Section Retesting: Do Students Perform as Expected?

Justine Radunzel, PhD, and Krista Mattern, PhD

Introduction

Beginning in September 2020, students will have the option to retake one or more sections of the ACT® test (referred to as section retesting, modular testing, or single-subject retesting), instead of needing to take the entire battery again. Section retests will only be available to students who have previously completed the full battery and only available to students retesting online. The section retest option is being made available to students because research conducted to date indicates that ACT scores combined across multiple administrations are valid; this includes results from the current study which suggest that students' performance when retesting in a single ACT subject area tends to be consistent with expected performance estimated from standard retesting with the full battery.

In addition to the findings from the current study to be described in detail in this report, there is other research providing evidence in support of offering ACT section retesting. First, based on decades of research, each individual ACT test is a valid and reliable measure of students' academic achievement level in the corresponding subject area (ACT, 2019; see chapters 10 and 11). Second, in a study (Mattern, Radunzel, Bertling, & Ho, 2018) that included over 277,000 students from 221 four-year postsecondary institutions, we evaluated the validity of using ACT scores obtained from various scoring methods across test administrations (average, highest, most recent, and superscoring) for identifying students who are likely to be successful in their first-year of college. The superscoring method combines the highest subject scores across test administrations into a new ACT Superscore. The study found that Superscores were as predictive – if not more predictive – of first-year grade point average and resulted in the least amount of differential prediction as compared to other scoring methods when statistically controlling for the number of times tested. More specifically, the differential prediction results suggested that first-year grades for examinees who tested more often tended to be underpredicted by ACT scores. That is, retesters performed better in college than what was expected based on their test scores. But, this prediction error was lower when Superscores were used instead of the other scoring methods (Mattern & Radunzel, 2019; Figure 1). This study provides evidence supporting section retesting as it suggests that selecting students' best scores from any test attempt (superscoring) results in the most predictive indicator of a student's preparedness for future success.

ACT.org/research

With section retesting, there is essentially a shift in which subject test is taken first, and therefore there is a need to ensure that scores are comparable, regardless of the order in which the subject tests were taken. A study conducted by ACT in 2016 examined whether the order in which a student takes the ACT subject tests impacts their scores (Andrews, 2019). This study included over 4,000 students who were randomly assigned to one of four order conditions. One condition administered the four subject tests in the standard order: English, math, reading, and science. For the other three conditions, a different subject test was administered first. All conditions took the ACT online. The study found that the average scale scores were similar, regardless of the order in which students took the ACT subject tests. The findings from the study provide additional support for offering section retesting.

More recently, a study by Mattern, Radunzel, and Andrews (2019) examined score gains for a convenience sample of nearly 100 students who had taken the four ACT subject tests over the course of four days (one test per day; Monday through Thursday) and had official ACT scores with the full battery either prior to or after participating in the study. The results indicated that the scores earned when taking one ACT subject test per day were consistent with expected score gains resulting from standard retesting using the entire ACT battery. These findings suggest that taking the ACT subject tests on different days does not lead to artificially inflated test scores.

Building on these prior studies, the focus of the current study was to examine student performance on single-subject retests. More specifically, the study objective was to examine whether section retesting results in larger score gains as compared to traditional retesting (taking the entire battery). That is, this study evaluated whether allowing students to take one subject test at a time resulted in students performing better than what is typically seen among students testing with the full battery, taking into account students' prior ACT scores and other testing characteristics. If we do not find higher scores with section retesting, then this study will help alleviate concerns that section retesting may lead to artificially inflated scores. If we find that ACT scores are higher with modular testing, this doesn't necessarily mean that the scores are less indicative of current academic preparation and future success. Section retesting may promote more effective learning strategies to better prepare students for testing. Therefore, the study also examined how performance on the single-subject retest relates to subsequent performance on the full-battery ACT to evaluate whether any improved performance is associated with true learning gains.

To conduct the study, we partnered with high schools from a state that not only offers students the opportunity to take the ACT during the spring of their junior year as part of the State and District testing program, but also provides the same opportunity for students to test again during October of their senior year (referred to as the fall senior retake). The state offers the fall senior retake to help increase college and scholarship opportunities for students. As part of the study, high schools were asked to administer a single-subject test to seniors in August or September 2019 as a practice session for students to help gauge readiness for their fall senior retake. Therefore, students included in this study would have ACT scores available from three testing events – junior spring testing with the entire battery, the single-subject retest from this study, and senior fall retesting with the entire battery – that would allow us to meet the study objectives. If larger score gains were found for section retesting, having the third testing event allowed

us to examine whether those gains were validated by higher scores on subsequent retakes with the entire ACT battery.

One limitation of the study design was that students did not receive college-reportable scores on their single-subject retake that was administered in August or September (more details about this test administration are provided in the Data and Methods section). Due to the low stakes nature of this testing event, there were concerns with student motivation and engagement on the single-subject test. To try to minimize these concerns, schools were strongly encouraged to identify eligible students interested in taking a practice test to prepare and gauge their readiness for the October senior retest. Students were to be informed about this testing opportunity well enough in advance so that they could take full advantage of the experience to prepare for their fall senior retake that would involve college-reportable scores and that could possibly open up more college and scholarship opportunities for them.
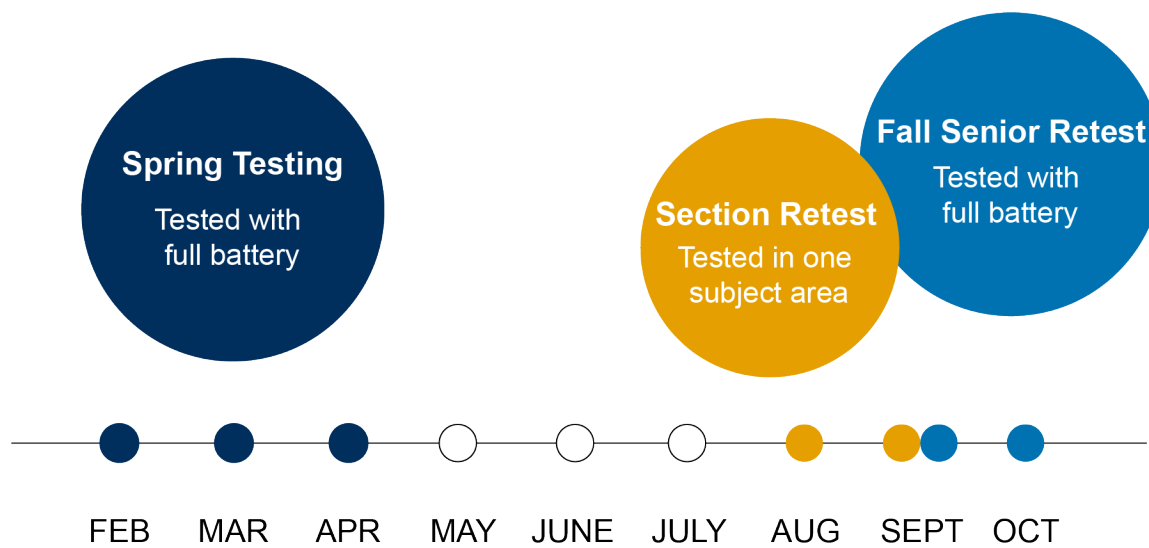
As will be discussed later in greater detail, there was evidence that some students tended to be less engaged in the study's single-subject retest than in their initial testing experience with the entire battery. We attempted to identify those students so that analyses could be conducted not only on the full sample of students but also on the engaged subsample. Results from the engaged subsamples suggested that section retesting did not lead to artificially inflated scores.

Data and Methods

Study Sample

High schools from a specific state that had administered the ACT to juniors in spring 2019 and planned to retest the same students as seniors in fall 2019 as part of the State and District testing program were invited to participate in this study. School participation in this study involved: (a) identifying students interested in taking a single subject test to prepare and gauge readiness for the October retest with the full battery, (b) administering a single subject test in August or September in paper format in a secure manner and under standard testing conditions to identified students, and (c) returning the completed answer documents to ACT. Each school tested in only one subject area on a first come, first included basis. Schools received a monetary incentive for participating based on the number of students tested. Score reports for the single subject test were not provided to schools or students because scores were used for research purposes only. However, students were allowed to keep the full practice test administered with the answer key to determine their score and to use it as an additional test preparation resource. In our recruitment efforts, we targeted 500 students for each subject.

Three to five public high schools participated per subject area. Table A1 in Appendix A provides detailed information on the number of seniors taking the section retake in August or September as part of the study protocol. For the analyses, we focused on the group of students who had previously taken the entire ACT test battery in the spring of their junior year (February to April, 2019) and then tested again on the full battery in early fall of their senior year (September or October 2019; Figure 1). Students were required to have taken all three assessments under standard testing time.

Figure 1. ACT Testing Timeline – 2019

Focusing on students who had taken the entire battery and received college reportable scores in both the spring and the fall provided a reference comparison to help interpret the study results on score gains associated with section retakes. For all but eight students (or 99.9% of students with all three testing events), the first testing event was in the spring of their junior year, and the second official testing event was in September or October 2019. The eight students that had taken the ACT prior to their spring test date or had retested with the full battery prior to their fall test date were excluded from the analyses presented, though their inclusion or exclusion had no impact on the results. The resulting sample size for analysis ranged from 402 students in reading to 596 students in math (Table A1). The samples included a significant percentage of the seniors at these high schools that had both fall and spring 2019 ACT scores; percentages ranged from 54% in reading to 90% in math. A more complete description of the samples is provided at the beginning of the Results section.

Measures

Outcomes. The primary outcome was students' scores on the single-subject retest from August or September; these scores were examined in relation to prior and subsequent performance, as well as in relation to expected performance. For descriptive analyses, the differences in scores from spring testing to section retesting were calculated for students by subtracting their spring test score on the full battery from their section retake score (Figure 1). A positive difference indicated a score gain, a difference of zero indicated no score change, and a negative difference indicated a score decline. Differences in scores were also evaluated from section retesting to fall testing and from spring testing to fall testing (Figure 1).

Another approach used to interpret and provide context around how students performed on the section retake in this study was the difference between students' "observed" and "expected" ACT scores. Utilizing a different data source than the study sample (referred to as the state reference sample), students' "expected" performance on the ACT was

estimated from other relevant characteristics such as prior test score and number of months between testing events. The state reference sample included students from the specific state who had participated in ACT state testing as a junior in spring of 2017, 2018, or 2019 and then subsequently took the ACT as their second testing event in September or October of the same year.[1] Students in the reference sample took the entire battery for both testing events. Given their additional testing experience with single subject retesting as part of this study, students from participating high schools that tested in 2019 were excluded from the state reference sample. The sample size per subject area for the state reference sample is provided in Table A1.

Non-engagement indicators. Incomplete tests or guessing patterns might indicate low motivation. Given that college-reportable scores were not provided on the section retake, we evaluated students' test item responses to help identify students who may have been less engaged in the testing event.[2] More specifically, the following three indicators were used to help identify non-engaged students:

- The omit indicator – this was set to 1 when a student did not respond to 25% or more of the test items and 0 otherwise.

- The long string indicator – this was set to 1 when a student had a string of 10 or more consecutive item responses of the same response option, such as "1111111111" and 0 otherwise.

- The rapid guessing indicator – this was set to 1 when a student had a repetitive response pattern, such as "123123123" and 0 otherwise. More than 100 repetitive response patterns that have been used in other ACT research studies (e.g., Allen & Mattern, 2019) were analyzed.

Incomplete tests or guessing patterns do not necessarily indicate low motivation; these could also be related to other factors such as students having difficulty working through the test items. However, later we provide evidence that these indicators were associated with fairly substantial score declines, on average, between testing events in the study sample, as well as in the state reference sample.

Analysis

Means and percentages were used to describe the outcomes and student characteristics. McNemar's test was used to determine if the percentage of students flagged as non-engaged differed between testing events. Additionally, linear regression models were developed that related ACT section retake scores to the non-engagement indicators after statistically controlling for prior ACT subject scores from the spring; these models were developed to evaluate the utility of the non-engagement indicators for identifying students who may not have been engaged in the section retake experience. Confidence intervals were estimated to determine whether the average gains in scores between testing events were significantly different from 0.

Utilizing the state reference sample, linear regression models were developed to compute "expected" performance on the ACT as a function of a student's prior ACT subject score, the number of months between testing events, the three non-engagement indicators for the prior testing event, the three non-engagement indicators for the retesting event, and the three two-way interactions between the same non-engagement

indicators from the two testing events. Other studies have found subsequent ACT scores to be related to prior ACT scores and the number of months between testing (e.g., Camara & Allen, 2017; Moore, Sanchez, & San Pedro, 2018). To determine whether student performance on the section retake was consistent with expected performance from standard retesting, differences between observed and expected ACT scores were computed and then evaluated to see whether the average value was significantly different from zero. A significance level of .05 was used in this study.

Engaged Sample

There was evidence supporting the use of the non-engagement indicators to help identify students who may have been less engaged in section retesting. First, in three out of the four subjects −English, math, and reading− there was a significantly higher percentage of students being flagged as non-engaged during the section retake event than during their prior spring testing (Table 1; 29.2% to 29.9% vs. 22.0% to 23.6%, respectively). In science, the percentage under section retesting was slightly higher (11.1% vs. 8.4%), though it was not significantly different from that seen from spring testing. Additionally, with the exception of science, the non-engaged percentages for the single-subject retests were higher than those observed for the state reference sample at spring testing (24.8% in English, 21.1% in mathematics, 19.4% in reading, and 14.1% in science).

Table 1. Percentage of Non-Engaged Students in Sample by Subject and Testing Event

Subject	Sample Size	Spring	Section Retake	Difference
English	487	22.0	29.2	7.2*
Math	596	22.2	29.5	7.3*
Reading	402	23.6	29.9	6.3*
Science	450	8.4	11.1	2.7

* indicates p value < 0.05 from McNemar's test for comparing the percentage of students flagged as non-engaged between spring testing (entire battery) and single-subject retesting.

Second, the non-engagement indicators from the single-subject retesting event were generally associated with fairly substantial score declines from spring testing to single-subject retesting, after controlling for students' prior subject scores from the spring (Table A2 in the Appendix). For example, in English, students who were flagged with the omit indicator, the long string indicator, or the rapid guessing indicator experienced a score decline of 6.8, 2.3, and 1.8 points, respectively, on their English section retake, on average. Results for the other subject areas are provided in Table A2. Although the long string indicator was not significantly associated with score declines for the single-subject retests, it was for the state reference sample to be discussed later.

Given these findings, analyses were not only conducted on the full sample of students who had participated in all three testing events (spring 2019, section retake, and early fall 2019), but also on the subset of students that were not flagged on any of the non-engagement indicators for their single-subject testing event. This subset of students was labeled as the engaged sample and represented about 71% of the full sample of students in English (*n* = 345), math (*n* = 420), and reading (*n* = 286), and 89% in science (*n* = 400).[3]

Results

Description of Samples

Table 2 provides average ACT scores for each testing event by subject and sample, as well as correlations between the scores. Compared to the state reference sample, students who participated in the study tended to earn higher scores on average during both the fall and spring testing events. Average scores were also generally higher for the engaged sample than for the full sample. The one exception to both of these findings was in science where scores were more comparable.

Table 2. Average ACT Scores and Correlations Between ACT Scores by Subject and Sample

Subject	Sample	N	Mean Score (SD)			Correlations Between ACT Scores		
			Spring (Feb/March/April)	Section Retake (Aug/Sept)	Fall (Sept/Oct)	Spring and Section Retake	Section Retake and Fall	Spring and Fall
English	Full	487	18.8 (5.8)	18.1 (6.9)	19.3 (6.1)	.80	.81	.86
	Engaged	345	19.8 (6.0)	20.1 (6.5)	20.2 (6.4)	.84	.85	.88
	State	87,587	17.5 (5.4)		17.9 (5.7)			.85
Math	Full	596	18.8 (4.2)	19.0 (4.2)	19.3 (4.2)	.81	.79	.84
	Engaged	420	19.1 (4.4)	19.3 (4.4)	19.4 (4.4)	.83	.83	.86
	State	87,588	17.8 (3.7)		17.9 (3.8)			.80
Reading	Full	402	19.3 (6.1)	18.9 (5.9)	19.7 (6.5)	.72	.74	.78
	Engaged	286	19.9 (6.6)	19.6 (6.2)	20.5 (6.8)	.73	.77	.79
	State	87,535	18.4 (5.5)		18.6 (5.8)			.76
Science	Full	450	18.4 (4.8)	18.5 (4.8)	18.4 (5.1)	.69	.71	.79
	Engaged	400	18.5 (4.9)	18.7 (4.9)	18.5 (5.2)	.69	.71	.79
	State	87,647	18.4 (4.5)		18.5 (4.6)			.71

Note. SD = standard deviation. All correlation coefficients were significantly different from 0 ($p < 0.0001$). Students completed the entire ACT battery during spring and fall testing.

Students' single-subject retest scores were positively correlated with both their spring and fall test scores from when they took the entire battery (Table 2). These correlations were relatively high ranging from .69 in science to .85 in English. With the exception of science, slightly higher correlations were observed in the engaged sample than in the full sample, though these differences were small. Between the official spring and fall testing events, the correlations in test scores were slightly higher in the study sample than in the state sample.

As shown in Table A3 in the Appendix, there was representation in the study samples across gender, racial/ethnic, annual family income, and parental education groups. When

comparing the full study sample to the engaged sample, the distributions for the student characteristics were generally comparable. Results for the state sample are also provided in the table. There were fairly substantial differences in the percentages of students missing data on the various characteristics between the study and state reference samples, making these comparisons less meaningful.

Differences in ACT Scores

Table 3 provides the average differences in ACT scores between testing events by subject and sample. Results based on the full sample support our initial concern that students may not have been as engaged in the single-subject testing experience as they were in their subsequent senior fall retake, especially in the English and reading samples. For this reason, we encourage the reader to focus on results for the engaged sample, even though results for both the engaged and the full sample are discussed.

Table 3. Average Gains in ACT Scores between Testing Events by Subject and Sample

| | | | Score Gains | | | | | |
| | | | Spring to Section Retake | | Section Retake to Fall | | Spring to Fall | |
Subject	Sample	N	Mean	95% CI	Mean	95% CI	Mean	95% CI
English	Full	487	**-0.7**	-1.1, -0.3	**1.2**	0.8, 1.5	**0.5**	0.2, 0.8
	Engaged	345	**0.3**	0.0, 0.7	0.1	-0.3, 0.4	**0.4**	0.1, 0.7
	State	87,587					**0.3**	0.3, 0.4
Math	Full	596	**0.2**	0.0, 0.4	**0.3**	0.1, 0.5	**0.5**	0.3, 0.7
	Engaged	420	0.2	-0.1, 0.4	0.2	-0.1, 0.4	**0.3**	0.1, 0.6
	State	87,588					**0.04**	0.02, 0.1
Reading	Full	402	**-0.4**	-0.8, -0.0	**0.9**	0.4, 1.3	**0.5**	0.0, 0.9
	Engaged	286	-0.3	-0.8, 0.3	**0.8**	0.3, 1.3	**0.6**	0.1, 1.1
	State	87,535					**0.2**	0.1, 0.2
Science	Full	450	0.1	-0.2, 0.5	-0.1	-0.5, 0.2	0.0	-0.3, 0.3
	Engaged	400	0.2	-0.1, 0.6	-0.2	-0.6, 0.2	0.0	-0.3, 0.4
	State	87,647					**0.1**	0.05, 0.1

Note. CI = confidence interval. Bolded means are significantly different from 0.

For example, students tended to experience a score decline in English from spring to single-subject retesting (by -0.7 on average) but then experience a score gain from spring to fall testing (by 0.5 on average). In contrast, among the engaged sample in English, students experienced a score gain on average on both their single-subject retake and senior fall retake; this finding suggests that the non-engagement indicators seemed to help identify students that were not engaged in the section retake experience for the English sample. The reading results suggest that this may not have been the case for that sample; for both the full and engaged reading samples, students experienced score declines on average from spring testing to single-subject retesting but experienced score gains from spring to fall testing.

In math, students tended to experience score gains on both of the subsequent testing events after spring testing, though the average difference from spring to single-subject retesting in the engaged sample was not significantly different from 0. The average math score gain was slightly higher from spring to fall retesting (0.3 for the engaged sample) than from spring to section retake (0.2 for the engaged sample); a similar result held in English and reading for the engaged sample. These results were expected as there was more instructional time between the two testing events. Additionally, the section retake administration provided students with a test prep opportunity for their fall senior retake. In science, although the average score gain was slightly higher at section retesting (0.2 for engaged sample) than at fall retesting (0.0), each average score gain was not significantly different from 0. Moreover, the corresponding confidence intervals overlapped between the two (-0.1 to 0.6 for section retesting compared to -0.3 to 0.4 for fall retesting for the engaged sample). From comparing gains in scores from section retesting to those from spring to fall testing, we did not see evidence that substantial score gains result from taking one subject test at a time as compared to taking the entire battery.[4]

Actual Performance Compared to Expected Performance

Next, how students performed on the section retest and their fall senior retake was compared to how they were expected to perform given their prior ACT score from the spring as well as other characteristics. Expected performance was derived from the models shown in Table A4 in the Appendix that were estimated from the state reference sample of students that had college-reportable scores from both the spring of their junior year and the fall of their senior year. According to these models, students' expected performance on the ACT was significantly related to their prior ACT score, combinations of the non-engagement indicators, and in some cases, the number of months between testing events.[5] The percentage of variance explained by the models ranged from 51% in science to 73% in English.

From Table 4, we see that in English for the engaged sample and in reading and science for both the full and engaged samples, students performed within expectations on their single subject retake (according to the confidence intervals of the differences including the value of 0). In English for the full sample, students tended to perform worse than expected which we attribute to the inclusion of students who were not as engaged during the section retake administration.

Math was the only subject where students tended to perform better than expected on their single-subject retake, and this was by only 0.5 points on average in the full sample and by only 0.4 points in the engaged sample. However, these students also performed better than expected in math on their senior fall retake with the full battery (by 0.7 and 0.6 points on average in the full and engaged samples, respectively).[6]

For both the full and engaged samples, students tended to perform within expectations on their senior fall retest with the full battery in English and science and better than expected in reading, though the lower bound of the confidence interval was near zero (0.0 for the full sample and 0.1 for the engaged sample). In comparison to the corresponding information for the single-subject retest where students' performance did not significantly differ from expected performance, the finding in reading may indicate that the non-engagement indicators considered in this study did not completely account for differences in engagement across test administrations.[7]

Table 4. Comparing Actual and Expected Performance on the ACT by Subject and Testing Event

Subject	Sample	N	Single-Subject Retest Score		Actual - Expected Performance on ACT Single-Subject Retest		Fall 2019 Score (entire battery)		Actual - Expected Performance on ACT Fall 2019 Retest	
			Actual	Expected	Mean	95% CI	Actual	Expected	Mean	95% CI
English	Full	487	18.1	18.7	**-0.5**	-0.8, -0.2	19.3	18.7	0.3	-0.0, 0.5
	Engaged	345	20.1	20.1	0.1	-0.3, 0.5	20.2	19.9	0.3	-0.0, 0.6
Math	Full	596	19.0	18.6	**0.5**	0.3, 0.6	19.3	18.6	**0.7**	0.5, 0.9
	Engaged	420	19.3	18.9	**0.4**	0.1, 0.6	19.4	18.8	**0.6**	0.4, 0.8
Reading	Full	402	18.9	19.2	-0.3	-0.7, 0.1	19.7	19.3	**0.4**	0.0, 0.8
	Engaged	286	19.6	20.1	-0.4	-0.9, 0.1	20.5	19.9	**0.6**	0.1, 1.1
Science	Full	450	18.5	18.4	0.1	-0.2, 0.4	18.4	18.5	-0.1	-0.4, 0.2
	Engaged	400	18.7	18.6	0.1	-0.2, 0.5	18.5	18.6	-0.1	-0.4, 0.3

Note. CI = confidence interval. Bolded means are significantly different from 0. Expected performance was derived from the model in Table A4 based on the state reference sample and the following predictors: prior ACT score from spring testing, non-engagement indicators, and number of months between testing events. The difference between the actual mean score and the expected mean score may not equal the mean difference in (actual – expected) scores due to rounding.

In supplemental analyses, we also compared actual and expected performance on the single-subject retake and senior fall retake by spring 2019 ACT score range to determine if similar patterns were seen among lower and higher scoring students. These results are shown in Table 5 for the engaged sample. The ACT score ranges were determined from the tertiles of the spring 2019 score distributions. In English, students in the lower two score ranges tended to perform consistently with expectations on both the section retake and the senior fall retake. In comparison, students in the upper English score range tended to perform greater than expected not only on the section retake (by 0.8) but also on the subsequent senior fall retake (by 0.8). In math, students tended to perform slightly better than expected on their single-subject retake for each of the three score ranges (by 0.3 to 0.5), a result that was also seen on their senior fall retake (by 0.4 to 0.9 point). In

reading and science, students generally performed as expected on their section retake. The one anomaly was in reading where students in the upper score range performed lower than expected (by 1.2 points on average).

Table 5. Comparing Actual and Expected Performance on the ACT by Subject and Testing Event for the Engaged Sample by ACT Score Range

Subject	Spring 2019 ACT Score Range	N	Single-Subject Retest Score		Actual - Expected Performance on ACT Single-Subject Retest		Fall 2019 Score (entire battery)		Actual - Expected Performance on ACT Fall 2019 Retest	
			Actual	Expected	Mean	95% CI	Actual	Expected	Mean	95% CI
English	1 to 16	114	14.1	14.2	-0.1	-0.8, 0.5	13.9	14.0	-0.1	-0.6, 0.4
	17 to 22	120	19.8	20.1	-0.3	-1.0, 0.4	20.2	20.0	0.2	-0.3, 0.7
	23 to 36	111	26.7	25.9	**0.8**	0.1, 1.4	26.8	25.9	**0.8**	0.2, 1.5
Math	1 to 16	163	16.1	15.8	**0.3**	0.01, 0.6	16.1	15.7	**0.4**	0.1, 0.6
	17 to 20	122	18.5	18.0	**0.5**	0.1, 0.9	18.8	18.0	**0.9**	0.4, 1.3
	21 to 36	135	23.8	23.4	0.4	-0.2, 0.9	24.1	23.4	**0.6**	0.2, 1.1
Reading	1 to 16	100	14.9	14.6	0.2	-0.4, 0.9	14.9	14.3	0.6	-0.1, 1.3
	17 to 22	91	19.3	19.6	-0.3	-1.2, 0.6	19.8	19.4	0.4	-0.5, 1.3
	23 to 36	95	25.0	26.2	**-1.2**	-2.2, -0.2	26.9	26.2	0.8	-0.1, 1.7
Science	1 to 16	139	15.3	14.9	0.4	-0.1, 0.9	14.5	14.8	-0.2	-0.8, 0.4
	17 to 20	122	18.5	18.4	0.2	-0.5, 0.8	17.7	18.3	**-0.6**	-1.2, -0.1
	21 to 36	139	22.3	22.6	-0.3	-0.9, 0.4	23.1	22.5	**0.6**	0.1, 1.1

Note. CI = confidence interval. Bolded means are significantly different from 0. Expected performance was derived from the model in Table A4 based on the state reference sample and the following predictors: prior ACT score from spring testing, non-engagement indicators, and number of months between testing events. The difference between the actual mean score and the expected mean score may not equal the mean difference in (actual – expected) scores due to rounding.

[1] ACT score ranges based on the tertiles of the spring 2019 score distributions.

Conclusions

In conclusion, findings from this study are consistent with those from an earlier study by Mattern et al. (2019) that found students taking ACT subject tests on different days did not earn artificially inflated test scores compared to what was expected. First, the average gains in ACT scores from spring testing with the full battery to single-subject retesting were relatively small, ranging from -0.3 in reading to 0.3 in English among the engaged sample. Second, when we estimated expected ACT performance based on prior scores from junior spring testing and other relevant testing characteristics, we found that performance on the section retest tended to be consistent with performance expectations from standard retesting with the full battery in English, reading, and science among engaged students. That is, students did not earn higher than expected scores when testing in a modular fashion for these three subject areas. On the math section retest, students performed slightly higher than expected on average (by 0.4 point with a 95% CI of 0.1 to 0.6 among engaged students; Table 4), but these students also

performed higher than expected on their fall senior retest with the full battery (by 0.6 point with a 95% CI of 0.4 to 0.8 among engaged students). A similar result in this subject area was observed when analyses were conducted by ACT score range; that is, this finding was not only seen among higher scoring students but also among lower and mid-range scoring students. While we do not have an explanation for why students performed better than expected in math, these findings suggest that the slightly higher than expected performance on the math section retesting may reflect true learning gains as the effect carried forward to the subsequent testing event.

This study was not without limitations. One limitation was the relatively small sample size per subject area, which tended to result in fairly wide confidence intervals for both average score gains (Table 3) and for differences in actual and expected performance (Tables 4 and 5). In order to make the statement that students performed "as expected" there should be reasonable statistical power to detect whether the average difference is significantly different from zero. Post-analysis power calculations suggested that there was sufficient power to detect average differences that were lower than a full score point within each subject area. Specifically, for the engaged sample, there was at least 80% power to detect a difference in actual and expected performance of 0.46, 0.31, 0.68, and 0.45 or higher in English, math, reading, and science, respectively, and at least 95% power for differences of 0.58, 0.40, 0.88, and 0.58 or higher, respectively. These detectable differences that are associated with sufficient power provide context around the interpretation of students' performance tending to be consistent with expectations or students performing "as expected."

Another limitation of the study was that students did not appear to be as engaged on the section retake as they were on their senior fall retake with the full battery. There were likely several reasons for lower engagement, but two that stand out include: (a) students did not receive college-reportable scores on the section retest, and (b) study recruitment was conducted at the high school level instead of at the student level. Participating high schools tended to have a majority of their students participate in the single-subject retesting experience as it provided a great opportunity for students to gauge their readiness for their upcoming senior retake. But, this appears to have resulted in the inclusion of many students that were not fully engaged in the section retake experience, especially in English and reading. Through the use of non-engagement indicators based on students' test item responses, we attempted to identify those who may not have been fully engaged and then conducted analyses on a subsample of students that excluded those who were flagged as non-engaged on the section retake. This method seemed to be somewhat effective though to a lesser extent for the reading sample. For this reason, results for the engaged sample rather than the full sample may provide a better estimate of the typical score gains associated with section retakes.

Despite these limitations, the results based on this study do not suggest that ACT scores will be artificially inflated through modular or section retesting. A follow-up research study currently underway will examine whether ACT scores obtained via a modular experience are as (or more) predictive of college success as ACT scores obtained via the traditional battery testing. Additionally, once the section retesting option becomes operational in September 2020, ACT is committed to continuing to monitor students' ACT score gains and investigate how students' retesting behaviors and strategies change and whether these changes have any impact on students' ACT performance.

References

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

Allen, J. M., & Mattern, K. (2019). *Validity considerations for 10th-grade ACT state and district testing*. Iowa City, IA: ACT.

Andrews, B. (2019). *Initial evidence in support of section retakes: The impact of administering the ACT subject tests in different orders on ACT scores.* Iowa City, IA: ACT.

Camara, W. J., & Allen, J. (2017). *Does testing data impact student scores on the ACT?* Iowa City, IA: ACT.

Mattern, K., & Radunzel, J. (2019). *Impact of superscoring on subgroup differences.* Iowa City, IA: ACT.

Mattern, K., Radunzel, J., & Andrews, B. (2019). *An initial look: Taking ACT subject tests on different days doesn't result in higher than expected scores.* 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.

Moore, R., Sanchez, E., & San Pedro, M. O. (2018). *Investigating test prep impact on score gains using quasi-experimental propensity score matching*. Iowa City, IA: ACT.

Notes

1. The inclusion of three years of data is consistent with the approach ACT often takes to develop normative results (e.g., percentile ranks).

2. At the end of the section retesting experience, students were asked to complete a short survey about their testing experience, test prep activities, and academic behaviors. For most subject areas, the percentage of students missing these survey responses was relatively high (51% in English, 37% in math, and 60% in science) either due to an entire school skipping the survey or students not responding to the items. In reading, the percentage missing was lower at 9%. One survey question asked students whether they agreed or disagreed with the following statement: *I was motivated to perform my best on today's ACT test*. This variable was not found to be helpful in identifying non-engaged students due to the relatively high missing rate in English, math, and science. It was used in supplemental analyses in reading (see note #7).

3. The engaged sample included students who would have been flagged as non-engaged during their spring or fall testing events (21.7%, 26.7%, 20.6%, and 12.8% of the engaged sample in English, math, reading, and science). For this reason, we kept students that would have been flagged as non-engaged during their spring or fall testing in the state comparison sample.

4. In all subject areas except science, the average score gain from spring to senior fall testing was greater for the study sample than for the state sample. This finding could be due to differences in the students and schools comprising the samples which would include the opportunity to participate in the section retake study as a test prep activity.

5. The lack of association between subsequent ACT scores and the number of months between testing events count be due to the narrow range of possible months between testing events (i.e., ranged from 5 to 8 months).

6. Supplemental analyses were conducted in which expected performance models were developed using data for students from only the participating high schools in 2017 and 2018, as opposed to using the entire state reference sample. In these analyses, students tended to perform better than expected not only in math but also in science; this was seen on both the single-subject retake and their fall senior retake in both subjects. However, the better than expected performance finding in science from these supplemental analyses need to be put into context: the average difference in scores from spring to fall testing for this alternative reference sample was negative (-0.3). That is, students from the participating high schools tended to experience a score decline in science in prior years. In comparison, students in 2019 experienced no change in science from spring to fall testing and a small increase from spring to single-subject retake, on average, that were not significantly different from 0 (Table 3).

7. Supplemental analyses were conducted for the reading sample that excluded from the engaged sample 88 students who strongly disagreed or disagreed that they were motivated to perform their best on the section retake in August/September (see note #2 for more details about this survey item). Even for this alternative engaged subsample (*n* = 198), though the average difference in ACT reading scores from spring to section retesting was 0.2 (95% CI = -0.5, 0.8), students' actual performance tended to be consistent with expected performance based on their prior ACT score and other testing characteristics (average actual = 20.16, average expected = 20.15; average difference between the two = 0.01).

Appendix

Table A1. Number of Students and High Schools by Sample and Subject Area

Sample Characteristics	English	Math	Reading	Science
Section retake study sample				
Took section retake under standard testing time in Aug/Sept 2019	678	695	540	523
Took full battery in spring 2019 (Feb/March/April) and section retake in Aug/Sept 2019 under standard testing time	555	617	451	473
Took full battery in spring 2019 (Feb/March/April), section retake in Aug/Sept 2019, and full battery in Fall 2019 (Sept/Oct)	487	596	402	450
Percentage of students participating in single-subject retake among those that had taken the entire ACT battery in the spring and fall of 2019 at high schools	80.0%	90.0%	53.7%	84.6%
Number of high schools	5	3	5	4
State sample to estimate expected performance				
Took full battery in spring (Feb/March/April), and in fall (Sept/Oct) of the same year under standard testing time*	87,587	87,588	87,535	87,647
Number of high schools	401	401	401	401

Note. Eight students who had taken the ACT prior to Spring 2019 were omitted (five in English, one in reading, and two in science) since an overwhelming majority of students (99.9%) completed their first ACT testing event in spring 2019 and their second official testing event in fall 2019.

*Based on three years of data (2017, 2018, and 2019) for state; sample size varies across subject areas because students from high schools that participated in the subject retake study in 2019 were removed.

Table A2. Regression Estimates Relating ACT Section Retest Scores to Non-Engagement Indicators by Subject

Model	English Est.	English 95% CI	Math Est.	Math 95% CI	Reading Est.	Reading 95% CI	Science Est.	Science 95% CI
Intercept	**19.3**	18.9, 19.7	**19.0**	18.8, 19.3	**19.2**	18.7, 19.7	**18.7**	18.3, 19.0
Spring subject score	**0.8**	0.8, 0.9	**0.8**	0.75, 0.85	**0.7**	0.6, 0.7	**0.7**	0.6, 0.7
Omit indicator	**-6.8**	-7.8, -5.8	**-2.2**	-3.9, -0.5	**-3.3**	-5.7, -0.9	**-2.5**	-4.9, -0.0
Long string indicator	**-2.3**	-3.2, -1.4	0.3	-0.2, 0.7	-0.6	-1.5, 0.4	-0.2	-1.7, 1.2
Rapid guessing indicator	**-1.8**	-3.3, -0.2	**-1.2**	-2.0, -0.3	**-2.2**	-4.2, -0.2	**-2.3**	-3.9, -0.7
Model fit R squared	0.74		0.66		0.53		0.48	

Note. The spring subject score was centered at the sample mean value. Bolded estimates were significantly different from 0. The non-engagement (omit, long string, and rapid guessing) indicators are measuring lack of engagement on the single-subject test administered as part of this study. Est. = estimate and CI = confidence interval.

Table A3. Student Characteristics by Subject and Sample

Student Characteristics	English Study Sample Full	English Study Sample Engaged	English State Sample	Math Study Sample Full	Math Study Sample Engaged	Math State Sample	Reading Study Sample Full	Reading Study Sample Engaged	Reading State Sample	Science Study Sample Full	Science Study Sample Engaged	Science State Sample
Sample size	487	345	87,587	596	420	87,588	402	286	87,535	450	400	87,647
Gender												
Female	55.2	58.3	44.8	55.4	57.1	44.8	50.8	47.6	44.8	50.2	50.2	44.8
Male	44.8	41.7	46.3	44.5	42.6	46.3	49.0	52.1	46.3	49.8	49.8	46.3
Missing	0.0	0.0	8.9	0.2	0.2	8.8	0.2	0.3	8.8	0.0	0.0	8.9
Race/ethnicity												
African American	35.5	29.9	18.9	15.6	13.6	19.1	27.1	26.6	19.0	12.9	11.8	19.1
Asian	1.4	1.5	1.1	3.2	4.3	1.1	0.5	0.4	1.1	1.3	1.5	1.1
Hispanic	9.2	9.6	9.2	9.6	9.1	9.2	9.2	7.7	9.2	5.6	6.0	9.2
White	48.7	54.2	54.4	66.6	68.3	54.4	55.0	57.7	54.4	74.9	76.3	54.3
Other	4.1	4.6	4.6	3.7	3.3	4.6	7.2	6.3	4.6	4.7	4.0	4.6
Missing	1.0	0.3	11.8	1.3	1.4	11.7	1.0	1.4	11.7	0.7	0.5	11.8
Annual family income												
Less than $36,000	28.7	27.0	24.6	25.8	27.4	24.6	8.0	8.0	24.8	11.8	11.0	24.7
$36,000 to $80,000	27.3	26.7	22.0	32.2	32.1	21.9	8.5	8.0	22.1	10.9	11.0	22.1
More than $80,000	25.5	27.5	11.1	24.3	24.3	11.1	3.7	3.9	11.2	8.9	9.2	11.1
Missing	18.5	18.8	42.3	17.6	16.2	42.3	79.9	80.1	42.0	68.4	68.8	42.1
Parental education												
Less than Bach degree	23.2	21.5	23.0	29.9	31.4	23.0	7.2	7.3	23.1	14.9	14.5	23.1
Bach degree	26.1	24.6	19.7	24.3	26.4	19.7	7.0	7.7	19.9	13.8	14.0	19.8
More than Bach degree	36.8	40.6	19.3	36.9	35.0	19.3	5.7	4.6	19.4	15.1	15.5	19.3
Missing	14.0	13.3	38.0	8.9	7.1	38.0	80.1	80.4	37.6	56.2	56.0	37.8
First language												
English	80.3	80.9	46.0	84.1	86.7	46.0	19.4	18.9	46.3	43.3	44.0	46.2
Other	3.9	4.9	3.0	3.7	2.4	2.9	0.2	0.3	3.0	1.3	1.5	3.0
English and Other	2.3	2.0	1.4	3.5	3.6	1.4	0.0	0.0	1.4	0.4	0.5	1.4
Missing	13.6	12.2	49.7	8.7	7.4	49.7	80.4	80.8	49.3	54.9	54.0	49.5
Educational goals												
Less than Bach degree	9.5	5.8	11.8	9.6	11.2	11.8	4.2	4.6	11.8	8.7	8.5	11.8
Bach degree	43.7	45.5	33.1	43.8	42.9	33.1	11.0	10.5	33.3	21.8	21.0	33.2
Beyond Bach degree	28.8	31.3	17.4	34.6	35.0	17.4	6.5	6.3	17.5	13.1	14.2	17.5
Other/missing	18.1	17.4	37.7	12.1	11.0	37.7	78.4	78.7	37.3	56.4	56.3	37.5
HSGPA												
Mean	3.18	3.28	3.03	3.29	3.29	3.03	3.34	3.32	3.03	3.07	3.07	3.03
SD	0.59	0.56	0.66	0.61	0.62	0.66	0.68	0.73	0.66	0.69	0.70	0.66
% missing HSGPA	9.5	9.0	30.1	3.9	1.9	30.2	77.6	78.0	29.8	48.9	48.0	30.0

Note. Bach = bachelor's. Other race/ethnicity includes American Indian students and multi-racial students.

Table A4. Regression Estimates Relating Fall Retest ACT Scores to Initial Spring ACT Scores and Non-Engagement Indicators by Subject for State Reference Sample

Model	English		Math		Reading		Science	
	Est.	95% CI	Est.	95% CI	Est.	95% CI	Est.	95% CI
Intercept	**18.8**	18.5, 19.1	**17.9**	17.7, 18.1	**19.3**	19.0, 19.7	**18.4**	18.1, 18.7
Initial ACT subject score	**0.89**	0.88, 0.89	**0.81**	0.81, 0.82	**0.79**	0.79, 0.80	**0.74**	0.73, 0.74
Omit indicator - first testing	**2.4**	2.3, 2.6	**1.0**	0.9, 1.1	**1.7**	1.4, 1.9	**1.7**	1.5, 2.0
Omit indicator - second testing	**-3.7**	-4.0, -3.5	**-2.3**	-2.5, -2.1	**-3.7**	-4.1, -3.4	**-4.4**	-4.8, -4.0
Interaction between Omit indicators	**0.6**	0.2, 0.9	**0.5**	0.1, 0.8	**0.8**	0.2, 1.4	**1.3**	0.6, 2.1
Long string indicator - first testing	**0.8**	0.8, 0.9	**0.2**	0.1, 0.2	**0.3**	0.2, 0.4	**0.3**	0.2, 0.4
Long string indicator - second testing	**-0.8**	-0.9, -0.8	**-0.2**	-0.3, -0.1	**-1.1**	-1.2, -1.0	**-0.6**	-0.7, -0.6
Interaction between long string indicators	**-0.3**	-0.4, -0.2	**-0.1**	-0.2, -0.0	**0.2**	0.0, 0.3	0.1	-0.1, 0.2
Rapid guessing indicator - first testing	0.0	-0.0, 0.1	**-0.1**	-0.2, -0.1	-0.1	-0.3, 0.0	-0.1	-0.2, 0.0
Rapid guessing indicator - second testing	**-0.9**	-1.0, -0.8	**-0.4**	-0.4, -0.3	**-1.8**	-1.9, -1.6	**-0.6**	-0.7, -0.5
Interaction between rapid guessing indicators	0.2	-0.0, 0.4	0.0	-0.3, 0.2	**0.9**	0.5, 1.3	-0.1	-0.4, 0.3
Number of months between testing	-0.0	-0.1, 0.0	**0.1**	0.0, 0.1	**-0.1**	-0.1, -0.0	0.02	-0.03, 0.1
Model fit *R* squared	0.73		0.65		0.58		0.51	

Note. Students took the entire ACT battery in the spring (initial score in February, March, or April) as part of the State and District testing program and again in the fall (second score in September or October of the same year). The spring ACT subject score was centered at the sample mean value. Bolded estimates were significantly different from 0. The reason for the inclusion of the non-engagement indicators at each testing event is that there could have been a lack of engagement during either event. The lack of an association between the number of months testing and performance on the second ACT test could be due to the fact that there was limited variability in the number of months between testing (ranged from 5 to 8 months). Est. = estimate and CI = confidence interval.

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.

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.