Research Report

2026-11



Evaluating Gender Differences in the Enhanced ACT Composite Score

HEATHER M. BUZICK

ACT®

© 2026 by ACT Education Corp. All rights reserved. | R2624

Conclusions

The analyses suggest that the transition from the legacy four-section ACT Composite score to the enhanced three-section Composite score was associated with, at most, a small change in the female–male Composite score gap. After I accounted for the modest pre-2026 narrowing of the gap, the estimated 2026 shift was 0.129 points on the 1–36 ACT score scale, which is a very small amount. A direct comparison of 2025 and 2026 produced a similarly small estimate of 0.087 points. These results indicate that removing science from the Composite score and administering the enhanced ACT may have slightly widened the female advantage in the Composite score, but the change is very small relative to the ACT score scale and the measurement precision of the Composite score.

So What?

The findings do not currently support the conclusion that the enhanced ACT Composite score creates a meaningful new threat to score interpretation for female or male students. The direction of the change matches what would be expected based on historical differences between subjects. But the change itself is very small, and only 1 year of data has been collected since the scoring change, which is too little to estimate it with much precision. This means that score users can continue to interpret the enhanced Composite score as a broad indicator of academic achievement across English, mathematics, and reading, while recognizing that gender differences should be monitored as part of ACT's broader validity and fairness evidence.

Now What?

Readers should interpret the enhanced ACT Composite score in the same way they interpreted prior Composite scores for female and male students, while recognizing that evidence is still preliminary. The observed change in the female–male gap is small, so the results do not call for changes in how Composite scores are used. Moreover, users should avoid overinterpreting small year-to-year differences and should consider future monitoring results as additional enhanced ACT data become available.

© 2026 by ACT Education Corp. All rights reserved. | R2624

About the Author

Heather M. Buzick is senior director of academic measurement research at ACT. She leads a team whose work supports the validity argument for college and career readiness assessments and the use of test scores and assessment data to inform education decision-making. Her expertise includes large-scale assessment, validity, and statistics.

Acknowledgments

The author would like to thank Bridget Feagler, Colin Dingler, Jeff Allen, and Xin Li for their helpful feedback on earlier versions of this report.

© 2026 by ACT Education Corp. All rights reserved. | R2624

Executive Summary

The enhanced ACT® test was introduced in phases beginning in 2025, with spring 2026 marking its first administration for the spring state school-day testing population examined in this report. ACT enhanced the test in several ways, including making the science test optional in national testing and letting states and districts choose whether to administer science during school-day testing. Because science is no longer a required subject for all testing populations, it was removed from the calculation of the ACT Composite score. The Composite score calculation for the legacy ACT averaged scores from four test sections (English, mathematics, reading, and science), while the Composite score calculation for the enhanced ACT averages scores from three test sections (English, mathematics, and reading). This report addresses a specific question raised by that change. Because male students have historically scored slightly higher than female students in mathematics and science, while female students have scored slightly higher in English and reading, removing science from the Composite score could unintentionally shift the balance of the overall score in favor of female students. The purpose of this study was to determine whether the shift to the three-subject Composite score changed the size of the gap between female and male students' ACT Composite scores and, if so, whether that change is large enough to matter for how ACT Composite scores should be interpreted.

This study includes an analysis of 6 consecutive years (2021 through 2026) of ACT Composite scores from 11th-grade students who took the ACT as part of statewide school-day testing programs. Each year's group of test takers was a new, independent set of hundreds of thousands of students, providing a very large set of data. For each year, the analysis included calculating the average difference between female and male students' Composite scores (referred to as the gender gap) and tracking how that gap changed from year to year, both before and after the 2026 change to the Composite score formula. To test whether the 2026 change in the scoring formula, as well as the other ACT enhancements, was responsible for any shift in the gender gap, the study used a statistical approach called difference-in-differences. This method compares the change in the gender gap after the enhancements with the change observed before the enhancements. Because the scoring formula changed for all students in 2026, the analysis could not rely on a contemporaneous comparison group tested under the prior Composite score formula. Instead, the study used additional statistical checks to evaluate whether the estimated shift was robust to the limited number of year-level observations available for analysis.

Across the study period, the female–male Composite score gap was small overall and changed only modestly after the ACT enhancements took effect. From 2021 to 2025, before the enhancements, female students' advantage on the Composite score had already been narrowing very slightly, indicating that the gap was not perfectly steady before the new scoring approach was introduced. After the existing trend was adjusted for, the estimated change in the female–male gap associated with the 2026 change was about one eighth of a point (0.129 points on the 1–36 ACT score scale; standardized effect size, $d = 0.02$). Comparing 2025 directly with 2026 produced a similar estimate (0.087 points; $d = 0.017$). Both figures represent a small change in practical terms. These findings are broadly consistent with an independent analysis by Allen and Cruce (2025), who recalculated Composite scores with and without the

© 2026 by ACT Education Corp. All rights reserved. | R2624

science test for the same group of students and found that removing science mechanically widened the female–male gap from about 0.2 points to about 0.4 points. This is somewhat larger than, but in the same direction as, the change identified in this study.

Currently, the estimated change in the ACT Composite score gender gap appears too small, and too statistically uncertain, to indicate a meaningful threat to valid score interpretation for female or male students. Nonetheless, because these results are based on a single year of data following the enhancements, a specific population was studied, and the scoring change occurred alongside several other simultaneous test changes, these findings should be treated as preliminary. The female–male Composite score gap should be monitored as additional years of enhanced ACT data become available to determine whether this small, statistically uncertain shift persists, grows, or fades over time.

Introduction

The enhanced ACT was introduced in phases beginning in 2025, with spring 2026 marking its first administration for the state school-day testing population examined in this report. The enhancements included changing the science test from required to optional in national testing and letting states and districts choose whether to administer science during school-day testing. As a result, the ACT Composite score calculation changed. The legacy ACT Composite score, referred to here as EMRS, was calculated by averaging the English, mathematics, reading, and science scale scores and rounding the result to the nearest integer on the 1–36 scale; the enhanced ACT Composite score, referred to here as EMR, is calculated by averaging the English, mathematics, and reading scale scores and rounding the result to the nearest integer on the same scale.

Several other ACT enhancements were introduced concurrently in spring 2026 for school-day testing, including a shorter overall test length, embedded field-test items, and content revisions across all four test sections. These changes provide important context, although they do not suggest an obvious directional mechanism for changing female–male Composite score differences in the way that removing science from the Composite might. Initial evidence also broadly supports continuity in score interpretations from the enhanced ACT (Allen & Cruce, 2025). Nonetheless, because in our analysis sample these enhancements occurred at the same time as the transition from EMRS to EMR, any post-2026 changes should be interpreted within this broader enhancement context.

The research question is whether removing science from the Composite score unintentionally altered gender differences in overall performance. Historically, males have tended to score slightly higher than females in science and mathematics, whereas females have tended to score higher in English and reading. Because EMR excludes science, this study evaluates whether the change produces systematically higher Composite scores for females relative to males, as well as whether any resulting differences exceed those found under EMRS.

To examine whether the transition from EMRS to EMR changed the gender gap in ACT Composite scores, this study uses repeated cross-sections of Grade 11 students who

© 2026 by ACT Education Corp. All rights reserved. | R2624

completed the ACT through spring state testing from 2021 through 2026. We compare female–male Composite score differences before and after the 2026 enhancements using descriptive trends and difference-in-differences models, with an event-study specification used to test the stability of pre-treatment trends. The goal is to determine whether any observed change in the gender gap after the ACT enhancements is statistically significant and large enough to be meaningful for understanding gender differences in academic achievement, including their potential relevance to decisions about admissions and scholarship eligibility.

Background

Previous ACT research suggests that gender differences in ACT scores are modest overall but differ by subject. In the 2025 National ACT Profile Report, females scored higher than males on the Composite (19.5 versus 19.4) and in English (19.0 versus 18.1) and reading (20.6 versus 19.6), whereas males scored higher in mathematics (19.4 versus 18.7) and science (19.8 versus 19.5; ACT, 2025). The 2024 National ACT Profile Report showed the same general pattern (ACT, 2024). Earlier ACT research similarly found very small Composite score differences by gender and showed that observed differences varied by testing population, with gaps diminishing or reversing for statewide test takers compared with those observed for self-selected national test takers (ACT, 2005).

This pattern is relevant to the transition from EMRS to EMR because the new Composite formula removes one subject area in which males have historically scored slightly higher. If science contributes positively to male Composite scores beyond what is reflected in mathematics, removing science could increase the female advantage on the Composite. The magnitude of any such change matters for valid score interpretations from the enhanced ACT.

Method

Sample

The sample comprised repeated cross-sections of Grade 11 students who completed the ACT through spring school-day state testing between 2021 and 2026. Because an independent cohort was drawn at each annual time point, no individual student contributed data across multiple years. Annual sample sizes were in the hundreds of thousands, providing ample statistical power for detecting small effects. Students were classified by a binary gender indicator (female versus male) as recorded in ACT administrative records at the time of testing. Students whose gender was recorded as something other than female or male, or whose gender data were missing, were excluded (19%). Average performance did not differ meaningfully between students classified as female or male and those who indicated another gender or had missing gender data. However, the percentage of students who did not indicate either female or male increased from 4% in 2021 to 20% in 2025 and was 18% in 2026, increases driven by increases in missing data. Because this excluded share grew substantially over the study period, the composition of the excluded group also shifted over time. The Discussion section addresses how that change could affect the observed change in the gap.

© 2026 by ACT Education Corp. All rights reserved. | R2624

Measures

ACT Composite score. The primary outcome was the ACT Composite score, reported on a 1–36 scale. According to ACT technical documentation (ACT, 2026), the Composite score's standard error of measurement (*SEM*) is approximately 0.90 for the legacy (EMRS) scale and approximately 1.3 (1.32–1.35 across the two enhanced forms) for the enhanced (EMR) scale.

Gender. A binary gender variable (female versus male) derived from ACT administrative records served as the grouping variable. The gender gap within each year was operationalized as the female mean ACT Composite score minus the male mean ACT Composite score. Positive values indicate relatively higher female performance, and negative values indicate relatively higher male performance.

Design

The study employed a difference-in-differences (DiD) quasi-experimental design to evaluate whether the ACT enhancements introduced in 2026 altered the gender gap in Composite scores. The years 2021 through 2025 made up the 5-year pre-treatment period, and 2026 served as the single post-treatment observation. The DiD framework identifies the treatment effect by comparing the pre- and post-treatment gender gaps. This approach controls for stable background factors that affect female and male students equally across time.

Data Analysis

All analyses were conducted to evaluate whether the transition from EMRS to EMR was associated with a change in the female–male ACT Composite score gap. The analysis proceeded in several steps. First, descriptive statistics were explored. Next, the stability of the pre-treatment gender gap was assessed by modeling the annual female–male ACT Composite score difference across the 2021–2025 pre-treatment period. This was done by fitting a simple linear regression of the standardized gap on a year variable centered at 2026 (see the appendix). An event-study model was also estimated using 2025 as the reference year to examine the pattern of pre-treatment differences (see the appendix). These analyses were used to evaluate whether the gender gap was approximately stable before the 2026 enhancements and to inform interpretation of the models.

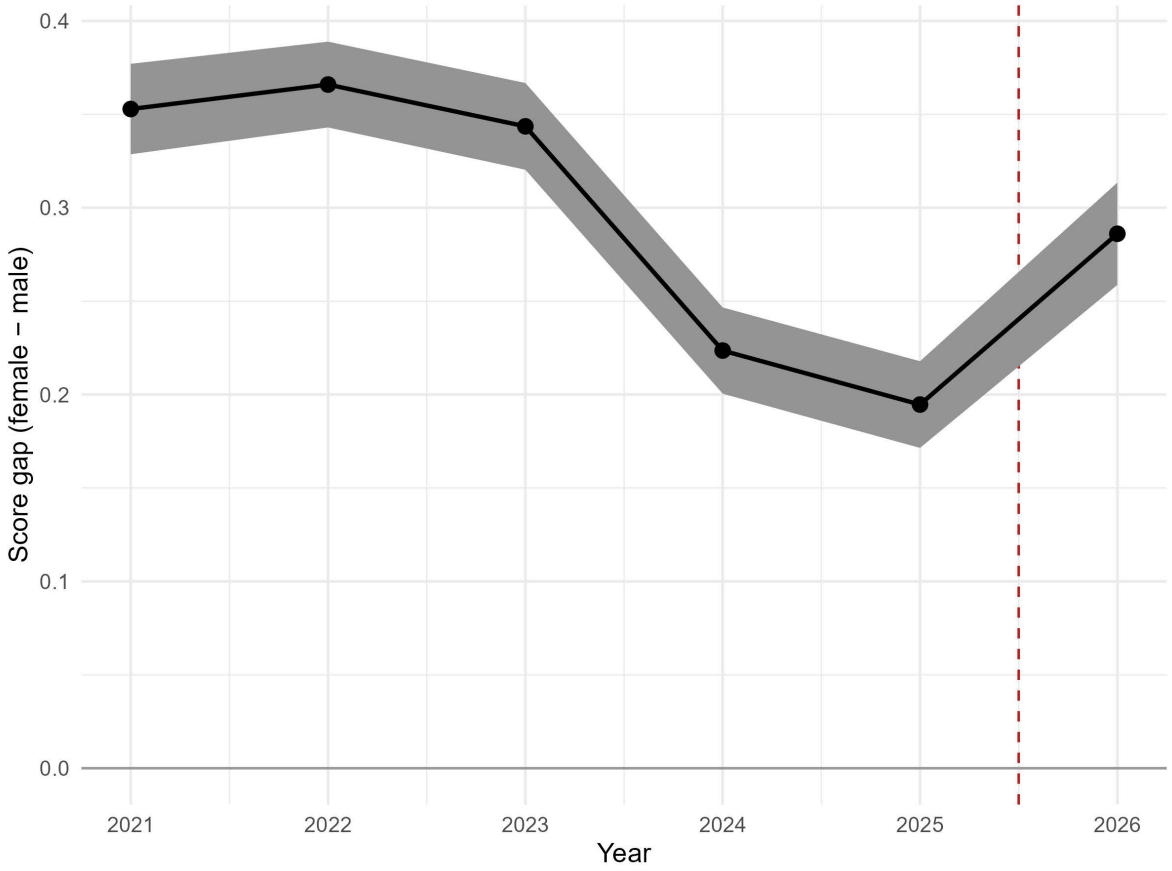
Two DiD models were specified. The primary trend-adjusted model included gender-specific linear time trends. A two-period DiD model comparing 2025 with 2026 was also reported as a sensitivity check. The primary estimand was the change in the female–male ACT Composite score gap in 2026 beyond what would have been expected based on the pre-treatment trend. Models were estimated in R (R Core Team, 2026) using the feols() function from the fixest package (Bergé et al., 2026). The primary trend-adjusted model used HC2 heteroskedasticity-robust standard errors; the two-period model used ordinary least squares standard errors, as described in the appendix.

© 2026 by ACT Education Corp. All rights reserved. | R2624

Results

Figure 1 displays the female–male gap in ACT Composite scores (female mean minus male mean) for each year from 2021 through 2026. The female–male Composite score gap was small across the study period and changed modestly following the transition from EMRS to EMR (and from legacy to enhanced ACT).

Figure 1. Observed Female–Male ACT Composite Score Difference, 2021–2026



Note. Shaded band = 95% confidence interval. Dashed line = introduction of enhanced ACT.

From 2021 to 2025, the female advantage in ACT Composite scores declined slightly, by about 0.009 standard deviations per year (β = −0.009, *SE* = 0.002, *p* = .030). Although small on the 1–36 ACT scale, this pre-2026 trend indicates that the gender gap was not completely stable before the enhanced ACT was introduced. An event-study specification, used as a diagnostic check, also showed modest year-to-year variation in the female–male gap during the pre-2026 period (see Figure A1). The main analyses therefore adjusted for the pre-2026 trend. After that adjustment, the estimated 2026 change in the female–male Composite score gap was 0.129 points. This is a small effect in practical terms, corresponding to a standardized effect size of about *d* = 0.02.

The primary interpretation is that the estimated 2026 change in the female–male score gap is practically small. A conventional standard error treating each of the hundreds of thousands of

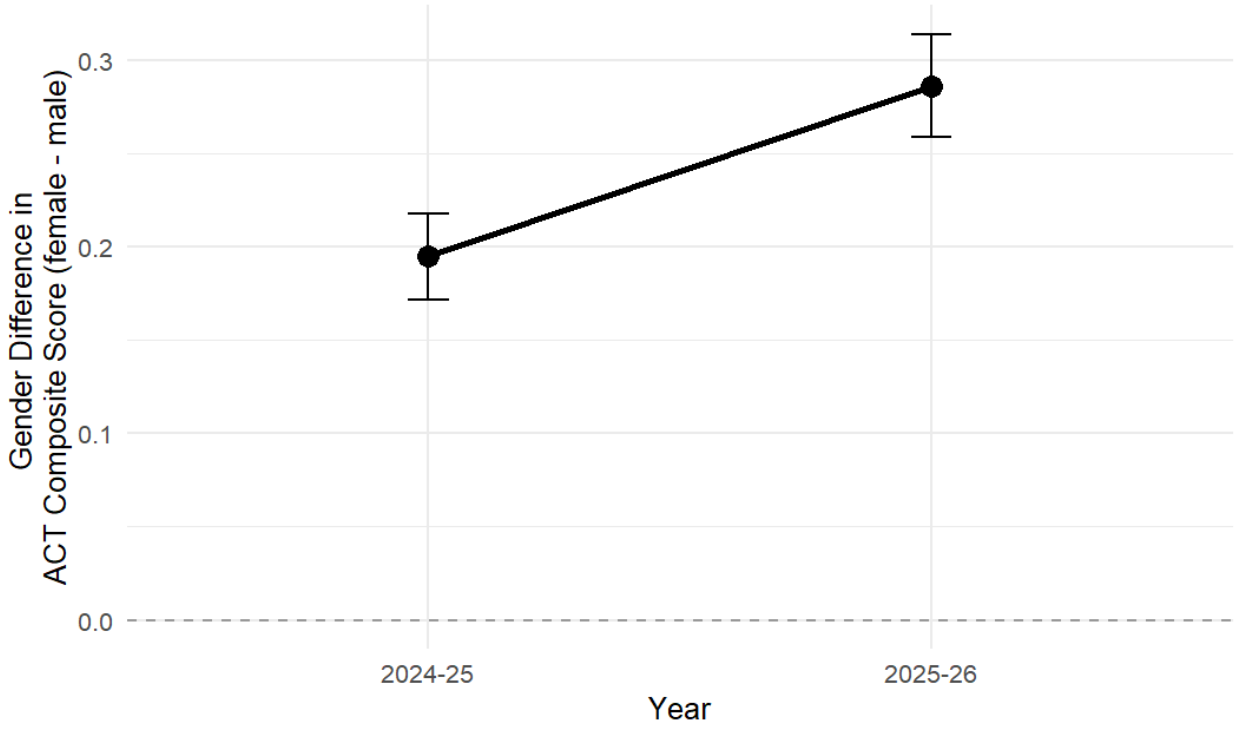
© 2026 by ACT Education Corp. All rights reserved. | R2624

student records as independent evidence suggested that the 2026 shift was statistically significant, but that approach likely overstates precision because the 2026 ACT enhancements and transition from the four-section EMRS Composite to the three-section EMR Composite occurred only once, at the year level, with only 6 years of data. Two approaches better suited to that limitation, a wild cluster bootstrap and a placebo permutation test, produced the same point estimate but found that the shift was not statistically significantly different from zero (p = .219 and p = .167, respectively). Table A1 and Figure A2 summarize these inferential checks and display the placebo distribution.

The two-period model comparing 2025 directly with 2026 produced an estimate of 0.087 points (d = 0.017), indicating that the gender gap was approximately 0.09 points wider in 2026 than in 2025. This effect size is similar in magnitude to the standardized effect size for the trend-adjusted estimate reported above (d = 0.02), showing that both specifications agree on a comparably small shift in the gender gap. Unlike the estimate presented earlier, this estimate does not adjust for the trend in the score gap that was evident from 2021 to 2025. Further, this estimate has not been subjected to the same cluster-robust inference procedures (wild cluster bootstrap and placebo permutation test) applied to the trend-adjusted estimate above. Because it also reflects treatment that varies only at the year level, its conventional OLS standard error likely overstates precision in the same way the unclustered HC2 estimate did for the trend-adjusted specification, and it should be interpreted with the same caution. Figure 2 shows the observed difference in mean test scores across the 2 most recent years.

© 2026 by ACT Education Corp. All rights reserved. | R2624

Figure 2. Female–Male ACT Composite Score Difference, 2025–2026



Note. Error bars = 95% CI.

Discussion

The estimated 2026 change in the female–male ACT Composite score gap was small in practical terms (b = 0.129 points, d = 0.02). This magnitude is broadly consistent with an independent, non-causal benchmark from Allen and Cruce (2025), who found that recomputing Composite scores without science widened the female–male gap from about 0.2 points under EMRS to about 0.4 points under EMR for the same students in the 2024 ACT-tested high school graduating cohort. Taken together, the benchmark and the estimate from the current study suggest that any shift in the Composite score gap is small relative to the scale and measurement precision of the ACT Composite score.

The main inferential caution is that the change to the enhanced ACT and the new Composite score calculation occurred once, at the year level, with only 6 years of data available. Conventional individual-level standard errors therefore overstate precision by treating student records as independent evidence about a year-level change. When the inference was assessed using methods better suited to this limitation, the estimated effect was not statistically distinguishable from zero (see Table A1 and Figure A2). In addition, causal interpretation is limited because the pre-2026 gender gap was not completely stable and because the scoring change occurred alongside other ACT enhancements, including a shorter test, embedded field-test items, and content revisions.

© 2026 by ACT Education Corp. All rights reserved. | R2624

Additional limitations concern the population represented in the analysis. The study relies on a binary female/male classification from administrative records and excludes students who identified as another gender or who had missing gender data. Because the number with missing data increased over time, changes in the composition of the included and excluded groups could have contributed to the observed pattern. The sample is also limited to Grade 11 students tested through spring state school-day programs; results may not generalize to national Saturday testers, international testers, or other ACT-tested populations.

Overall, the available evidence suggests that any change in the female–male ACT Composite score gap is small. The gap is consistent with a broader pattern of modest gender differences in academic achievement rather than a new source of unfairness in ACT score use for female or male students. However, because the findings are based on a single post-enhancement year and a specific testing population, they should be viewed as provisional. Continued monitoring is recommended as additional years of enhanced ACT data become available to determine whether the small 2026 shift persists, grows, or fades over time.

© 2026 by ACT Education Corp. All rights reserved. | R2624

References

ACT. (2026). *Design framework for the ACT enhancements*. https://www.act.org/content/dam/act/unsecured/documents/R2519-Design-Framework-for-the-ACT-Enhancements-2026-02.pdf

ACT. (2025). *National ACT profile report: Graduating class 2025*. https://www.act.org/content/dam/act/unsecured/documents/2025-act-profile-report-us.pdf

ACT. (2024). *National ACT profile report: Graduating class 2024*. https://www.act.org/content/dam/act/unsecured/documents/2024-act-national-graduating-class-profile-report.pdf

ACT. (2005). *Gender fairness using the ACT*. https://www.act.org/content/dam/act/unsecured/documents/gender.pdf

Allen, J., & Cruce, T. (2025). *Initial evidence supporting interpretations of scores from the enhanced ACT test* (Research Report R2425). ACT. https://www.act.org/content/dam/act/unsecured/documents/R2425-Initial-Evidence-Supporting-Interpretations-of-Scores-from-the-Enhanced-ACT-Test-2025-06.pdf

Bergé, L. R., Butts, K., & McDermott, G. (2026). *Fast and user-friendly econometrics estimations: The R package fixest* [Computer software]. https://arxiv.org/html/2601.21749v1

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275. https://doi.org/10.1162/003355304772839588

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414–427. https://doi.org/10.1162/rest.90.3.414

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *The Econometrics Journal*, *21*(2), 114–135. https://doi.org/10.1111/ectj.12107

© 2026 by ACT Education Corp. All rights reserved. | R2624

R Core Team. (2026). *R: A language and environment for statistical computing* (Version 4.61). R Foundation for Statistical Computing. https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf

© 2026 by ACT Education Corp. All rights reserved. | R2624

Appendix

Analysis Approach

To estimate the stability of the pre-treatment gender gap, for each year *t*, a standardized gender gap in test scores was computed as Cohen's *d*:

$$d_t = \frac{M_{t,female} - M_{t,male}}{SD_t}$$

where d_t is the standardized gender gap in year *t*, $M_{t,\text{female}}$ and $M_{t,\text{male}}$ represent the mean scores for female and male students in year *t*, respectively, and SD_t is the standard deviation of scores across all students in year *t*.

To evaluate whether the standardized gender gap changed over time, year was centered at 2026 (i.e., $\text{YearRel}_t = t - 2026$), restricting the sample to years prior to 2026. A simple linear regression was then estimated:

$$d_t = \beta_0 + \beta_1(YearRel_t) + \varepsilon_t$$

In this model, β_0 represents the predicted standardized gender gap at the reference year (2026), and β_1 represents the estimated annual linear change in the standardized gap. The term ε_t is the residual error. A statistically significant β_1 would indicate that the gender gap in test scores has been widening ($\beta_1 > 0$) or narrowing ($\beta_1 < 0$) over the observed period.

As an exploratory robustness check, we examined whether year-to-year changes in the overall mean Composite score explained the pre-2026 instability in the standardized female–male gap. The pre-trend model was extended to include the year-over-year change in the pooled mean Composite score as an additional predictor (β_2). Because the model could use only 4 of the 5 pre-treatment years (the year-over-year change is not available for 2021) and included two predictors plus an intercept, it had only one residual degree of freedom and was not suitable for formal inference. The pooled-mean change coefficient was not statistically significant (β_2 = 0.036, *SE* = 0.069, *p* = .695), and including it did not attenuate the trend coefficient (β_1 = −0.014, compared with −0.009 in the unadjusted model). Thus, this check did not provide evidence that overall-score volatility accounted for the observed pre-2026 trend, although the limited degrees of freedom make the result inconclusive. The unadjusted linear trend implies a cumulative pre-2026 change in the female–male gap of approximately −0.24 points over 5 years (5 × β_1 × the mean of SD_t across 2021–2025), roughly twice the magnitude of the estimated 2026 shift of 0.129 points. This comparison supports adjusting for the existing linear trend in the primary analysis rather than relying on a simple two-period comparison.

To examine how the female–male gap in scores evolved around the ACT enhancements introduced in 2026 for spring state testing, defined here as the transition from the four-section EMRS Composite to the three-section EMR Composite together with other concurrent changes to the ACT, an event-study model was estimated using ordinary least squares. Relative event

© 2026 by ACT Education Corp. All rights reserved. | R2624

time was defined as $YearRel_t = t - 2026$, where t is the year in which a score was observed, indexing each observation's distance in years from 2026, when the enhanced ACT was first administered in spring state testing. The model was specified as

$$Score_{it} = \sum_{\substack{k=-5 \\ k\neq -1}}^{0} \delta_k \mathbf{1}(YearRel_t = k) \times Female_i + \alpha_g + \sum_{k=-5}^{0} \gamma_k \mathbf{1}(YearRel_t = k) + \varepsilon_{it}$$

where $Score_{it}$ is the ACT Composite score of student i in year t, $Female_i$ is an indicator for female students, α_g represents fixed effects for gender group g, and γ_k represents fixed effects for each relative event-time period k. The term ε_{it} is the residual error, and $1(YearRel_t = k)$ is an indicator equal to 1 when an observation is k years from 2026 and 0 otherwise. The index k runs from −5 (2021) to 0 (2026), excluding −1 (2025). Gender has two levels (female and male), so α_g is equivalent to absorbing a single female indicator. The year 2025 serves as the omitted reference category, such that each coefficient δ_k is equal to the difference in the female–male score gap at k years relative to the 2026 ACT enhancements, benchmarked against the gap in 2025 (the year just before the enhancements). Heteroskedasticity-robust standard errors (HC2) were used, without clustering.

To assess the plausibility of the parallel-trends assumption required for DiD analyses, a joint Wald test was conducted on the pre-2026 interaction coefficients (δ_k for relative years −5 to −2), testing the null hypothesis that these coefficients were jointly equal to zero. This null hypothesis was rejected—$F(4, 4{,}456{,}894) = 45.9$, $p < .001$—indicating that the female–male gap in scores was not stable in the years preceding the 2026 ACT enhancements, a violation of the parallel-trends assumption.

Two complementary difference-in-differences (DiD) specifications were estimated to assess whether the 2026 enhancements were associated with a change in the female–male score gap.

Trend-Adjusted Model. To account for preexisting differential trends between female and male students, this specification incorporated a gender-specific linear trend in relative time, estimated over the full panel of years surrounding the 2026 ACT enhancements:

$$Score_{it} = \beta_0 + \beta_1 Female_i + \beta_2 Post_t + \beta_3(Female_i \times Post_t) + \beta_4 YearRel_t + \beta_5(Female_i \times YearRel_t) + \varepsilon_{it}$$

where $Post_t$ is an indicator for the post-enhancement period (1 for 2026 and 0 for 2021–2025), $YearRel_t$ denotes years relative to the 2026 ACT enhancements, and $Female_i \times YearRel_t$ allows female and male students to follow separate linear trajectories in relative time. The intercept β_0 is the expected score for male students when all predictors equal 0, β_1 is the baseline female–male difference, β_2 is the post-period change for male students, β_4 is the linear trend in relative time for male students, β_5 is the female–male difference in that trend, and ε_{it} is the residual error. In this model, β_3 represents the change in the gender gap associated with the 2026 enhancements net of any preexisting linear divergence between female and male students,

© 2026 by ACT Education Corp. All rights reserved. | R2624

providing a more conservative test of whether the enhancements were associated with a shift in the gap beyond what would have been expected from prior trends alone.

Two-Period Model. A simpler model was estimated as a sensitivity check. This specification compared outcomes in 2025 (pre-enhancement) and 2026 (post-enhancement) using ordinary least squares:

$$Score_{it} = \beta_0 + \beta_1 Female_i + \beta_2 Post_t + \beta_3 (Female_i \times Post_t) + \varepsilon_{it}$$

where $Post_t$ is an indicator equal to 1 for 2026 and 0 for 2025 (with $Female_i$ defined as above), β_0 is the intercept, β_1 is the female–male difference in 2025, β_2 is the change from 2025 to 2026 for male students, and ε_{it} is the residual error. The coefficient of interest, β_3, captures the change in the female–male score gap from 2025 to 2026, assuming the gap would otherwise have remained stable across the two periods.

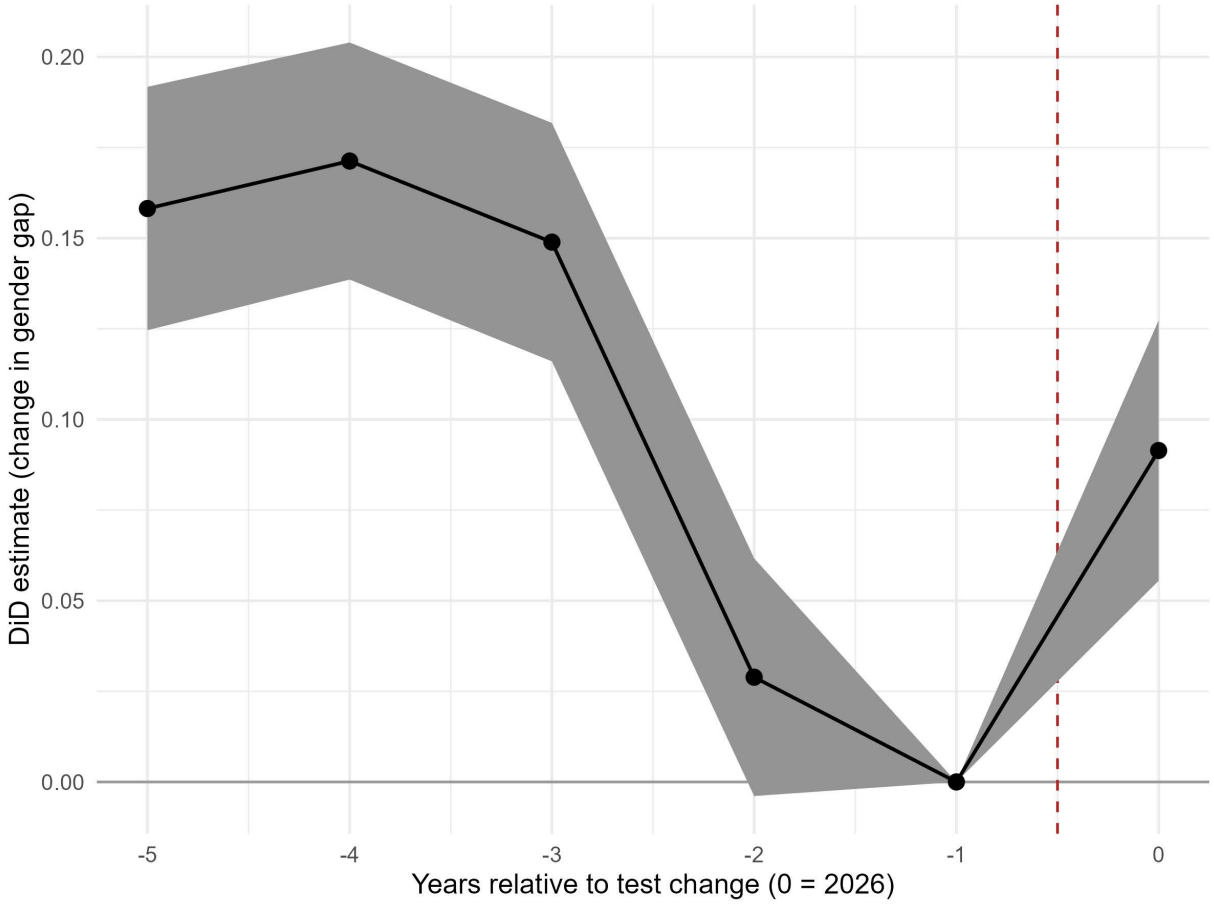
Comparing β_3 across the two specifications assesses the DiD estimate's sensitivity to the parallel-trends assumption. Consistent estimates would strengthen confidence that any change in the gap reflects the 2026 enhancements rather than existing trends.

Because the 2026 enhancements apply at the year level, individual-level heteroskedasticity-robust standard errors can overstate precision by treating each student observation as independent evidence about a change that occurred once, in a single year. We therefore supplemented the primary trend-adjusted model with two inference procedures intended to assess sensitivity in a setting with few year-level clusters and only one treated cluster (Bertrand et al., 2004). First, we applied a wild cluster bootstrap (Cameron et al., 2008), clustering by year with six clusters and 9,999 Rademacher draws. Because 2026 is the only treated year, the design effectively has a single treated cluster, a case in which both cluster-robust *t* tests and the wild cluster bootstrap are known to perform poorly (MacKinnon & Webb, 2018). Accordingly, the bootstrap results are interpreted as evidence concerning the fragility of the inference rather than as a definitive test. Second, we conducted a placebo permutation test by reassigning the post-enhancement label to each of the 5 non-2026 years in turn, re-estimating the gender × post interaction (β_3) under each reassignment, and comparing the observed 2026 estimate with this null distribution. Because only 6 years of data are available, the exact permutation *p* value is bounded below at 1/6 ≈ .167, calculated as (1 + number of placebo estimates at least as extreme as the observed estimate) / (1 + total number of placebo estimates).

© 2026 by ACT Education Corp. All rights reserved. | R2624

Supplemental Results

Figure A1. Event Study: Change in Gender Gap Over Time



Note. The DiD estimate at −1 (reference year) is fixed at 0. Shaded band = 95% CI.

Table A1. Comparison of Inference Methods for the Gender × Post-Treatment Interaction

Method	*b*	*SE* / 95% CI	*p*
HC2, unclustered	0.129	[0.092, 0.166]	<.001
Wild cluster bootstrap, clustered by year	0.129	[−0.591, 0.484]	.219
Permutation test, exact, 6 years	0.129	—	.167

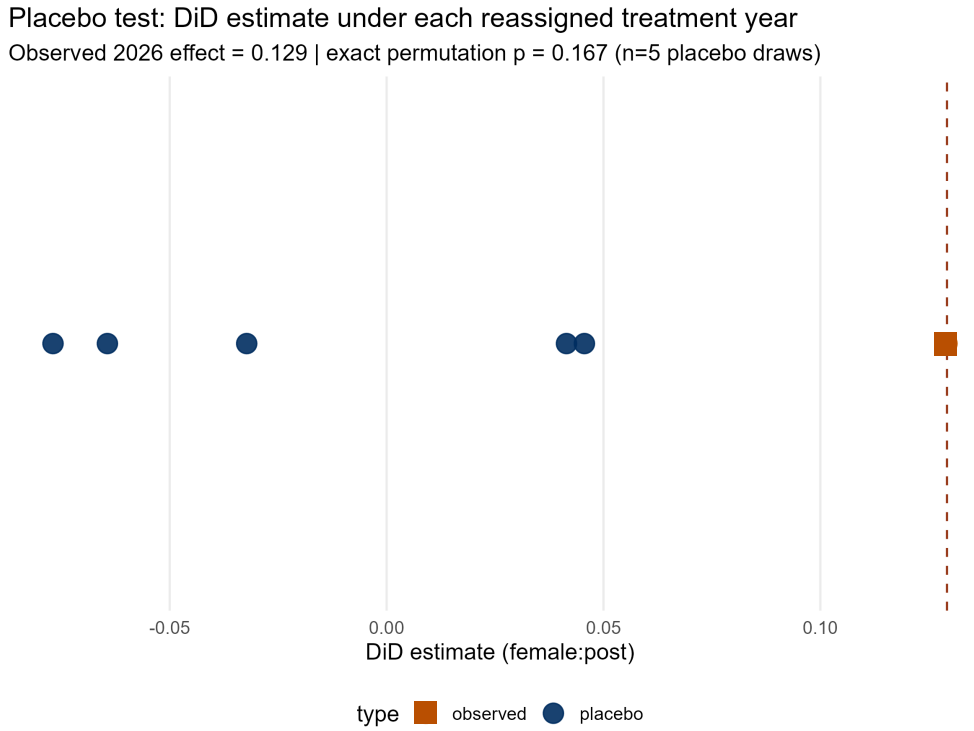
Note. Here, *b* is the estimate of the female × post interaction coefficient (β_3) in the trend-adjusted model. All three methods use the same point estimate but differ in how standard errors and inferential tests are computed. The wild cluster bootstrap and permutation test provide sensitivity checks for inference when treatment varies at the year level with only 6 years of data.

Figure A2 presents the placebo distribution. The observed 2026 estimate was larger than all five placebo estimates, indicating that the 2026 estimate was atypical relative to estimates obtained under alternative treatment-year assignments. However, because only six year-level treatment

© 2026 by ACT Education Corp. All rights reserved. | R2624

assignments were possible, this result should be interpreted as descriptive evidence rather than as a conventional test of statistical significance.

Figure A2. Placebo Permutation Test of the 2026 Difference-in-Differences Estimate

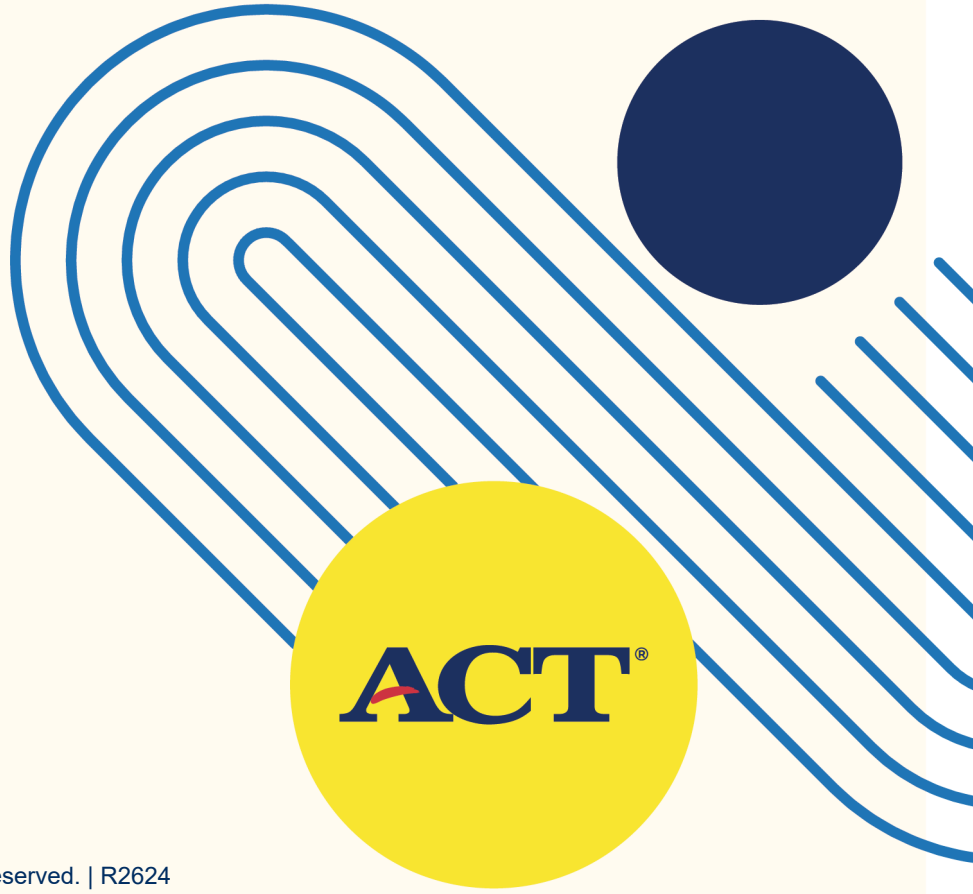


Note. Navy blue squares represent the female × post coefficient obtained when each of the 5 non-treatment years (2021–2025) is alternately assigned as the post period, holding all other model specifications constant. The dashed line and orange square indicate the observed effect when 2026 is correctly specified as the post period (*b* = 0.129).

© 2026 by ACT Education Corp. All rights reserved. | R2624

ABOUT ACT

ACT is transforming college and career readiness pathways so that everyone can discover and fulfill their potential. Grounded in more than 65 years of research, ACT's learning resources, assessments, research, and work-ready credentials are trusted by students, job seekers, educators, schools, government agencies, and employers in the U.S. and around the world to help people achieve their education and career goals at every stage of life. Visit us at www.act.org.



© 2026 by ACT Education Corp. All rights reserved. | R2624