The widening achievement divide during COVID-19

November 2022

Karyn Lewis, Megan Kuhfeld, Meredith Langi, Scott Peters, and Erin Fahle
KEY FINDINGS

• Math and reading test scores are more variable in spring 2022 than spring 2019, indicating that students—based on their current achievement levels—are more different from one another now than before the pandemic.

• Increased variability is primarily driven by a widened distance between low and high achievers given students with lower achievement experienced larger initial impacts at the onset of the pandemic and less improvement during 2021-22.

The disruptions associated with the COVID-19 pandemic have impacted students’ academic achievement in unprecedented ways. Numerous studies have documented that students’ math and reading test scores were significantly lower in the 2021-22 school year compared to a prepandemic year (e.g., Betthäuser et al., 2022; Kuhfeld & Lewis, 2022a). Furthermore, the impacts of the pandemic have worsened pre-existing educational inequities. Specifically, the pandemic was more harmful for marginalized students of color and students attending high-poverty schools, who spent more time in remote instruction on average, and fell even farther behind their white, advantaged peers (Cohodes et al., 2022; Goldhaber et al., 2022). Additionally, recent results from the NAEP Long-Term Trend (LTT) assessment show that test score decreases for 9-year-olds between 2020 and 2022 were greater for students in the lower end of the score distribution compared to the upper end (U.S. Department of Education, 2022).

The uneven impacts of the pandemic suggest that test scores may be more variable now compared to prepandemic trends, which implies teachers and schools will need to be equipped to address greater diversity in learning needs.⁠¹ In this brief, we examine how variability in students’ math and reading test scores changed across the pandemic, including whether increased dispersion was concentrated at one part of the test score distribution.

In addition to examining whether test scores are more variable, we expand on recent findings from the NAEP LTT suggesting lower-scoring students have lost more ground and explore whether this is true for other age ranges. To do this, we compared growth patterns across the three COVID-interrupted school years relative to prepandemic trends separately for students with the highest and lowest test scores at the start of the pandemic. Our prior work shows different impacts of the pandemic depending on initial achievement status: high achievers made gains between fall 2019 and fall 2021 that were more consistent with prepandemic normative trends whereas lower-achieving students were more likely to fall short of normative growth projections (Lewis & Kuhfeld, 2021). Here, we continue to examine growth trends before and during the pandemic and disaggregate these trends according to initial achievement status to further show when and to what extent the distance between low and high achievers has widened.

---

¹ While we do not directly examine whether variability has widened in classrooms, we explore changes in within- and between-school variation in Table 4 and Figure A1 in the technical appendix.
This research brief addresses two questions:

1) To what degree have students’ reading and math test scores become more variable as a result of the COVID-19 pandemic?

2) How do achievement gains across the pandemic compare to prepandemic trends for students who were low- or high-achieving at the start of the pandemic?

To address our first question, we examined how variability in test scores changed during the pandemic by comparing the test scores of students in grades 3–8 in spring 2022 with a same-grade sample of students tested in spring 2019. In total, this sample included data from 8 million students across 24,000 public schools in grades 3-8 who took MAP® Growth™ assessments in reading and math (see Table 1 in the technical appendix for a description of the sample).

To address our second question, we followed a subset of students longitudinally for three school years to compare growth trends for students in the upper and lower ends of the achievement distribution. Specifically, we focused on a subset of 1 million students whose prepandemic achievement placed them in the top or bottom decile (90th percentile and above or 10th percentile and below, respectively) of the achievement distribution relative to national prepandemic norms. We followed these students across the three COVID-interrupted school years (2019–20, 2020–21, and 2021–22) and compared these data to a roughly comparable sized sample of students who tested between 2016–17 and 2018–19 and also had starting achievement in the top or bottom decile (see Table 2 in the technical appendix for a description of the longitudinal sample).

**Framing**

In this brief, we continue to document how the pandemic has harmed some students more than others. We specifically examine whether initial achievement status was related to achievement gains across the pandemic and focus on the extremes of the distribution to compare the experiences of students with the highest and lowest test scores at the start of the pandemic. For these analyses, we refer to students as “low achievers” or “high achievers” but recognize that this phrasing can be problematic as it may be seen as adopting a deficit-based perspective that perpetuates victim blaming. We want to explicitly acknowledge that a student’s achievement is a function of the complex education system they are a part of and the opportunities that they have benefited from or been denied. Our intention in categorizing students as “low achievers” is not to imply failure on the part of students, but rather to continue to shine light on the students with the greatest unmet needs.

**Terminology**

*Achievement gains:* Changes in students’ test scores between fall and spring.

*Achievement gap:* Difference between the COVID and pre-COVID samples in a grade level, reported as standardized differences in average test scores.

*Decile:* Test scores were converted to percentile rankings based on NWEA 2020 MAP Growth norms (Thum & Kuhfeld, 2020). Students were categorized into achievement deciles according to their percentile ranking. For instance, the bottom decile includes students with test scores that place them in the 10th percentile or below and the top decile includes students with test scores that place them in the 90th percentile or above.

*Rebounding:* Patterns of achievement gains that mirror or exceed prepandemic trends. “Rebounding” is not interchangeable with “recover”; rather, the former describes progress toward the latter.

*Recovery:* We define recovery as—at minimum—reaching equivalence with prepandemic achievement levels. However, we note that significant educational inequities predate the pandemic, and a goal of returning to a prepandemic status quo will be insufficient to ameliorate these inequities.

*Variability:* Variability refers to how spread out or clustered together test scores are—low variability indicates scores are more similar to one another; high variability indicates scores are more different from one another. One way to index variability is by examining the standard deviation of a distribution, which captures the typical distance from average.
Elementary students’ test scores are more variable in spring 2022 relative to pre-pandemic test scores

We compared the standard deviations (SDs) of the test score distributions observed in grades 3–8 in spring 2022 with those observed prior to the pandemic (in spring 2019) to assess changes in test score variability. In grades 3–5, test score variability in both subjects is about 5–10% higher in spring 2022. In grades 6–8, test score variability in reading also increased, though to a smaller extent (3–4%); math test score variance in these grades is largely consistent with prepandemic trends. An implication of this increased variability is that we are likely to observe more students who are not performing on grade level in reading, a finding that has been observed in other datasets as well (Curriculum Associates, 2022; Amplify, 2022).

Increased variability is primarily driven by larger test score decreases in the lower end of the distribution

Subsequent analysis showed the increased variability is due to larger test score decreases among lower-achieving students. Figure 1 shows changes in math and reading test scores at five selected percentiles to show how changes differ among lower- (10th and 25th percentiles), middle- (50th percentile), and higher- (75th and 90th percentiles) achieving students. In both subjects, average test scores for lower-achieving students decreased more than those for higher-achieving students, widening the distance between students at the top and bottom of the test score distribution. For instance, the third grade reading test score associated with the 90th percentile in a given year dropped by only 0.3 RIT points, whereas the test score associated with the 10th percentile dropped by 5.2 RIT points. This pattern is fairly consistent across subjects and grades and suggests that historically low-achieving students fell further behind during the pandemic.

---

2 See Table 4 in the technical appendix for the SD estimates and change in estimates between spring 2019 and spring 2022.

3 Note that these are observed percentiles based on how students ranked relative to one another within a given test season and not based on how students ranked relative to national norms for the MAP Growth assessment.

4 The one exception is math test scores for eighth grade, where higher-achieving students showed larger declines than lower-achieving students, resulting in less variability across students. Figures for the non-depicted grades are available in the technical appendix that accompanies this report.
Figure 1. Score changes between spring 2019 and spring 2022 in reading (left panel) and math (right panel) at five selected achievement percentiles for students in grades 3, 5, and 7.

Note. The boxes depict the interquartile range (25th to 75th percentile) of test scores for spring 2019 (dark blue) and spring 2022 (light blue), and the vertical bars extending from the boxes indicate the score associated with 10th and 90th percentile in each term. These percentiles are based on the ranking of observed scores within a test season (and not based on NWEA MAP Growth norms). The horizontal lines depict the change in the test score at each percentile between spring 2019 and spring 2022.
Students who started in the bottom decile experienced larger initial impacts of the pandemic during 2020–21 and less improvement during 2021–22

To better understand when and how disparities between low and high achievers have widened, we used a longitudinal sample to examine achievement patterns across the pandemic-affected school years (following cohorts of students from 2019–20 to 2021–22) separately for the lowest and highest achievers and compared gains for these COVID samples against the achievement gains for pre-COVID samples (cohorts of students who tested between 2016–17 to 2018–19).\(^5\) We compared the COVID and pre-COVID samples separately for students whose initial achievement status placed them in the bottom decile (at or below the 10th percentile) versus the top decile (at or above the 90th percentile) based on the 2020 MAP Growth norms (Thum & Kuhfeld, 2020).\(^6\)

Figure 2 compares average fall and spring achievement (shown as points) as well as gains during each school year and summer (solid and dashed lines, respectively, connecting the points) over a three-year period for the COVID sample relative to the pre-COVID sample, separately for students with starting achievement in the top and bottom deciles. For simplicity, Figure 2 depicts average gains across years for one cohort of students (those who were third graders in 2019–20 and fifth graders by 2021–22).\(^7\)

Consistent with our prior research (Kuhfeld & Lewis, 2022a), the pandemic resulted in diminished achievement gains during the 2020–21 school year relative to prepandemic trends for both student groups (the light and dark lines fan out that year). However, there are differences between the groups: the divergence is more extreme for students in the bottom decile than it is for top decile students. As a result, students who started in the bottom decile fell further behind during the 2020–21 school year (indicated by a larger gap between the pre-COVID sample and the COVID sample in spring 2021), which widened the pre-existing disparities between the top and bottom deciles.

Compared to 2020–21, average increases in test scores in 2021–22 were more consistent with prepandemic trends (the light and dark lines are closer to parallel) which shows some evidence of rebounding. Here, too, however there are differences between the groups. Students in the bottom decile showed test score increases in 2021–22 that still lagged prepandemic trends across subjects in most grades, although this divergence was not as extreme as it was in 2020–21. In contrast, test score increases in 2021–22 for students in the top decile actually exceeded prepandemic trends in most grades, especially in math.\(^8\)

The numbers reported below the circles in Figure 2 reflect the achievement gap between the COVID sample and the pre-COVID sample calculated as standardized mean differences within a term. For instance, in the spring of 2022, fifth grade reading test scores for students who started in the bottom decile were 0.22 SD below peers in the pre-COVID sample who also started in the bottom decile. In contrast, fifth grade reading test scores for students who started in the top decile were only 0.04 SD below top decile peers in the pre-COVID sample.\(^9\)

---

\(^5\) This parallels analyses included in our prior work (Kuhfeld & Lewis, 2022a) examining overall achievement patterns across the pandemic-affected school years with the exception that here we examine a three-year panel (2019–20 to 2021–22) instead of a four-year panel (2018–19 to 2021–22). We did not include the 2018–19 school year so that starting achievement status was as proximal as possible to the start of the pandemic.

\(^6\) Initial achievement status was based on fall 2019 achievement for the COVID sample and fall 2016 for the pre-COVID sample. We restricted the sample to students with observed test scores in this first fall (2016 or 2019), but student data could be missing at any other timepoint and still be included in the longitudinal sample. Thus, our results may underestimate the impacts of the pandemic, given students with missing test scores were also likely the most vulnerable and hardest hit (Johnson & Kuhfeld, 2020).

\(^7\) Results for the other cohorts are consistent with trends depicted in Figure 3 and are available in the technical appendix that accompanies this report.

\(^8\) These patterns are most evident when comparing the ratios of test score gains in the COVID sample relative to the pre-COVID sample for top and bottom decile students (see Table 7 in the technical appendix).

\(^9\) One way to contextualize the achievement gaps due to the pandemic is to compare them to the estimated impacts from other large-scale school disruptions, such as when math and reading scores dropped 0.17 SDs in the year following Hurricane Katrina (Sacerdote, 2012). For further information on understanding the magnitude of COVID gaps relative to typical educational interventions, see a recent Brookings blog (Kuhfeld et. al, 2022) and work by Matthew Kraft (Kraft, 2019).
Figure 2. Average MAP Growth achievement for grade 3–5 cohort by starting achievement decile percentile in reading (top panel) and math (bottom panel)

Note. Average test scores for top decile students are shown in orange and bottom decile students are shown in blue. The lighter shade represents the pre-COVID sample and the darker shade represents the COVID sample. Decile group is determined by starting achievement status. The top decile includes students who scored at the 90th percentile or above and the bottom decile includes students who scored in the 10th percentile or below. Percentiles are calculated based on the 2020 MAP Growth norms (Thum & Kuhfeld, 2020). Spring 2020 means are asterisked because they are based on approximately 5% of the students relative to other terms due to testing interruptions during COVID school closures. Standardized mean differences between the COVID and the pre-COVID sample within achievement decile are shown below the COVID sample lines, with negative values indicating that the COVID sample scored lower than the pre-COVID sample.
Table 1 summarizes the size and the change in the achievement gaps in test scores between the pre-COVID and COVID sample separately by initial achievement decile for all six cohorts. Despite some signs of rebounding across groups in 2021–22, we see less evidence of improvements for bottom decile students compared to top decile students. Across cohorts and subjects, top decile students have consistently improved (i.e., the gaps between the COVID and pre-COVID samples are smaller in spring 2022 than they were in spring 2021). Results for bottom decile students were mixed across cohorts—achievement gaps decreased, held steady, and even increased in some cases. As a result, the achievement gaps that remain in spring 2022 are significantly larger for bottom decile students and top decile students are much closer to reaching recovery.

Table 1. Difference in achievement gaps between spring 2021 and spring 2022 in reading and math by initial achievement decile

<table>
<thead>
<tr>
<th>Subject</th>
<th>Cohort</th>
<th>Students who started in bottom decile</th>
<th></th>
<th>Students who started in top decile</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Achievement Gap</td>
<td></td>
<td>Achievement Gap</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring 21</td>
<td>Spring 22</td>
<td>Change</td>
<td>Spring 21</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Reading</td>
<td>1–3</td>
<td>0.24</td>
<td>0.29</td>
<td>-0.05</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>2–4</td>
<td>0.24</td>
<td>0.22</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>3–5</td>
<td>0.24</td>
<td>0.22</td>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>4–6</td>
<td>0.28</td>
<td>0.21</td>
<td>0.07</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>5–7</td>
<td>0.25</td>
<td>0.26</td>
<td>-0.01</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>6–8</td>
<td>0.26</td>
<td>0.26</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Math</td>
<td>1–3</td>
<td>0.38</td>
<td>0.45</td>
<td>-0.07</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>2–4</td>
<td>0.43</td>
<td>0.40</td>
<td>0.03</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>3–5</td>
<td>0.41</td>
<td>0.38</td>
<td>0.03</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>4–6</td>
<td>0.37</td>
<td>0.24</td>
<td>0.13</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>5–7</td>
<td>0.19</td>
<td>0.22</td>
<td>-0.03</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>6–8</td>
<td>0.20</td>
<td>0.22</td>
<td>-0.02</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Note. The achievement gaps reported in this table are the standardized mean difference between the pre-COVID and COVID samples in a given grade/term.

Implications for equitable gifted and talented identification

Our prior research has consistently shown that Hispanic, Black, and American Indian/Alaska Native (AIAN) students have been disproportionately impacted by the pandemic, showing the largest achievement declines compared to white and Asian students (Kuhfeld et al., 2021). These disparities are evident across the achievement spectrum. Despite top decile students showing less impacts of the pandemic, here too we see differences by race and ethnicity. As a result, the highest achieving group of students now includes even fewer marginalized students of color than in the past. This implies that applying pre-pandemic national normative cutoffs (e.g., MAP Growth scores at or above the 95th percentile) to identify students for advanced learning opportunities will result in fewer marginalized students of color receiving access to these services—exacerbating the pre-existing underrepresentation of these groups among students identified as gifted (Peters et al., 2019a). Schools should consider applying alternative methods, such as using local norms, that have been shown to increase the number of marginalized students identified for these opportunities (Peters et al., 2019b).
Summary

Students’ math and reading test scores are more variable in spring 2022 than they were prior to the pandemic. Further research is needed to unpack the degree to which shifts in variability may be due to shifting sample compositions across time, but our initial findings show students have a wider range of achievement levels now than would be expected in a typical school year.

Across both subjects, the increased variability in test scores is primarily driven by a spreading out at the bottom end of the distribution, with test scores for lower-achieving students dropping more than the declines observed for higher-achieving students. Similar to (and overlapping with) how the pandemic has widened pre-existing disparities by race/ethnicity and school-poverty level, we see that the distance between students at the extremes of the achievement distribution has widened over the last two years. High achievers were also harmed by the pandemic, but the widening distance between the groups is largely explained by greater impacts on lower-achieving students who were both harder hit during 2020–21 and made less improvement during 2021–22. The cumulative result is that there is greater unfinished learning for students who were already struggling before the pandemic.

These results highlight a key challenge for teachers who must now meet the needs of students who may be in the same grade but differ even more than they did prior to the pandemic in terms of learning needs. Schools and teachers will require increased or updated professional development and instructional resources to meet the wider range of learning needs. Focusing on providing supports to help the lowest achieving students catch up is especially critical to address the achievement divide that has widened significantly over the last two years.
References


Details on the methodology behind these analyses can be found in:

Suggested citation:
ABOUT THE AUTHORS

Dr. Karyn Lewis is director of the Center for School and Student Progress at NWEA where she leads a team of researchers who operate at the intersection of K-12 education research, practice, and policy. Her research interests focus on the interplay between students’ academic achievement and growth, their social-emotional development and well-being, and how they experience their school’s climate. Prior to joining NWEA, she was a senior researcher at Education Northwest/REL Northwest where she led a diverse portfolio of applied research, technical assistance, and evaluation projects centered around social-emotional learning. Dr. Lewis is a former data fellow with the Strategic Data Project at the Harvard Center for Education Policy Research. She completed a National Science Foundation funded postdoctoral fellowship at the University of Colorado Boulder and earned a PhD from the University of Oregon in social psychology.

Dr. Megan Kuhfeld is a senior research scientist for the Collaborative for Student Growth at NWEA. Her research seeks to understand students’ trajectories of academic and social-emotional learning (SEL) and the school and neighborhood influences that promote optimal growth. Dr. Kuhfeld completed a doctorate in quantitative methods in education and a master’s degree in statistics from the University of California, Los Angeles (UCLA).

Dr. Meredith Langi is a research scientist for the Center for School and Student Progress at NWEA. She specializes in educational measurement and latent variable modeling and is particularly interested in how assessments can support classroom instruction. Prior to joining NWEA, she supported psychometrics for the English Language Proficiency Assessment (ELPA21). She also served as a teacher and resource specialist with the Peace Corps in Tonga. Dr. Langi holds a PhD in social research methodology and an MS in statistics from the University of California, Los Angeles, as well as an EdM in international education policy from the Harvard Graduate School of Education.

Dr. Scott Peters is a senior research scientist for the Center for School and Student Progress at NWEA. He specializes in educational assessment and data use, gifted and talented student identification, equity within advanced educational opportunities, and effectiveness of educational policy. His research focuses on how schools can leverage assessment data for maximum school and student benefit. His ongoing projects relate to balancing cost, sensitivity, and equity in gifted and talented student identification, how to proactively screen students for advanced learning opportunities, examining growth trajectories for advanced learners, and how to ensure all students have access to advanced learning opportunities. Prior to joining NWEA, Dr. Peters served as a professor of assessment and research methodology at the University of Wisconsin - Whitewater. He received his PhD from Purdue University in educational psychology and applied research methodology.

Dr. Erin Fahle is a research scientist for the Collaborative for Student Growth at NWEA, where she investigates how social and school context affects gender, racial/ethnic, and socioeconomic inequalities in students’ access to educational opportunities and subsequent achievement. Dr. Fahle’s goal is to help states, districts, and schools identify areas for policy and practice interventions that can improve the educational circumstances of children across the US. She believes deeply that this work must be done in partnership with school leaders and is committed to designing research that reflects their perspectives. Dr. Fahle also serves as the codirector of the Educational Opportunity Project. This initiative, led by Stanford University, aims to make high-quality data and research on US educational opportunities widely accessible to policymakers, practitioners, and parents. Prior to joining NWEA, Dr. Fahle was an assistant professor at St. John’s University. She holds a PhD in education policy from Stanford University and an MS in applied mathematics and statistics from Georgetown University.
ABOUT NWEA

For more than 40 years, NWEA® has been a pioneer in educational research and assessment methodology with a focus on improving learning outcomes for every student. NWEA continues this discovery through dedicated research that explores foundational issues in education, practical challenges in today’s schools, and the evolving role of technology in the lives of students. As a mission-based not-for-profit educational research organization, NWEA’s research agenda reflects our commitment to attacking big challenges in education and measurement and empowering education stakeholders with actionable insights.

ABOUT THE CENTER FOR SCHOOL AND STUDENT PROGRESS

The Center for School and Student Progress (CSSP) engages directly with NWEA partner schools to influence education practices and policies that promote student success. The CSSP focuses on issues that impact the daily work of educators and the students they serve, such as achievement and growth patterns for traditionally underserved students, the integrity of testing systems, supporting college and career readiness, and school accountability. CSSP researchers also serve as consultative partners, offering advanced technical support, custom research projects, and analysis to school leadership, educators, and policymakers.

ABOUT THE COLLABORATIVE FOR STUDENT GROWTH

The Collaborative for Student Growth at NWEA is devoted to transforming education research through advancements in assessment, growth measurement, and the availability of longitudinal data. The work of our researchers spans a range of educational measurement and policy issues, including achievement gaps, assessment engagement, social-emotional learning, and innovations in how we measure student learning. Core to our mission is partnering with researchers from universities, think tanks, grant-funding agencies, and other stakeholders to expand the insights drawn from our student growth database—one of the most extensive in the world.