
Understanding the Impact of the COVID-19 Pandemic on Student Achievement in Arkansas: Comparing On-Site, Hybrid, and Virtual Modes of Learning

Jeff Allen

Introduction

The COVID-19 pandemic caused widespread disruptions to the educational system in Arkansas and across the United States. At the onset of the pandemic in March 2020, schools in Arkansas were forced to replace on-site instruction with virtual instruction. During the 2020–2021 academic year, there were three student instructional options: on-site/traditional learning, hybrid/blended learning, or virtual/remote learning.

Arkansas has administered the ACT[®] Aspire Assessment System (3rd–10th grades) and the ACT[®] test (11th grade) statewide since spring 2016. ACT Aspire testing was canceled in spring 2020 because of the pandemic. Using ACT Aspire and ACT results from spring 2019 and spring 2021, the research summarized in this paper examines impacts of the pandemic on student achievement in Arkansas.

The paper is organized into two parts:

- **Comparing Performance from 2019 and 2021.** We present a formal analysis of performance in spring 2019 (before the pandemic) and spring 2021 (during the pandemic) and account for changes in test participation and the tested population across years.
- **Comparing On-Site, Hybrid, and Virtual Modes of Learning.** We examine whether performance in spring 2021 varied by mode of learning and compare results for each mode-of-learning group to their performance in spring 2019.

Because the same primary test forms were used in 2019 and 2021, we can be confident that differences between the two cohorts are not confounded by test form effects. The analysis focused on the change in average (mean) test scores and was conducted for each test section and grade level (3rd–11th grades).



Comparing Performance from 2019 and 2021

Comparing Performance from 2019 and 2021: Data and Methods

ACT Aspire and 11th-grade ACT test scores from 2019 and 2021 were matched to student enrollment and demographic data provided by the Arkansas Department of Education. For 5th–11th grades, the data were also matched to prior ACT Aspire Summative test scores (from spring 2017 for the 2019 cohort and from spring 2019 for the 2021 cohort). Finally, the data were matched to instructional option data provided by the Arkansas Department of Education. As described later, the instructional option data were used to group students according to their mode of learning (on-site, hybrid, or virtual) during the 2020–2021 academic year.

Table 1. Sample Sizes for 2021 Cohort for Analysis of COVID-19 Impacts on Composite Scores

Grade Level	Population	Analysis Sample		
		Status	Growth	Mode of Learning
3	35,491	34,628	—	30,146
4	35,502	34,664	—	29,957
5	36,133	35,261	32,454	28,366
6	37,421	36,248	33,428	28,726
7	38,566	37,219	34,264	29,281
8	38,798	37,355	34,342	28,845
9	36,884	34,868	31,721	26,748
10	36,066	33,492	30,663	25,191
11	33,854	26,966	22,726	18,918

At each grade level (3rd–11th), analysis was conducted for each test section score as well as English language arts (ELA), science, technology, engineering, and mathematics (STEM), and ACT and ACT Aspire Composite scores. Table 1 shows the sample sizes for the analysis of Composite scores for the 2021 cohort. “Population” represents the number of students in the Arkansas population with demographic data. “Status” represents the number of students in the Arkansas population with a Composite score. “Growth” represents the subset of students in the status sample who also had a prior ACT Aspire Composite score. As described later, “Mode of Learning” is the subset of students who also had mode-of-learning data (on-site, hybrid, or virtual) reported for the 2020–2021 academic year.

To ensure that the two cohorts were similar in terms of demographic characteristics, we used propensity score weighting (Austin, 2011) to weight each group to be similar to the pooled data set, which combined the 2019 and 2021 data. The procedure used

logistic regression to estimate each student's propensity for being in each cohort based on their gender, race/ethnicity, disability status, economic disadvantage status, and English learner status (i.e., covariates). For 5th–11th grades, prior ACT Aspire Summative test scores (from spring 2017 for the 2019 cohort and from spring 2019 for the 2021 cohort) were used to ensure that the two cohorts (2019 and 2021) did not differ on prior academic achievement.¹

Using the weighted data, we fit hierarchical linear regression models in order to estimate the difference in average test scores from 2019 to 2021. The models included a random intercept for the school effect and included the same covariates that were used in the propensity score weighting model. The general form of the regression model was:

test score = cohort + covariates + school effect

The regression model estimated the difference in average test scores between 2019 and 2021, adjusted for changes in the tested population. We refer to this as the **adjusted score difference**. The adjusted score difference can be expressed in three ways:

1. On the original test score scale (e.g., 400–452 for the ACT Aspire Composite score)
2. In standard deviation units (d_{sd} = adjusted score difference divided by the pooled standard deviation of the test score)
3. In average annual gain units (d_{gain} = adjusted score difference divided by the estimated score gain normally observed over one academic year)²

The hierarchical linear regression models were also used to estimate the adjusted score differences for different student groups, including

- male and female students;
- racial/ethnic groups: African American, Asian, Hispanic, Native American/Native Hawaiian, White, and two or more races;
- students with disabilities and students without disabilities;
- students from low-income households and students from higher-income households; and
- English learners and non-English learner students.

Group-specific estimates were obtained by fitting the hierarchical linear regression model with interactions between the cohort and the group indicator.

Comparing Performance from 2019 and 2021: Total Group Results

The adjusted Composite score differences provide an overall estimate of the pandemic's impact on ACT Aspire and ACT test scores. Figure 1 provides the estimated Composite score declines from 2019 to 2021 by grade level. The estimate based on the Growth sample, which accounts for prior test scores, student demographics, and school effects, is reported by the orange line. The estimate based on the Status sample, which accounts for student demographics and school effects, is reported by the blue line. The Growth sample estimate is available only for 5th–11th grades, whereas the Status sample estimate is available for 3rd–11th grades. For 11th grade, note that the score declines are expressed on the ACT Composite score scale; therefore, the lines are not connected between 10th grade and 11th grade.

For 6th–11th grades, the declines in Composite scores are more severe for the Growth sample, suggesting that not accounting for prior test scores (Status model) may lead to understating the pandemic's impact on ACT Aspire and ACT test scores. Figure 1 shows that Composite score declines were observed for all grade levels.

Figure 1. Decline in Average Composite Scores from 2019 to 2021

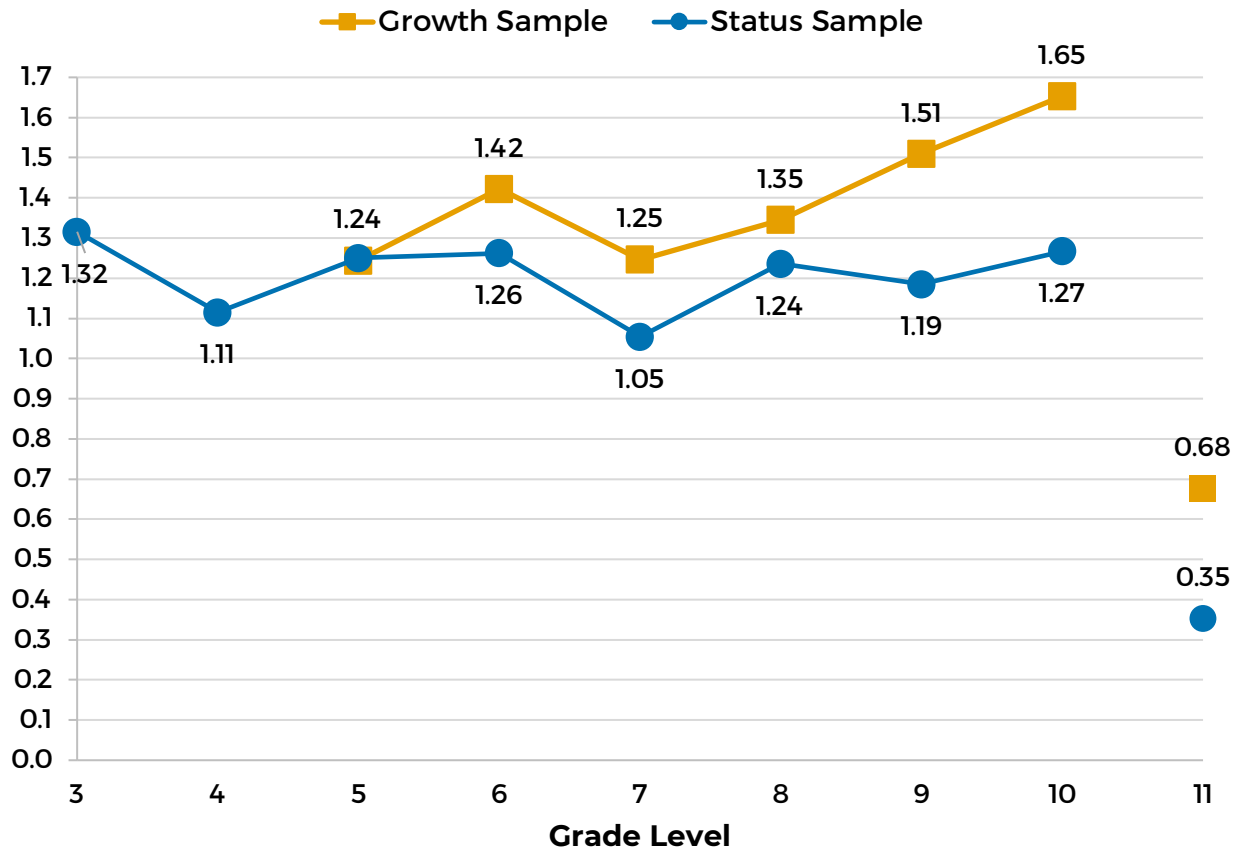


Figure 2 provides the estimated Composite score declines, in standard deviation units, by grade level. The estimated score declines in Figure 2 can be interpreted as standardized effect sizes. Score declines reported in this fashion allow us to interpret the score declines relative to the distribution of test scores at each grade level. For example, $d_{sd} = 0.10$ is comparable to a decrease of four percentile rank units (e.g., moving from the 50th to the 46th percentile). Generally, the standardized score declines were larger for lower grade levels. For 3rd grade, Composite scores declined by 0.25 standard deviation units, which is like students' overall performance dropping from the 50th to the 40th percentile.

Figure 2. Decline in Average Composite Scores, in Standard Deviation Units, from 2019 to 2021

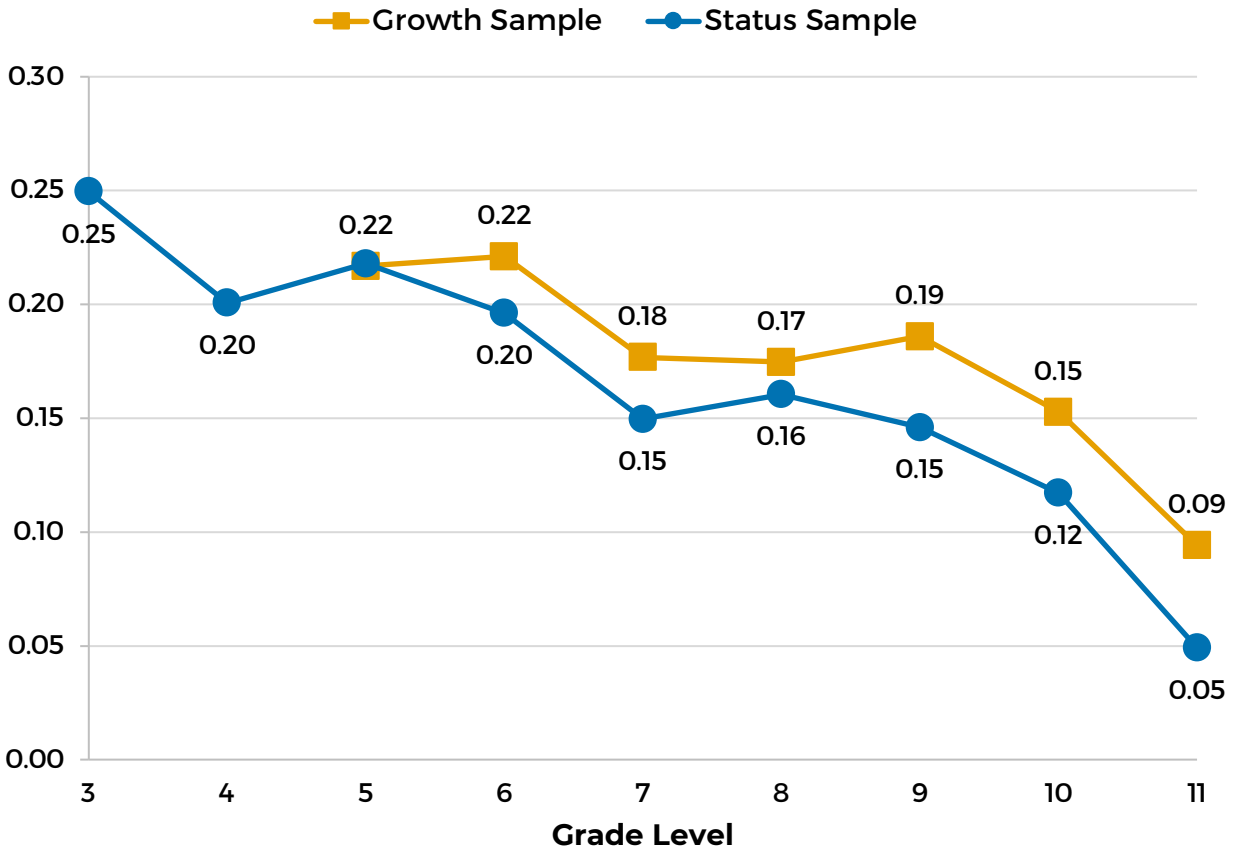


Figure 3 provides the estimated Composite score declines, in average annual gain units, by grade level. The estimated score declines in Figure 3 allow us to interpret the score declines relative to the annual score gain that is normally observed for each grade level. For example, $d_{gain} = 0.50$ indicates that the score decline is comparable to half of the score gain that is normally observed in one year. The most extreme result was observed for 9th-grade Composite scores ($d_{gain} = 1.07$), suggesting that the decline in scores from 2019 to 2021 is comparable to approximately one year of typical score gain. It is worth noting that d_{gain} is based on estimates for both the numerator (the adjusted score difference) and the denominator (the average score gain from 8th to 9th grade), so it is especially prone to being overestimated or underestimated. Reflecting this uncertainty, a 90% confidence interval for d_{gain} in this case is [0.78, 1.70].

Relative to the average annual gains, the score declines were smaller for lower grade levels, with the exception of the 10th-grade ACT Aspire and 11th-grade ACT test

results. Contrasting Figures 2 and 3, we see that the relationship between grade level and score decline depends on which score decline metric is used. Relative to the distribution of test scores at each grade level, the score declines are less severe for higher grade levels. But relative to annual score gains, the score declines are generally more severe for higher grade levels. This paradox is caused by the variability (standard deviation) of test scores increasing with grade level, while average score gains tend to decrease with grade level.

Figure 3. Decline in Average Composite Scores, in Average Annual Gain Units, from 2019 to 2021

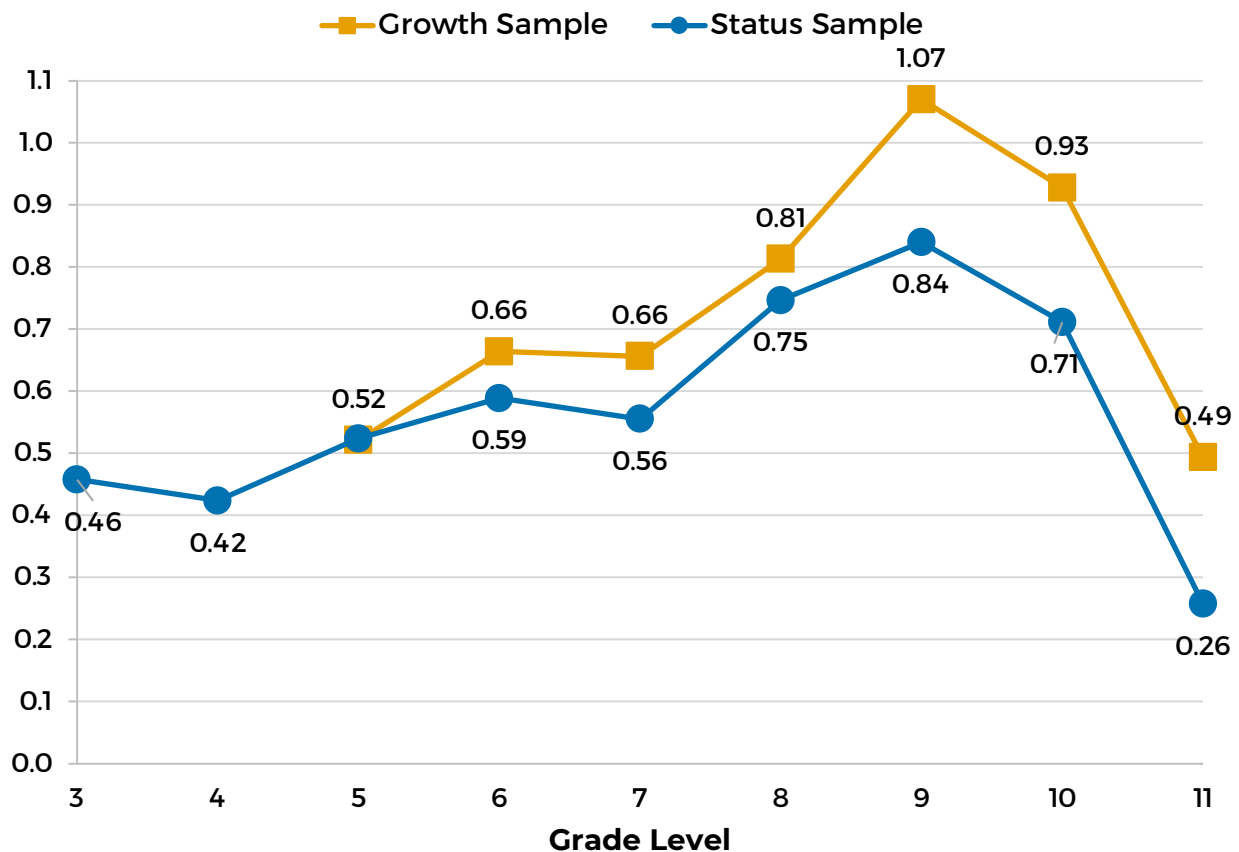
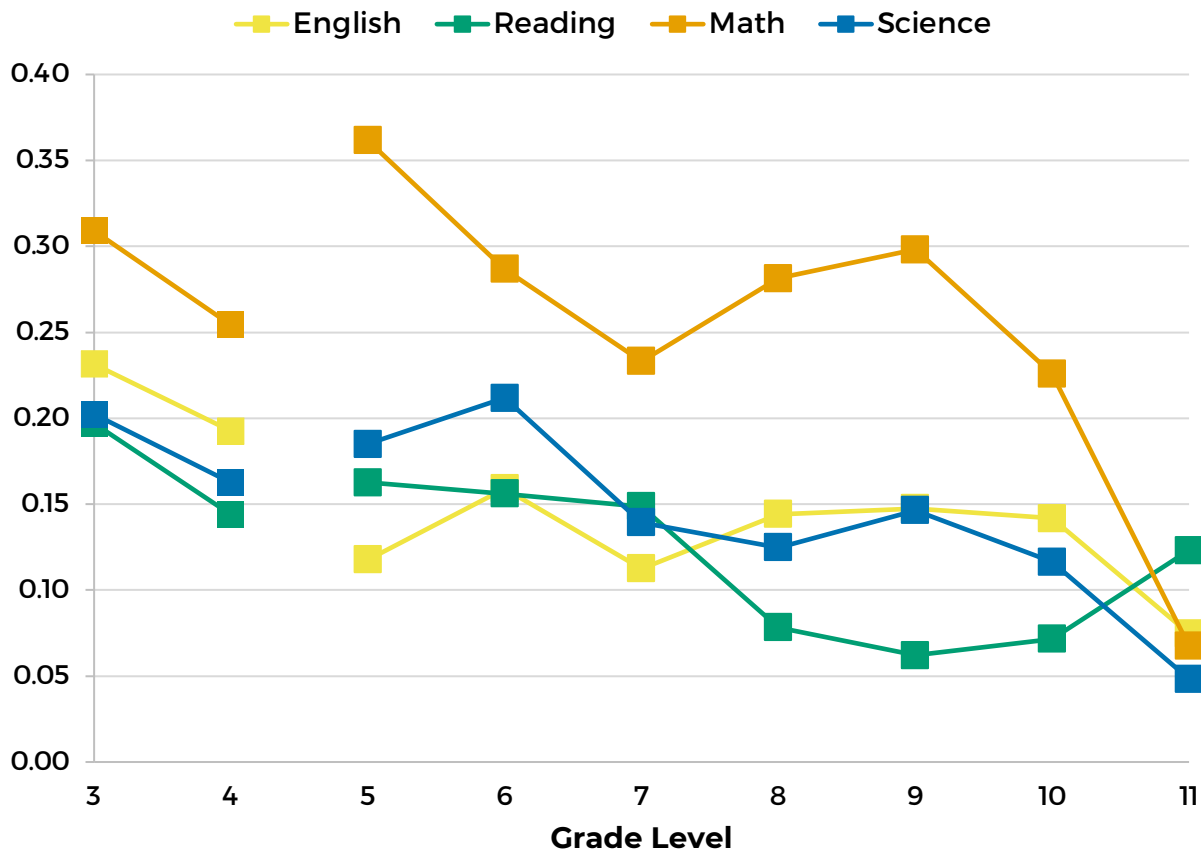


Figure 4 provides the estimated score declines, in standard deviation units, by section and grade level. Note that the estimates are based on the Status sample for 3rd and 4th grades and the Growth sample for 5th–11th grades. Therefore, the results for 3rd and 4th grades may understate the pandemic’s impact. Because they are based on different statistical models (the Growth model uses prior test scores as covariates, whereas the Status model does not), the lines between 4th and 5th grade are not connected. For 3rd–10th grades, we see that the relative score declines were

much larger in math relative to the other test sections. The score declines were generally more severe for English and science than for reading. For all sections and grade levels, average scores declined from 2019 to 2021.

Figure 4. Decline in Average Section Test Scores, in Standard Deviation Units, from 2019 to 2021



Comparing Performance from 2019 and 2021: Results by Student Group

The decline in ACT Aspire and ACT test scores from 2019 to 2021 was also examined for different student groups. Figure 5 shows the decline in average Composite scores, by grade level, for African American, Hispanic, and White racial/ethnic groups. The 2021 sample sizes used for the analysis varied across grade levels: The number of students who are African American ranged from 3,846 for 11th grade to 6,870 for 7th grade; the number of students who are Hispanic ranged from 2,738 for 11th grade to 5,029 for 4th grade; and the number of students who are White ranged from 14,916 for 11th grade to 20,735 for 8th grade. For 5th–11th grades, the score declines were

less severe for the African American and Hispanic groups relative to the White group. For 3rd and 4th grades, the estimated score declines do not account for prior test scores and are based on the Status sample rather than the Growth sample. For 3rd and 4th grades, the estimated score declines are more severe for the African American group relative to the Hispanic and White groups.

Figure 5. Decline in Average Composite Scores by Race/Ethnicity, in Standard Deviation Units, from 2019 to 2021

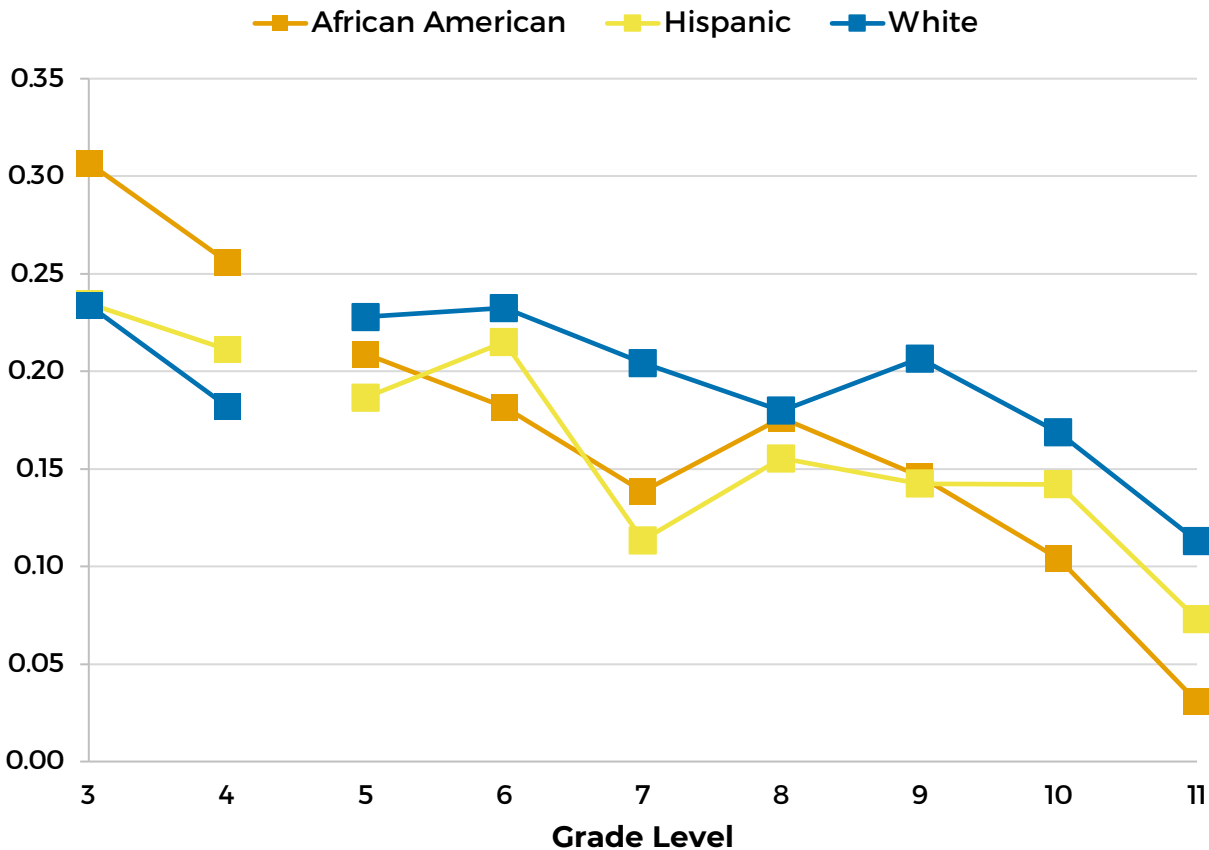


Figure 6 shows the decline in average Composite scores by grade level and disability status. The 2021 sample sizes used for the analysis varied across grade levels: The number of students with disabilities ranged from 1,720 for 11th grade to 5,291 for 3rd grade. The score declines were consistently less severe for students with disabilities relative to students without disabilities.

Similarly, Figure 7 shows that the score declines were consistently less severe for English learners relative to non-English learners. Note that the 2021 sample sizes for English learners ranged from 1,124 for 11th grade to 3,243 for 3rd grade.

While the results show that the score declines were less severe for students with disabilities and English learners, it is not clear if this is due to scale “floor” effects, range restriction, or true differences in the pandemic’s impact across different student groups. For example, higher-performing student groups had more room to go down, and so their test scores may have been more sensitive to disruptions in schooling. Lower-performing groups may have experienced the same or larger educational impacts, but they may have resulted in smaller score changes because of the floor effects or restricted range of scores.

Figure 6. Decline in Average Composite Scores by Disability Status, in Standard Deviation Units, from 2019 to 2021

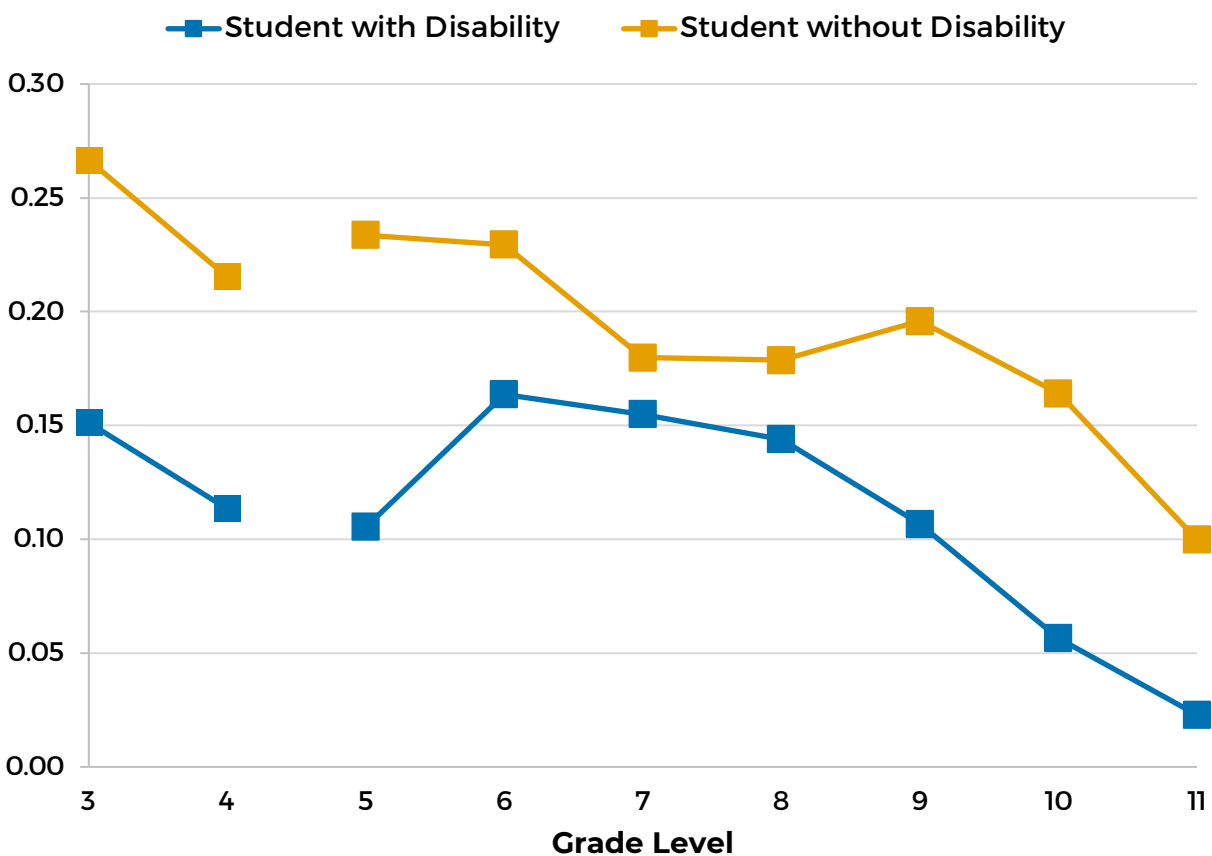
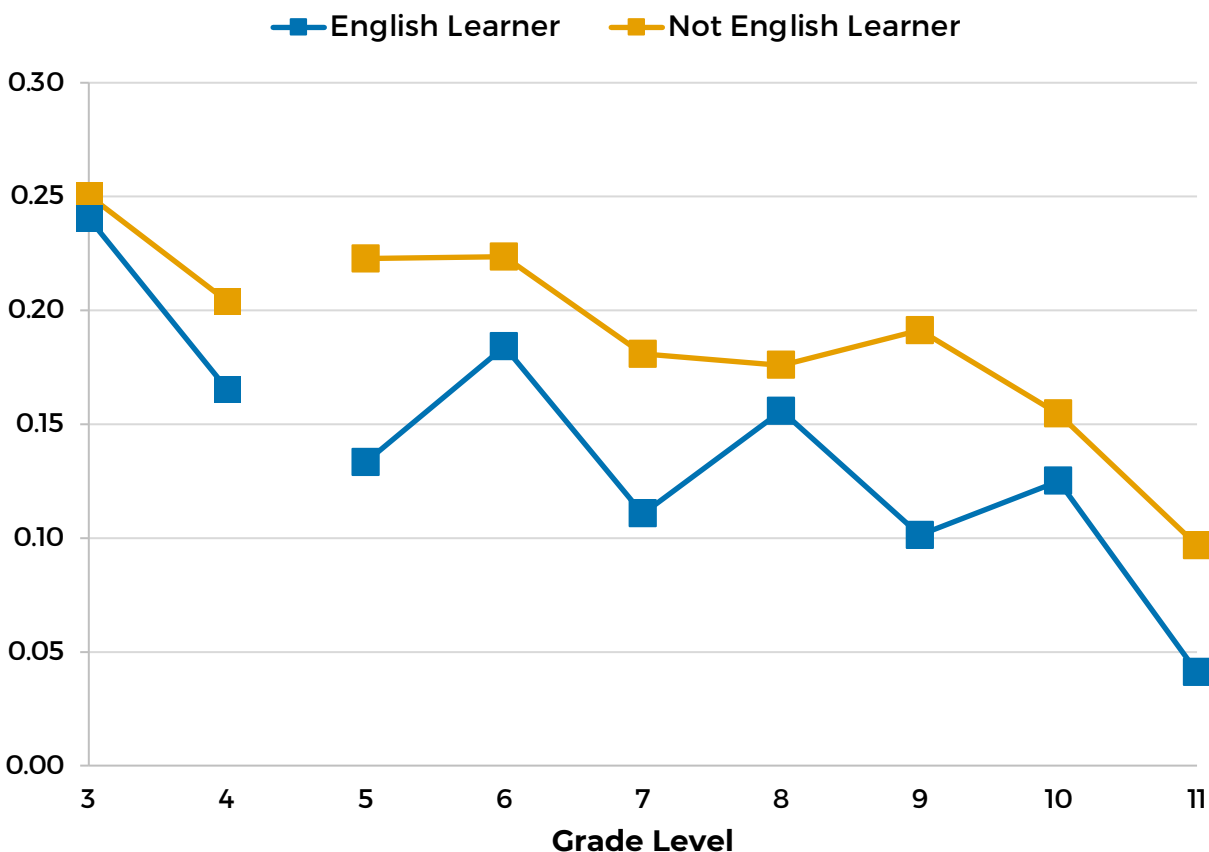


Figure 7. Decline in Average Composite Scores by English Learner Status, in Standard Deviation Units, from 2019 to 2021



Score declines were also examined by gender and by economic status (results not presented). Score declines were generally larger for female students than male students, but the differences were very small. The differences in score declines by economic status were small and inconsistent across grade levels.

Comparing Performance from 2019 and 2021: Summary

- After controlling for changes in the tested population from 2019 to 2021, we found that average ACT Aspire and ACT scores declined for all test sections and grade levels, suggesting that the COVID-19 pandemic has had a negative impact on student achievement.
- Score declines expressed relative to the standard deviation of test scores suggest that the severity of the pandemic's impact decreases with grade level; score declines expressed relative to average annual score gains suggest that the severity of the pandemic's impact increases with grade level.

- The pandemic-related score declines are most severe for math and least severe for reading.
- For 5th–11th grades, the score declines were less severe for African American and Hispanic students relative to White students.
- The score declines for students with disabilities and English learners were less severe than those for students without disabilities or students who are not English learners.

Comparing On-Site, Hybrid, and Virtual Modes of Learning

Comparing On-Site, Hybrid, and Virtual Modes of Learning: Data and Methods

The data used for this analysis are a subset of the Growth sample (for 5th–11th grades) and the Status sample (for 3rd and 4th grades). Table 1 shows the Composite score sample sizes by grade level. For students tested in 2021, the Composite sample size for the mode-of-learning analysis ranged from 30,146 for 3rd grade to 18,918 for 11th grade.

During the 2020–2021 academic year, data were collected on students' instructional option: on-site, hybrid, or virtual. This set of data was collected at four time points: November 15, 2020 (Cycle 3); February 15, 2021 (Cycle 5); April 15, 2021 (Cycle 6); and June 15, 2021 (Cycle 7). Approximately 95% of students had mode-of-learning data reported for all four cycles.

Students were categorized as follows:

- All on-site: Students who were classified as on-site at all four cycles
- All hybrid: Students who were classified as hybrid at all four cycles
- All virtual: Students who were classified as virtual at all four cycles
- Hybrid, then on-site: Students who were classified as hybrid for Cycle 3 or Cycles 3 and 5 and then classified as on-site for the later cycles
- Virtual, then on-site: Students who were classified as virtual for Cycle 3 or Cycles 3 and 5 and then classified as on-site for the later cycles
- Other - 75% or more in-person: Students who did meet criteria for the other categories and who had a weighted average of 75% or more in-person across

cycles (For purposes of calculating the weighted average, on-site is considered 100% in-person, hybrid is 50% in-person, and virtual is 0% in-person.)

- **Other - 38–63% in-person:** Students who did meet criteria for the other categories and who had a weighted average of 38–63% in-person across cycles
- **Other - 25% or less in-person:** Students who did meet criteria for the other categories and who had a weighted average of 25% or less in-person across cycles
- **Missing:** Students who had missing instructional option data for one or more cycles

Note that the accuracy of the classifications is limited because of the nature and timing of the mode-of-learning data collected. For example, a student who began the 2020–2021 academic year learning in the virtual mode but switched to on-site before the first data collection date (November 15) could be classified as “all on-site,” even though they spent some time learning virtually. Moreover, the only instructional option at the onset of the pandemic (spring 2020) was virtual, so no students were on-site for the duration of the study period during the pandemic (i.e., March 2020 through spring 2021).

Table 2 provides the relative frequency of each mode-of-learning category by grade level. The percentage of students learning purely on-site decreased with grade level, from a high of 65% for 3rd grade to a low of 48% for 11th grade. Conversely, the percentage of students learning purely in hybrid mode increased from 3rd grade (9%) to 11th grade (12%), and the percentage of students learning purely in virtual mode increased from 3rd grade (12%) to 11th grade (21%).

Table 2. Percentage in Each Mode-of-Learning Category, by Grade Level

Category	Grade level									
	3	4	5	6	7	8	9	10	11	
All on-site	65	64	64	61	60	58	56	50	48	
All hybrid	9	9	9	9	10	10	10	12	12	
All virtual	12	13	13	14	14	15	16	18	21	
Hybrid then on-site	1	1	1	1	1	2	1	3	3	
Virtual then on-site	4	4	4	5	4	5	4	4	4	
Other: $\geq 75\%$ in-person	1	1	1	2	2	2	3	3	4	
Other: 38–63% in-person	1	1	1	2	2	2	3	3	3	
Other: $\leq 25\%$ in-person	0	1	1	1	1	1	1	1	1	
Missing	6	6	5	5	5	5	5	5	4	

Table 3 summarizes participation in each mode-of-learning category by student group. African American (27%) and Asian (34%) students were more likely to participate in virtual instruction relative to other racial/ethnic groups, while Hispanic (65%), Native American/Hawaiian (64%), and White (61%) students were more likely to participate in on-site instruction. English learners (67%) were also more likely to participate in on-site instruction.

Table 3. Percentage in Each Mode-of-Learning Category, by Student Group

Group	Mode-of-learning category				
	All on-site	All hybrid	All virtual	Other	Missing
Total	59	10	15	11	5
African American	46	7	27	15	5
Asian	44	8	34	11	3
Hispanic	65	8	12	10	5
Native American/Hawaiian	64	10	9	6	11
Two or more races	59	8	14	12	7
White	61	12	12	10	5
Student with disability	60	10	15	9	6
Student from low-income household	60	9	16	12	3
English learner	67	8	9	9	7

An analysis of mode-of-learning differences is simplified when there are fewer groups and the groups are well-defined. Because the majority of students were classified as all on-site, all hybrid, or all virtual (ranging from 87% for 3rd grade to 81% for 10th and 11th grades), and because the percentage of students not classified into these groups was similar across student groups, the analysis was limited to a comparison of those three groups.

To compare test scores for each mode-of-learning group against a pre-pandemic baseline group, we included the total 2019 sample as the reference group. Propensity score weighting was used to weight each of the four groups to be similar to the pooled data set, which combined the 2019 data and the three mode-of-learning groups from 2021. Logistic regression was used to estimate each student's propensity for being in each group based on their gender, race/ethnicity, disability status, economic status, and English learner status. For 5th–11th grades, prior ACT Aspire Summative test scores were used to ensure that the four groups did not differ on prior academic achievement.

Using the weighted data, we fit hierarchical linear regression models to estimate the difference in average test scores between the four groups. The models included a

random intercept for the school effect and employed the same covariates that were used for the propensity score weighting model. The general form of the regression model was:

test score = group + covariates + school effect

The regression model produced an estimate of the difference in average test score between each mode-of-learning group and the 2019 reference group, adjusted for changes in the tested population. As with the earlier analysis of pandemic-related score declines, adjusted score differences can be expressed on the original test score scale, in standard deviation units (d_{sd} = adjusted score difference divided by the standard deviation of the test score), or in average annual gain units (d_{gain} = adjusted score difference divided by the estimated score gain normally observed over one academic year).

The hierarchical linear regression models were also used to estimate the adjusted score difference for different student groups, including

- male and female students;
- racial/ethnic groups: African American, Asian, Hispanic, Native American/Native Hawaiian, White, and two or more races;
- students with disabilities and students without disabilities;
- students from low-income households and students from higher-income households;
- and English learner students and non-English learner students.

Group-specific estimates were obtained by fitting the hierarchical linear regression model with interactions between group and the student group indicator.

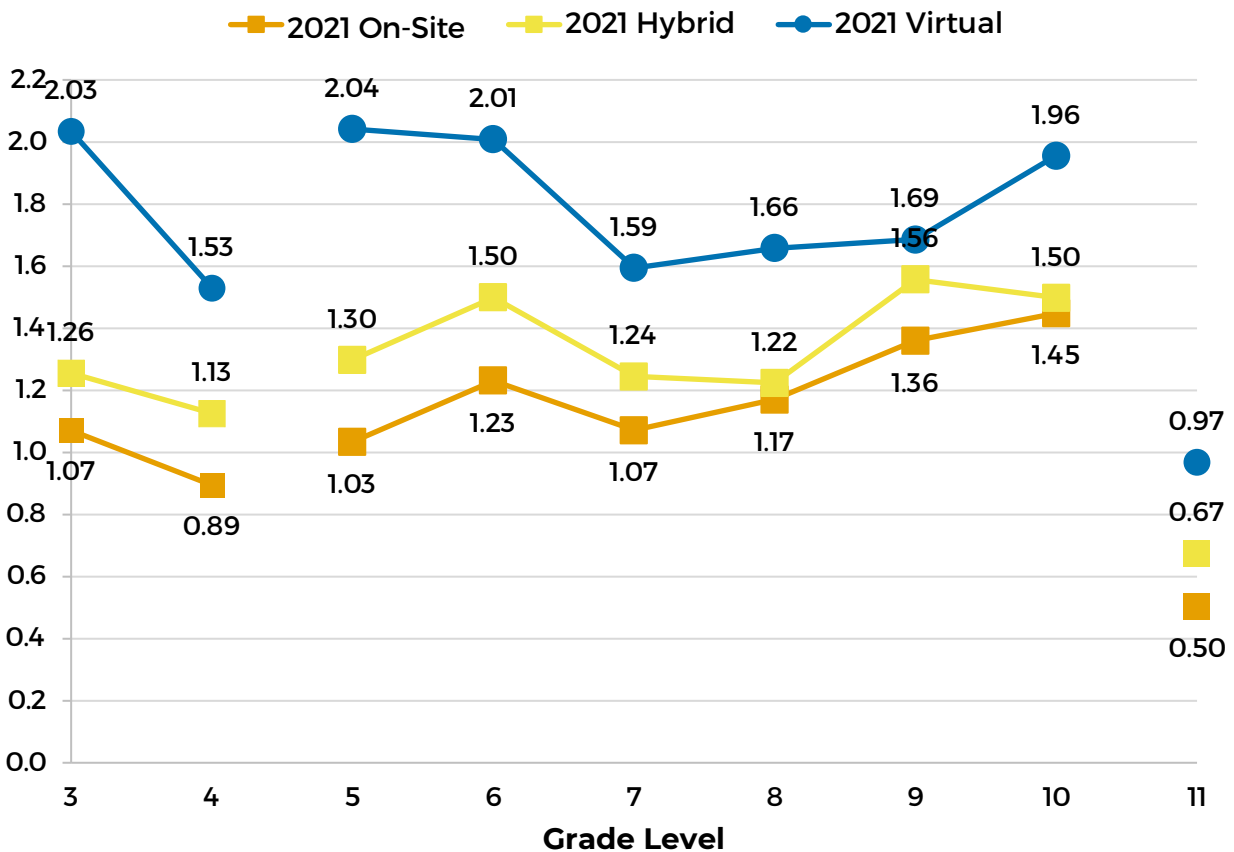
Comparing On-Site, Hybrid, and Virtual Modes of Learning: Total Group Results

The adjusted Composite score differences between each mode-of-learning group and the 2019 reference group are provided in Figure 8. For 5th–11th grades, the estimates account for prior test scores, student demographics, and school effects. For 3rd and 4th grades, the estimates account for student demographics and school effects. Because different models are used for 3rd–4th grades and 5th–11th grades, there are no lines connecting 4th and 5th grades. For 11th grade, the score declines are expressed on the ACT Composite score scale, which is different than the ACT

Aspire score scales; therefore, there are no lines connecting results from 10th grade to 11th grade.

For all grade levels, average scores for the three 2021 mode-of-learning groups were below the scores for the 2019 pre-pandemic group. Also, the score declines were most severe for the virtual group, followed by the hybrid group, and finally the on-site group. The differences between virtual and hybrid are generally larger than the differences between hybrid and on-site. The differences between the on-site group and the reference group are large, showing that even students in the on-site group suffered score declines relative to the pre-pandemic group. Generally, the performance gap for students who learned virtually is smaller for higher grade levels.

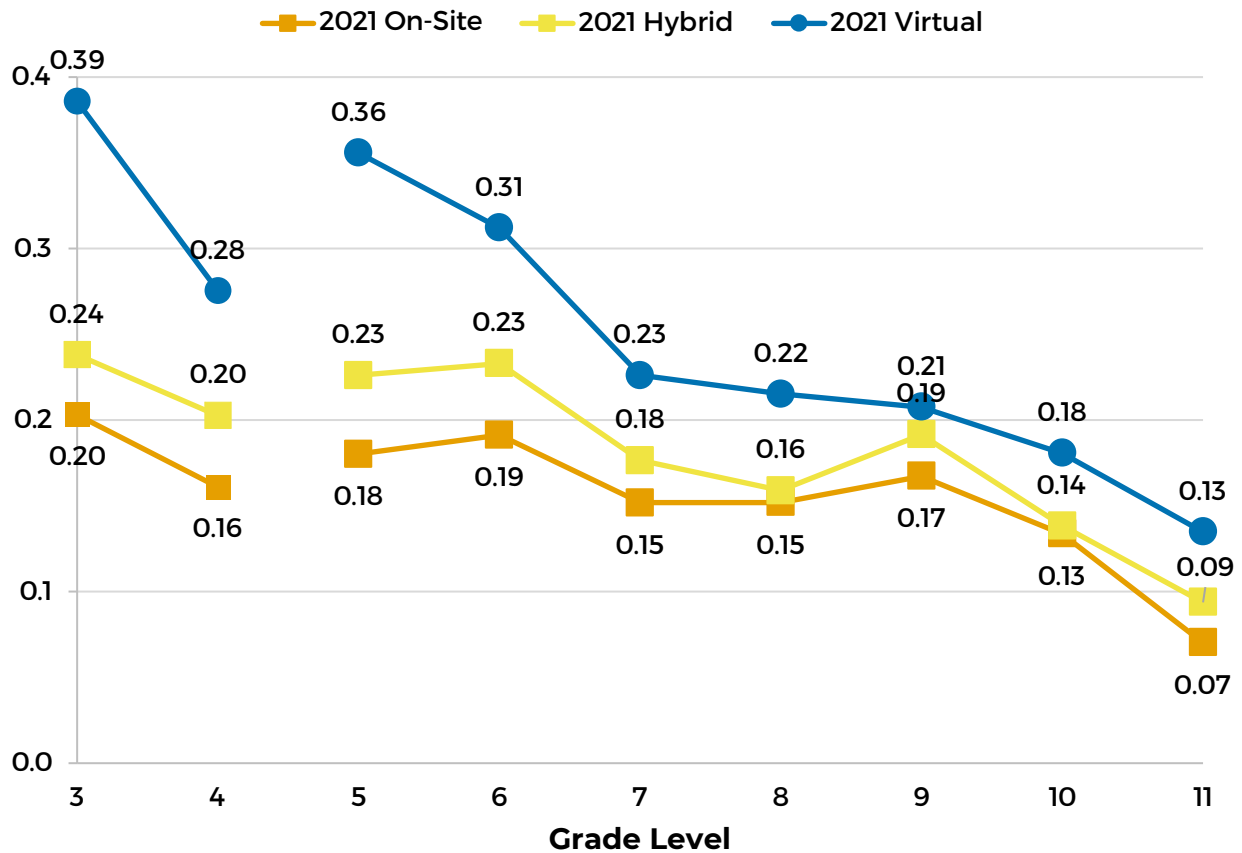
Figure 8. Decline in Average Composite Scores Relative to 2019, by Mode-of-Learning Group



The adjusted Composite score differences between each mode-of-learning group and the 2019 pre-pandemic group, expressed in standard deviation units, are provided in Figure 9. Relative to the variability of test scores at each grade level, we

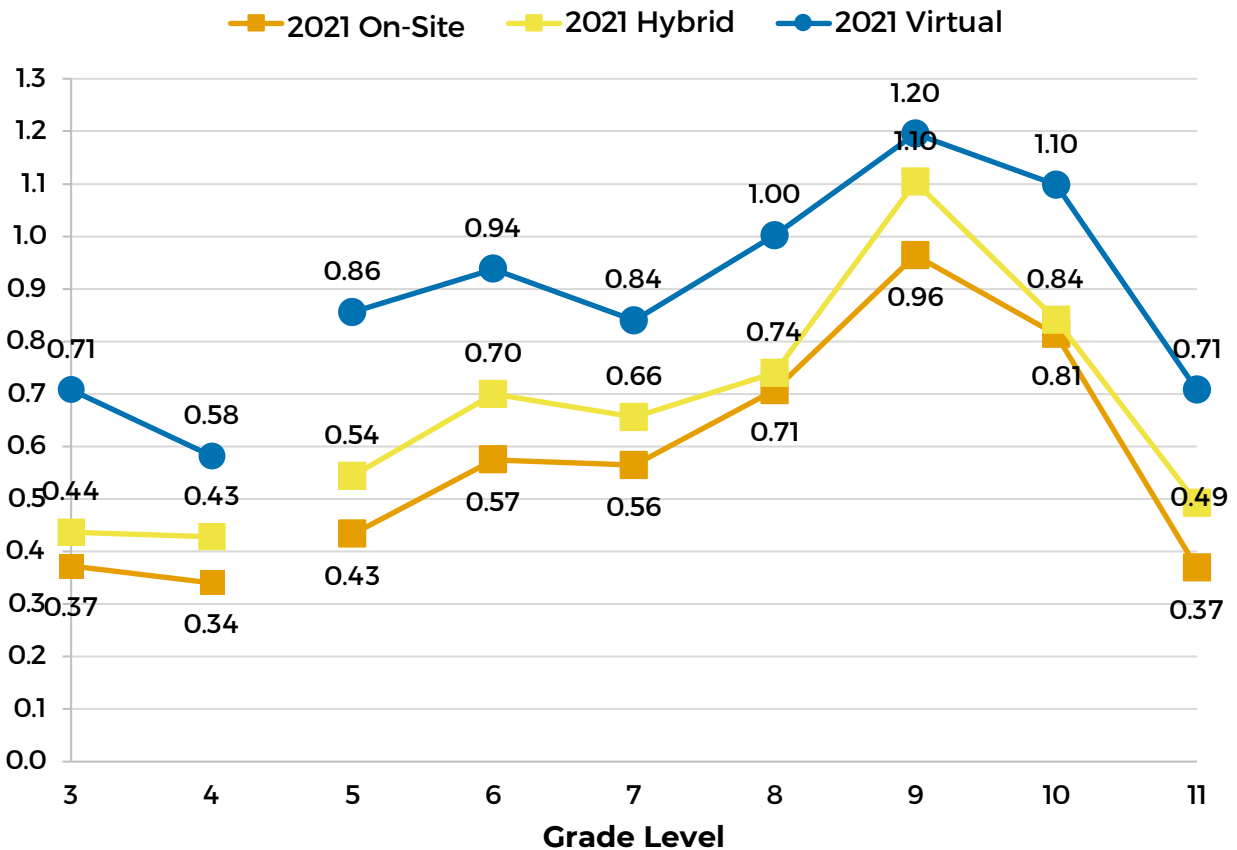
see that the differences between the 2021 mode-of-learning groups and the 2019 group tend to become smaller for higher grade levels.

Figure 9. Decline in Average Composite Scores in Standard Deviation Units Relative to 2019, by Mode-of-Learning Group



The adjusted Composite score differences between each mode-of-learning group and the 2019 pre-pandemic group, expressed in average annual gain score units, are provided in Figure 10. When expressed in average annual gain score units, the differences between the 2021 mode-of-learning groups and the 2019 group tend to become larger for higher grade levels. Note that the 11th grade results do not follow this trend but are based on the ACT Composite score, which is not on the same scale as the ACT Aspire Composite score. It's possible that the drops from 10th grade to 11th grade are due to scale differences.

Figure 10. Decline in Average Composite Scores in Average Annual Gain Score Units Relative to 2019, by Mode-of-Learning Group

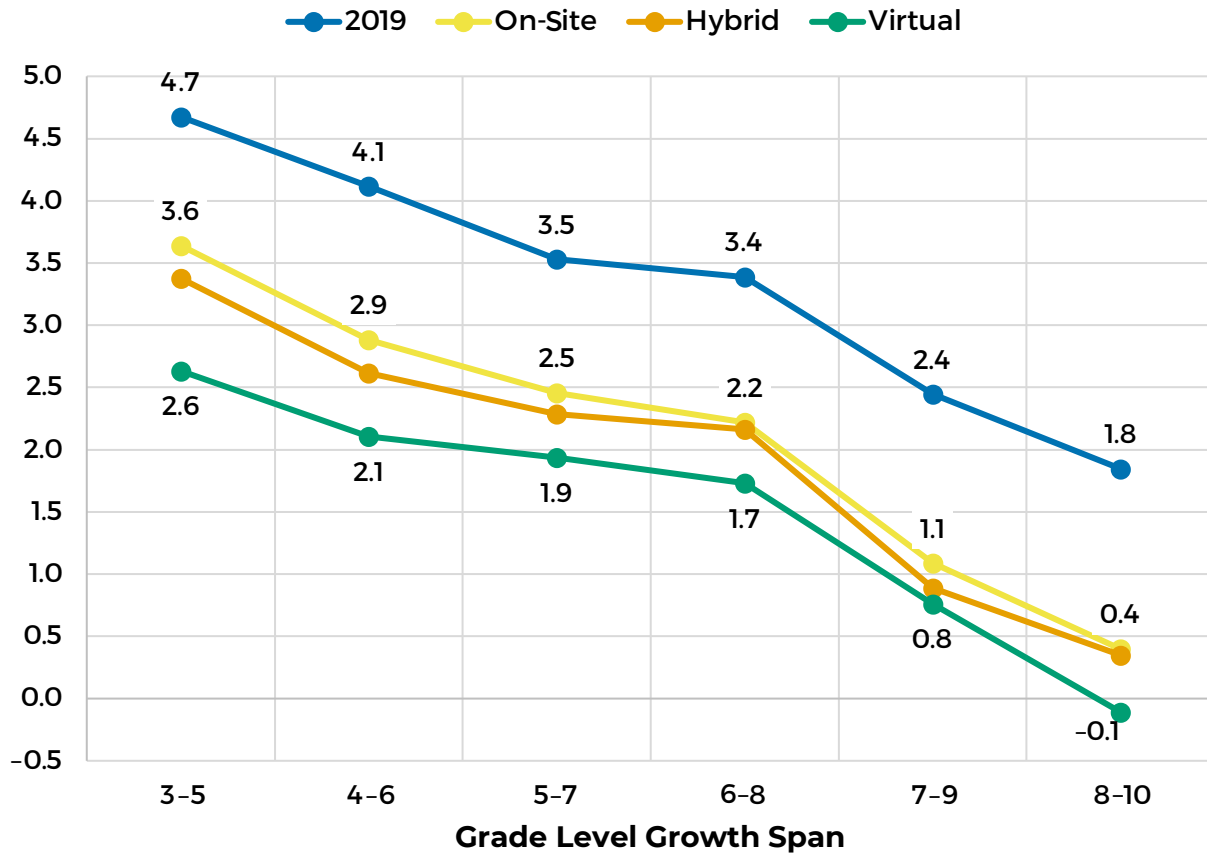


The grade level trends shown in Figures 8, 9, and 10 are similar to the trends seen in the overall analysis of score declines (Figures 1, 2, and 3). The estimates for the three mode-of-learning groups represent a disaggregation of the estimates for the overall 2021 group. For example, 5th-grade ACT Aspire Composite scores decreased by 1.24 points in 2021 relative to 2019 (Figure 1). The score decreases varied by mode-of-learning group (Figure 8) with average score declines of 1.03 for on-site, 1.30 for hybrid, and 2.04 for virtual.

Another way to understand differences across groups is to compare average gain scores. Gain scores are only available for 5th–10th grades because prior ACT Aspire scores are required to calculate gain scores. Note that gain scores are not available for 11th grade because ACT test scores are not reported on the same vertical scale as ACT Aspire test scores. Figure 11 shows the average gain in Composite test scores for each mode-of-learning group relative to the 2019 pre-pandemic group (blue line). For each grade level and mode-of-learning group, the gain scores are substantially

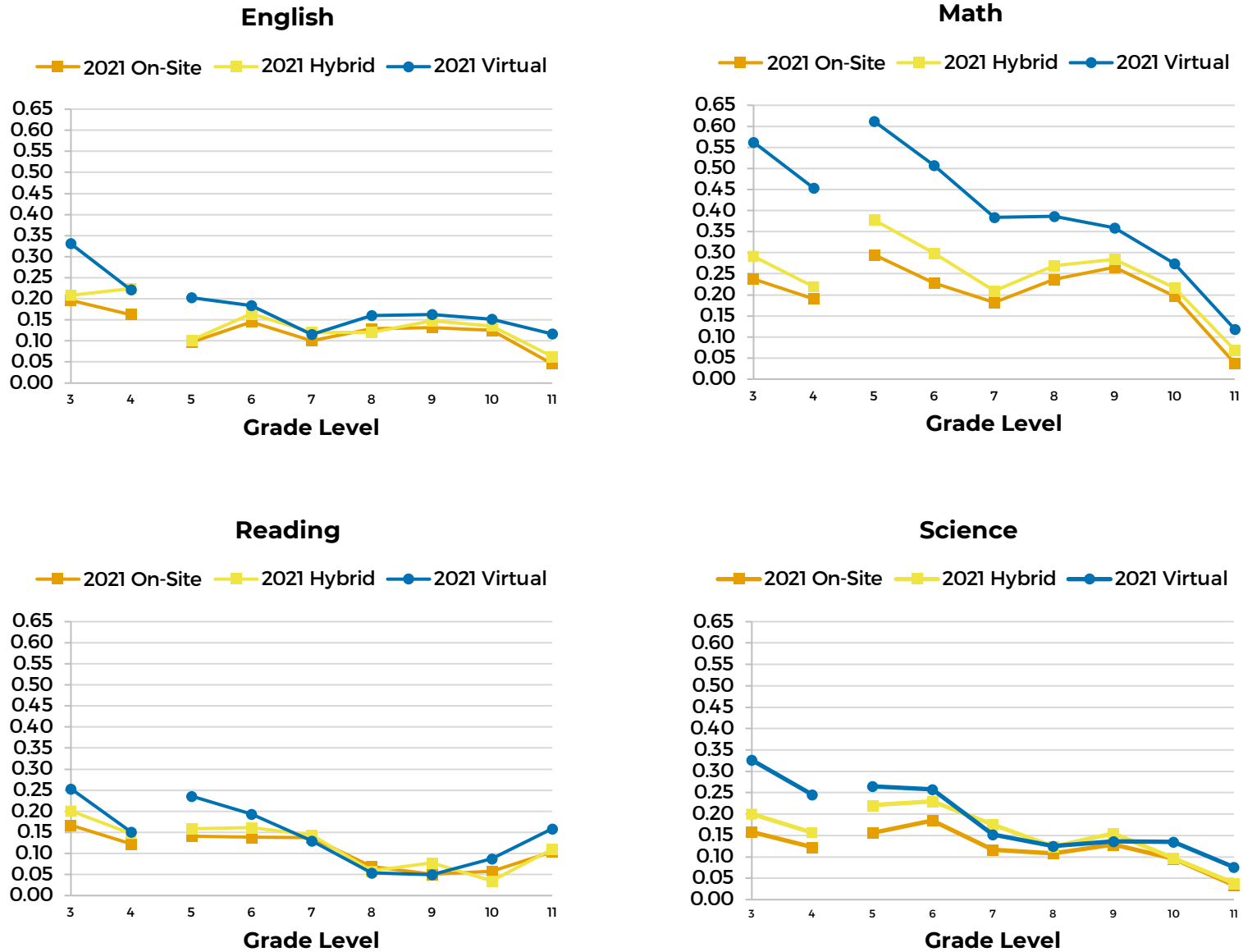
lower than those observed for the 2019 group. The differences are most pronounced for students in the virtual group.

Figure 11. Average Gain in Composite Scores, by Group



Next, we examine score declines by mode-of-learning group, test section, and grade level (Figure 12). For all three mode-of-learning groups, score declines (in standard deviation units) were largest for math relative to the other sections, with a few exceptions. The gap between virtual and other groups was clearly largest for math, but the gap decreased with grade level. For English, reading, and science, the gap between virtual and the other learning modes was smaller. For all sections, grade levels, and modes of learning, scores declined relative to the 2019 group.

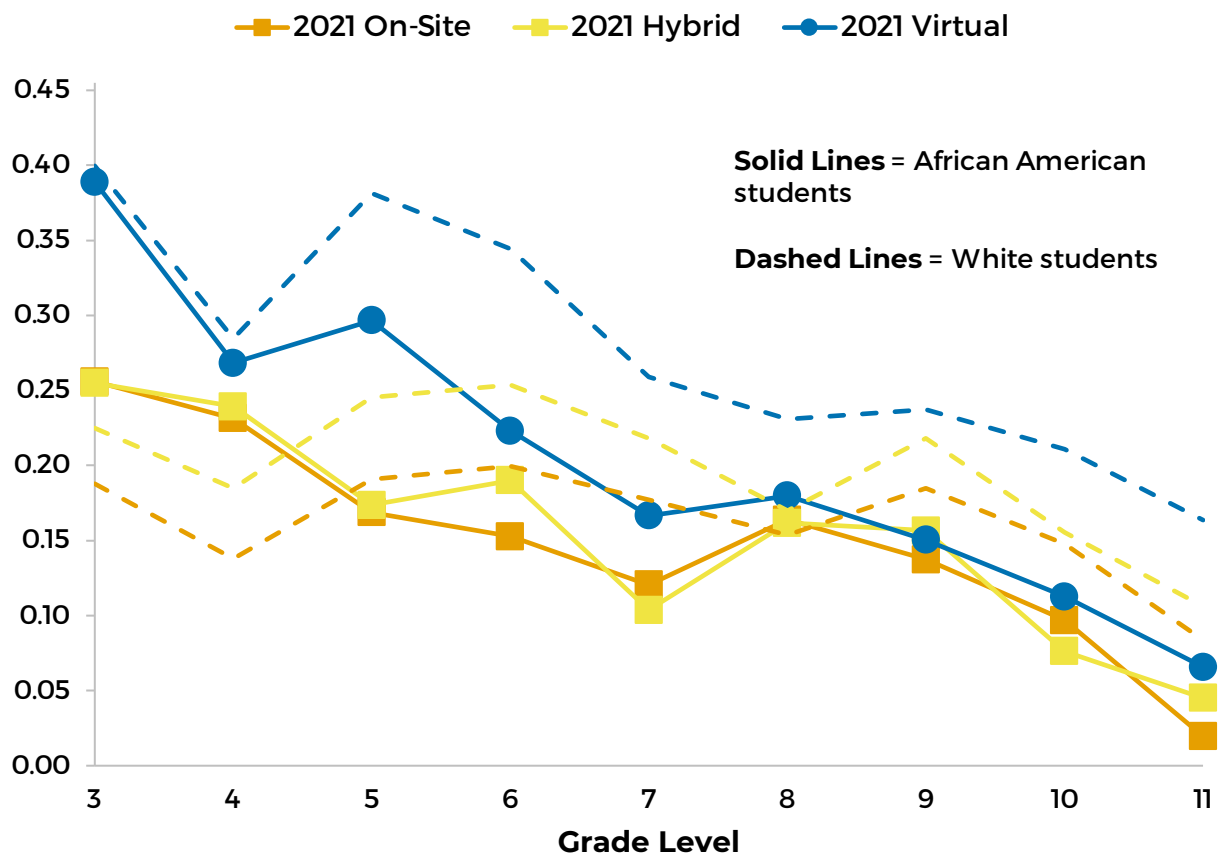
Figure 12. Decline in Average Test Scores in Standard Deviation Units by Group, Section, and Grade Level



Comparing On-Site, Hybrid, and Virtual Modes of Learning: Results by Student Group

Mode-of-learning differences were also examined for different student groups. In this section, we highlight Composite score results for two groups that had mode-of-learning differences that varied from the patterns observed for the total group. Figure 13 shows the decline in average Composite scores for African American and White students by mode of learning and grade level. The mode-of-learning differences were substantially smaller for African American students, especially for higher grade levels. In particular, the virtual gap for 8th–11th grade was much smaller for African American students relative to White students. For African American students, the differences between on-site and hybrid were particularly small.

Figure 13. Decline in Average Composite Test Scores in Standard Deviation Units by Mode of Learning, for African American and White Students



While the mode-of-learning differences were smaller for African American students, this does not imply that hybrid and virtual modes of learning were more effective for

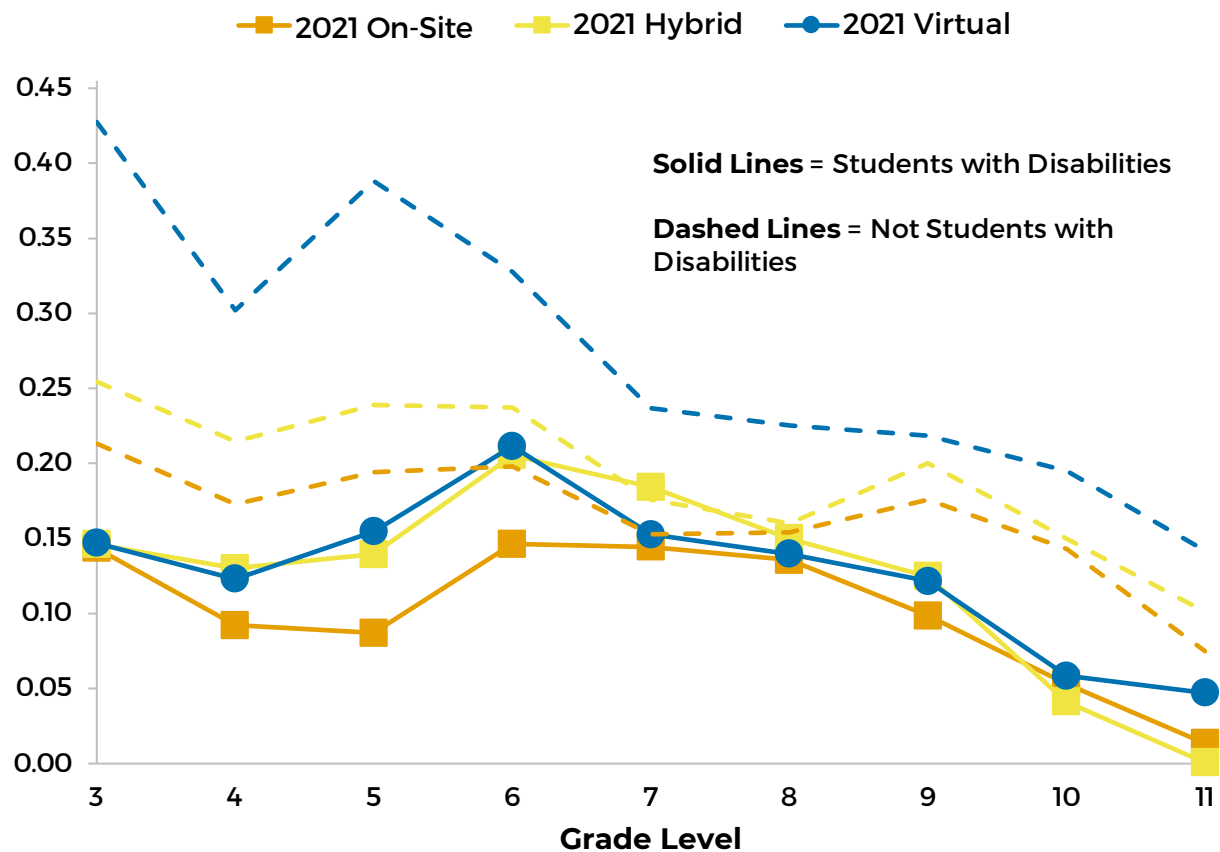
African American students relative to others. To illustrate this point, we examined the average Composite gain scores for African American students relative to students in other racial/ethnic groups (Table 4). Results are averaged across 5th–10th grades. We see that the baseline (2019) gain was 2.50 for African American students and 3.61 for other racial/ethnic groups, for a difference of –1.11 score points. For each 2021 mode-of-learning group, the differences were less severe. However, in all cases, students in other racial/ethnic groups still had higher average gain scores than African American students, suggesting that hybrid and virtual modes were not more effective for African American students.

Table 4. Composite Gain Scores for African American Students and Students in Other Racial/Ethnic Groups, by Mode-of-Learning Group

Group	African American	Other racial/ethnic groups	Difference
2019	2.50	3.61	–1.11
2021 on-site	1.40	2.31	–0.91
2021 hybrid	1.62	2.11	–0.49
2021 virtual	1.16	1.63	–0.47

Figure 14 shows the decline in average Composite scores for students with disabilities and students without disabilities by mode-of-learning group and grade level. For all grade levels, the differences across mode-of-learning groups were smaller for students with disabilities (solid lines). For 8th and 10th grades, the differences were extremely small. On the other hand, the mode-of-learning differences for students without disabilities were large and consistent across grade levels, with the on-site group scoring higher than the hybrid group, and the hybrid group scoring higher than the virtual group.

Figure 14. Decline in Average Composite Scores in Standard Deviation Units by Mode of Learning and Disability Status



Comparing On-Site, Hybrid, and Virtual Modes of Learning: Summary

- Based on data collected at four time points during the 2020–2021 academic year, the mode-of-learning classifications with the highest percentages were all on-site (59%), all hybrid (10%), or all virtual (15%).
- Relative to 2019, ACT Aspire and ACT test scores declined in 2021 for all three mode-of-learning groups.
- Score declines were most severe for students who learned virtually, among the three mode-of-learning groups.
- Score declines were less severe for students who learned on-site relative to those who learned in a hybrid mode.

- Mode-of-learning differences were small for reading.
- The virtual gap was most severe for math and for lower grade levels.
- African American and Asian students were less likely to learn on-site.
- Mode-of-learning differences were smaller for African American students and students with disabilities.

Future Research Directions

We found that scores declined during the pandemic, at least through spring 2021. Additional research should examine whether the negative impacts of the pandemic persisted beyond spring 2021. Will the decline in scores continue in spring 2022 and spring 2023? Or will scores rebound? Addressing these questions will help education leaders understand any lingering effects of the pandemic and the implications for the college and career readiness of students in Arkansas.

The study found that score declines were less severe for traditionally lower-performing student groups, including students with disabilities and English learners. However, it is not clear if those results are due to “floor” effects or true differences in the pandemic’s impact across different student groups. Future research can seek to address this question.

While we found evidence that students who learned virtually experienced less academic growth (especially in math), we should keep in mind that the virtual programs were mostly new offerings, and so educators and students may need time and experience to make those programs more effective. Future research can also examine whether the effectiveness of virtual learning programs improves over time.

Acknowledgements

Thanks to the Arkansas Department of Education for providing the data needed for this study and for their collaboration and feedback on this research. Thank you to Yi-Fang Wu and Jeffrey Conway for their reviews and feedback on an earlier version of this paper.

Notes

1. ACT Aspire scale scores are comparable across test forms because of rigorous equating processes. Further, because the same primary test forms were administered in both 2019 and 2021, the changes in average test scores are not likely attributed to test form differences. For 5th–11th grades, prior ACT Aspire test scores from 2017 and 2019 were also used in the analysis, with a post-hoc adjustment applied to these prior test scores to reduce possible random errors. Because the same ACT Aspire Interim test forms were used across years, Interim test scores can be used as anchors to adjust average Summative test scores across years. The post-hoc adjustment quantity was derived using linear regression with ACT Aspire Interim test scores used as predictors of ACT Aspire Summative test scores from 2017 and 2019. For example, if Summative test scores were 0.2 points lower in 2017 relative to 2019 (after adjusting for Interim test scores), the scores from 2017 were adjusted +0.2 score points.
2. To estimate the score gain normally observed over one academic year, we considered pre-pandemic data from Arkansas; specifically, for 4th–10th grades, the one year gains from 2016 to 2017, 2017 to 2018, and 2018 to 2019 were used to estimate the annual gain unit for 4th–10th grades. For each test section, a linear regression model was used to smooth out differences in mean gains across grade levels and to extrapolate the mean gain for 3rd grade. For the ACT test (11th grade), the average pre-pandemic gain is based on students who took the ACT as part of school-day testing in 10th grade and took the ACT again in 11th grade (10–14 months later).

Reference

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399–424. doi: 10.1080/00273171.2011.568786.