

Reconciling Recent Evidence on Academic Recovery in the US from 2022 to 2023

Dan Dewey, Harvard University
Erin Fahle, Stanford University
Tom Kane, Harvard University
Megan Kuhfeld, NWEA
Sean F. Reardon, Stanford University

29 August, 2024

Background

In January 2024, the [Education Recovery Scorecard \(ERS\) results](#) showed that students' math and reading scores improved from Spring 2022 to Spring 2023, making up about a third of their pandemic-era decline from 2019-2022 in math and a quarter of their decline in reading. These gains were large relative to historical changes in math and reading achievement on the National Assessment of Educational Progress. However, in July 2023, [NWEA](#) reported that average student learning during the 2022-23 school year was slower than in pre-pandemic years, indicating stalled progress towards academic recovery. **The two reports used different samples, different methods, and different tests, making it difficult to understand what might cause the differences in the reported recovery trends.**

After investigating differences in estimates for overlapping samples of districts, we come to the following conclusions:

First, in math, **we believe that the Education Recovery Scorecard estimate of a .045 standard deviation improvement in achievement between 2022 and 2023 accurately characterizes the improvement in achievement in the 28 states included in the ERS sample.** When limited to the same sample of districts in those states, the NWEA and ERS estimates of improvement in math are similar. The slower growth in NWEA scores in those states seems to be due to lower test participation rates and the non-representativeness of NWEA-participating districts.

Second, after comparing NWEA and ERS results in reading, **we believe that the ERS results are overstated due to anomalously large increases in reading achievement in Ohio and Illinois and that the NWEA finding of little change in reading scores is closer to the truth.** Without Ohio and Illinois, both the NWEA and ERS estimates imply little change in scores between 2022 and 2023. Thus, we believe that there was little change in reading achievement in the remaining 24 states with both ERS and NWEA scores.

Below, we describe our analyses in more detail.

A Comparison of the NWEA and ERS Samples

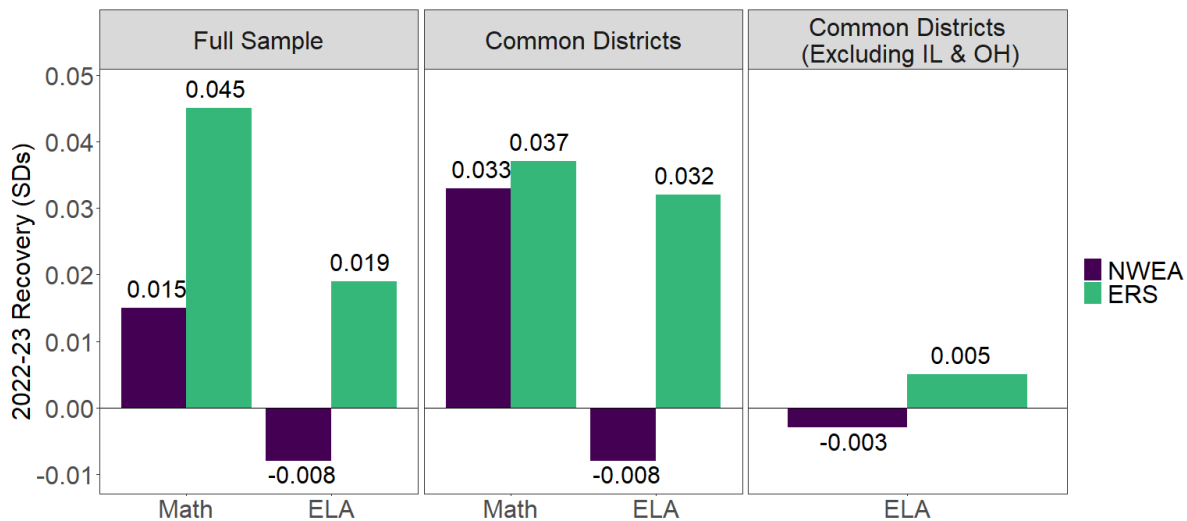
To try to reconcile the two different 2022 to 2023 recovery stories between ERS and NWEA, researchers at Stanford, Harvard, and NWEA sought to replicate the ERS sample and methodology within the NWEA sample. The two studies used different methods of measuring academic recovery and used different samples of students. The ERS used data from almost all districts in 29 states (28 states in math; 26 states in ELA), and compared 2022 and 2023 average scores among 3-8th graders within each district. The NWEA report measured within-student learning rates during the 2022-23 school year and compared those to average within-student learning rates prior to the 2019-20 school year. The NWEA sample

included data from the non-random sample of approximately 20,000 schools that used the MAP Growth assessments. Thus, neither the methods nor samples were the same between the two studies.

To better compare the results of the two studies, we used a common method and sample of school districts. First, we applied the ERS methodology of estimating recovery (for more details, see [Fahle et al. 2024](#)) based on districts’ average test scores (aggregated across grades 3-8) to the full sample of approximately 2,300 NWEA districts. We then compared those estimates to the ERS estimates from approximately 5,300 ERS districts (see left panel in the figure below). This comparison—using the same method, but different samples—still showed significantly different results in the two studies. The NWEA data show a far smaller recovery in math than the ERS data (.015 compared to .045 SDs); NWEA data show a negative recovery in ELA (-.008 SDs), while ERS data show a small positive ELA recovery trend (.019 SDs).

Second, we imposed stricter inclusion rules on the NWEA sample to more closely match the ERS sample inclusion rules. Specifically, we required that each district must have (a) tested 80% of enrolled students in a grade/year/subject to be included in the aggregation across grades and (b) tested in all three years (2019, 2022, and 2023) to be included in the final analysis, so that the analyses include the same set of district over time. We then further restricted both the ERS and NWEA samples to include only the approximately 1,150 districts that were common to both samples. This provides an apples-to-apples comparison of the two studies: using the same measure of recovery and the same set of school districts. The comparison based on this sample is shown in the middle panel of the figure. In math, both studies yield very similar estimates of recovery in the set of common districts (.033 compared to .037 SDs). This implies that **the difference between the two reports in math results is largely driven by differences in their samples of districts.**

Figure 1. Estimated Change in Average Test Scores, 2022-2023, Education Recovery Scorecard and NWEA



Note: Figure 1 shows the average change in grade 3-8 math and ELA scores, measured in within-grade standard deviations of the national student distribution. “Full Sample” includes all districts included in the ERS and NWEA studies, respectively. “Common Districts” includes all districts that were in both the ERS and NWEA data and that had data for all three years (Spring 2019, 2022, and 2023) and where at least 80% of students were tested in each year. “Common Districts (Excluding IL and OH)” is the “Common Districts” sample excluding all districts from IL and OH. ERS: Education Recovery Scorecard.

The same pattern does not hold in ELA. In fact, the difference between the two studies' recovery estimates was even wider when we restricted the sample to common districts. Upon further investigation, this difference appears to be largely driven by results in two states: Illinois and Ohio. In both states, the set of districts common to both studies show a much larger recovery in the ERS data (based on state assessments) than in the NWEA data (based on the NWEA MAP assessment). When we exclude the data from these two states (third panel of the figure), both ERS and NWEA point to essentially zero recovery in ELA. Thus, **the difference between the ERS and NWEA results in ELA appears to be driven largely by the divergent results on state assessments and the MAP assessment in Illinois and Ohio.**

Which should we believe?

Knowing what drives the differences is useful, but still leaves open the question of which results best reflect the actual post-pandemic trend in student achievement. Each study has some strengths and weaknesses:

The ERS data represent a near complete sample of students in 28 states in math and 26 states in ELA, but tell us nothing about recovery in the other states. The NWEA data are not a representative sample, but approximately one in three public schools serving grades 3-8 are included in the full report sample. On these grounds, the ERS results are preferable, at least with respect to reporting trends from the included states.

The ERS results rely on state assessments, which may not be strictly comparable from year to year; the NWEA results rely on the MAP assessments, which are common to all districts in their sample and are comparable over time. Even though the states in ERS were included because they reported that their assessments had not changed, there may still have been unreported changes in the assessments or the scoring that lead to inaccurate reported trends in the ERS results. On these grounds, the NWEA results are preferable, as they are much more clearly comparable over time.

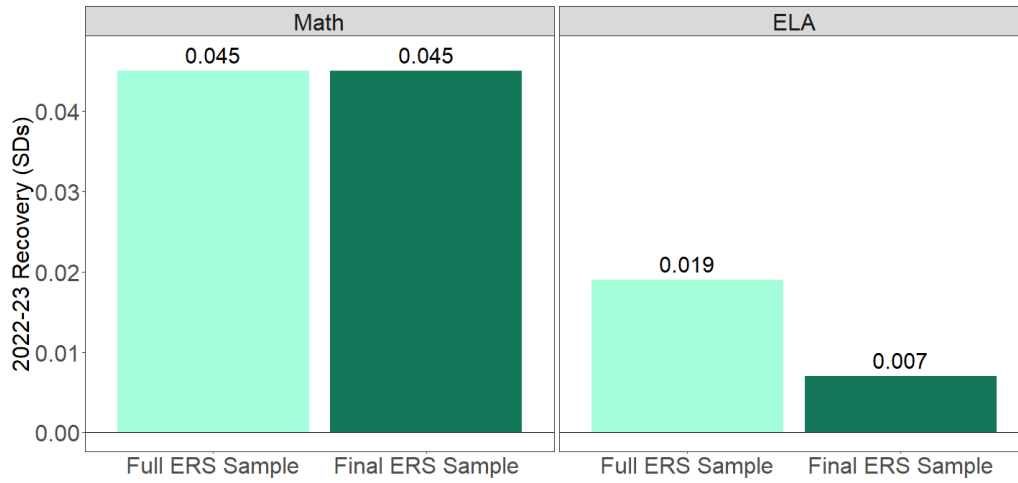
Conclusions

Given that the discrepancy in math results appears driven by sample differences, **we believe the .045 SD recovery ERS estimate is a better depiction of the recovery within the 28 states included in the math analyses in the ERS report** (see Figure 2). We cannot make any valid conclusion about math recovery in the other states.

In ELA, it appears that two states were in large part driving the ERS recovery estimate. After excluding Illinois and Ohio from the full ERS sample, the ERS data indicate an average ELA recovery in the remaining 24 states of .007 SDs (much lower than .019 reported when IL and OH are included). **Therefore, we conclude there was essentially no recovery in ELA from 2022 to 2023 in these 24 states.** We cannot make any other valid conclusion about ELA recovery in the other states.

Finally, it is important to note that **much better estimates of academic recovery will be available in January 2025, when the NAEP results from Spring 2024 results are available.** The NAEP data will be more informative than the ERS report because they rely on the same NAEP assessment in 2022 and 2024, and they will be more informative than the NWEA data because they are based on large, random samples of students that are representative of each state and of the nation as a whole.

Figure 2. 2022-2023 Change in Test Scores, Education Recovery Scorecard Revised Estimates



Note: Figure shows average change in grade 3-8 math and ELA scores, measured in within-grade standard deviations of the national student distribution. "Full ERS Sample" includes all districts included in the ERS studies. For Math, the "Final ERS Sample" is identical to the full 28-state sample, while in ELA it is a 24-state sample (excluding all districts from IL and OH). ERS: Education Recovery Scorecard.