern, PhD, and Justine Radunzel, PhD

When applicants take the ACT® more than once, how do colleges and universities reconcile and make sense out of the multiple scores? In terms of validity, fairness, and impact on subgroup differences, are certain score-use polices better than others? Given that the proportion of students retaking the ACT has increased over time (Harmston & Crouse, 2016), answers to these questions have become increasingly relevant and pressing. The focus of this technical brief is to summarize empirical evidence on the validity and fairness of various score-use policies with an emphasis on superscoring. Specifically, findings from a study that examined the differential validity and predictions of different score-use policies that was published in 2018 will be reviewed. Additionally, new analyses demonstrating the impact of superscoring on subgroup differences will be presented. Finally, responses to ACT's Higher Education Score Use Survey are presented to help contextualize these findings. The intent is to arm higher education professionals with the most recent evidence to help support informed decision making on their own campus.

Background

A survey of the current landscape of college admissions found that there isn't a "one-size-fits-all" solution to how institutions of higher education treat multiple test records from the same applicant. Some postsecondary institutions use a student's most recent score. Others "pick and choose," selecting the best scores a student has earned in each content area over the course of multiple test administrations and forming a combined highest Composite score (superscoring). And still others have different and sometimes multiple policies in place. In particular, based on responses from 115 higher education professionals who completed ACT's Higher Education Score Use Survey in April of 2019, 33% indicated that they superscore the ACT, whereas 49% indicated that they superscore the SAT (ACT, 2019). The percentage that superscore both tests was 32%. Of those that superscore the ACT, all but one said they also superscore the SAT with the one exception indicating that they do not use the SAT. Among those that superscore the SAT, a different pattern emerges: 66% superscore the ACT, 32% use the highest score from a single administration on the ACT, and 2% do an in-depth review of all the ACT scores. This finding raises two issues that need further attention. First, given that superscoring is a fairly common practice among postsecondary institutions, what are the implications of this score-use policy in terms of validity and fairness? Second, given that many institutions have inconsistent score-use policies depending on the test (ACT vs. SAT), what are the validity and fairness implications for ACT test-takers? This report will focus on the first issue. ACT has provided the following recommendations as it relates to the second issue:

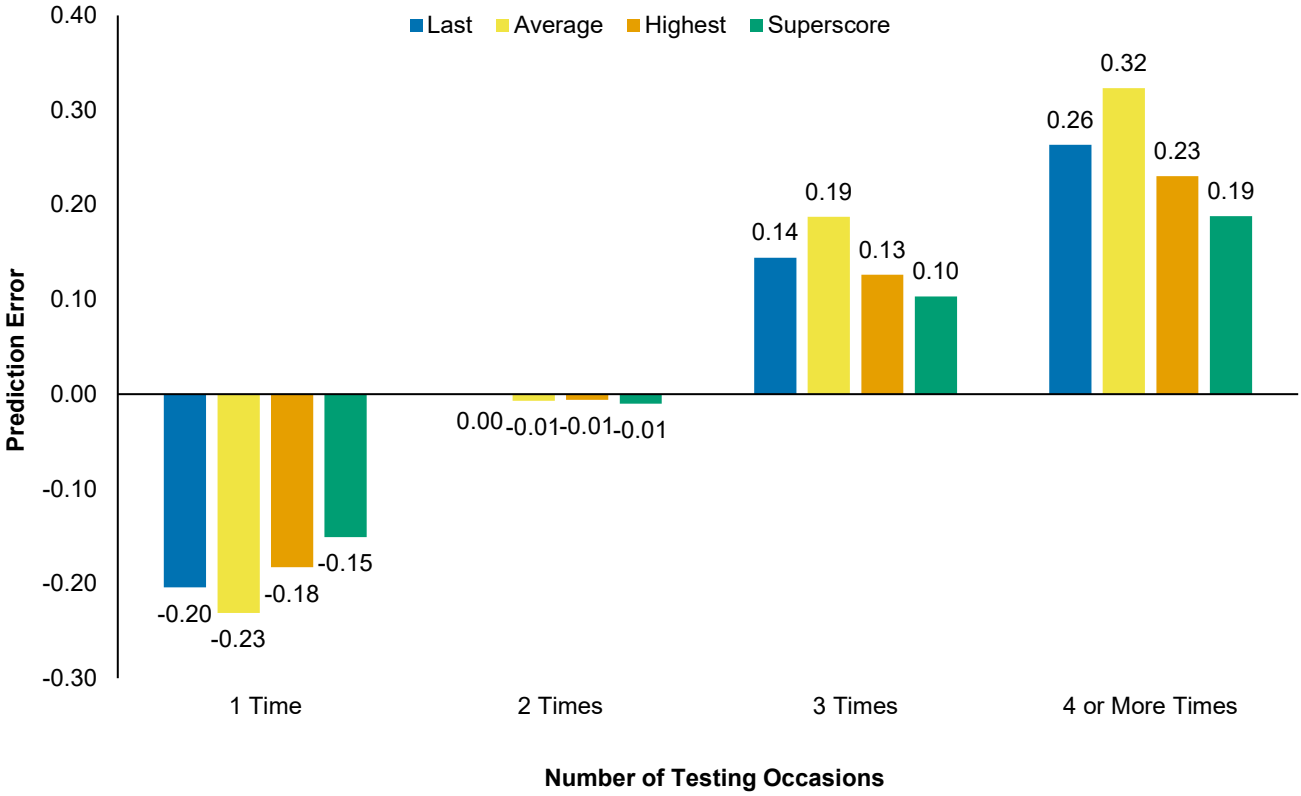
1. ***Consistency.*** Whatever score-use policy an institution chooses, that policy should be applied consistently to all applicants. Concerns of fairness arise if one score-use policy (most recent score) is applied to some groups of applicants (e.g., females, ACT test-takers) and a different score-use policy (superscore) is applied to other groups of applicants (e.g., males, SAT test-takers).

2. ***Concordance.*** For institutions that receive both ACT and SAT scores from applicants, the 2018 ACT-SAT concordance should be used to convert SAT scores to ACT scores and vice versa (The College Board & ACT, 2018). Given the change in the score scale for the 2016 SAT, using the previous ACT-SAT concordance puts ACT test-takers at an unfair disadvantage (ACT, 2009).

Validity and Fairness of Superscoring

ACT has been examining the validity and fairness of different scoring practices over the last several years. Results from an initial study were first made publicly available as both an ACT Working Paper with an overview provided in the *2017 Higher Education Research Digest* (Mattern, Radunzel, Bertling, & Ho, 2017; ACT, 2017) and later published in a peer-reviewed journal (Mattern, Radunzel, Bertling, & Ho, 2018). Contrary to expectations, the results showed that scores based on the superscoring method (referred to as superscores) were just as predictive (actually slightly more predictive) of first-year grades as compared to other scoring methods (recent, average, highest). Moreover, superscoring resulted in the least amount of differential prediction by the number of times a student tests. Interestingly, we found that first-year grades for students who tested more often were underpredicted even when prediction models were based on superscores (see Figure 1).[1]

Figure 1. Magnitude of Differential Prediction by Number of Testing Occasions and Four Composite Scoring Methods when ACT Composite Score is Held Constant at the Sample Mean of 23



Note: Prediction error is calculated by subtracting one's expected FYGPA based on the overall model from the expected value based on the model that includes retesting subgroup indicators and the interaction between the ACT Composite score and retesting indicators. Negative values indicate overprediction; positive values indicate underprediction.

As shown in Figure 1, retesters performed better in college than what was expected based on their test scores. And this prediction error was minimized when superscores were used, as compared to the other scoring methods. If superscores reflected positive measurement error—that is, an overestimate of one's true achievement level—then superscores would predict students to earn higher grades in college than what they actually earned, and this overprediction would increase as the number of retests increases. However, the results of the study suggested exactly the opposite.

Why is this the case? One hypothesis is that superscores and number of retests reflect not only academic preparation but also a motivational component. Specifically, students who are willing to forgo multiple Saturdays to sit for a multiple-hour test with the hope of maybe increasing their score are also the students who are likely to ask questions in their college courses, visit their professor during office hours, and take advantage of any extra credit opportunities to ensure the best possible grade. Future research should explore these hypotheses.

Another contribution of this study is the evaluation of the diversity implications of employing one scoring method versus another. Interestingly, despite the fact that underserved students are less likely to retest (Harmston & Crouse, 2016), the superscoring method did not result in a less diverse admitted class as compared to the other three methods. In fact, the gender, racial, and parental income distributions of a simulated admitted class were identical across the four scoring methods.

Current Study

The focus of the current study is to extend the previous research with an emphasis on further exploring the diversity implications of superscoring. As mentioned above, underserved students are less likely to retest as compared to their affluent peers. For students who test only once, superscoring has no impact on their ACT Composite score. Only students who retest have the potential to increase their ACT Composite score through superscoring, and the magnitude of this difference should be related to the number of times the student retests, in general. With that in mind, one potential concern or unintended consequence of superscoring is that subgroup differences will be exacerbated under this scoring policy. The focus of the current study is to investigate the extent to which superscoring increases, decreases, or has no impact on subgroup differences.

Method

Using data on the 2018 ACT-tested graduating class, we compared the average ACT Composite score for various student subgroups based on their most recent ACT Composite score as well as a superscore ACT Composite score. We estimated subgroup performance differences in terms of both:

1. Mean differences or unstandardized differences (USTD): the difference between the mean value in two groups

2. Standardized differences (STD): the difference between the mean value in two groups, divided by the overall standard deviation.

We estimated performance differences by the following student characteristics: race/ethnicity, gender, household income, and parental education. These characteristics were self-reported by students at the time they registered to take the ACT.

Results

Retesting Rates. Samples sizes and retesting rates for the 2018 ACT-tested graduating high school class are summarized in Table 1. Results are presented for the overall sample and by the student subgroups of interest. The total group consisted of over 1.9 million students. Of those students, 44% took the ACT more than once. As previously documented, we find that minority students and students from lower socio-economic households are less likely to retest.

For example, the retest rates for African American and Hispanic students were 43% and 34%, respectively, as compared to 49% for both White and Asian students. Differences by family income and highest parental education level were more pronounced. Students whose parents did not attend college had a retest rate of 36% compared to 62% for students whose highest parental education level was more than a bachelor's degree.

Table 1. Proportion Retesting by Student Demographic Characteristics

Group	N	Proportion Retesting
Total	1,914,814	0.44
Gender		
Male	893,609	0.41
Female	991,973	0.48
Missing	29,232	0.37
Race/ethnicity		
African American	243,077	0.43
American Indian	15,449	0.36
White	996,712	0.49
Hispanic	307,358	0.34
Asian	91,899	0.49
Native Hawaiian/Pacific Islander	5,753	0.27
Multiracial	85,316	0.42
Missing	169,250	0.38
Annual family income		
Less than $36,000 (Low)	353,315	0.40
$36,000 to $80,000 (Mid)	382,947	0.47
More than $80,000 (High)	498,300	0.60
Missing	680,252	0.34
Parental education level		
No college	308,539	0.36
Some college	354,574	0.45
Bachelor's degree	418,863	0.55
Beyond bachelors	353,896	0.62
Missing	478,942	0.27

Note: The total number of students in the 2018 national ACT-tested cohort that are reported here differs from that previously reported (e.g., ACT Condition of College and Career Readiness Report and National Profile Report) due to a small number of students since then being identified as being included more than once.

Table 2. Subgroup Unstandardized (USTD) and Standardized (STD) Differences in ACT Composite Scores by Scoring Method

	Most recent score			Superscore			Most recent score - Superscore	
Group	**M**	**USTD**	**STD**	**M**	**USTD**	**STD**	**ΔUSTD**	**ΔSTD**
Gender								
Male	20.8	-0.1	-0.02	21.3	-0.2	-0.03	0.10	0.02
Missing	17.6	-3.3	-0.57	18.1	-3.4	-0.58	0.10	0.01
Female	20.9			21.5				
Race/ethnicity								
African American	16.9	-5.3	-0.91	17.4	-5.4	-0.92	0.10	0.00
American Indian	17.3	-4.9	-0.84	17.7	-5.1	-0.86	0.20	0.02
Hispanic	18.8	-3.4	-0.59	19.2	-3.6	-0.61	0.20	0.02
Asian*	24.5	2.3	0.40	25.1	2.3	0.39	0.00	-0.01
Native Hawaiian/Pacific Islander	18.2	-4.0	-0.69	18.5	-4.3	-0.73	0.30	0.04
Multiracial	21.1	-1.1	-0.19	21.6	-1.2	-0.20	0.10	0.01
Missing	19.8	-2.4	-0.41	20.2	-2.6	-0.44	0.20	0.03
White	22.2			22.8				
Annual family income								
Less than $36,000 (Low)	18.2	-5.7	-0.98	18.7	-5.9	-1.00	0.20	0.02
$36,000 to $80,000 (Mid)	20.7	-3.2	-0.55	21.3	-3.3	-0.56	0.10	0.01
Missing	20.0	-3.9	-0.67	20.4	-4.2	-0.71	0.30	0.04
More than $80,000 (High)	23.9			24.6				
Parental education level								
No college	17.9	-7.0	-1.21	18.3	-7.3	-1.24	0.30	0.03
Some college	19.8	-5.1	-0.88	20.4	-5.2	-0.88	0.10	0.00
Bachelor's degree	22.5	-2.4	-0.41	23.2	-2.4	-0.41	0.00	-0.01
Missing	18.9	-6.0	-1.03	19.2	-6.4	-1.08	0.40	0.05
Beyond bachelors	24.9			25.6				

Note: Δ refers to differences in unstandardized (USTD) and standardized (STD) subgroup differences for most recent score - superscore. Positive Δ values indicate that superscoring increases subgroup differences. Negative Δ values indicate that superscoring reduces subgroup differences. Referent group is italicized.

* For Asian students, the subgroup differences were positive; therefore, the signs for ΔUSTD and ΔUSD were reversed to maintain consistency in directionality.

Subgroup Differences. Table 2 provides the means, USTDs, and STDs in ACT Composite scores based on students' most recent ACT Composite score and their superscore ACT Composite score for the student subgroups. The difference in USTDs and STDs for the two scoring methods—most recent versus superscoring—was calculated to directly test whether superscoring resulted in larger, smaller, or similar subgroup differences as compared to using students' most recent score. Those results are presented in the last two columns of Table 2. The results indicate that superscoring increased subgroup differences marginally. On average, USTDs are 0.17 larger (on a 1 to 36 scale) for superscores as compared to the most recent scores (differences range from 0.00 to 0.40 among the subgroups examined; refer to ΔUSTD column).

For example, the average ACT Composite score for African American students is 5.3 points lower than White students (16.9 versus 22.2) when based on the most recent ACT Composite score. Comparatively, the average ACT Composite score for African American students is 5.4 points lower than White students (17.4 versus 22.8) when based on a superscore ACT Composite score, resulting in a difference in USTD of 0.10 (5.4 minus 5.3).

In terms of STDs, superscoring increases subgroup differences by 0.02, on average (differences ranging from -.01 to .05; refer to ΔSTD column), representing a very small effect. The largest effects were found for students with missing data for either household income or parental education level. Building off the previous example comparing African American students' performance to White students' performance, the results indicate no change in STD with the implementation of superscoring.[2]

Subgroup Differences by Number of Testing Occasions. The next set of analyses explored subgroup differences by scoring method controlling for the number of testing occasions. The rationale for these additional analyses was to be able to tease apart subgroup differences from differential retesting rates, given that undeserved students are less likely to retest. Table 3 provides the distribution of students in terms of the number of testing occasions along with the average ACT Composite score based on most recent score and superscoring. For this sample, 56% of the sample took the ACT only once. Among retesters (44%), the breakdown by the number of testing occasions was as follow: 24% tested twice, 11% tested three times, and 9% tested four or more times.

Table 3. Mean ACT Composite Scores by Scoring Method and Number of Times Tested

			Most recent score		Superscore	
Number of times tested	N (%)		Mean	SD	Mean	SD
1	1,064,222	(56%)	19.3	5.5	19.3	5.5
2	465,650	(24%)	22.0	5.7	22.9	5.6
3	215,527	(11%)	23.3	5.5	24.7	5.3
4 or more	169,415	(9%)	23.9	5.1	25.6	4.8
Total	1,914,814	(100%)	20.8	5.8	21.3	5.9

Note: SD = standard deviation. The total number of students in the 2018 national ACT-tested cohort that are reported here differs from that previously reported (e.g., ACT Condition of College and Career Readiness Report and National Profile Report) due to a small number of students since then being identified as being included more than once.

Students who tested more often tended to have higher ACT Composite scores. Moreover, the difference between the average ACT Composite score based on the most recent score as compared to superscoring increased as the number of testing occasions increased, as expected. For example, for students who tested once, the average ACT Composite score was 19.3. For students who tested twice, the difference in the average Composite score between the two methods was 0.9 (22.9 minus 22.0). For students who tested four or more times, this difference increased to 1.7. The results broken out by student subgroups are provided in the Appendix.

Table 4. Scoring Method Differences in USTD and STD in ACT Composite Score by Number of Times Tested

	Number of Times Tested							
	One		Two		Three		Four or more	
Group	ΔUSTD	ΔSTD	ΔUSTD	ΔSTD	ΔUSTD	ΔSTD	ΔUSTD	ΔSTD
Gender								
Male*	0.00	0.00	0.00	0.00	0.10	0.02	0.10	0.02
Missing	0.00	0.00	-0.30	-0.04	-0.40	-0.06	-0.30	-0.05
Female								
Race/ethnicity								
African American	0.00	0.00	-0.10	0.00	-0.10	0.02	-0.30	0.01
American Indian	0.00	0.00	-0.10	0.00	-0.10	0.01	-0.20	0.00
Hispanic	0.00	0.00	0.00	0.01	0.00	0.02	-0.10	0.01
Asian*	0.00	0.00	0.00	0.01	0.00	0.01	0.10	0.04
Native Hawaiian/Pacific Islander	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01
Multiracial	0.00	0.00	-0.10	-0.01	0.00	0.01	-0.10	-0.01
Missing	0.00	0.00	0.00	0.01	0.00	0.00	-0.10	-0.01
White								
Annual family income								
Less than $36,000 (Low)	0.00	0.00	-0.10	0.00	-0.10	0.02	-0.20	0.01
$36,000 to $80,000 (Mid)	0.00	0.00	-0.10	-0.01	-0.10	0.00	0.00	0.03
Missing	0.00	0.00	-0.20	-0.03	0.00	0.01	0.00	0.01
More than $80,000 (High)								
Parental education level								
No college	0.00	0.00	-0.10	0.00	-0.20	0.01	-0.10	0.04
Some college	0.00	0.00	-0.10	0.00	0.00	0.03	-0.10	0.03
Bachelor's degree	0.00	0.00	-0.10	-0.01	0.00	0.01	0.00	0.02
Missing	0.00	0.00	-0.10	0.00	0.00	0.02	-0.10	0.00
Beyond bachelors								

Note: Δ refers to differences in unstandardized (USTD) and standardized (STD) subgroup differences for most recent score - superscore. Positive Δ values indicate that superscoring increases subgroup differences. Negative Δ values indicate that superscoring reduces subgroup differences. Referent group is italicized.

* For Asian students and males, the subgroup differences were positive; therefore, the signs for ΔUSTD and ΔSTD were reversed to maintain consistency in directionality.

Table 4 provides the difference (Δ) in USTD and STD in ACT Composite score for the two scoring methods by the number of testing occasions. The directionality of the subgroup differences were the same as those suggested in Table 2, except for gender where ACT scores were higher on average for males than for females among students who took the ACT two, three, or four or more times. As discussed, the most recent and superscore results are identical for students who tested only once. When the results are disaggregated by the number of testing occasions, we see even smaller increases attributed to superscoring than previously described and often the results reverse where superscoring results in smaller subgroup differences (as indicated by negative values). For students who tested twice, USTDs are 0.09 smaller on average (ranging from -.30 to .00) and STDs are the same (0.00 on average; differences ranging from -.04 to .01) when based on superscoring as compared to when based on the most recent score. For students who tested three times, USTDs are 0.06 smaller (ranging from -.40 to .01) and STDs are 0.01 higher, on average (differences ranging from -.06 to .03). For students who tested four or more times, USTDs are 0.09 smaller (ranging from -.30 to .01) and STDs are 0.01 higher, on average (differences ranging from -.05 to .04).

Discussion

In sum, the results indicate that subgroup differences are largely unaffected by the two scoring policies examined in the current study—most recent versus superscoring. Given that students tend to improve their scores through retesting and the high reliability of ACT scores, it is not surprising that results based on a student's most recent test record are quite similar to those based on superscoring. Also contributing to the finding of small to no differences based on superscoring is the relatively low frequency (less than half of students) of retesting overall and retesting more than once.

The results also suggest that the slight increases in USTDs and STDs can be attributed to differences in retest rates among subgroups. Analyses controlling for the number of times a student retests indicated that subgroup differences were more likely to decrease rather than increase when superscoring was applied. These results are very promising. If we improve retesting rates among groups who are less likely to retest, such as underserved students, through programs and initiatives, these results suggest that superscoring may help reduce subgroup differences.

For example, broader awareness that ACT provides two fee waivers to low-income students for national test administrations may help promote retesting among underserved students. However, awareness may not be sufficient. ACT data indicate that low-income students who register for the ACT with a fee waiver have higher no-show rates than students who pay to take the ACT—22% for fee waiver students versus 6% for students who pay for their registration (Cruce, Hahn, & Metcalfe, 2017). Future research should explore ways to not only promote registering for the ACT via the fee waiver program but also encourage fee waiver students to show up for the test.

The results of the current study provide a preview into the impact of superscoring on subgroups differences. Superscoring had little to no effect on subgroup differences and in some cases, resulted in smaller subgroup differences when the number of retests was held constant. Despite these positive findings, the results may change if retesting behavior changes significantly in the future in terms of who retests and how often.

Notes

1. Research on the SAT found similar findings pertaining to superscoring (Boldt, Centra, & Courtney, 1986).
2. The reason why there is a slight increase in the USTD but no change in the STD is due to the fact that the standard deviation of superscores is larger than the standard deviation of the most recent scores (5.9 vs. 5.8, respectively).

References

ACT, Inc. (2009). *ACT-SAT Concordance Tables.* Iowa City, IA: ACT. Retrieved from the ACT website: https://www.act.org/content/dam/act/unsecured/documents/ACTCollegeBoardJointStatement.pdf

ACT, Inc. (2017). *2017 Higher Education Research Digest*. Iowa City, IA: ACT.

ACT, Inc. (unpublished manuscript). 2019 ACT's Higher Education Score Use Survey. Iowa City, IA: ACT.

Boldt, R. F., Centra, J. A., & Courtney, R. G. (1986). *The validity of various methods of treating multiple SAT® scores* (College Board Report No. RR-86-4). New York, NY: The College Board.

The College Board & ACT. (2018) *Guide to the 2018 ACT/SAT Concordance.* Retrieved from the ACT website: https://www.act.org/content/dam/act/unsecured/documents/ACT-SAT-Concordance-Information.pdf

Cruce, T. M., Hahn, R. W., & Metcalfe, R. D. (2017). *Nudging to No Effect: Attempts to Improve the College Entrance Exam Attendance Rates of Low-Income Students.* Paper presented at the annual meeting of the Association for Education Finance and Policy. Washington: D.C.

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

Mattern, K., Radunzel, J., Bertling, M., & Ho, A. D. (2017). *How should colleges treat multiple admissions test scores?* ACT Working Paper 2017-4. Iowa City, IA: ACT.

Appendix

Table A1. Mean ACT Composite Scores by Scoring Method, Number of Times Tested, and Student Subgroup

	One Time			Two Times			Three Times			Four or More Times		
Group	**%**	**Recent**	**Superscore**	**%**	**Recent**	**Superscore**	**%**	**Recent**	**Superscore**	**%**	**Recent**	**Superscore**
All students	56%	19.3	19.3	24%	22.0	22.9	11%	23.3	24.7	9%	23.9	25.6
Gender												
Male	59%	19.2	19.2	23%	22.3	23.2	10%	23.6	25.0	8%	24.1	25.9
Female	52%	19.4	19.4	26%	21.9	22.8	12%	23.2	24.5	10%	23.8	25.5
Missing	63%	16.5	16.5	25%	18.4	19.6	6%	21.1	22.8	5%	22.7	24.7
Race/ethnicity												
African American	57%	16.1	16.1	25%	17.6	18.6	11%	18.3	19.7	8%	19.1	21.0
American Indian	64%	16.2	16.2	21%	18.1	19.1	8%	20.1	21.5	7%	21.6	23.4
White	51%	20.6	20.6	25%	23.3	24.2	13%	24.3	25.6	11%	24.7	26.3
Hispanic	66%	17.9	17.9	23%	20.2	21.1	7%	21.8	23.1	4%	22.5	24.2
Asian	51%	23.2	23.2	26%	25.5	26.4	13%	26.3	27.6	10%	26.4	28.1
Native Hawaiian/Pacific Islander	73%	17.0	17.0	18%	20.3	21.2	6%	22.7	24.0	3%	23.5	25.1
Multiracial	58%	19.8	19.8	24%	22.3	23.3	11%	23.5	24.8	7%	23.9	25.6
Missing	62%	18.1	18.1	22%	21.7	22.6	9%	23.6	24.9	7%	24.1	25.8
Annual family income												
Less than $36,000	60%	17.4	17.4	25%	19.0	19.9	9%	19.7	21.1	6%	20.5	22.4
$36,000 to $80,000	53%	19.7	19.7	26%	21.4	22.3	12%	22.2	23.6	10%	22.9	24.6
More than $80,000	40%	22.6	22.6	28%	24.5	25.3	17%	25.0	26.3	15%	25.1	26.8
Missing	66%	18.5	18.5	20%	21.8	22.8	8%	23.9	25.2	5%	24.6	26.3
Parental education level												
No college	64%	17.2	17.2	23%	18.7	19.7	8%	19.5	21.0	5%	20.6	22.4
Some college	55%	19.0	19.0	26%	20.4	21.4	11%	21.2	22.5	8%	21.9	23.7
Bachelor's degree	45%	21.3	21.3	28%	23.0	24.0	15%	23.8	25.1	13%	24.1	25.8
Beyond bachelor's	38%	23.7	23.7	29%	25.5	26.4	17%	25.9	27.2	15%	25.7	27.4
Missing	73%	17.9	17.9	18%	20.7	21.7	6%	23.0	24.3	4%	23.8	25.6

Acknowledgement

The authors thank Wayne Camara and Jeff Allen for their feedback on earlier versions of this brief.

Krista Mattern, PhD
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. Also known for work in evaluating the efficacy of learning products to help improve intended learner outcomes.

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