# An Examination of Same-Form Effects on ACT Aspire Interim Assessments

**Wei Tao, PhD, and Yong He, PhD**

## Introduction

The ACT® Aspire® Interim Early High School (EHS) Assessments are designed with content appropriate for students in grades 9 and 10. Because the test forms can be administered at both grades, it is possible for grade 10 students to take the same form they took in grade 9. This document presents a research study examining whether students tend to score higher if they took the same form at both grade levels compared to students who took different forms.

## Study Sample

The student sample in this study included students who took the ACT Aspire Interim EHS Assessments at grade 9 in the academic year of 2017–2018 and in the academic year of 2018–2019 at grade 10. Students taking the accommodation forms were excluded from the study. In each year, there were four Interim forms available to choose from for administration. Teachers could choose any of the four forms in any sequence to test their students anytime during an academic year and as often as teachers thought necessary. A small number of students may take the same form within a year, although teachers would typically administer different forms in an academic year.

We use the following terms to describe students' repeated test attempts:

- **within-year repeats**—students took the same test (e.g., English grade 9) more than once in a given academic year;
- **cross-year repeats**—students took the same test more than once in two consecutive years at two grade levels;
- **same-form repeats**—students took the same form (i.e., same set of items) when retaking a test; and
- **different-form repeats**—students took a different form when retaking the test.

Table 1 presents the sample size and demographic characteristics of students who took the test in two years, as well as the number of within-year repeats. There were around 17,000–19,000 students who had cross-year repeats between 2017–2018 and 2018–2019. For the cross-year repeats, students may have had same-form repeats, different-form repeats, or both. (Later, we provide breakdowns of same-form and different-form repeats.) The sample was demographically diverse with about 50% White, 10% African American, 5% Hispanic, and 3% Asian students; 30% of the students were economically disadvantaged, 2–5% tested with disability status, and 2–4% were English-language learners. Within each year, 95–97% of the students took the same test three times or less. Most of the within-year repeats (> 99%) involved different forms.

**Table 1.** Cross-Year Repeats Between 2017–2018 and 2018–2019

| | English | Math | Reading | Science |
|---|---|---|---|---|
| Number of students | 17,149 | 19,054 | 19,431 | 18,505 |
| Gender (%) | | | | |
| Female | 48.7 | 49.9 | 48.5 | 49.9 |
| Male | 51.3 | 50.1 | 51.5 | 50.1 |
| Race/Ethnicity (%) | | | | |
| African American | 10.8 | 9.9 | 10.3 | 10.5 |
| Asian | 2.9 | 2.8 | 2.9 | 3.1 |
| Hispanic | 5.1 | 7.7 | 7.9 | 5.8 |
| White | 47.1 | 49.7 | 48.8 | 52.9 |
| Other race/ethnicity | 3.3 | 3.3 | 3.4 | 3.5 |
| Missing | 30.8 | 26.7 | 26.7 | 24.1 |
| Economically disadvantaged (%) | 30.6 | 29.4 | 30.5 | 31.0 |
| English-language learner (%) | 1.9 | 3.1 | 3.8 | 1.9 |
| Disability status (%) | 5.6 | 2.2 | 6.1 | 2.4 |
| Percentage of students, by the number of tests taken in 2017–2018 | | | | |
| Once | 12% | 11% | 11% | 11% |
| Twice | 46% | 44% | 43% | 46% |
| Three times | 38% | 40% | 42% | 40% |
| Four times | 5% | 5% | 5% | 4% |
| Percentage of students, by the number of tests taken in 2018–2019 | | | | |
| Once | 19% | 19% | 18% | 19% |
| Twice | 50% | 53% | 51% | 53% |
| Three times | 27% | 24% | 28% | 25% |
| Four times | 3% | 3% | 3% | 4% |

*Note*. Less than 0.1% of the sample took the test more than four times in each year.

# Analyses

The study examined, among cross-year test repeaters, whether students who took the same form tend to score higher than those who took different forms. Three pieces of information were used in evaluating same-form effects, the first two of which are overall indices:

- **difference in gain score,** computed using the average gain score of the same-form repeaters minus the average gain score of the different-form repeaters (Large positive values indicate strong same-form effects);
- **percentage in gain score,** computed as **dividing the difference in gain score** by the average gain score of the different-form group (For example, if the average gain score of the different-form group is 0.8 and the average gain score of the same-form group is 1.0, the percentage in gain score is 25%: (1.0 – 0.8)/0.8. A percentage of 100% indicates that the average gain score of the same-form group was twice as much as the average gain score of the different-form group. Note that this index is sensitive to the magnitude of score gain: the smaller of the gain score, the larger the percentage will be); and
- **visual presentation** of the gain score distributions for the two groups.

Since ACT Aspire Interim Assessments can be administered anytime during an academic year, the time elapsed between pair-wise cross-year test repeats varied. To ensure a valid comparison, time elapsed was controlled when examining same-form effects.

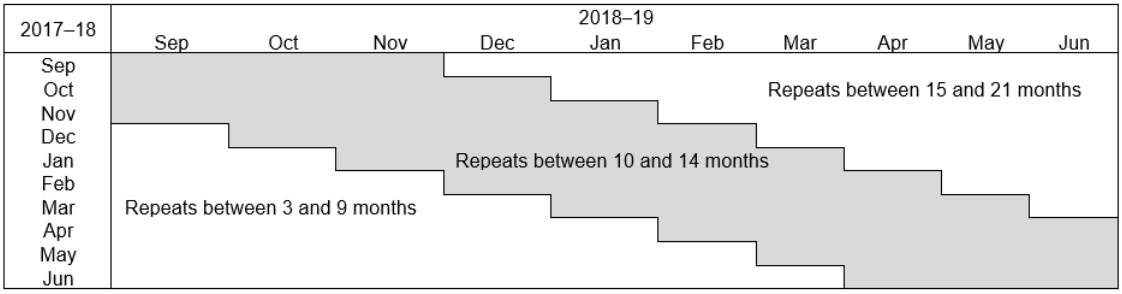
# Results

## Time Elapsed

The time elapsed between cross-year repeats could be as short as 3 months (i.e., tested in June 2018 at grade 9 and again in September 2018 at grade 10) or as long as 21 months (i.e., tested in September 2017 at grade 9 and again in June 2019 at grade 10). To facilitate the analysis, time elapsed was grouped into three categories as illustrated in Figure 1:

- between 3 and 9 months (i.e., spring-to-fall repeats);
- between 10 and 14 months (i.e., about one year apart; fall-to-fall repeats or spring-to-spring repeats); and
- between 15 and 21 months (i.e., fall-to-spring repeats).

**Figure 1.** Time Elapsed (in Months) Between Cross-Year Test Repeats




For the same student, if there were any within-year repeats, there would be multiple pair-wise cross-year repeat comparisons. For example, if a student took the Interim English test twice in both 2017–2018 and 2018–2019, in total there were four different combinations of cross-year repeat records: 1st 2017–2018 test record versus 1st 2018–2019 test record, 1st 2017–2018 record versus 2nd 2018–2019 record, 2nd 2017–2018 record versus 1st 2018–2019 record, and 2nd 2017–18 versus 2nd 2018–2019 record.

Table 2 presents the number and percentage of cross-year repeats by subject and time elapsed. Since we considered all possible combinations of repeats for a single student, the number of records in Table 2 is larger than the number of students shown in Table 1. For all subjects, over half of the cross-year repeats were between 10 and 14 months, around a quarter between 3 and 9 months, and another quarter between 15 and 21 months.

**Table 2.** Number and Percentage of Cross-Year Repeats by Subject and Time Elapsed

| | | Time Elapsed | | |
|---|---|---|---|---|
| Subject | No. of Records | 3–9 months | 10-14 months | 15-21 months |
| English | 90,897 | 23% | 54% | 24% |
| Math | 102,353 | 23% | 54% | 23% |
| Reading | 105,087 | 23% | 53% | 23% |
| Science | 98,493 | 22% | 54% | 24% |

Since most of the test repeats occurred between 10 and 14 months, we present results under this category first, followed by the results under the other two categories.

## Repeats Between 10 and 14 Months

Table 3 presents the same-form effects for repeats between 10 and 14 months. Whether or not students took the same form, the average gain score from 2017–2018 to 2018–2019 is positive. The differences in gain scores and percentage in gain scores (in parentheses) are shown in the same-form effect column.

There were roughly equal numbers of records for same-form repeaters and different-form repeaters. The same-form effects in gain score difference were close to zero for all four subjects, and the same-form effects in gain score percentage were also small (within 10%). Figure 2 plots the gain score distributions for both groups. As shown, the two groups were very similar. Results based on repeats elapsed between 10 and 14 months did not reveal same-form effects.

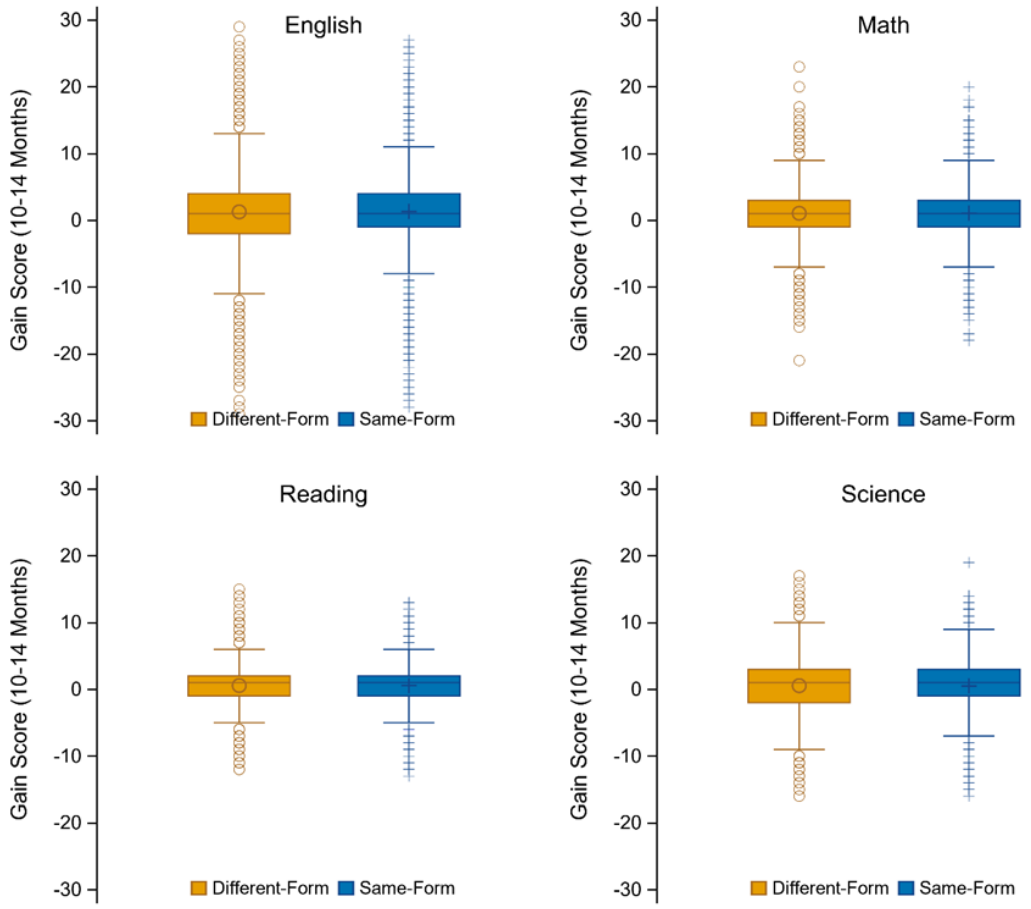
Under the category of time elapsed between 10 and 14 months, the repeats could be from fall to fall or from spring to spring. To further examine whether potential same-form effects exist by season, we separated the records into the fall sample and spring sample: students taking Interim in the fall of 2017–2018 (September to December) were defined as the fall sample; students taking Interim in the spring of 2017–2018 (January to June) were defined as the spring sample. Since the time elapsed was between 10 and 14 months, in essence, the fall sample included fall-to-fall repeats, and the spring sample included spring-to-spring repeats.

**Table 3.** Same-Form Effects: Repeats Between 10 and 14 Months

| Subject (HOSS) | Same Form | No. of Records | No. of Students | 2018 Scale Score Mean (SD) | 2019 Scale Score Mean (SD) | Average Gain Score | Same-Form Effect |
|---|---|---|---|---|---|---|---|
| English (186) | No | 25,124 | 12,877 | 171.09 (7.14) | 172.36 (7.65) | 1.27 | 0.06 (5%) |
| | Yes | 23,525 | 13,107 | 170.41 (7.17) | 171.73 (7.74) | 1.33 | |
| Math (180) | No | 27,401 | 14,172 | 158.34 (3.85) | 159.39 (4.54) | 1.06 | 0.01 (1%) |
| | Yes | 27,868 | 15,399 | 158.29 (3.90) | 159.36 (4.63) | 1.07 | |
| Reading (165) | No | 27,539 | 14,667 | 157.59 (3.47) | 158.16 (3.64) | 0.57 | –0.01 (–2%) |
| | Yes | 28,629 | 15,421 | 157.47 (3.48) | 158.03 (3.67) | 0.56 | |
| Science (171) | No | 26,588 | 14,076 | 161.56 (3.92) | 162.11 (4.15) | 0.55 | –0.04 (–7%) |
| | Yes | 26,410 | 14,509 | 161.36 (3.96) | 161.87 (4.23) | 0.51 | |

*Note*. HOSS = highest obtainable scale score; the Interim lowest obtainable scale score is 150 for all subjects.

**Figure 2.** Distribution of Gain Scores for Time Elapsed Between 10 and 14 Months




Tables 4 and 5 present the fall sample and spring sample results, respectively. In the fall sample, same-form repeaters scored slightly higher than different-form repeaters for English with a difference in gain score of 0.15 scale score points (13% higher) but slightly lower for science with a difference in gain score of –0.08 (15% lower). In the spring sample, the same-form repeaters scored slightly lower than the different-form repeaters for math with a difference in gain score of –0.14 (13% lower). Neither of the two indices was large enough to suggest a strong same-form effect.

Figure 3 presents the gain score distributions by season. In most cases, the variability of the same-form group was noticeably less than that of the different-form group, suggesting that repeating the same form removed some of the variability in gain scores. No pronounced difference was observed between the same-form and different-form groups for either the fall or spring sample.

**Table 4.** Same-Form Effects: Fall-to-Fall Repeats Between 10 and 14 Months

| Subject (HOSS) | Same Form | No. of Records | No. of Students | 2018 Scale Score Mean (SD) | 2019 Scale Score Mean (SD) | Average Gain Score | Same-Form Effect |
|---|---|---|---|---|---|---|---|
| English (186) | No | 12,444 | 9,625 | 170.91 (7.10) | 172.04 (7.62) | 1.13 | 0.15 (13%) |
| | Yes | 15,108 | 12,176 | 170.39 (7.11) | 171.67 (7.74) | 1.28 | |
| Math (180) | No | 13,262 | 10,410 | 158.16 (3.75) | 159.21 (4.47) | 1.04 | 0.10 (10%) |
| | Yes | 18,482 | 14,343 | 158.17 (3.88) | 159.31 (4.62) | 1.14 | |
| Reading (165) | No | 13,915 | 11,038 | 157.58 (3.54) | 158.07 (3.64) | 0.49 | 0.05 (10%) |
| | Yes | 18,732 | 14,513 | 157.48 (3.47) | 158.02 (3.65) | 0.54 | |
| Science (171) | No | 13,215 | 10,301 | 161.53 (3.92) | 162.08 (4.15) | 0.55 | –0.08 (–15%) |
| | Yes | 16,950 | 13,564 | 161.40 (3.91) | 161.87 (4.20) | 0.47 | |

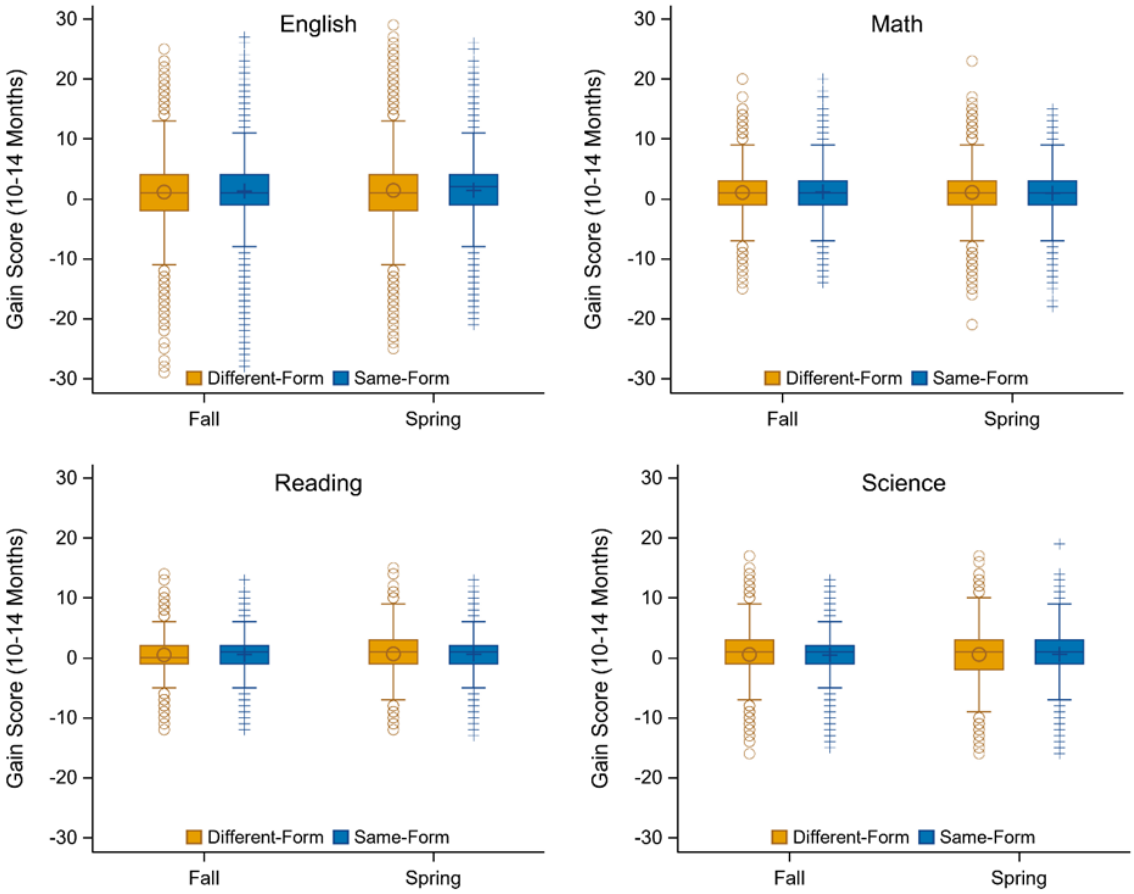
*Note*. HOSS = highest obtainable scale score; the Interim lowest obtainable scale score is 150 for all subjects.

**Table 5.** Same-Form Effects: Spring-to-Spring Repeats Between 10 and 14 Months

| Subject (HOSS) | Same Form | No. of Records | No. of Students | 2018 Scale Score Mean (SD) | 2019 Scale Score Mean (SD) | Average Gain Score | Same-Form Effect |
|---|---|---|---|---|---|---|---|
| English (186) | No | 12,680 | 9,807 | 171.27 (7.17) | 172.67 (7.66) | 1.41 | –0.01 (–1%) |
| | Yes | 8,417 | 7,191 | 170.44 (7.28) | 171.85 (7.74) | 1.40 | |
| Math (180) | No | 14,139 | 10,821 | 158.50 (3.93) | 159.56 (4.59) | 1.07 | –0.14 (–13%) |
| | Yes | 9,386 | 8,043 | 158.51 (3.94) | 159.45 (4.64) | 0.93 | |
| Reading (165) | No | 13,624 | 10,750 | 157.60 (3.40) | 158.25 (3.65) | 0.65 | –0.05 (–8%) |
| | Yes | 9,897 | 8,632 | 157.45 (3.49) | 158.04 (3.70) | 0.60 | |
| Science (171) | No | 13,373 | 10,390 | 161.59 (3.91) | 162.15 (4.15) | 0.56 | 0.02 (4%) |
| | Yes | 9,460 | 8,168 | 161.28 (4.04) | 161.86 (4.30) | 0.58 | |

*Note*. HOSS = highest obtainable scale score; the Interim lowest obtainable scale score is 150 for all subjects.

**Figure 3.** Distribution of Gain Scores for Time Elapsed Between 10 and 14 Months, by Season




## Repeats Between 3 and 9 Months and Between 15 and 21 Months

Table 6 and Figure 4 present similar information as Table 3 and Figure 2, respectively, but with time elapsed between 3 and 9 months. Math and science showed slightly larger same-form effects under this category. Table 7 and Figure 5 present the results for time elapsed between 15 and 21 months. Reading and science showed slightly larger same-form effects.

Under these two categories of time elapsed, the number of students taking the same form was substantially lower than the number of students taking different forms. It turned out that the same-form group was only 10–13% the size of the different-form group for the samples between 15 and 21 months and 5–8% for the samples between 3 and 9 months. Although relatively larger gain scores were observed for some subjects by students taking the same form, this observation could be attributed to factors other than the same-form effects. Our analysis did not account for all possible student and school characteristics that could contribute to the observed differences.

**Table 6.** Same-Form Effects: Repeats Between 3 and 9 Months

| Subject (HOSS) | Same Form | No. of Records | No. of Students | 2018 Scale Score Mean (SD) | 2019 Scale Score Mean (SD) | Average Gain Score | Same-Form Effect |
|---|---|---|---|---|---|---|---|
| English (186) | No | 19,194 | 14,011 | 171.03 (7.29) | 171.75 (7.74) | 0.73 | –0.04 (–5%) |
| | Yes | 1,568 | 1,451 | 172.83 (6.88) | 173.52 (7.55) | 0.69 | |
| Math (180) | No | 22,445 | 15,939 | 158.55 (4.06) | 159.17 (4.56) | 0.62 | 0.34 (55%) |
| | Yes | 1,180 | 1,097 | 159.85 (4.50) | 160.81 (5.08) | 0.96 | |
| Reading (165) | No | 22,982 | 16,273 | 157.67 (3.47) | 158.00 (3.62) | 0.33 | 0.01 (3%) |
| | Yes | 1,685 | 1,584 | 158.68 (3.53) | 159.01 (3.65) | 0.34 | |
| Science (171) | No | 20,837 | 15,274 | 161.54 (4.03) | 161.96 (4.19) | 0.42 | 0.19 (45%) |
| | Yes | 1,187 | 1,088 | 162.40 (4.03) | 163.01 (4.25) | 0.61 | |

*Note.* HOSS = highest obtainable scale score; the Interim lowest obtainable scale score is 150 for all subjects.

**Figure 4.** Distribution of Gain Scores for Time Elapsed Between 3 and 9 Months

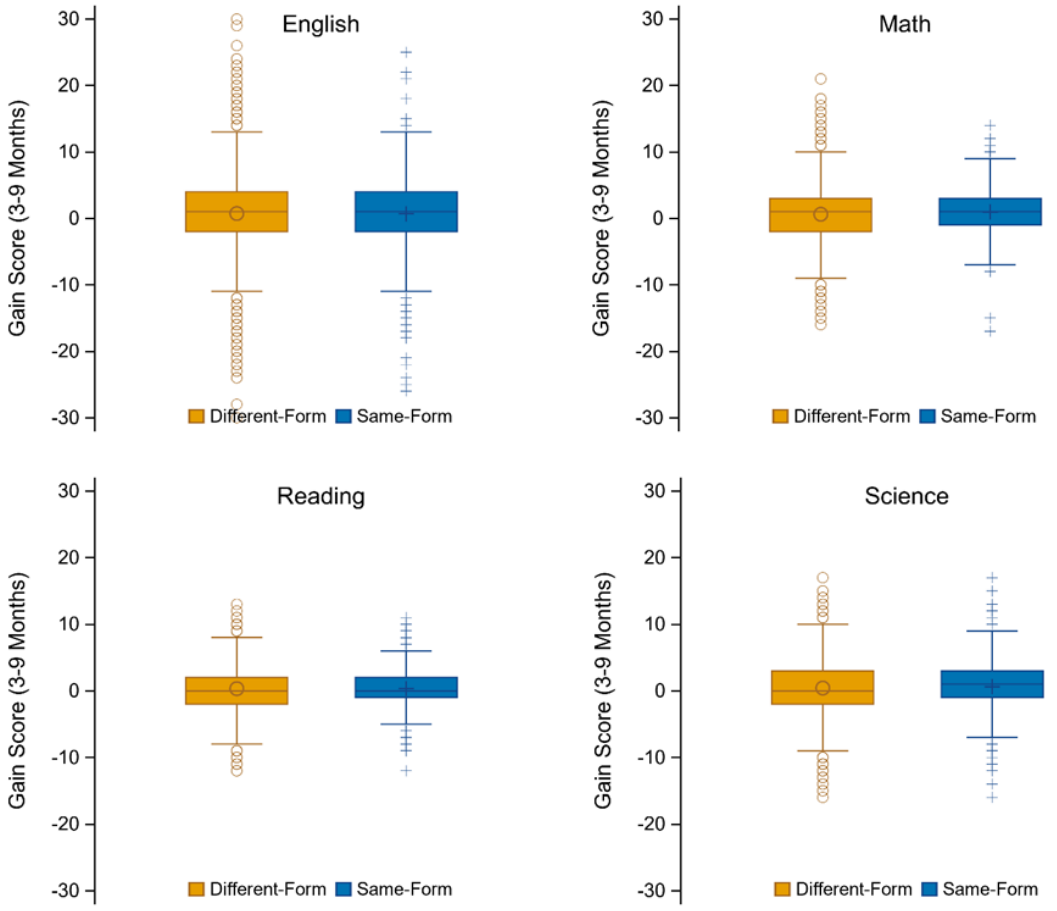
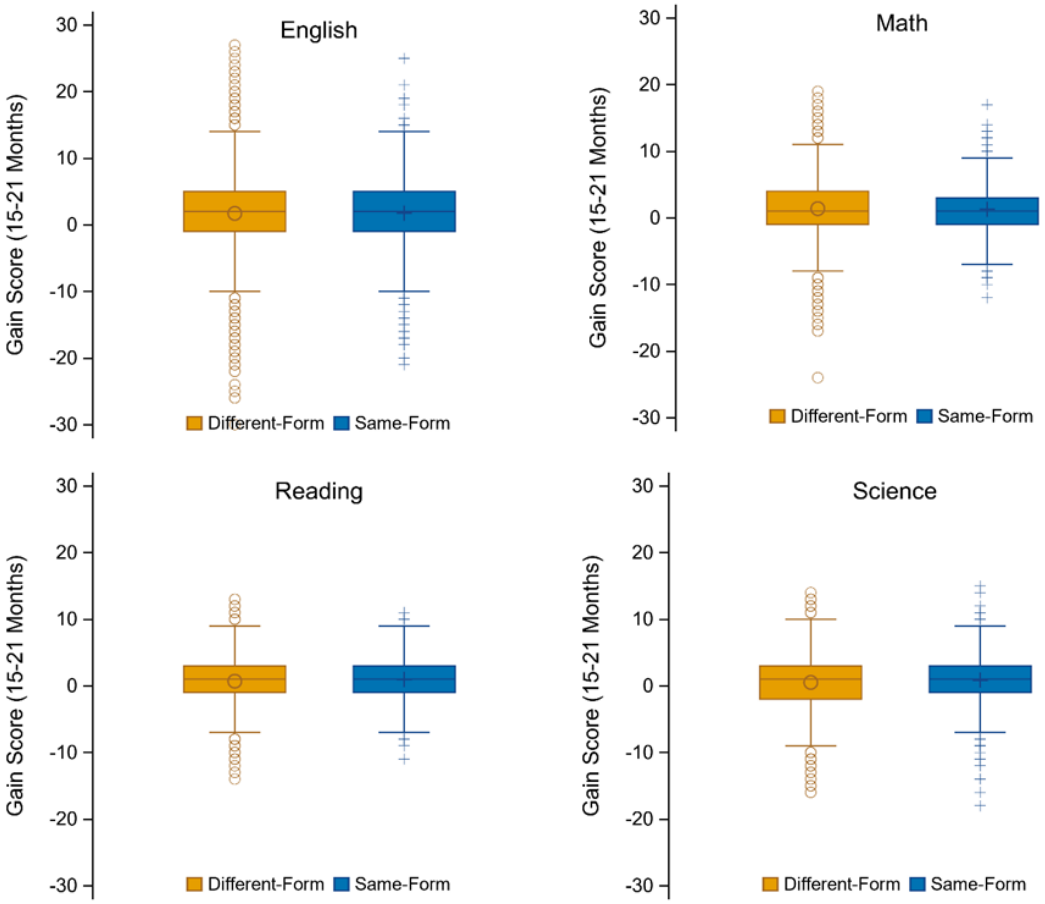


**Table 7.** Same-Form Effects: Repeats Between 15 and 21 Months

| Subject (HOSS) | Same Form | No. of Records | No. of Students | 2018 Scale Score Mean (SD) | 2019 Scale Score Mean (SD) | Average Gain Score | Same-Form Effect |
|---|---|---|---|---|---|---|---|
| English (186) | No | 19,030 | 13,171 | 170.59 (7.09) | 172.31 (7.71) | 1.72 | 0.04 (2%) |
| | Yes | 2,456 | 2,230 | 170.52 (7.11) | 172.28 (7.81) | 1.76 | |
| Math (180) | No | 21,052 | 14,828 | 158.21 (3.88) | 159.59 (4.73) | 1.38 | –0.09 (–7%) |
| | Yes | 2,407 | 2,168 | 157.88 (3.40) | 159.17 (4.27) | 1.29 | |
| Reading (165) | No | 22,017 | 15,106 | 157.57 (3.44) | 158.26 (3.68) | 0.69 | 0.26 (38%) |
| | Yes | 2,235 | 2,027 | 157.21 (3.49) | 158.17 (3.74) | 0.95 | |
| Science (171) | No | 20,804 | 14,202 | 161.53 (3.90) | 162.04 (4.26) | 0.51 | 0.34 (67%) |
| | Yes | 2,667 | 2,502 | 161.19 (3.93) | 162.04 (4.25) | 0.85 | |

*Note*. HOSS = highest obtainable scale score; the Interim lowest obtainable scale score is 150 for all subjects.

**Figure 5.** Distributions of Gain Scores for Time Elapsed Between 15 and 21 Months



## Discussion and Conclusion

In this technical brief, we examined whether students taking the same form tend to score higher on ACT Aspire Interim EHS tests than those who took a different form between grades 9 and 10. Because Interim tests can be administered at any time and in any order (including administering the same form repeatedly) within an academic year, students are not randomly assigned to groups (same-form or different-form). Therefore, differences between groups cannot be attributed only to same-form effects. Based on the factors examined in this study, there is no strong evidence showing same-form effects on the ACT Aspire Interim EHS tests. At most, average test scores were about 0.3 scale score points higher for students who took the same form, relative to students who took a different form. Averaged across all subjects and categories of time elapsed, the ACT Aspire Interim test scores were only about 0.06 points higher for students who take the same form. Considering that the standard deviation of the test scores is typically around four points, the average same-form effect is very small.

**Wei Tao, PhD**
Wei Tao is a lead psychometrician in Psychometric Research specializing in statistical and psychometric methodologies involved in test equating, scaling, and scoring.

**Yong He, PhD**
Yong He is a senior psychometrician in Psychometric Research at ACT specializing in test equating and statistical analyses.