Can Repeated Aggregate Cross-Sectional Data Be Used to Measure Average Student Learning Rates? A Validation Study of Learning Rate Measures in the Stanford Education Data Archive

AUTHORS

Sean F. Reardon
Stanford University

John P. Papay
Brown University

Tara Kilbride
Katharine O. Strunk
Joshua Cowen
Michigan State University

Lily An
Kate Donohue
Brown University

ABSTRACT

In this paper we compare two approaches to measuring the average rate at which students learn in a given school or district. One type of measure—longitudinal growth measures—relies on student-level longitudinal data. A second type—cohort growth measures—relies only on repeated aggregated, cross-sectional data.

Because student-level data is often not readily available, cohort growth measures are sometimes the only type available. The estimated school and district learning rates reported in the Stanford Education Data Archive (SEDA), for example, are cohort growth measures based on aggregated data. Understanding how much researchers and policymakers can rely on these cohort growth estimates requires one to know how well, and under what conditions, the estimates obtained from this approach align with those based on longitudinal data.

In this report we address these questions. We do so by using longitudinal student data from three states (Massachusetts, Michigan, and Tennessee) to construct both average gain score measures (longitudinal growth) and change-in-average measures (cohort growth) for each public school district and school serving students in any of grades 3-8 in the three states. We then compare the two sets of estimates in order to assess how well the latter replicates the former. We do this separately for districts and schools.

We find that the longitudinal and cohort growth measures are generally highly correlated in these three states. On average, the cohort growth measures largely rank schools and districts similarly to longitudinal growth measures. The correlations at the district-level (r=0.87) are somewhat higher than the school-level correlations (r=0.80), which reflects the fact that there is less student mobility among districts than among schools. Additionally, in cases where student mobility in and out of schools or districts is high, the measures are less well aligned. Mobility rates are higher, on average, in small schools and districts, schools with long grade spans, and in charter schools. As a result the alignment of the two measures is weaker in these cases. We conclude that the cohort growth measures are useful proxies for longitudinal growth measures in most, but not all cases.

Can Repeated Aggregate Cross-Sectional Data Be Used to Measure Average Student Learning Rates?
A Validation Study of Learning Rate Measures in the Stanford Education Data Archive

Non-Technical Summary

November, 2019

Sean F. Reardon
John P. Papay
Tara Kilbride
Katharine O. Strunk
Joshua Cowen
Lily An
Kate Donohue

Over the past decade, a consensus has emerged among both researchers and policymakers that average test scores (or academic proficiency rates) are poor proxies for school quality. Students' average scores reflect not only the inputs of a school, but also out-of-school factors that shape students' opportunities to learn. Instead, policymakers have begun relying more heavily on student growth, seeking to measure the effectiveness of a school by assessing how quickly its students are learning new material.

In principle, one could measure average student learning rates by a) testing students each year; b) computing each students' annual learning rate (the rate of change of their test scores from one grade to the next); and then c) averaging these learning rates among all students in a school in a given grade and year. Such a measure (which we refer to as a "longitudinal growth measure") would provide useful information about how much students in a given school learn, on average, each year.

Computing individual students' learning rates and then averaging them requires individual student records containing past and current scores, which are often not available to the general public or even to researchers. More readily available are average student test scores within each school in each grade and year. From such data we can compute an approximate measure of average student learning rates: the difference between average scores of all students in a specific grade in a school and the average scores of students in the previous grade in the prior year. This value indicates, for example, how much student test scores changed, on average, from 3rd grade in one year to 4th grade in the following year. We can average this measure over multiple grades to obtain an average learning rate measure for a school or district. This measure, which we refer to as a "cohort growth measure," will not be exactly the same as the longitudinal growth measure if there are students who enter or leave the cohort over time and these students perform differently, on average, than other students in the cohort.

The estimated school and district learning rates reported in the Stanford Education Data Archive (SEDA) are cohort growth measures based on aggregated data, not individual data. Understanding how much researchers and policymakers can rely on these cohort growth estimates requires us to know how different the estimates obtained from this approach will be from those we would get if we had longitudinal data tracking individual students across the same time-span. How much should we trust the cohort growth measures available in SEDA (or other measures like them) that are constructed as differences in average scores? Under what conditions do the SEDA estimates align poorly to what we would get if we had access to longitudinal student test score records?

In this report we address these questions. We do so by using longitudinal student data from three states (Massachusetts, Michigan, and Tennessee) to construct both average gain score measures (longitudinal growth) and change-in-average measures (cohort growth) for each public school district and school serving students in any of grades 3-8 in the three states. We then compare the two sets of estimates in order to assess how well the latter replicates the former. We do this separately for districts and schools.

We focus on two central comparisons. First, we estimate correlations between longitudinal and cohort growth measures. High correlations indicate that the two measures rank schools the same, on average, and that researchers can generally use the SEDA cohort growth measures in analyses that compare growth rates among schools or districts. Second, we compute the average discrepancy between longitudinal and cohort growth measures. Small discrepancies indicate that cohort growth measures are accurate proxies for longitudinal growth and that users can use a given district's/school's estimate as a reasonable measure of learning rates in that specific district.

We find that the longitudinal and cohort growth measures are generally highly correlated in these three states. This suggests that, on average, the SEDA-style cohort growth measures largely rank schools and districts the same as longitudinal growth measures do. The correlations at the district-level (r=0.87) are somewhat higher than the school-level correlations (r=0.80), which reflects the fact that there is less student mobility among districts than among schools.

In addition, we find that for most districts, the discrepancy between the two types of estimates is relatively small, suggesting that the cohort growth measure is a good proxy for the longitudinal growth measure in the district. However, for about a quarter of districts, the discrepancy is large. The differences are larger, on average, for schools, so there are more schools where the discrepancy is large enough to warrant concern.

Because users without longitudinal data cannot know which districts have large discrepancies and which have small ones, we examine what kind of schools and districts have more accurate cohort growth measures. In general, the cohort growth measures are more accurate in schools and districts with lower rates of student in- and out-mobility, as we would expect (because student mobility is what drives the discrepancy between the two measures).

However, student mobility rates are not readily or publicly available for most states. Instead, public users of cohort growth measures can rely on proxies for student mobility to predict in which schools or districts cohort growth measures accurately reflect longitudinal student growth rates. Student mobility rates tend to be higher in schools with longer grade spans (because there are more grades in which students can enter or leave a school). Moreover, small schools and districts are more likely to have high mobility rates, because the mobility of a few students can substantially change the mobility rate. We find quite strong correlations (over 0.85) between longitudinal and cohort growth measures in all but the smallest districts and schools (those with fewer than 40 students in a given grade in a given year). We also find stronger correlations among schools with 4 or fewer tested grades than those with longer spans of tested grades.

Finally, because of a concern that student mobility may be higher in public charter schools than in traditional public schools in general—and because charter schools are often smaller in size and have longer grade-spans—we also examine the relationship between cohort and longitudinal growth measures separately for charter and traditional public schools.

In charter schools, we find that the cohort growth measures are systematically larger than the longitudinal measures, and the absolute discrepancies tend to be much larger in charter schools than in traditional public schools. This is a result of the fact that charter schools have more *systematic* differential student mobility than traditional public schools: students who leave charter schools have lower test

scores, on average, than the students who enter charter schools in the three states we study. As a result, the cohort growth measures for charter schools are, on average, biased upwards relative to the estimates one would get from longitudinal student records. Nonetheless, the correlation between longitudinal growth rates and cohort growth rates among charter schools is still quite high (r=0.81), indicating that the cohort growth rates still generally rank charter schools against one another similarly as the longitudinal growth measures.

In traditional public schools, in contrast, the cohort growth measures do not systematically overstate or understate growth rates. This is because the students who leave traditional public schools have, on average, similar test scores to those who enter. Moreover, among traditional public schools, the correlation between longitudinal and cohort growth rates is the same is it is for charter schools (r=0.81).

To be clear, the report does not evaluate the effect of charter or traditional public schools (or any set of schools) on student achievement. Nor does it compare the average effectiveness of charter and traditional public schools to one another. The report instead assesses whether cohort growth measures can be used for such purposes. We conclude that, although the cohort growth measures may be useful to determine how particular charter schools compare *to one another*, and how particular traditional public schools compare *to one another*, the cohort growth measures are inappropriate for comparisons of growth rates *between* charter and traditional public schools. Because the cohort growth estimates are systematically too high for charter schools (and are not systematically too high for traditional public schools), any such comparison would be biased in favor of charter schools.

In summary, we conclude that:

- On average, SEDA-style cohort growth measures are useful proxies for longitudinal growth measures;

- The SEDA-style cohort growth measures provide useful estimates of longitudinal growth in all but the smallest schools and districts and/or in schools with a grade span of more than four tested grade levels;

- SEDA-style cohort growth measures may overstate charter school growth in the three states we examine, suggesting that these estimates should not be used to draw comparisons between charter and traditional public school sectors.