Stanford cepa | Center for Education Policy Analysis

CEPA Working Paper No. 19-08

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.

VERSION

November 2019

Suggested citation: Reardon, S.F., Papay, J.P., Kilbride, T., Strunk, K.O., Cowen, J., An, L., & Donohue, K. (2019). 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. (CEPA Working Paper No.19-08). Retrieved from Stanford Center for Education Policy Analysis: http://cepa.stanford.edu/wp19-08

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

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

November, 2019

This work would not have been possible without the collaboration of our partners in the Massachusetts Department of Elementary and Secondary Education (MDESE); the Michigan Department of Education (MDE); Michigan's Center for Educational Performance and Information (CEPI); the Tennessee Department of Education (TDE); and the Tennessee Education Research Alliance (TERA). This research result used data collected and maintained by these agencies and organizations. However, all results, information, and opinions included here represent the analysis, information, and opinions of the authors and are not endorsed by, nor reflect, the views or positions of MDESE, MDE, CEPI, TDE, TERA, or any employee thereof. All errors are our own.

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

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

1. Introduction

One of the features of the modern US educational system is the extent to which students' skills are assessed using standardized test scores. State standards-based reform efforts and the federal No Child Left Behind (NCLB) Act led to widespread interest in how to use such test scores to hold schools accountable for student performance. Over the past decade, a consensus has emerged among both researchers and policymakers that measuring the level of student academic proficiency in a given school is a poor proxy for school quality because average test scores reflect not only the inputs of a school, but also out-of-school factors that shape students' opportunities to learn. Thus, policymakers have begun relying more heavily on student growth, seeking to measure the effectiveness and quality of a school by assessing how quickly its students are learning new material.

Ideally, we would measure the test-score growth for all students in a school in a given grade and year. The presence of out-of-school factors affecting student learning implies that a test-score growth measure would not provide an unbiased estimate of the *causal* effect of the school itself. That test-score growth measure could, however, provide a measure of how much students in that school were learning over the course of the year.

Unfortunately we can only rarely observe current and prior year test scores for all students in a school; student mobility across schools, districts, and states complicates efforts to develop estimates of student learning growth. Given such limitations, there are several ways to operationalize a measure of average learning rates. The best possible feasible approach is to compare end of year scores from a given year to scores from the previous year for students who took both tests. There are several ways to construct this type of measure, but each of them relies on repeated measures of the same students over time. For example, we can compute the average change in scores of the individual students in a given school in each grade and year. By averaging this measure over multiple grade-years, we can get a *longitudinal growth measure* of the average rate at which the students in a given school learn the tested material.

Constructing such measures, however, requires longitudinal student data, which are often not readily available. It is easier to obtain from public sources aggregate data of the average student test scores within a school-grade-year. From such data we can compute a different 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 estimate provides a *cohort growth measure* that documents how much student test scores changed, on average, from 3rd grade in one year to 4th grade in the following year, for example. If the exact same sets of students are tested in both years in sequential grades, the longitudinal and cohort growth estimates are exactly the same. However, if some students were tested in one year but not the other, the difference in average scores obtained from the cohort growth approach may not match the average gain score provided by the longitudinal measure.

When longitudinal data are not available, the cohort growth estimate may be the only feasible approach. Indeed, this is the approach used in the Stanford Education Data Archive (SEDA). SEDA is based on the data collected by the US Department of Education as part of the ED*Facts* data system. The ED*Facts* data include aggregate test score data from virtually every public elementary and middle school in the US from 2008-09 to 2015-16. The aggregated scores are available at the school-grade-year-subject-subgroup level and represent over 300 million individual test scores. However, these data are not available as longitudinal student files, making it impossible to compute individual student gain scores or longitudinal growth measures. Instead, changes in average scores, the only measure of student growth that can be computed from these data is the change in average scores, which we refer to in this report as a "cohort growth" measure. SEDA estimates are particularly valuable as they enable us to draw comparisons about test score levels and cohort growth across states, while estimates of longitudinal growth are only available in certain states and any inferences are only valid in comparison to other schools in the same state (Fahle, et al., 2018).

Thus, there are clear opportunities presented by the cohort growth estimates presented in SEDA. However, 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 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?

Our goal in this report is to answer these questions. We do so by using longitudinal student data from 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 any of grades 3 to 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 assess how similar these estimates are in two ways, addressing two distinct questions that policymakers and researchers might have. First, a user might be interested in whether rankings of districts or schools according to their growth rates are consistent across cohort and longitudinal measures. To assess this, we examine the correlation between the two types of estimates. High correlations suggest that, on average, knowing the cohort growth measure will be a good proxy for understanding longitudinal growth.

Second, a user might be interested in understanding how much a cohort growth estimate may differ from the average longitudinal growth of students in a particular school or district. Here, we focus on both the size and direction of the discrepancy between cohort and longitudinal growth estimates. To determine whether the cohort growth measure systematically overstates or understates longitudinal growth, we examine whether discrepancies tend to be positive or negative. We use two measures to assess the magnitude of these discrepancies: the mean absolute deviation and the root mean square error. We can think of the former as how far off (in absolute value) the cohort growth measure is from the longitudinal measure on average, while the latter is a measure of how spread out the discrepancies are around this average.

Intuitively, we expect the two growth measures to align well so long as the groups of students in each cohort do not change much from year to year. However, the two measures may differ in schools and districts with higher mobility rates or larger gaps in performance across mobile and non-mobile students. Because there is generally more mobility in and out of schools than districts, we might expect that the cohort and longitudinal growth measures will align better for districts than schools. The effects of mobility may compound over grades, so schools that span more grade levels may be particularly susceptible to large discrepancies between cohort and longitudinal growth estimates. Finally, variations in state context and policy have implications for the amount of mobility students might experience, suggesting that cohort growth measures may more accurately capture average individual student growth in some contexts than in others.

We find that the correlations between longitudinal and cohort growth measures are generally high. This suggests that, on average, the SEDA-style cohort growth measures largely rank schools and districts consistently with longitudinal growth measures. For most districts, the discrepancy between the two types of estimates are very small, suggesting that cohort growth is a good proxy for longitudinal growth. However, for about a quarter of districts, the discrepancy is large enough to warrant concern. We find slightly stronger correlations at the district-level (r=0.87) than the school-level (r=0.80), and that the absolute magnitudes of differences between the two estimates are smaller, on average, for districts than schools.

Furthermore, as expected, the correlations in districts and schools with higher student mobility are somewhat weaker than in those with lower mobility (r=0.84 for districts and r=0.75 for schools with more than 15% annual mobility). Across our validation sample, mobility rates tend to be higher schools with longer grade spans and there is greater variation in test score gaps between mobile and non-mobile students in smaller schools and districts. Because mobility is not observable to many end users, we focus on size and grade span as proxies for mobility rates and mobility-related test score gaps, respectively. 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 relatively weaker correlations among schools with 5 or 6 tested grades (e.g., K-8 schools) than those with shorter spans of tested grades.

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. We find that the correlations are similar in charter schools and traditional public schools, but that the cohort growth measures are systematically larger than the longitudinal measures (suggesting a bias that overstates charter growth relative to traditional schools) and the absolute discrepancies tend to be much larger in charter schools. Thus, although the cohort growth measures may be useful to determine how particular charter schools compare to *each other*, the cohort growth measures are inappropriate for comparisons *between* charter and traditional public school growth.

Thus, we conclude that:

- On average, SEDA-style cohort growth measures are useful proxies for longitudinal growth measures;
- The SEDA-style cohort measures provide useful estimates of longitudinal growth in all but the smallest schools and districts 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.

This report is laid out as follows: first, we provide a more detailed description of how the cohort and longitudinal growth measures are constructed and a formal description of the conditions under which they differ. We then describe the data we use for this analysis, and the methods we use to assess the alignment of the two measures. Third, we report the results of the analysis of the alignment of the measures for school districts, followed by a description of the results of the school-level analysis. We conclude with some practical recommendations for users of the SEDA data or similar measures of average student growth.

2. Student growth and mobility

For most research and potentially for many accountability purposes, we want to assess the degree to which student achievement increases in a given school or district over the course of a year. In this section, we discuss how researchers might create two measures of achievement growth: 1) longitudinal growth that tracks individual students' achievement growth over time and 2) cohort growth that measures changes in cohorts of students' average achievement from one year to the next. We also discuss how natural patterns of student mobility may impact the accuracy of each of these growth measures.

Our target parameter is the average achievement gain for all students in a given educational unit (such as a school or district) during a particular grade. The growth of an individual student i during the school year t_2 is determined by subtracting the initial year assessment score ($y1_i$, measured at the end of the previous school year, t_1) from the following year's assessment score ($y2_i$, measured at the end of t_2), represented by $\Delta_i = y2_i - y1_i$.

Given a set of students P of size N_P, the average change in scores of all individual students in P is equal to the change in the average score of students in P, as shown in Equation 1. If P is the entire population of students in a particular educational unit (such as a school or a district) during year t_2, then the target parameter is $\overline{\Delta}_P$, is equivalent to the average change in scores among those who were in the unit during t_2.

$$\overline{\Delta}_P = \frac{1}{N_P}\sum_{i \in P} \Delta_i = \frac{1}{N_P}\sum_{i \in P}(\mathrm{y}2_i - \mathrm{y}1_i) = \frac{1}{N_P}\sum_{i \in P} \mathrm{y}2_i - \frac{1}{N_P}\sum_{i \in P} \mathrm{y}1_i = \overline{y2}_P - \overline{y1}_P \tag{1}$$

In a world with no mobility, we could calculate this target parameter directly because the same set of students would be in the unit in time t_1 and t_2 and we could observe all of the students' scores at both times. However, when there is mobility across units, then the sets of students who contribute to the observed values of $\overline{y2}$ and $\overline{y1}$ differ between the two years; this can lead to differences between $\overline{\Delta}_P$ and the difference in observed means, $\overline{y2} - \overline{y1}$.

When longitudinal data are available, an empirically feasible approximation of the target parameter is the average change in test scores between two testing periods across all students in a unit who were also tested in the previous grade level in the previous testing period. We call this *longitudinal growth*. When only aggregate data are available, an empirically feasible approximation of the target parameter is the difference in average achievement across students in the same cohort relative to the average achievement of the same cohort in the previous year. SEDA growth measures are estimated with aggregate ED*Facts* data using this *cohort growth* approach. These two types of empirically feasible growth measures are equivalent to each other and equivalent to the target parameter when the cohort of students in a unit includes the exact same individual students in two consecutive years t_1 and t_2. These measures may differ from each when there is mobility of students in and/or out of the unit between t_1 and t_2, particularly when the performance of students who enter or leave the unit differs from the performance of those who stay.

Because student mobility has the potential to cause important differences between the two empirically feasible measures of growth (and between these and the target parameter), it is useful to formalize the relationship between student mobility and each type of growth measure. We first define several categories of students based on patterns of mobility between two years t_1 and t_2: students who are present in the unit at both t_1 and t_2 ("stayers"); students in the unit at t_1 but not at t_2 ("leavers"); and those in the unit at t_2 but not at t_1 ("enterers"). We split enterers into two subsets: "movers" (enterers for whom we have t_1 test score data but who were in a different unit at t_1) and "new" students (enterers who appear in the data for the first time at t_2 and therefore have no test score data for the previous year t_1). In the context of our analyses here using state longitudinal data systems, the "new" students are students who moved into the state's public education system (either from out of state or from a private school) between t_1 and t_2. Students in the first year and/or the earliest tested grade level in our state data are also considered "new" in the sense that they have no prior score; these students are excluded from enterer counts. We do not distinguish between leavers who move to other units in the system and leavers who exit the system entirely, as the t_2 scores of leavers do not influence either type of growth measure.

We also define ratios of students in each mobility category and quantify the overall level of mobility in and out of a unit. The leaver ratio, r_l, is the proportion of students in $t_1$who leave the unit (i.e., the number of leavers divided by the sum of leavers and stayers). The enterer ratio, r_e, is the proportion of students in t_2 who entered the unit (i.e., the number of enterers divided by the sum of enterers and stayers). We partition the enterer ratio into a mover ratio and new ratio (r_m and r_n, respectively). The ratio of these two groups combined (r_{m+n}) is equivalent to the enterer ratio (r_e). We also define a "total mobility ratio" as the proportion of all observations across $t_1$and t_2 that are not stayers (all movers, leavers, and new students).

Table 1 outlines the notation we use to describe student counts, assessment scores, and student growth for students in each of these mobility categories, as well as ratios that describe the proportions of all students in a school or district that belong to each mobility category. Note that we cannot observe all of the terms in the table, specifically measures of leaver scores in the year after they exit the data and their change in scores, new students in the year before they enter the data and their change in scores, and all enterers in the year they enter the data and their change in scores. We denote each of these quantities that we do not observe with an asterisk (*).

Table 1. Notation for mobility categories and related terms

Category	Subscript	Description	Student count	t_1 score	t_2 score	$\boldsymbol{y2 - y1}$
Stayer	*s*	In unit at t_1 and t_2	n_s	$y1_s$	$y2_s$	Δ_s
Leaver	*l*	In unit at t_1 only	n_l	$y1_l$	$y2_l$ *	Δ_l *
Mover	*m*	In different unit at t_1	n_m	$y1_m$	$y2_m$	Δ_m
New	*n*	Missing at t_1	n_n	$y1_n$ *	$y2_n$	Δ_n *
Enterer	*m+n*	In unit at t_2 only	n_{m+n}	$y1_{m+n}$ *	$y2_{m+n}$	Δ_{m+n} *

Variable name	Description	Equation
Leaver Ratio	Ratio of leavers to total t_1 observations	$r_l = \dfrac{n_l}{n_s + n_l}$
Mover Ratio	Ratio of movers to total t_2 observations	$r_m = \dfrac{n_m}{n_s + n_m + n_n}$
New Ratio	Ratio of new to total t_2 observations	$r_n = \dfrac{n_n}{n_s + n_m + n_n}$
Enterer Ratio	Ratio of enterers to total t_2 observations	$r_{m+n} = \dfrac{n_m + n_n}{n_s + n_m + n_n}$
Total mobility ratio	Ratio of non-stayers to total observations	$r_t = \dfrac{n_l + n_m + n_n}{2n_s + n_l + n_m + n_n}$

*We do not observe scores for new students at t_1 or leavers at t_2.

The *target parameter,* $\overline{\Delta}_P$ is the average change in performance of all students in a unit at t_2, i.e. the average change in performance across stayers and enterers (both movers and new students), shown in **Equation 2**.

$$target\ parameter = \frac{n_s\overline{\Delta}_s + n_m\overline{\Delta}_m + n_n\overline{\Delta}_n}{n_s + n_m + n_n} \tag{2}$$

Longitudinal growth (*LG*) indicates the average change in performance of all students in a unit at t_2 who have t_1 scores in the data (i.e. the average change in performance across stayers and movers), shown in **Equation 3**. When n_n= 0, longitudinal growth is equivalent to the target parameter. In other words, if we can observe t_1 test scores for all individuals (regardless of whether they were in the same school/district at t_1 or not), our longitudinal growth measure would fully reflect the average growth of students in the unit in t_2.

$$LG = \frac{n_s\overline{\Delta}_s + n_m\overline{\Delta}_m}{n_s + n_m} \tag{3}$$

Cohort growth (CG) refers to the change in cohort means from t_1 to t_2, shown in Equation 4. Because individual students are not tracked longitudinally in these measures, the mean performance of students in a cohort at t_1 includes both stayers and leavers, and the mean performance of students in the cohort at t_2 includes both stayers and enterers (which include both movers and students who are new to the data). When there is no mobility (i.e., $n_m = n_n = n_l = 0$), the cohort growth measure equals the target parameter.

$$CG = \frac{n_s\overline{\mathrm{y2}}_s + n_m\overline{\mathrm{y2}}_m}{n_s + n_m} + \frac{n_s\overline{\mathrm{y2}}_s + n_n\overline{\mathrm{y2}}_n}{n_s + n_n} - \frac{n_s\overline{\mathrm{y1}}_s + n_l\overline{\mathrm{y1}}_l}{n_s + n_l} \tag{4}$$

Equations 2, 3, and 4 can also be expressed in terms of the mobility ratios defined in Table 1 (shown in Equations 5, 6, and 7, respectively). The term $\overline{\Delta}_s$, which appears in all three equations, represents the average growth of stayers. The remaining terms highlight differences in how mobility contributes to the three types of growth measures. In addition to the growth of stayers, the target parameter captures growth of both types of enterers, while *LG* captures growth of movers but not new-to-system students, and *CG* captures scores of different mobility groups in the two time periods (leavers at t_1 only and both types of enterers at t_2 only).

$$target\ parameter = \overline{\Delta}_s + r_m(\overline{\Delta}_m - \overline{\Delta}_s) + r_n(\overline{\Delta}_n - \overline{\Delta}_s) \tag{5}$$

$$LG = \overline{\Delta}_s + \frac{r_m}{1 - r_n}(\overline{\Delta}_m - \overline{\Delta}_s) \tag{6}$$

$$\begin{aligned} CG &= \overline{\Delta}_s + r_m\left(\overline{\mathrm{y2}}_m - \overline{\mathrm{y2}}_s\right) + r_n\left(\overline{\mathrm{y2}}_n - \overline{\mathrm{y2}}_s\right) - r_l\left(\overline{\mathrm{y1}}_l - \overline{\mathrm{y1}}_s\right) \\ &= \overline{\Delta}_s + r_e\left(\overline{\mathrm{y2}}_e - \overline{\mathrm{y2}}_s\right) - r_l\left(\overline{\mathrm{y1}}_l - \overline{\mathrm{y1}}_s\right) \end{aligned} \tag{7}$$

We define equations for the bias in each empirically feasible growth measure by subtracting the formula for the target parameter from the *LG* and *CG* formulas. The degree of bias in the *LG* measure, shown in Equation 8, depends on the ratios of movers and new-to-system students and differences in the growth rates of these students relative to stayers. Bias in the *CG* measure, shown in Equation 9, depends on the ratios of movers, new-to-system students, and leavers, and differences between the t_1 scores of movers and new-to-system students relative to stayers and between the t_2 scores of leavers relative to stayers.

$$LG\ bias = \frac{r_m r_n}{1-r_n}(\overline{\Delta}_m - \overline{\Delta}_s) - r_n(\overline{\Delta}_n - \overline{\Delta}_s). \tag{8}$$

$$\begin{aligned} CG\ bias &= r_m(\overline{y1}_m - \overline{y1}_s) + r_n(\overline{y1}_n - \overline{y1}_s) - r_l(\overline{y1}_l - \overline{y1}_s) \\ &= r_e(\overline{y1}_e - \overline{y1}_s) - r_l(\overline{y1}_l - \overline{y1}_s) \end{aligned} \tag{9}$$

Thus, the bias in longitudinal growth depends on whether the *growth rates* of enterers and stayers differ, while the bias of the cohort growth measure depends on whether stayers have different *test-score levels* than non-stayers. Differences in growth across students tend to be much smaller than differences in levels (e.g., Kane & Staiger, 2002).[1] Thus, the bias in longitudinal growth measures should be substantially smaller than bias in cohort growth measures.

Although we cannot observe either of these biases directly (because the baseline achievement scores and growth rates of new-to-system students are not observed), we expect the longitudinal growth measures to be a much closer approximation of the target parameter. In the remainder of the report, then, we focus our analysis on measuring the degree of difference between cohort and longitudinal growth measures. Technically, we can quantify the difference by subtracting **Equation 6** from **Equation 7**. This discrepancy, given by **Equation 10**, is affected by the mover, new-to-system, and leaver ratios, as well as differences between stayers and leavers at t_1, differences between stayers and new-to-system students at t_2, and differences between movers and stayers in both years.

$$CG - LG = \frac{r_m}{1-r_n}(\overline{y1}_m - \overline{y1}_s) - \frac{r_m r_n}{1-r_n}(\overline{y2}_m - \overline{y2}_s) + r_n(\overline{y2}_n - \overline{y2}_s) - r_l(\overline{y1}_l - \overline{y1}_s) \tag{10}$$

Appendix Table A-1 separates the biases in the cohort and longitudinal growth measures defined in **Equations 8** and **9**, respectively, as well as the discrepancy between these two measures, defined in **Equation 10**, into several different components that each describe the effect of the test scores from students in a particular mobility category (leavers, movers, or new-to-system students) in a particular school year (t_1 or t_2) on these quantities. The *CG bias* and *LG bias* equations both include a combination of observed and unobserved components. Every component of the *CG-LG discrepancy* equation, however, are observable in our data.

3. Data and Methods

3.1 Source data

We use data from all public schools and school districts in Massachusetts, Michigan, and Tennessee as our validation sample.[2] We chose these three states for two main reasons. First, this is a convenience

[1] We find that this is empirically true in the SEDA data and in the three state datasets we analyze in this study.
[2] More precisely, we focus on districts that appear in both the SEDA and state datasets, which is >99% of districts in these three states, and schools that appear in both the SEDA crosswalk and state datasets. Comparisons of the SEDA dataset with each of the source datasets from the three validation states can be found in **Appendix Table A-3.** We outline additional details of these datasets in **Appendix Table A-4**, including state-by-state differences in the source

sample of states; members of the research team have access to longitudinal student-level data that covered the timeframe available in SEDA and state-level partners were amenable to the study. Second, each of these three states is empirically interesting in different ways; Massachusetts has the highest test-score levels of all states in the SEDA data, while Tennessee has particularly high test-score growth rates, and Michigan has high levels of student mobility and one of the largest percentages of students attending charter schools.

In **Table 2**, we provide means and standard deviations of district characteristics from the SEDA covariate dataset, first for all districts, then all districts in the validation sample, and finally for all districts in each individual validation state. These comparisons illustrate two main points. First, the validation sample is fairly reflective of national averages, although our states have somewhat fewer Hispanic and ELL students and somewhat more students in charter schools. Second, the three validation states vary considerably from each other along nearly all dimensions, including average district size, school resources, charter landscape, and student composition. These different properties enable us to offer guidance to SEDA users that is generalizable beyond the three validation states and also to study how patterns vary across different state contexts.

Table 2. Characteristics of school districts nationwide and in validation states.

	All States	Validation States			
		All	MA	MI	TN
Average per-grade enrollment	306 (1170)	281 (549)	257 (347)	223 (421)	552 (1043)
Total number of schools/ districts	7.8 (24.9)	7.6 (14.6)	6.3 (10.4)	6.9 (12.4)	13.3 (25.0)
Number of charter schools/district	0.5 (4.0)	0.5 (3.3)	0.2 (1.3)	0.6 (4.2)	0.3 (2.7)
Average student-teacher ratio	15.5 (12.2)	16.4 (2.8)	13.6 (1.9)	18.2 (1.9)	15.6 (2.0)
Average per-pupil expenditure ($)	12,819 (4354)	12,123 (3612)	16,281 (3345)	10,653 (1566)	8,778 (921)
Percent White	73.9 (27.5)	83.3 (19.9)	84.5 (17.2)	82.5 (21.9)	83.7 (17.3)
Percent Black	7.9 (16.6)	7.9 (15.8)	3.6 (6.4)	8.9 (18.9)	10.6 (15.3)
Percent Hispanic	13.3 (20.3)	6.0 (9.2)	7.6 (12.1)	5.4 (8.1)	4.8 (4.8)
Percent Asian	2.1 (4.9)	2.2 (4.1)	4.1 (5.4)	1.5 (3.2)	0.8 (1.0)

datasets, and decision rules applied while replicating the SEDA dataset and constructing each of the different growth measures.

Percent Native American	2.8	1.0	0.3	1.7	0.2
	(10.7)	(4.4)	(0.5)	(5.8)	(0.2)
Percent free/reduced-price lunch eligible (FRL)	39.0	36.8	19.7	42.4	52.1
	(20.1)	(20.5)	(16.7)	(18.0)	(11.0)
Percent English language learners (ELL)	4.3	2.2	2.7	2.1	1.7
	(8.2)	(4.5)	(4.4)	(4.9)	(2.2)
Percent special education	13.8	14.2	16.7	13.0	13.5
	(5.5)	(3.6)	(2.9)	(3.5)	(2.5)
Percent of students in charter schools	1.3	2.8	2.0	3.9	0.1
	(6.0)	(8.4)	(6.5)	(10.1)	(0.8)
Number of districts	12,052	945	291	519	135

We stratify our sample by a few key characteristics, as outlined in **Table 3**. We categorize both schools and districts by grade size (calculated as the average number of students per grade-year-subject cell) and by total mobility (calculated as the percent of all observations in two consecutive years that are not from stayers). There are far more "high-mobility" schools than districts, as school-level mobility rates include students that move between different schools in the same district. Massachusetts has higher proportions of low-mobility schools and districts than either of the other states. Michigan has the highest proportion of high-mobility districts, while Tennessee has the highest proportion of high-mobility schools. This is likely because Tennessee has the highest proportion of very large districts, resulting in more within-district mobility and less between-district mobility, compared to Michigan, which has the highest proportions of small districts.

We also stratify districts by the number of grade-year-subject cells (out of a possible 84 cells for 6 grades, 7 years, and 2 subjects) that are used to calculate the growth measures. Schools generally do not have students in every tested grade, so the total number of possible grade-year-subject cells varies depending on the grade span of a school. Rather than the number of cells, we stratify schools by grade span and number of years in the data. We also differentiate schools by sector (charter vs. traditional public school) and level (elementary vs. middle vs. combined), as we suspect there may be greater mobility in charter schools and schools spanning both elementary and middle school grade levels.

Table 3. Sample stratification categories, abbreviations, descriptions, and percentages

			Percent of Districts				*Percent of Schools*			
Variable	*Category*	*Description*	MA *N*=296	MI *N*=512	TN *N*=135	All *N*=943	MA *N*=1,386	MI *N*=2,544	TN *N*=1,417	All *N*=5,347
Mobility	Low	< 10% non-stayers	85.8	43.2	54.8	58.2	56.5	26.1	8.6	29.3
	Mid	10%-15% non-stayers	11.5	39.3	36.3	30.1	17.5	29.0	37.4	28.3
	High	> 15% non-stayers	2.7	17.6	8.9	11.7	26.0	44.9	54.0	42.4
Average enrollment per grade	Very small	<40 students per grade-year cell	11.8	12.5	3.0	10.9	14.6	17.0	14.5	15.7
	Small	40-99 per grade-year cell	15.9	30.7	14.1	23.7	51.8	58.7	51.3	54.9
	Medium	100-199 per grade-year cell	27.4	24.6	23.7	25.3	23.5	14.4	25.9	19.8
	Large	200-399 per grade-year cell	30.4	20.5	28.9	24.8	9.5	9.3	8.1	9.0
	Very large	≥400 per grade-year cell	14.5	11.7	30.4	15.3	0.6	0.7	0.2	0.5
Cell count	Low	≤40 grade-year-subject cells	7.8	2.2	0.0	3.6	-----	-----	-----	-----
	Mid	41-80 grade-year-subject cells	18.9	11.3	3.7	12.6	-----	-----	-----	-----
	High	81-84 grade-year-subject cells	73.3	86.5	96.3	83.8	-----	-----	-----	-----
School type	Elementary	Lowest tested grade ≤4, highest ≤6	-----	-----	-----	-----	60.9	58.4	57.4	58.8
	Middle	Lowest tested grade ≥5	-----	-----	-----	-----	27.6	26.3	26.2	26.6
	Combined	Lowest tested grade ≤4, highest ≥7	-----	-----	-----	-----	11.5	15.3	16.5	14.7
Sector	TPS	Traditional public school	-----	-----	-----	-----	95.2	89.8	96.5	93.1
	Charter	Charter school	-----	-----	-----	-----	4.8	10.2	3.5	7.0
Grade span	2-4	2-4 tested grades	-----	-----	-----	-----	88.5	84.8	83.7	85.5
	5-6	5-6 tested grades	-----	-----	-----	-----	11.5	15.2	16.3	14.5
Number of years	1-3	In data 1-3 years	-----	-----	-----	-----	5.1	11.0	5.9	8.1
	4-6	In data 4-6 years	-----	-----	-----	-----	6.3	10.5	9.1	9.0
	All 7	In data all 7 years	-----	-----	-----	-----	88.7	78.5	85.1	82.9

3.2 Estimation strategy

We first construct longitudinal growth measures using each state's student-level longitudinal panel. The three state datasets span the 2008-2009 through 2014-2015 school years and include math and ELA standardized assessment scores and demographic variables for all students in grades 3 through 8.

We first standardize student-level scores on state math and ELA assessments within states, grades, years, and subjects, and then compute unit (school or district)-grade-year-subject mean scores, prior year scores, and student counts separately for stayers, leavers, movers, and new-to-system students. We then use **Equation 3** to compute a longitudinal growth estimate for each unit-grade-year-subject. The mean longitudinal growth for each unit-subject is computed as the average *LG* estimate across grades and years, as shown in **Equation 11**.

$$\widehat{LG}_{db} = \frac{1}{totcells_{db}} \sum_{gy} \frac{n_{s\cdot dgyb}\bar{\Delta}_{s\cdot dgyb} + n_{m\cdot dgyb}\bar{\Delta}_{m\cdot dgyb}}{n_{s\cdot dgyb} + n_{m\cdot dgyb}} \tag{11}$$

where $n_{s\cdot dgyb}$ is the number of students in who were assessed in subject *b* in unit *d* both in grade *g* in year *y* and in grade *g*-1 in year *y*-1 (stayers); $\bar{\Delta}_{s\cdot dgyb}$ is the average difference between scores on these two assessments among this group of students; $n_{m\cdot dgyb}$ is the number of students who were assessed in subject *b* in grade *g* in year *y* in unit *d* and assessed in a different unit (before moving to unit *d*) in subject *b* in grade *g-1* in year *y*-1; $\bar{\Delta}_{m\cdot dgyb}$ is the average difference between scores on these two assessments among this group of students; $totcells_{db}$ is the total number of grade-year cells where growth in subject *b* can be observed in unit *d*.

The average of the two subject-specific *LG* estimates gives an overall measure of growth for each unit (shown in **Equation 12**).

$$\widehat{LG}_{d\cdot overall} = \frac{\widehat{LG}_{d\cdot math} + \widehat{LG}_{d\cdot ELA}}{2} \tag{12}$$

Next, to allow us to compare the two different types of growth measures, we must create our *SEDA-style cohort growth* measures slightly differently than those found in the SEDA dataset available to the public. First, we construct our SEDA-style cohort growth measures using the same restricted state datasets to ensure that when we are comparing estimates, they are based on identical raw data. Second, in SEDA's data, mean scores in a grade-year-subject are estimated from a heteroskedastic ordered probit (HETOP) model from proficiency counts, while we estimate them by averaging finer-grained test scores. Thus, our comparisons are not confounded by differences in how the means are computed.

We first restructure each state dataset to match the format and contents of the SEDA version 2.1 long state-scaled data file. To do this, we again standardize student-level scores on state math and ELA assessments within states, grades, years, and subjects. Then we collapse to a single observation per unit-grade-year-subject containing a mean standardized assessment score, the standard error of the mean score, and the number of student observations used to calculate each school or district mean score.

We replicate SEDA subject-specific cohort growth measures using the model shown in **Equation 13**.

$$\begin{gathered}
\hat{y}_{dygb} = [\beta_{0md} + \beta_{1md}(cohortc_{dygb}) + \beta_{2md}(gradec_{dygb})]math_b \\
+ [\beta_{0ed} + \beta_{1ed}(cohortc_{dygb}) + \beta_{2ed}(gradec_{dygb})]ela_b + u_{dygb} + e_{dygb} \\
\begin{aligned}
\beta_{0md} &= \gamma_{0m0} + v_{0md} & \beta_{0ed} &= \gamma_{0e0} + v_{0ed} \\
\beta_{1md} &= \gamma_{1m0} + v_{1md} & \beta_{1ed} &= \gamma_{1e0} + v_{1ed} \\
\beta_{2md} &= \gamma_{2m0} + v_{2md} & \beta_{2ed} &= \gamma_{2e0} + v_{2ed} \\
\beta_{3md} &= \gamma_{3m0} + v_{3md} & \beta_{3ed} &= \gamma_{3e0} + v_{3ed}
\end{aligned} \\
e_{dygb} \sim N(0, \omega^2_{dygb});\; u_{dygb} \sim N(0, \sigma^2);\; \begin{bmatrix} v_{0md} \\ \vdots \\ v_{2ed} \end{bmatrix} \sim MVN(0, \tau^2)
\end{gathered} \tag{13}$$

The variable $\hat{y}_{dygb}$ is the mean standardized assessment score for unit *d,* in year *y,* for grade *g,* in subject *b*. The variable *cohort* is defined as the year when a cohort of students began kindergarten and is calculated by subtracting *grade* from *year*. These variables are each centered around the midpoint of their highest and lowest possible values (i.e. *cohortc* is centered on 2006.5 and *gradec* is centered on 5.5). The binary variable *math* equals 1 for math observations and 0 for ELA observations. The coefficients for *gradec*, β_{2md} and β_{2ed}, represent the cohort growth estimates for unit *d* for math and ELA, respectively. These estimates are compared to the subject-specific *LG* estimates that were computed using **Equation 11.**

Next, we replicate SEDA overall cohort growth measures using the model shown in **Equation 14**.

$$\begin{gathered}
\hat{y}_{dygb} = \beta_{0d} + \beta_{1d}(cohortc_{dygb}) + \beta_{2d}(gradec_{dygb}) + \beta_{3d}(mathc_b) + u_{dygb} + e_{dygb} \\
\begin{aligned}
\beta_{0d} &= \gamma_{00} + v_{0d} \\
\beta_{1d} &= \gamma_{10} + v_{1d} \\
\beta_{2d} &= \gamma_{20} + v_{2d} \\
\beta_{3d} &= \gamma_{30} + v_{3d}
\end{aligned} \\
e_{dygb} \sim N(0, \omega^2_{dygb});\; u_{dygb} \sim N(0, \sigma^2);\; \begin{bmatrix} v_{0d} \\ v_{1d} \\ v_{2d} \\ v_{3d} \end{bmatrix} \sim MVN(0, \tau^2)
\end{gathered} \tag{14}$$

In this model, the variable *mathc* is defined by centering *math* around the midpoint of its highest and lowest possible values (0.5). The coefficient for *mathc*, β_{3d}, gives the difference between math and ELA performance for unit *d*. The coefficient for *gradec*, β_{2d}, gives an overall cohort growth estimate for unit *d*, which is compared to the overall longitudinal growth estimate from Equation 12. Fahle, Shear, Kalogrides, Reardon, DiSalvo, & Ho (2018) provides more details on the approaches used to estimate Equations 13 and 14.

We compare estimates from corresponding *CG* and *LG* measures across all districts and schools, as well as within each of the categories outlined in Table 3. We focus our comparisons on four main quantities of interest (described in Table 4). The first is the correlation coefficient (*r*), which describes the general strength of the relationships between the two types of measures. The root mean square error (RMSE), similarly, describes how closely the *CG* measure corresponds to the *LG* measure. We consider both of these quantities because, when the standard deviation of *LG* is high, it is possible for the RMSE to be large (indicating large differences between the *CG* and *LG* measures) even when the correlation is strong. We also examine the mean *CG-LG* discrepancy to determine whether *CG* estimates are systematically higher or lower than *LG* estimates and the mean absolute *CG-LG* discrepancy to determine how far apart the two estimates are on average.

Table 4. Quantities of interest for comparing cohort and longitudinal growth measures.

Correlation coefficient *(r)* $\frac{\sum(CG-\overline{CG})(LG-\overline{LG})}{\sqrt{\sum(CG-\overline{CG})^2\sum(LG-\overline{LG})^2}}$	Strength of relationship between *CG* and *LG* measures - Informative about how consistently the two measures rank districts or schools
Mean discrepancy $\overline{CG-LG}$	Average difference between *CG* and *LG* estimates - Can be positive or negative - Informative about directional biases
Mean absolute discrepancy $\lvert\overline{CG-LG}\rvert$	Average magnitude of *CG-LG* discrepancy - Always positive - Informative about the typical size of discrepancies
Root mean square error (RMSE) $\sqrt{\overline{(CG-LG)^2}}$ $=\beta\cdot SD_{LG}\cdot\sqrt{\frac{1-r^2}{r^2}}$	Square root of the mean squared *CG-LG* discrepancy - Informative about the spread of discrepancies - Weights large discrepancies more than does the mean absolute discrepancy - Related to the correlation coefficient (*r*), slope from the regression of *CG* on *LG* (β), and standard deviation of *LG*

4. Findings

We present results separately for districts and schools. In each section, we provide overall results and then show differences by state and by subject. We separate the district- and school-level analyses for several reasons. First, policymakers and users may be separately interested in understanding how well SEDA-style estimates perform for districts and for schools. Second, districts tend to be much larger than schools, so there is substantially less sampling variation in estimates of test-score levels and growth at the district level. Third, districts tend to have much lower in or out-of-unit mobility than schools, suggesting that longitudinal and cohort growth measures should be more comparable at the district level. In addition, we examine differences between growth measures for traditional public and charter schools.

4.1 District-level analysis

4.1.a Measures of growth, discrepancy, mobility and mobility score gaps

Before delving into the results, we first provide summary statistics for the different growth measures, discrepancies between the measures, mobility ratios, and score gaps across all districts in our three states in **Table 5** to help interpret our findings. In the top panel, we show the characteristics of our growth measures, overall and by subject. Growth is measured in standard deviations of the state test scores, i.e., a growth estimate of 0.1 indicates the average student in the district moved up by one tenth of a state standard deviation of achievement on the state test, relative to other students in the state. Because we are comparing students within the same state, average growth measures are by definition close to zero – this does not mean that students are not learning, just that our measures are relative to the average growth for all students in the state. Overall longitudinal and cohort growth estimates are very similar. However, estimates of growth in math are slightly larger and more variable than estimates of ELA growth across both the cohort and longitudinal measures. In the last row of the top panel, we present the discrepancy between cohort and longitudinal growth measures. The variance of the discrepancy is approximately one quarter as large as the variance of longitudinal and cohort growth for both overall and subject-specific measures (the standard deviation is half as large).

The middle panel describes the proportions of students in each mobility category. Empirically, we find that approximately 4% of observations in our three-state sample are "new-to-system." This means that we cannot calculate longitudinal growth for 4% of students who experienced school in the unit in year t_2 and suggests that longitudinal growth measures may be slightly biased estimates of the target parameter, especially if "new-to-system" students are somehow different than the other students for whom we can observe test scores in both years t_1 and t_2. Overall, we find that approximately 10% of students leave their district in a given year and 10% enter it, suggesting a greater scope for bias in cohort growth measures than longitudinal measures. That said, on average, districts are gaining a similar proportion of students (enterer ratio) as they are losing (leaver ratio).

The last panel includes estimates of the gaps in test scores between students who enter vs. stay and students who leave vs. stay. The Math and ELA enterer and leaver gaps show that mobile students tend to be lower performing than the stayers. These relationships imply that, on average, the students who leave a district are quite similar in test-score levels to the students who enter, but different than those who remain in their districts.

Table 5. Distribution of district-level mobility ratios, test score gaps, and growth estimates.

	Math	ELA	Overall
Growth estimates			
stayer growth ($\overline{\Delta}_s$)	0.004 (0.066)	-0.004 (0.050)	-0.000 (0.051)
cohort growth (CG)	0.003 (0.062)	-0.001 (0.047)	0.001 (0.053)
longitudinal growth (LG)	0.003 (0.063)	-0.003 (0.051)	-0.000 (0.050)
discrepancy (CG-LG)	0.000 (0.032)	0.001 (0.028)	0.001 (0.026)
Mobility groups			
leaver ratio	0.101 (0.053)	0.101 (0.053)	0.101 (0.053)
mover ratio	0.062 (0.039)	0.062 (0.039)	0.061 (0.039)
new-to-system ratio	0.038 (0.018)	0.039 (0.018)	0.039 (0.018)
Test score gaps			
leaver-stayer at t_1	-0.286 (0.142)	-0.277 (0.153)	-0.281 (0.143)
mover-stayer at t_1	-0.287 (0.200)	-0.259 (0.172)	-0.273 (0.182)
mover-stayer at t_2	-0.306 (0.177)	-0.266 (0.150)	-0.286 (0.161)
new-stayer at t_2	-0.357 (0.215)	-0.342 (0.235)	-0.350 (0.217)

Note: Standard deviations in parentheses.

4.1.b: District-level Results

Overall, district-level estimates of longitudinal and cohort growth are quite similar. As is shown in Table 6 (below), the correlation between the two is very strong (r=0.87 overall). This means that the cohort growth measures explain ¾ of the variation in longitudinal growth measures. It suggests that the two measures rank districts quite similarly. For example, 73% of districts ranked in the top quartile on longitudinal growth rank in the top quartile on cohort growth.

Table 6. Comparison of district-level CG and LG estimates.

	Math	ELA	Overall
CG/LG Correlation	0.869	0.842	0.866
Mean CG-LG Discrepancy	0.000	0.001	0.001
Mean Absolute Discrepancy	0.021	0.018	0.018
Root Mean Square Error	0.032	0.028	0.026
Standard Deviation of LG	0.063	0.051	0.050

In addition, we see that the average discrepancy (column 2) between the two measures is very close to zero (mean=0.001) suggesting that in the average district the cohort growth measure will be an unbiased estimate of longitudinal growth.

However, just because these estimates are right "on average" does not mean that they are reliable enough to make judgements about growth in individual districts. Because the discrepancy can be either positive or negative, we also examine the mean absolute value of the difference between the two measures. We find that, in the average district, the cohort measure differs from the longitudinal growth measure by 0.018 (column 3). NAEP data suggest the average student gains 0.33 SD per year on vertically equated tests, so the mean discrepancy we observe is about +/-5% of a year's growth.

We represent visually the high level of consistency between cohort growth (CG) and longitudinal growth (LG) in **Figure 1**, where we plot district-level CG estimates against LG estimates. The CG and LG estimates for the majority of districts fall along the 45-degree line. In **Figure 2**, we show the CG-LG discrepancy against the LG measure. Here, again, most CG-LG discrepancies fall near the horizontal line, which is drawn at the value of CG-LG=0. However, both figures show that there are a few outlier districts in which there is substantial discrepancy between the two measures of growth. One reasonable metric for assessing how well the CG measure compares to the LG measure is to note that we want the error variance to be no more than 25% of the true variance, which corresponds to a reliability of 0.8. This corresponds to a discrepancy of +/- 0.025. In Figure 2, we include horizontal lines at +/-0.025, showing that 78% of districts fall within this range.

These figures show the strong relationship between cohort and longitudinal growth measures and document the magnitudes of the discrepancies across districts. To illustrate these magnitudes more clearly, we present the cumulative distribution function (CDF) of the discrepancy in **Figure 3**. Here, we see that 10% of districts have discrepancies below -0.025, while 13% have discrepancies above 0.025.

Figure 1. Cohort vs Longitudinal District-Level Growth Estimates

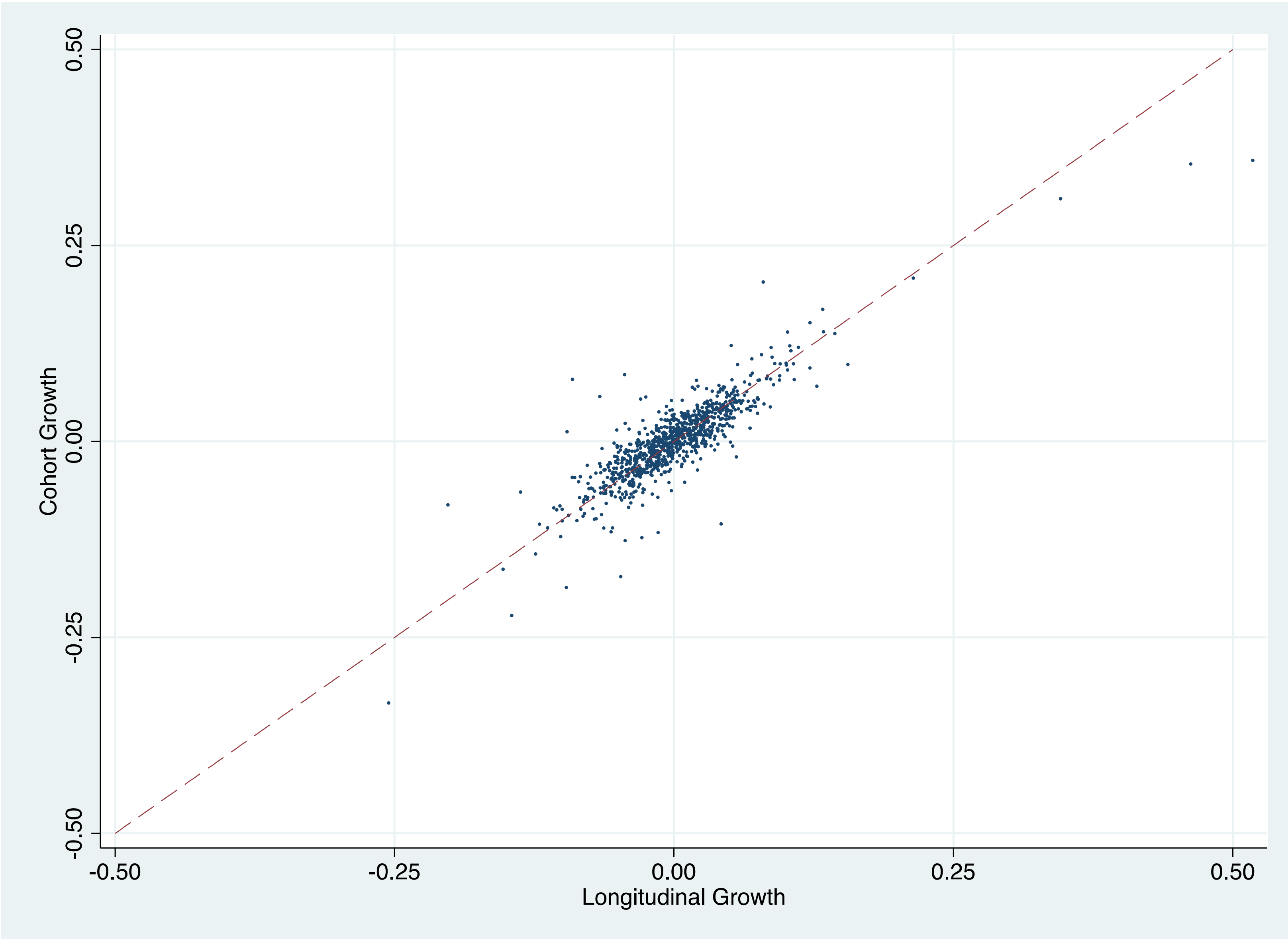


Figure 2. District-Level Cohort-Longitudinal Growth Discrepancies vs Longitudinal Growth

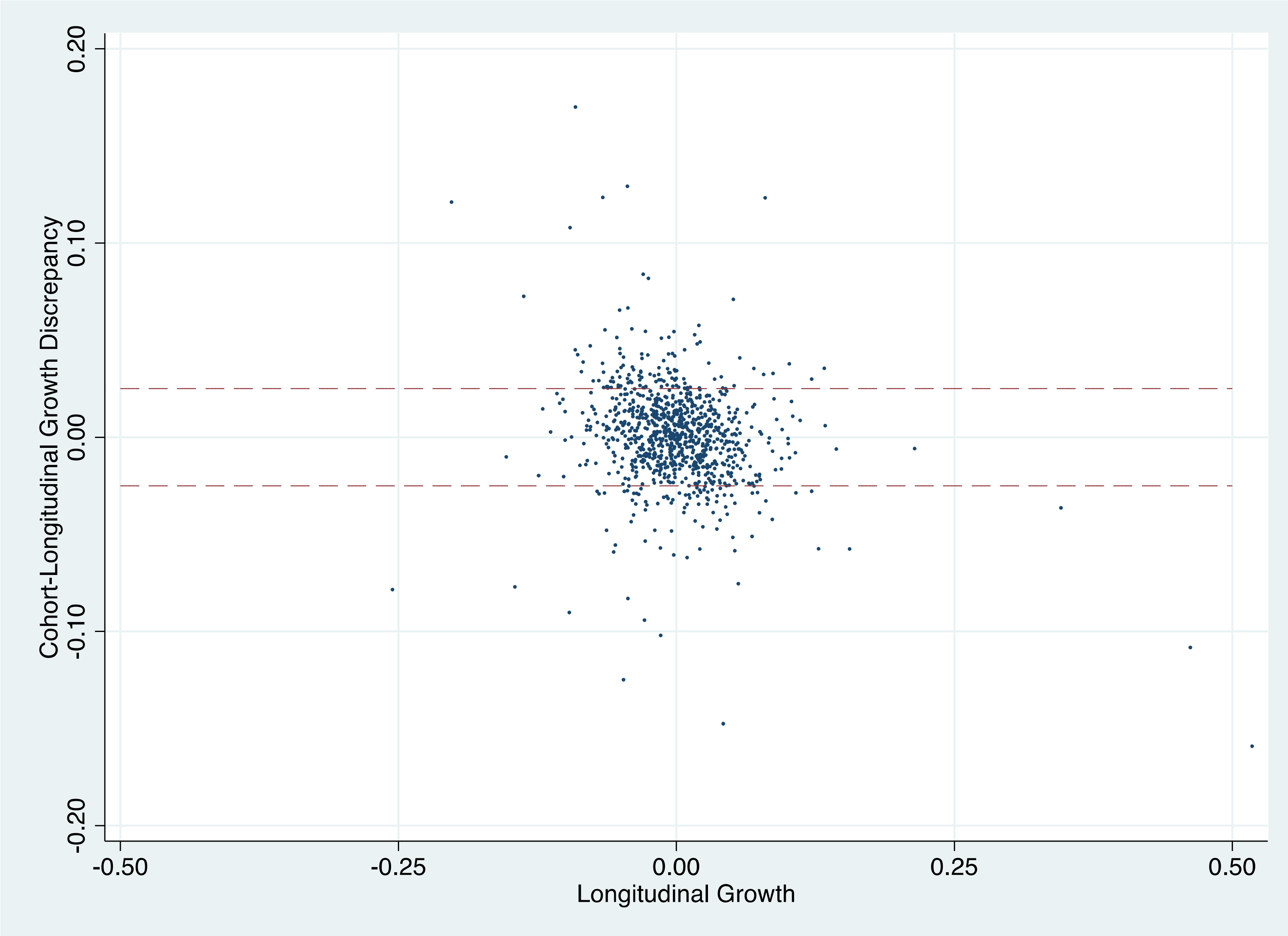
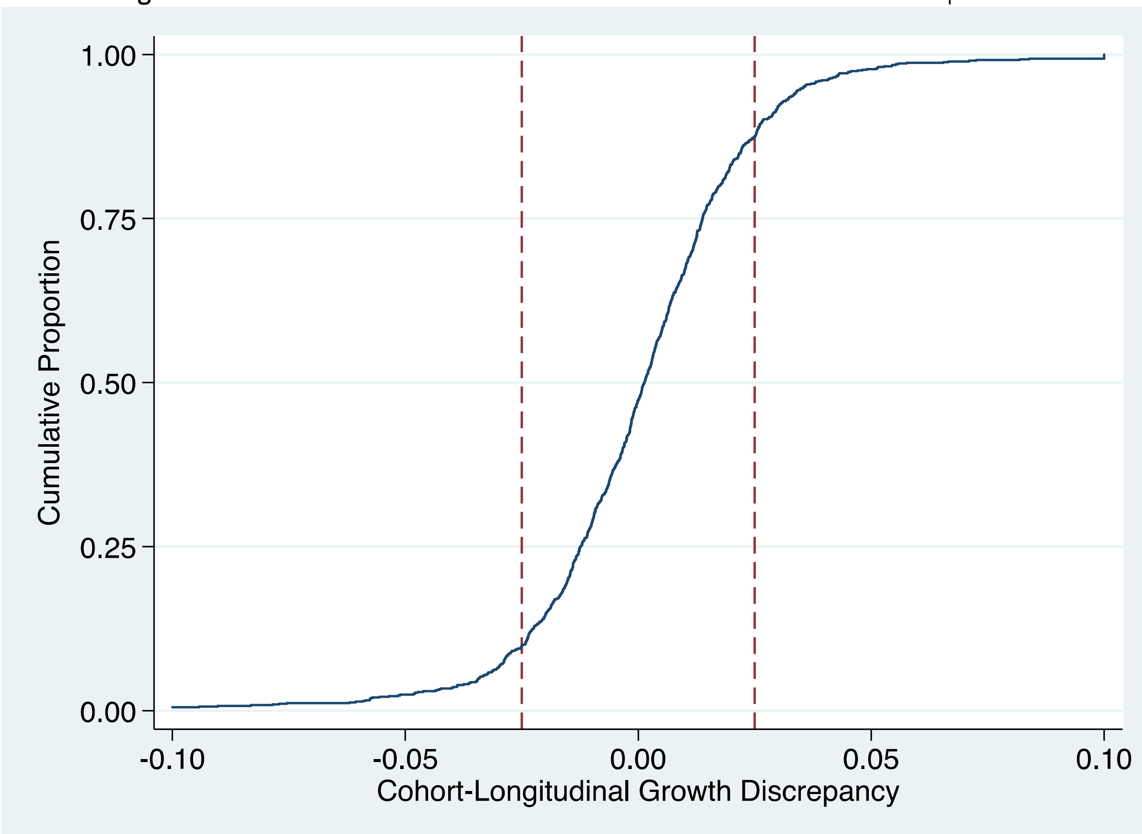


Figure 3. Cumulative distribution function of district-level CG-LG discrepancies



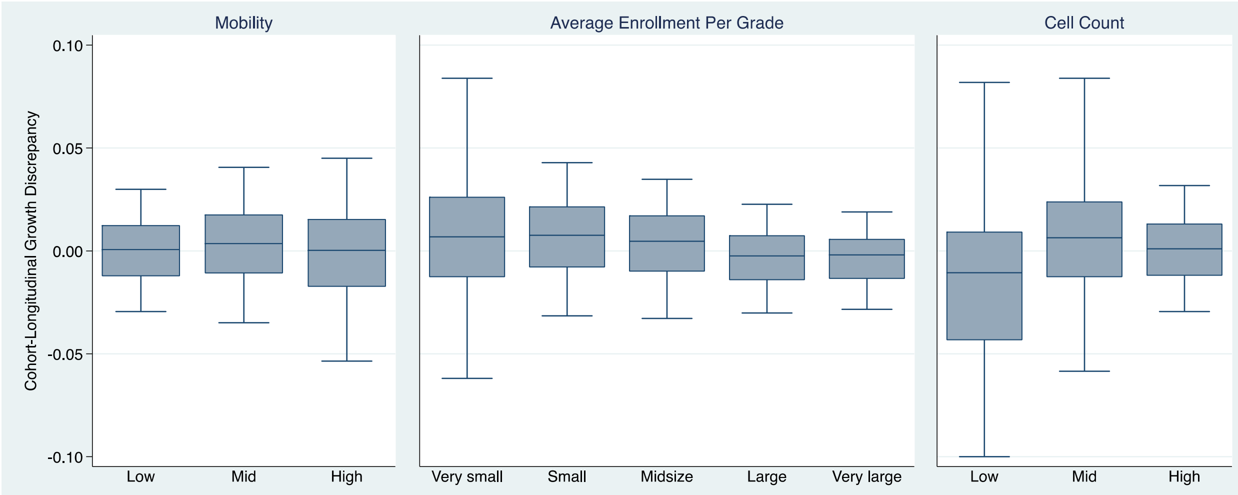
We expect mobility to be a primary driver of the CG-LG discrepancy; as such, we look at the relationship between the discrepancy and mobility as well as proxies that are readily available in the SEDA data such as average grade enrollment and the number of grade-year-subject cells that contribute to each growth estimate. **Table** 7 shows the CG-LG correlations, mean absolute deviations, RMSEs, and mean discrepancies of the overall growth measures for each category of districts. Smaller sizes, higher mobility, and lower cell counts are associated with slightly lower correlations and considerably larger mean absolute deviations and RMSEs. In other words, growth measures vary more in places with less information; this is most evident in districts that are very small or have lower cell counts.

Table 7. Comparison of overall district-level CG and LG estimates by category

	CG/LG Correlation	Mean CG-LG Discrepancy	Mean Absolute Discrepancy	Root Mean Square Error
Mobility				
Low (<10%)	0.885	0.000	0.016	0.025
Mid (10-15%)	0.825	0.003	0.019	0.026
High (>15%)	0.835	-0.001	0.023	0.031
Average enrollment per grade				
Very small (<40)	0.830	0.005	0.033	0.050
Small (40-99)	0.866	0.006	0.019	0.024
Medium (100-199)	0.912	0.002	0.017	0.023
Large (200-399)	0.870	-0.002	0.013	0.018
Very large (≥400)	0.886	-0.006	0.014	0.017
Cell count				
Low (≤40)	0.900	-0.020	0.048	0.065
Mid (41-80)	0.854	0.007	0.028	0.041
High (81-84)	0.862	0.001	0.015	0.019

In **Figure 4**, we present box-and-whiskers plots that show the median, interquartile range, 5th percentile, and 95th percentile of discrepancies for districts across size, mobility, and cell count categories. These plots reveal several key insights. First, the median discrepancy is very close to zero in all categories except for districts with low cell counts (recall, these are the 3.6% of districts that have less than half of the possible grade-subject-year combinations because they opened or closed in the middle of our sample timeframe or because they only serve elementary or middle grades students). Second, we see that smaller average grade sizes, higher mobility, and lower cell counts are associated with larger discrepancies. For example, for districts with fewer than 40 grade-year-subject cells, the interquartile range of the discrepancy is about 0.05, while it is half as large for districts with 41-80 cells. We also looked at other proxies, such as student demographics, and found no systematic patterns.

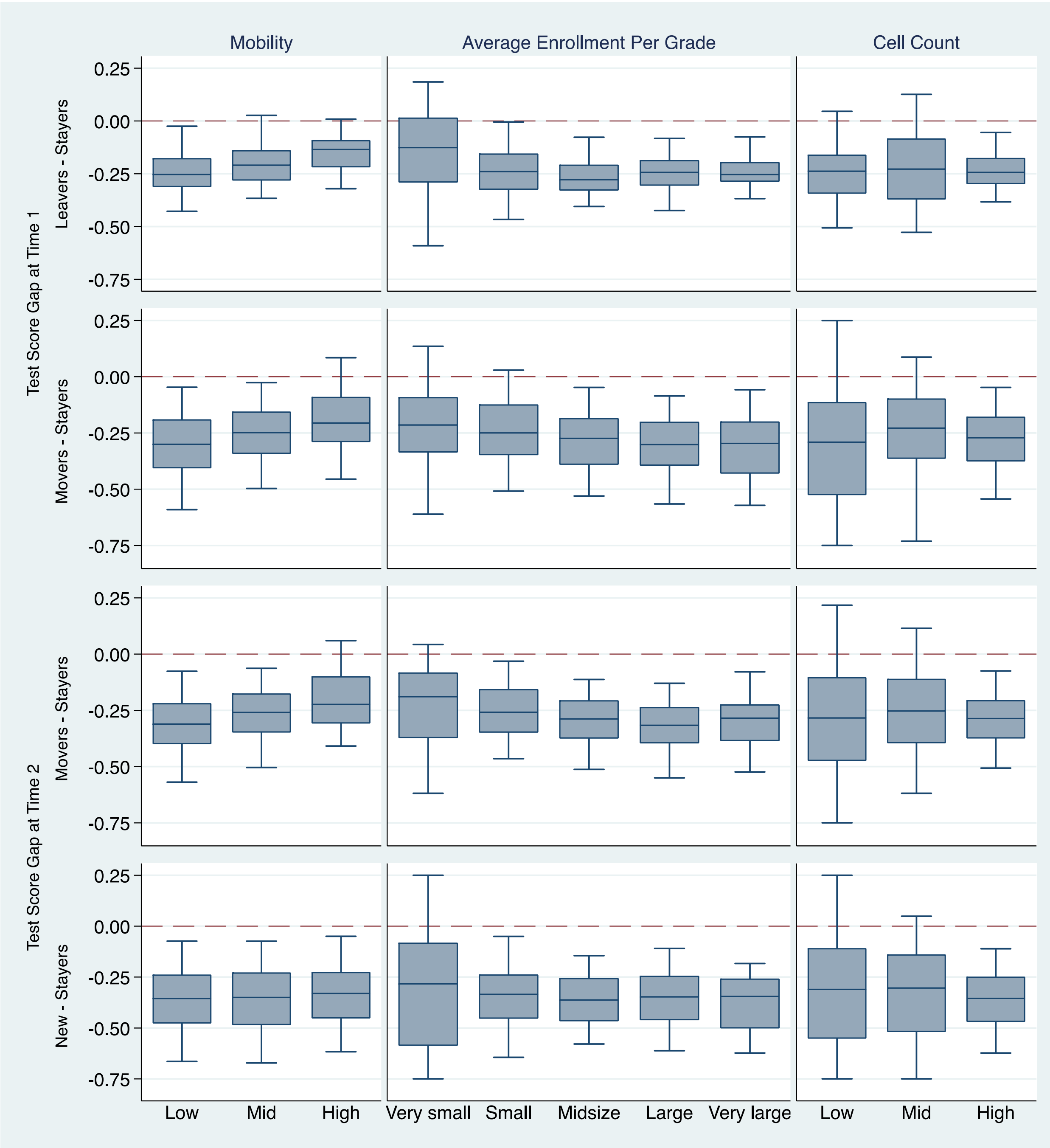
Figure 4. Distribution of district-level cohort-longitudinal growth discrepancies by category



**Note: boxes represent IQR, whiskers represent the 5th and 95th percentiles*.

We also examine the gaps in test scores among mobile students (leavers, movers, or new) and non-mobile students (stayers), as these differences are also related to the discrepancies between cohort and longitudinal growth estimates. Gaps for movers and leavers, relative to stayers, tend to be smaller in high-mobility districts, while gaps for new-to-system students are approximately the same across mobility categories. All of these gaps have wider distributions among very small districts and districts with low cell counts; this is likely why discrepancies are larger in these districts even when mobility is low.

Figure 5. Distribution of district-level test score gaps by category.



**Note: boxes represent IQR, whiskers represent the 5th and 95th percentiles.*

<u>*4.1.c State-by-state differences*</u>:

As demonstrated in **Table 3**, the distributions of districts across categories differ between the three states. Michigan districts tend to be moderately small while Tennessee districts tend to be large. Massachusetts districts are more likely and Tennessee districts are less likely to be missing one or more grade-year-subject cells. The proportion of low-mobility districts is highest in Massachusetts, while the proportion of mid-mobility districts is highest in Tennessee, and the proportion of high-mobility districts is highest in Michigan.

Table 8 provides additional insight into differences in mobility across states. In all three states, the ratio of students entering the state data system for the first time (r_n) is lower than that of any other mobility group. The new-to-system ratio is somewhat higher on average in Tennessee than either other state (implying that there is a larger proportion of students for whom growth cannot be observed). However, Tennessee also has the smallest test score gaps between mobility groups.

Table 8. Distribution of district-level mobility ratios, test score gaps, and growth estimates by state.

	MA	MI	TN
Overall growth estimates			
stayer growth ($\bar{\Delta}_s$)	0.008 (0.067)	-0.007 (0.042)	0.008 (0.036)
cohort growth (CG)	0.006 (0.065)	-0.004 (0.041)	0.010 (0.038)
longitudinal growth (LG)	0.010 (0.067)	-0.008 (0.039)	0.007 (0.038)
discrepancy (CG-LG)	-0.004 (0.032)	0.003 (0.023)	0.003 (0.019)
Mobility groups			
leaver ratio	0.068 (0.031)	0.119 (0.057)	0.106 (0.034)
mover ratio	0.036 (0.025)	0.075 (0.040)	0.056 (0.027)
new-to-system ratio	0.034 (0.020)	0.039 (0.016)	0.046 (0.020)
Overall test score gaps			
leaver-stayer at t_1	-0.238 (0.143)	-0.320 (0.125)	-0.226 (0.102)
mover-stayer at t_1	-0.311 (0.197)	-0.258 (0.159)	-0.252 (0.150)
mover-stayer at t_2	-0.329 (0.176)	-0.268 (0.144)	-0.265 (0.106)
new-stayer at t_2	-0.390 (0.197)	-0.360 (0.204)	-0.239 (0.110)

Note: Standard deviations in parentheses.

Despite these differences, we generally find similar patterns in growth measures across states. There are a few slight differences in their distributions: all measures tend to be slightly positive in Massachusetts and Tennessee and slightly negative in Michigan, discrepancies are slightly negative in Massachusetts and slightly positive in Michigan and Tennessee, and standard deviations are highest in Massachusetts.

We observed similar relationships between cohort and longitudinal growth measures within categories of districts across the three states, as shown in **Table 9**. However, there are a few minor differences. Mean discrepancies are slightly negative across nearly all categories in Massachusetts and slightly positive across nearly all categories in the other two states. This suggests that cohort growth measures slightly understate longitudinal growth in Massachusetts and slightly overstate it in the other two states. Furthermore, the mean absolute discrepancies and RMSEs are larger for Massachusetts than the other two states, which makes sense considering that the standard deviations of the growth measures are larger in Massachusetts and the state tends to have smaller districts.

Within-state patterns in correlations and discrepancies across categories of districts are similar to the patterns observed across the entire validation sample (shown in **Table 7**). For instance, correlations are higher and mean absolute discrepancies/RMSEs are lower for low-mobility districts than for mid-mobility districts. Similarly, correlations are lower and mean absolute discrepancies/RMSEs are higher for very small districts than for districts in any other size category in Massachusetts and Michigan. There are only four Tennessee districts in the very small category, so although the correlation is very high for these districts, there are too few observations to make any meaningful inferences about this group.

Table 9. District-level results by state

	CG/LG Correlation*			Mean CG-LG Discrepancy			Mean Absolute Discrepancy			Root Mean Square Error		
	MA	MI	TN	MA	MI	TN	MA	MI	TN	MA	MI	TN
All districts	0.882	0.840	0.875	-0.004	0.003	0.003	0.021	0.017	0.015	0.032	0.023	0.019
Math	0.874	0.862	0.890	-0.004	0.002	0.001	0.024	0.021	0.019	0.039	0.028	0.025
ELA	0.855	0.810	0.857	-0.004	0.004	0.004	0.023	0.017	0.014	0.037	0.024	0.017
Mobility												
Low (<10%)	0.895	0.848	0.898	-0.003	0.002	0.005	0.019	0.014	0.012	0.030	0.019	0.015
Mid (10-15%)	0.785	0.836	0.881	-0.012	0.006	0.002	0.031	0.017	0.017	0.045	0.022	0.021
High (>15%)	[0.924]	0.821	[0.763]	-0.012	-0.001	-0.004	0.019	0.023	0.025	0.025	0.032	0.028
Enrollment per grade												
Very small (<40)	0.843	0.770	[0.989]	0.002	0.006	0.014	0.050	0.024	0.029	0.070	0.035	0.031
Small (40-99)	0.865	0.855	[0.940]	-0.004	0.009	0.008	0.024	0.017	0.015	0.031	0.022	0.021
Medium (100-199)	0.941	0.892	0.847	0.000	0.003	0.004	0.016	0.017	0.019	0.024	0.022	0.024
Large (200-399)	0.908	0.865	0.798	-0.007	-0.002	0.006	0.014	0.013	0.014	0.018	0.019	0.016
Very large (≥400)	0.904	0.852	0.909	-0.011	-0.005	-0.004	0.016	0.009	0.010	0.021	0.017	0.015
Cell count												
Low (≤40)	0.943	[0.651]	----	-0.024	-0.008	----	0.042	0.060	----	0.060	0.073	----
Mid (41-80)	[0.819]	0.881	[0.939]	0.010	0.006	-0.015	0.033	0.023	0.023	0.051	0.029	0.026
High (81-84)	0.876	0.858	0.869	-0.005	0.003	0.004	0.015	0.015	0.015	0.019	0.019	0.019

**Correlations for groups with fewer than 20 districts are shown in brackets.*

4.1.d Differences across subjects

Patterns in subject-specific growth measures are relatively similar to each other and consistent with those in the overall growth measures. The correlations between subject-specific cohort and longitudinal growth estimates are high, as seen in Table 10. As seen in the overall patterns, we see lower correlations, higher mean absolute discrepancies/RMSEs, and more variation in both math and ELA in high mobility and small grade size districts.

Table 10. District results by subject and category

	CG/LG Correlation		Mean CG-LG Discrepancy		Mean Absolute Discrepancy		Root Mean Square Error	
	Math	ELA	Math	ELA	Math	ELA	Math	ELA
Mobility								
Low (<10%)	0.884	0.863	-0.001	-0.001	0.020	0.016	0.029	0.025
Mid (10-15%)	0.879	0.775	0.003	0.005	0.021	0.019	0.029	0.028
High (>15%)	0.771	0.861	-0.001	-0.002	0.029	0.026	0.047	0.038
Enrollment per grade								
Very small (<40)	0.793	0.803	0.005	0.003	0.043	0.037	0.066	0.057
Small (40-99)	0.885	0.857	0.005	0.007	0.023	0.019	0.029	0.025
Medium (100-199)	0.920	0.891	0.001	0.003	0.019	0.018	0.026	0.023
Large (200-399)	0.903	0.858	-0.003	-0.002	0.016	0.014	0.020	0.019
Very large (≥400)	0.895	0.862	-0.007	-0.006	0.016	0.011	0.021	0.017
Cell count								
Low (<40)	0.889	0.962	-0.019	-0.011	0.055	0.047	0.076	0.056
Mid (41-80)	0.839	0.833	0.005	0.003	0.035	0.032	0.059	0.055
High (81-84)	0.883	0.867	0.000	0.002	0.018	0.015	0.023	0.019

5.1 School Level Analysis

We now examine the relationship between cohort and longitudinal growth measures at the school level. The average district in our sample has 5.9 elementary and middle schools. Given that there is more mobility across schools than there is across school districts and that schools are necessarily smaller (on average) than districts, cohort growth measures do not replicate longitudinal growth estimates as well for schools as they do for districts. Nonetheless, we find strong correlations (r=0.81) between cohort and longitudinal growth measures, suggesting that cohort measures rank schools similarly to longitudinal measures, on average. These correlations are almost the same magnitude as the district-level correlations. However, we see that in the average school, discrepancies tend to be larger in magnitude, suggesting that cohort measures do a worse job of estimating an individual school's longitudinal growth than they do for districts. We structure this section analogously to the district results analysis.

5.1.a Measures of growth, discrepancy, mobility and mobility score gaps

To provide context for our findings, we present comparisons of the different growth measures and discrepancies between cohort and longitudinal growth estimates in the top panel of Table 11. Descriptively, cohort and longitudinal measures are quite similar. On average, both measures are slightly greater than zero. Again, because growth is measured in standard deviations of the state test scores, an estimate of zero for a school does not imply that their students did not learn, but rather that these students' relative rankings within the state stayed the same over time. We should note that there is more variation in growth across schools than across districts; the standard deviation of our longitudinal growth estimates is nearly twice as large for schools (0.09) as it was for districts (0.05).

The discrepancy between the cohort and longitudinal growth estimates are close to zero on average. The variance of the discrepancy is approximately 40% as large as the variance of longitudinal growth (the standard deviation is about 60% as large).

The bottom two panels of Table 11 present information about aspects of student mobility related to discrepancies between the different growth measures. As discussed in the district-level section, approximately 4% of observations in our three-state sample are "new-to-system" and therefore were not observed at t_1. This suggests that longitudinal growth measures may be slightly biased because they cannot capture the growth of the "new-to-system" students in a school. Cohort growth measures are also susceptible to bias when some students are tested in a school at either t_1 or t_2, but not both. We find that approximately 16% of students leave their school in a given year and 15% enter it (compared to 10% for districts), suggesting a greater scope for bias in cohort growth measures than longitudinal measures. These proportions are significantly larger than those for districts, as the school-level growth measures are also affected by within-district mobility.

Cohort and longitudinal growth estimates tend to be slightly smaller than the average growth across stayers (top panel), suggesting that the growth rates of mobile students are lower, on average, than those of stayers. The gaps in mean test scores shown in the bottom panel highlight that the achievement levels of leavers, movers, and new-to-system students are generally lower than those of stayers as well. On average, the students who leave a school are quite similar in test-score levels to the students who enter from a different school in the state. Students who enter from outside of the state system tend to have the lowest test scores.

Table 11. Distribution of school-level growth estimates, mobility ratios, and test score gaps.

	Math	ELA	Overall
Growth estimates			
stayer growth ($\overline{\Delta}_s$)	0.014 (0.125)	0.012 (0.092)	0.013 (0.096)
cohort growth (CG)	0.010 (0.122)	0.009 (0.091)	0.009 (0.094)
longitudinal growth (LG)	0.010 (0.120)	0.008 (0.088)	0.009 (0.091)
discrepancy (CG-LG)	-0.000 (0.067)	0.001 (0.057)	0.000 (0.057)
Mobility groups			
leaver ratio	0.162 (0.106)	0.162 (0.106)	0.162 (0.106)
mover ratio	0.113 (0.088)	0.113 (0.088)	0.113 (0.088)
new-to-system ratio	0.043 (0.027)	0.044 (0.027)	0.044 (0.026)
Test score gaps			
leaver-stayer at t_1	-0.244 (0.203)	-0.235 (0.202)	-0.240 (0.192)
mover-stayer at t_1	-0.257 (0.297)	-0.237 (0.274)	-0.246 (0.268)
mover-stayer at t_2	-0.271 (0.257)	-0.239 (0.250)	-0.254 (0.239)
new-stayer at t_2	-0.320 (0.307)	-0.307 (0.337)	-0.316 (0.294)

Note: Standard deviations in parentheses.

5.1.b: School-level Results

School-level growth estimates are generally quite consistent across cohort and longitudinal measures. As is shown in Table 12 (below), the correlation between the two measures is lower for schools than districts, but still very strong (r=0.81 overall, indicating that the cohort growth measure explains 65% of the variation in the longitudinal growth measure). This suggests that the two measures rank schools similarly. For example, 73% of schools ranked in the top quartile on longitudinal growth also rank in the top quartile on cohort growth. In addition, we see that the average discrepancy (row 2) between the two measures is very close to zero (mean=0.000), suggesting that in the average school, the cohort growth measure will provide an unbiased estimate of longitudinal growth.

Table 12. Comparison of school and district-level CG and LG estimates.

	Schools			Districts		
	Math	**ELA**	**Overall**	**Math**	**ELA**	**Overall**
CG/LG Correlation	0.848	0.797	0.808	0.869	0.841	0.866
Mean CG-LG Discrepancy	-0.000	0.001	0.000	0.000	0.001	0.001
Mean Absolute Discrepancy	0.041	0.037	0.036	0.021	0.018	0.018
Root Mean Square Error	0.067	0.057	0.057	0.032	0.028	0.026
Standard Deviation of LG	0.120	0.088	0.091	0.063	0.051	0.050

However, just because these estimates are right "on average" does not mean that they are reliable enough to make judgements about growth in individual schools. Because the discrepancy can be either positive or negative, we also examine the mean absolute value of the difference between the two measures. We find that, in the average school, the cohort measure differs from the longitudinal growth measure by 0.036 (row 3), which is twice as large as the discrepancy in the average district. NAEP data suggest the average student gains 0.33 SD per year from grade 4-8 on vertically equated tests, so the mean absolute discrepancy we observe is about +/-10% of a year's growth.

We represent visually the high level of consistency between cohort growth (CG) and longitudinal growth (LG) in **Figure 6**, where we plot school-level CG estimates against LG estimates. The CG and LG estimates for the majority of schools fall along the 45-degree line. In **Figure 7**, we show the CG-LG discrepancy against the LG measure. As with the district-level analysis, most school-level CG-LG discrepancies fall near the horizontal line, which is drawn at the value of CG-LG=0. However, both figures show a few outlier schools in which the two measures of growth differ substantially. One reasonable metric for assessing how well the CG measure compares to the LG measure is to note that we want the error variance to be no more than 25% of the true variance, which corresponds to a reliability of 0.8. This corresponds to a discrepancy of +/- 0.045. In Figure 2, we include horizontal lines at +/-0.045, showing that roughly 75% of schools fall within this range.

Figure 6. Cohort vs Longitudinal School-Level Growth Estimates

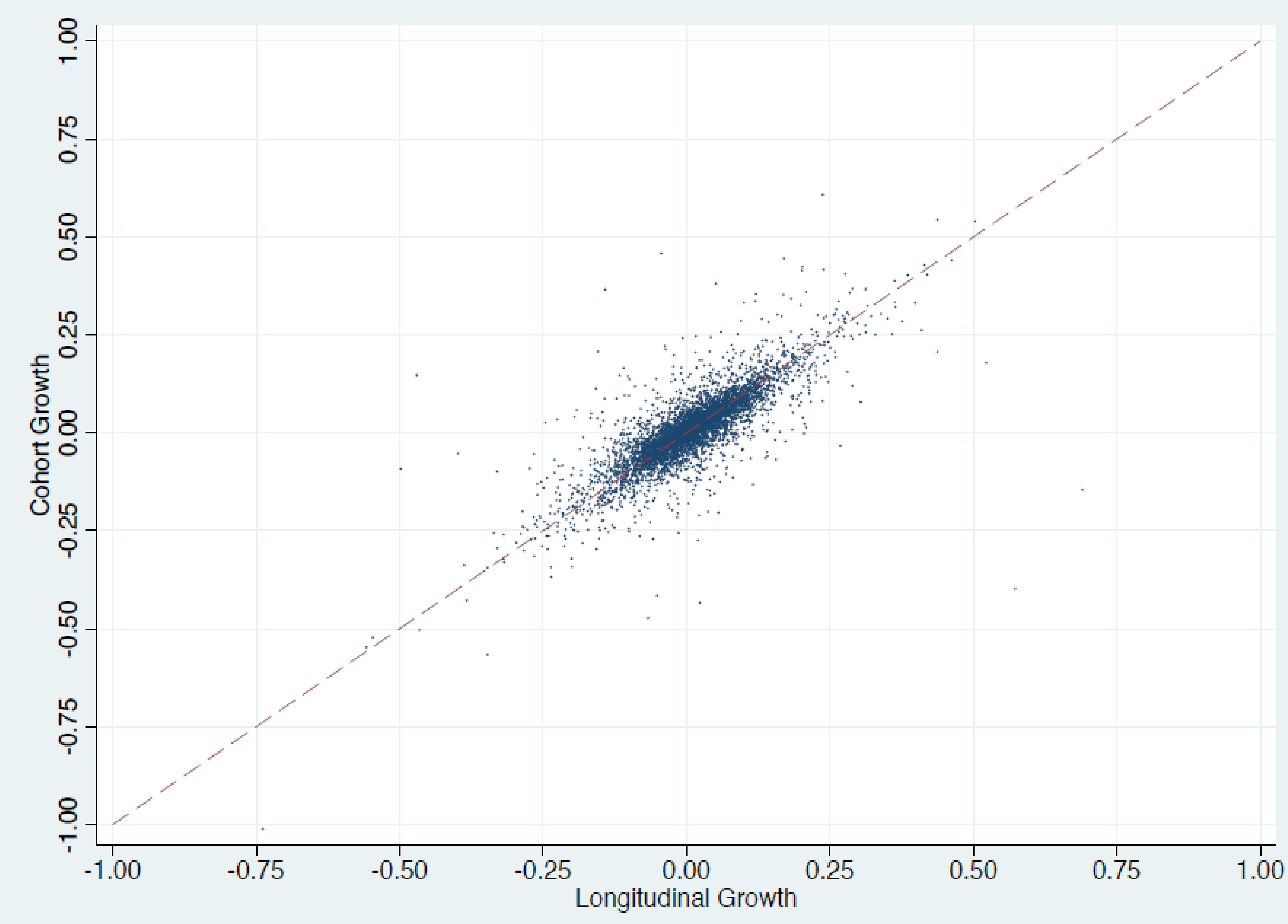
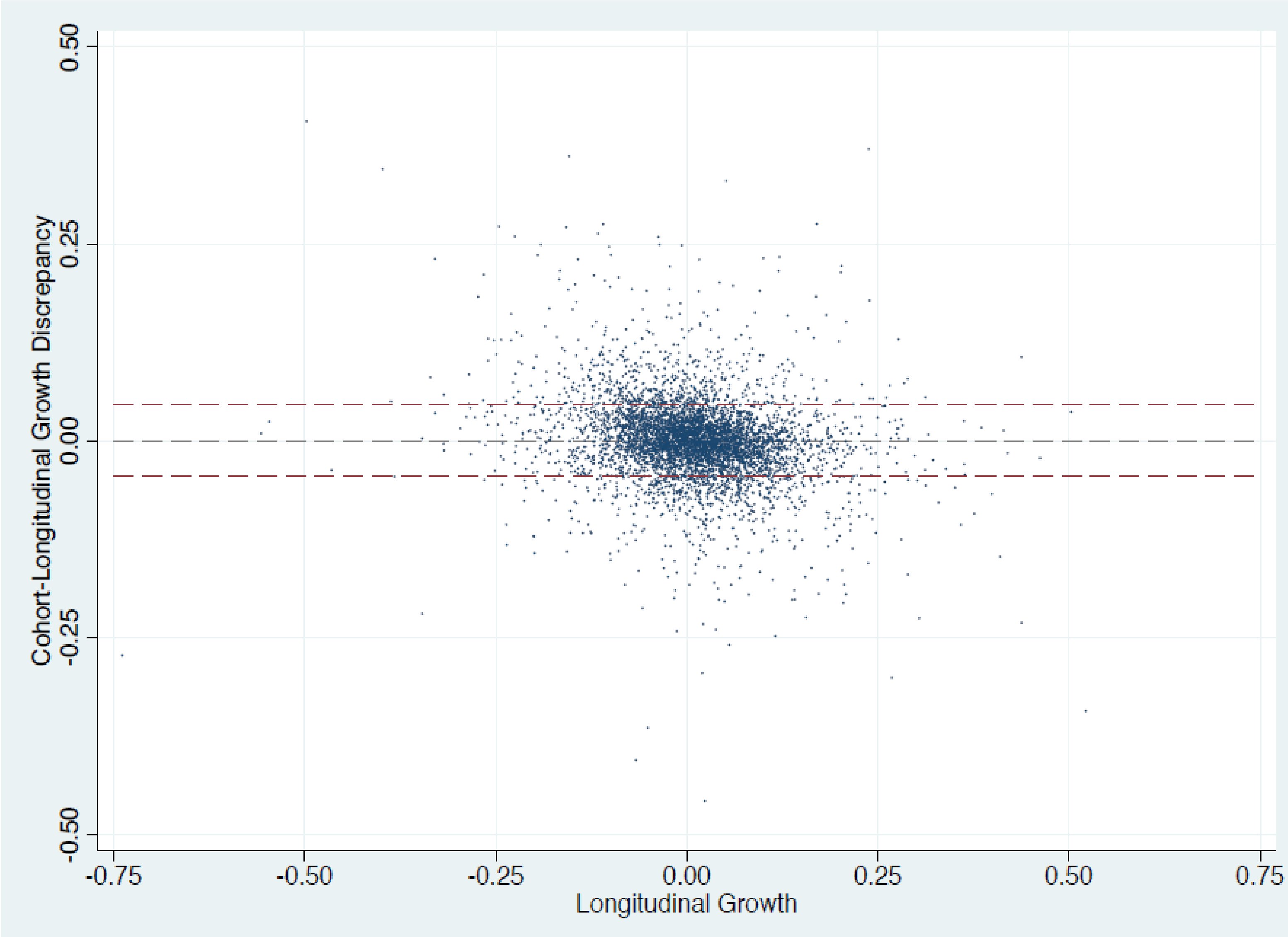


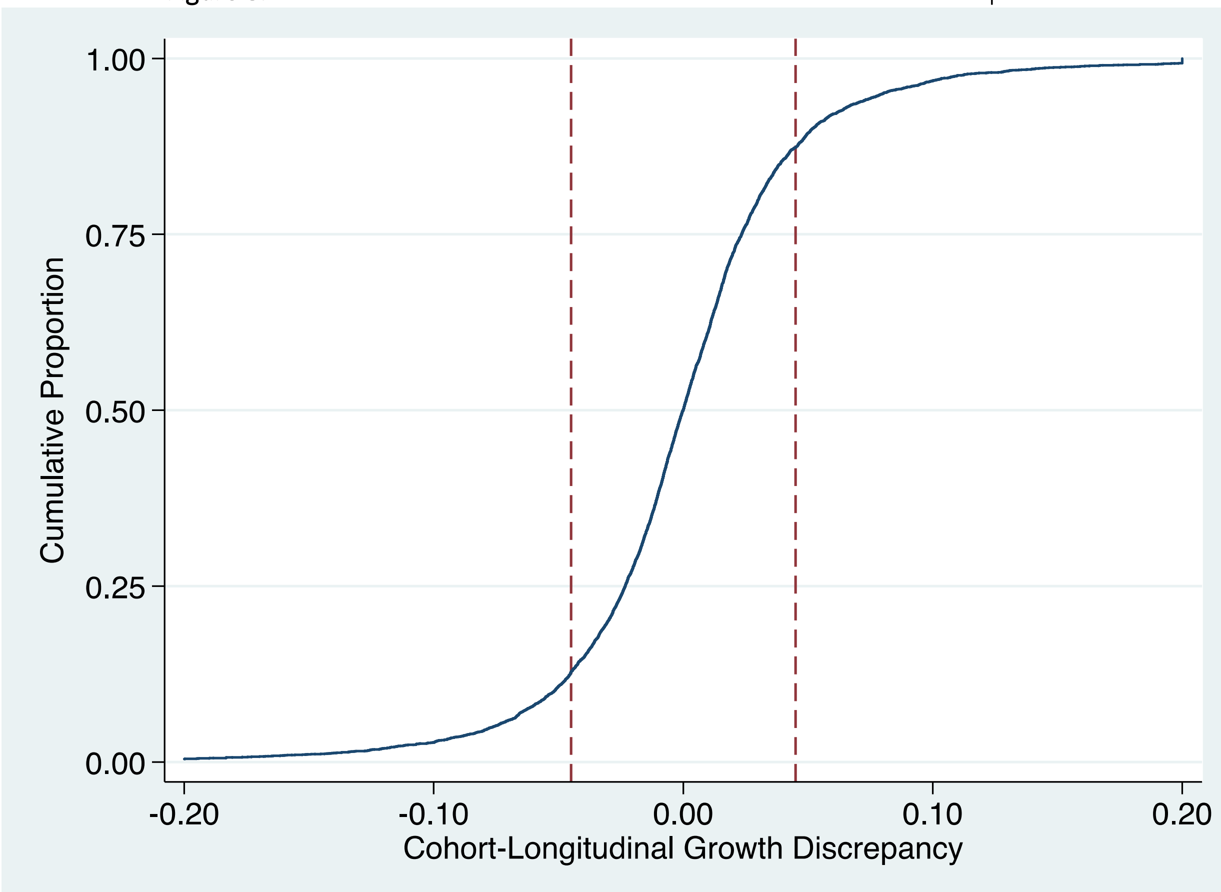
Figure 7. School-Level Cohort-Longitudinal Growth Discrepancies vs Longitudinal Growth



**Five extreme outliers were omitted to improve readability.*

These figures show the strong relationship between cohort and longitudinal growth measures and document the magnitudes of the discrepancies across schools. To illustrate these magnitudes more clearly, we present the cumulative distribution function (CDF) of the discrepancy in **Figure 8**. Here, we see that 13% of schools have discrepancies below -0.045, while 13% have discrepancies above 0.045.

Figure 8. Cumulative distribution function of school-level CG-LG discrepancies.



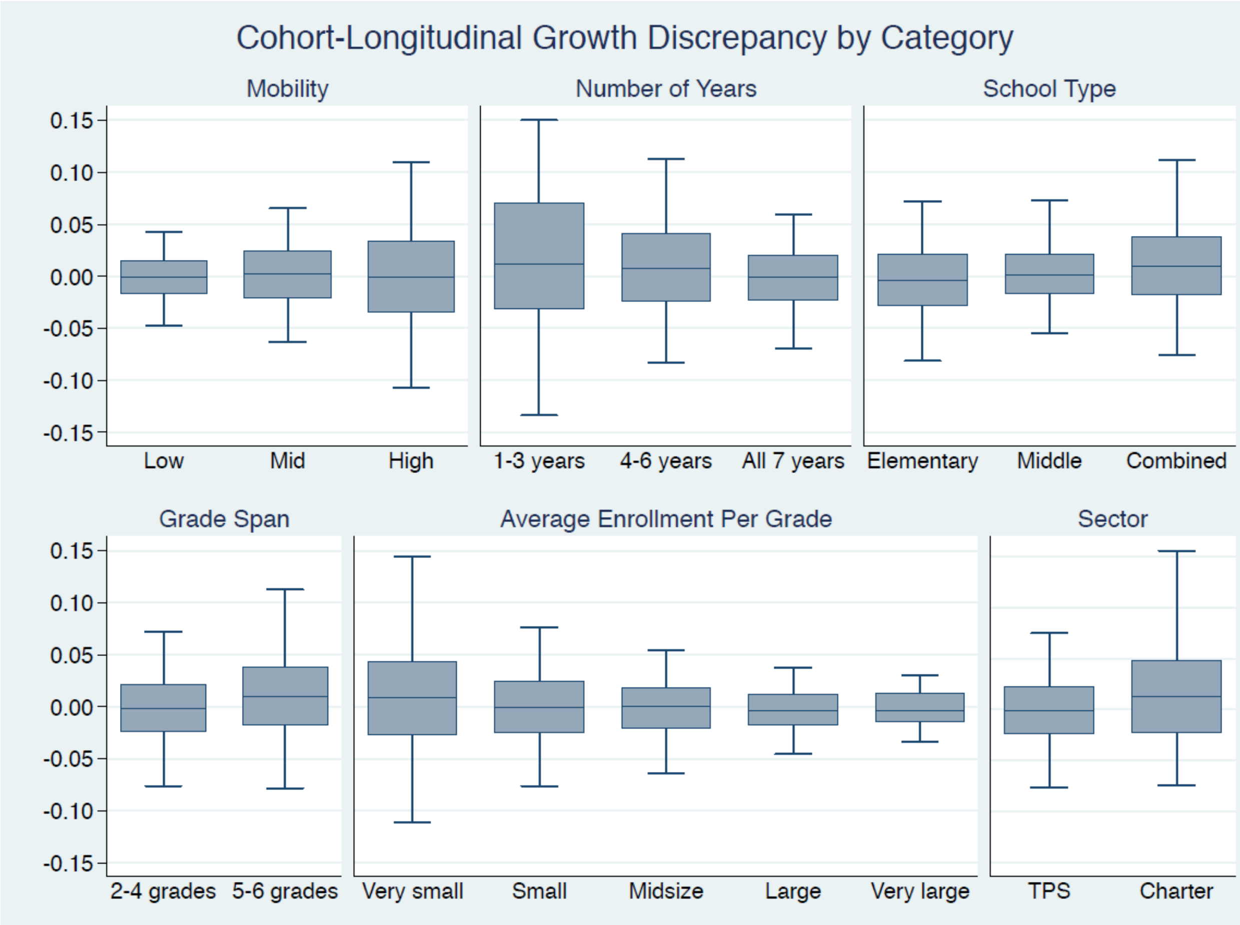
We expect mobility to be a primary driver of the CG-LG discrepancy; as such, we look at the relationship between the discrepancy and mobility as well as proxies that are readily available in the SEDA data such as average grade enrollment, and the number of years and grades that contribute to a school's growth estimate. We also examine difference by sector (charter or traditional public school). **Table 13** shows the CG-LG correlations, mean absolute deviations, RMSEs, and mean discrepancies for each category of districts. Smaller sizes, higher mobility, and fewer years of data are associated with slightly lower correlations and considerably larger mean absolute deviations and RMSEs. In other words, discrepancies in growth measures tend to be larger in magnitude in schools that are smaller, have greater mobility rates, and are in the data only for a few years; this is most evident in very small schools. In schools that span more tested grade levels, there is more opportunity for mobility; as a result, we see lower correlations and larger RMSEs in these schools.

Table 13. Comparison of school-level CG and LG estimates by category

	CG/LG Correlation	Mean CG-LG Discrepancy	Mean Absolute Discrepancy	Root Mean Square Error
Mobility				
Low (<10%)	0.927	-0.001	0.021	0.031
Mid (10-15%)	0.846	0.001	0.030	0.046
High (>15%)	0.742	0.001	0.050	0.076
Average enrollment per grade				
Very small (<40)	0.695	0.009	0.058	0.096
Small (40-99)	0.846	-0.001	0.035	0.051
Medium (100-199)	0.868	-0.002	0.028	0.041
Large (200-399)	0.875	-0.004	0.020	0.032
Very large (≥400)	0.941	-0.003	0.016	0.019
School Type				
Elementary	0.850	-0.004	0.036	0.054
Middle	0.758	0.004	0.031	0.058
Combined	0.684	0.012	0.044	0.068
Grade Span				
2-4 grades	0.827	-0.002	0.034	0.055
5-6 grades	0.683	0.012	0.044	0.068
Number of years				
1-3	0.776	0.017	0.075	0.112
4-6	0.772	0.012	0.048	0.071
All 7	0.832	-0.003	0.031	0.047
Sector				
TPS	0.809	-0.001	0.035	0.055
Charter	0.814	0.018	0.052	0.079

Finally, we see interesting differences by sector. Here, the correlation between cohort and longitudinal growth is quite similar across sectors, but there is substantially more variability in discrepancies in charter schools. In the average charter school, the absolute CG-LG discrepancy is nearly 50% larger than in the average traditional public school (mean absolute discrepancy = 0.052 vs. 0.035). Finally, we see that the mean discrepancy is positive and rather large in charter schools; this suggests that, in the average charter school, cohort growth tends to overstate longitudinal growth by 0.018 standard deviations. This is not a trivial amount; it is one-fifth of a standard deviation of longitudinal growth (0.09; Table 11). Figure 9 illustrates these findings, presenting box-and-whiskers plots that show the median, interquartile range, 5^{th} percentile, and 95^{th} percentile of cohort-longitudinal growth discrepancies by school size, mobility, school type, grade span, number of years of data available, and sector.

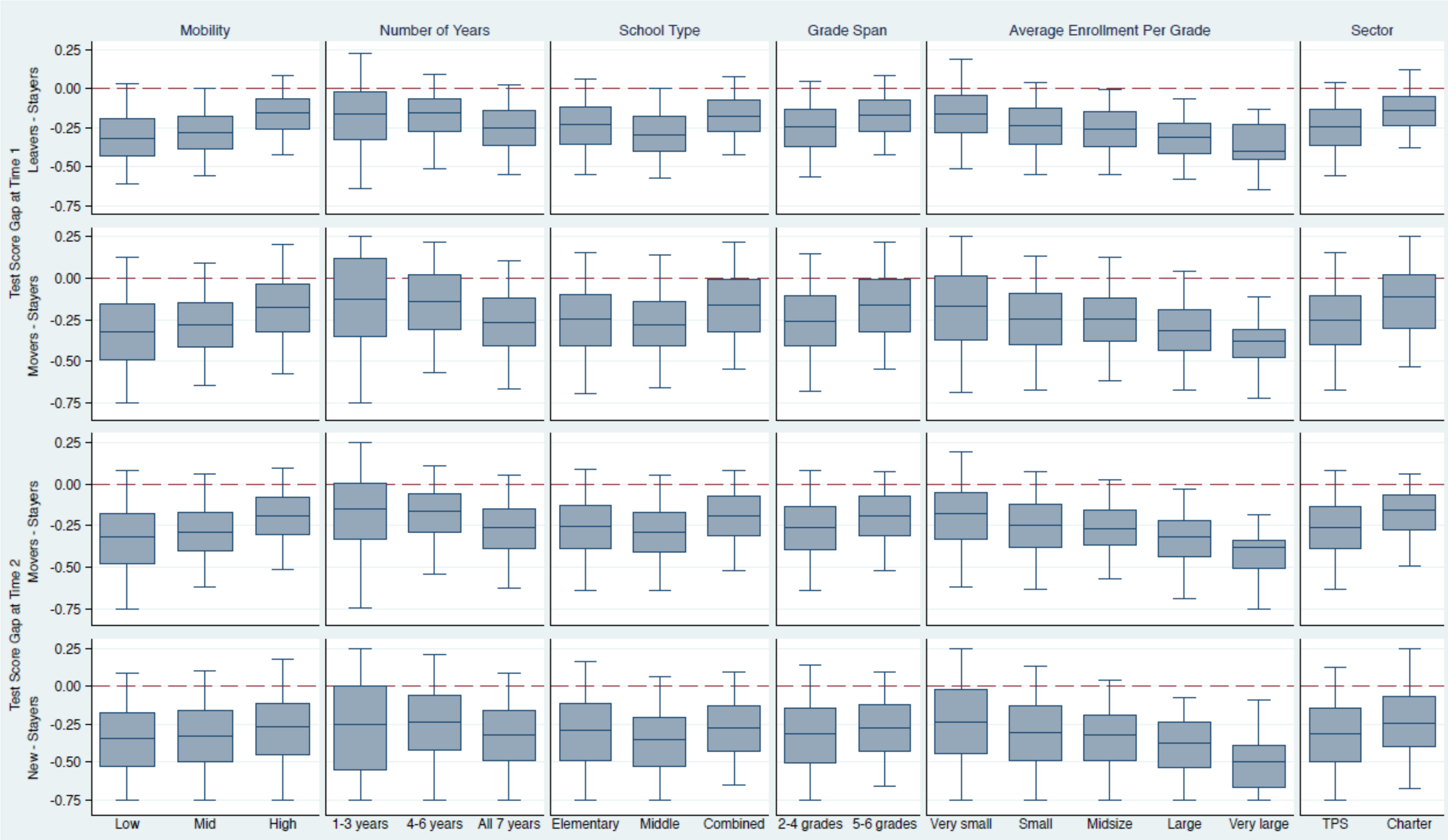
Figure 9. Distribution of school-level cohort-longitudinal growth discrepancies by category



**Note: boxes represent IQR, whiskers represent the 5th and 95th percentiles.*

We also examine the gaps in average test scores among mobile students (leavers, movers, or new) and non-mobile students (stayers), as these differences are also related to the discrepancies between cohort and longitudinal growth estimates. Gaps between the mean test scores of leavers, movers, and new-to-system students and the mean test scores of stayers are closer to zero, on average, in charter schools and schools with fewer students per grade, long grade spans, high mobility, or fewer years of data. However, very small schools and schools with fewer years of data also have the widest distributions of test score gaps. This suggests that gaps tend to be larger in magnitude, but less uniform in direction, for schools in these categories, and helps illustrate why we see larger discrepancies in these types of schools.

Figure 10. Distribution of school-level test score gaps by category.



**Note: boxes represent IQR, whiskers represent the 5th and 95th percentiles.*

5.1.c. Sector differences:

As noted above, we see important differences across charter and traditional public schools. Specifically, cohort growth measures appear to overstate longitudinal measures somewhat substantially in charter schools (but not in traditional public schools), and there is much more variability in these discrepancies in charter schools. **Table 14** documents the potential sources of these differences, replicating Table 11 for charter and traditional public schools. Recall from equation (10) that the discrepancy between cohort and longitudinal growth measures depends on the differences in test scores for stayers, movers, and leavers and the proportion of mobile students in the district.

Table 14. Distribution of school-level growth estimates, mobility ratios, and test score gaps by sector.

	TPS	Charter	Overall
Mobility groups			
leaver ratio	0.154 (0.100)	0.272 (0.125)	0.162 (0.106)
mover ratio	0.106 (0.081)	0.203 (0.121)	0.113 (0.088)
new-to-system ratio	0.043 (0.026)	0.051 (0.033)	0.044 (0.026)
Test score gaps			
leaver-stayer at t_1	-0.247 (0.191)	-0.137 (0.160)	-0.240 (0.192)
mover-stayer at t_1	-0.255 (0.268)	-0.132 (0.253)	-0.246 (0.268)
mover-stayer at t_2	-0.261 (0.241)	-0.172 (0.177)	-0.254 (0.239)
new-stayer at t_2	-0.323 (0.292)	-0.233 (0.309)	-0.316 (0.294)

Note: Standard deviations in parentheses.

Interestingly, we see much smaller gaps between mobile students and stayers in charter schools than traditional public schools. All else equal, this suggests that the discrepancies should be smaller. However, we see much higher rates of mobility in charters. In particular, the fact that approximately 27% of charter school students leave the school (compared to just 15% in traditional public schools) and that these students have test scores substantially lower than those of stayers drives much of the discrepancy.

In **Table 15**, we examine whether observable differences between charter and traditional public schools explain these patterns. In Model 1, we show the raw difference in both the mean discrepancy (top panel) and the mean absolute discrepancy (bottom panel) by sector. In Model 2A {NOTE: THIS IS M3 BELOW}, we control for the predictors listed in Table 13: grade size, school type, grade span, and number of years, along with interactions between these variables, while in Model 3A, we add a set of predictors from the

Common Core of Data, such as school demographics and urbanicity. We also add mobility rate to the regressions in Model 2B and 3B; while the other predictors in these models are variables that would be available to analysts using SEDA data, school-level mobility rates would not necessarily be.

This table reveals several lessons. First, these controls account for some of the difference in mean discrepancy by sector, but even controlling for our full set of covariates cohort growth in charter schools continues to overstate longitudinal growth. Second, we see that these covariates do not explain much of the variation in mean discrepancies – controlling for our full set of predictors only explains 9.4% of the variation in mean discrepancies. Third, we see that the covariates are stronger predictors of the mean absolute discrepancy; the predictors in Table 13 explain nearly 25% of the variation in this measure. Thus, we can do a better job of explaining the variability in these estimates than we can in explaining whether they will be biased. Fourth, mobility is an important predictor of both mean discrepancy and mean absolute discrepancy, even after we control for the other predictors in Table 13. Thus, our proxies for mobility are not perfect and do not capture all of the important variation in this predictor.

Table 15. Differences in mean CG-LG discrepancy in charter and traditional public schools, uncontrolled and with SEDA and CCD controls.

	Model 1	Model 2		Model 3	
		A	B	A	B
Charter	0.020***	0.007*	0.004	0.012***	0.010**
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Mobility			0.034***		0.098***
			(0.009)		(0.012)
R-sq	0.008	0.069	0.071	0.081	0.094
Controls from Table 13	X		X	X	X
CCD Data				X	X
Mobility			X		X

5.1.d State-by-state differences:

Table 16 describes the distributions of growth estimates, mobility rates, and test score gaps among schools in each state. We see similar patterns at the school level as we do at the district level. The mean growth across stayers, mean cohort growth, and mean longitudinal growth estimates are all approximately the same in Tennessee, resulting in a mean discrepancy close to zero. Cohort growth has a lower mean than the other two measures in Massachusetts, corresponding to a slightly negative mean discrepancy, while longitudinal growth has a lower mean than the other two measures in Michigan, corresponding to a slightly positive mean discrepancy. This suggests that cohort growth measures slightly understate longitudinal growth in Massachusetts and slightly overstate it in Michigan.

In all three states, the ratio of students entering the state data system for the first time (r_n) is lower than that of any other mobility group. The new-to-system ratio is somewhat higher on average in Tennessee than either other state (implying that there is a larger proportion of students for whom growth cannot be observed). However, Tennessee also has the smallest test score gaps between mobility groups.

Table 16. Distribution of school-level growth estimates, mobility ratios, and test score gaps by state.

	MA	MI	TN
Growth estimates			
stayer growth ($\overline{\Delta}_s$)	0.029 (0.109)	0.007 (0.084)	0.008 (0.100)
cohort growth (CG)	0.015 (0.107)	0.007 (0.084)	0.008 (0.098)
longitudinal growth (LG)	0.027 (0.101)	-0.000 (0.081)	0.008 (0.094)
discrepancy (CG-LG)	-0.012 (0.062)	0.007 (0.053)	0.001 (0.058)
Mobility groups			
leaver ratio	0.110 (0.080)	0.177 (0.115)	0.185 (0.092)
mover ratio	0.071 (0.071)	0.128 (0.093)	0.127 (0.079)
new-to-system ratio	0.039 (0.030)	0.042 (0.024)	0.050 (0.025)
Test score gaps			
leaver-stayer at t_1	-0.224 (0.225)	-0.269 (0.186)	-0.202 (0.153)
mover-stayer at t_1	-0.299 (0.354)	-0.231 (0.237)	-0.236 (0.243)
mover-stayer at t_2	-0.298 (0.328)	-0.245 (0.212)	-0.239 (0.196)
new-stayer at t_2	-0.375 (0.302)	-0.345 (0.297)	-0.207 (0.251)

Note: Standard deviations in parentheses.

Table 17 shows relationships between cohort and longitudinal growth within categories of schools by state. While most patterns are consistent across states, there are a few minor differences. Mean discrepancies are slightly negative across nearly all categories in Massachusetts and slightly positive in Michigan, while the sign of Tennessee's discrepancy varies by category. Across all three states, correlations are higher and mean absolute discrepancies/RMSEs are lower for low-mobility schools than for mid-mobility schools. Similarly, correlations are lowest and mean absolute discrepancies/RMSEs are highest for very small schools.

Table 17. School-level results by state

	CG/LG Correlation*			Mean CG-LG Discrepancy			Mean Absolute Discrepancy			Root Mean Square Error		
	MA	MI	TN	MA	MI	TN	MA	MI	TN	MA	MI	TN
All Schools	0.825	0.794	0.816	-0.012	0.007	0.001	0.036	0.036	0.035	0.063	0.054	0.058
Math	0.854	0.833	0.865	-0.012	0.005	0.001	0.038	0.042	0.042	0.066	0.064	0.072
ELA	0.845	0.776	0.749	-0.012	0.008	0.000	0.039	0.037	0.034	0.064	0.054	0.054
Mobility												
Low (<10%)	0.938	0.909	0.930	-0.006	0.002	0.008	0.022	0.021	0.022	0.032	0.029	0.033
Mid (10-15%)	0.887	0.861	0.811	-0.012	0.006	0.001	0.034	0.031	0.028	0.046	0.041	0.052
High (>15%)	0.703	0.727	0.809	-0.027	0.011	-0.001	0.069	0.049	0.042	0.110	0.070	0.065
Enrollment per grade												
Very small (<40)	0.733	0.724	0.589	-0.010	0.016	0.013	0.066	0.056	0.054	0.115	0.081	0.104
Small (40-99)	0.863	0.814	0.880	-0.014	0.004	0.001	0.035	0.036	0.035	0.053	0.050	0.051
Medium (100-199)	0.896	0.854	0.865	-0.011	0.011	-0.006	0.027	0.027	0.029	0.041	0.041	0.041
Large (200-399)	0.868	0.900	0.886	-0.010	0.001	-0.006	0.019	0.019	0.023	0.043	0.026	0.030
Very large (≥400)	0.932	0.958	0.949	-0.014	0.001	-0.001	0.022	0.013	0.015	0.026	0.016	0.017
School Type												
Elementary	0.876	0.824	0.855	-0.013	0.001	-0.004	0.036	0.036	0.035	0.056	0.052	0.057
Middle	0.673	0.839	0.770	-0.009	0.011	0.004	0.032	0.027	0.037	0.076	0.042	0.063
Combined	0.847	0.643	0.629	-0.017	0.024	0.011	0.043	0.051	0.034	0.066	0.074	0.056
Sector												
TPS	0.810	0.799	0.829	-0.013	0.006	-0.001	0.036	0.035	0.033	0.063	0.052	0.053
Charter	0.937	0.756	0.784	0.003	0.017	0.049	0.040	0.046	0.100	0.064	0.066	0.139
Grade Span												
2-4 grades	0.826	0.826	0.832	-0.012	0.004	-0.002	0.035	0.034	0.035	0.063	0.049	0.059
5-6 grades	0.846	0.643	0.628	-0.017	0.024	0.011	0.043	0.051	0.034	0.067	0.075	0.056
Number of Years												
1-3 years	0.779	0.772	0.803	-0.030	0.023	0.039	0.089	0.067	0.088	0.154	0.093	0.126
4-6 years	0.831	0.766	0.736	-0.008	0.019	0.009	0.049	0.047	0.050	0.083	0.063	0.079
All 7 years	0.845	0.816	0.844	-0.011	0.003	-0.003	0.032	0.030	0.030	0.051	0.044	0.047

5.1.e Differences across subjects

Patterns in relationships between cohort and longitudinal measures of subject-specific growth (shown in **Table 18**) are similar to the patterns in the overall growth measures and consistent across the two subjects. Correlations are slightly higher for math than ELA across nearly all categories of schools. However, because there is much more variation across schools in math longitudinal growth measures (see Table 11), the absolute discrepancies and RMSE's are also generally larger for math than ELA, despite the higher correlations.

Table 18. School results by subject and category

	CG/LG Correlation		Mean CG-LG Discrepancy		Mean Absolute Discrepancy		Root Mean Square Error	
	Math	ELA	Math	ELA	Math	ELA	Math	ELA
Mobility								
Low (<10%)	0.942	0.919	-0.003	0.000	0.025	0.023	0.036	0.033
Mid (10-15%)	0.876	0.822	0.000	0.003	0.036	0.031	0.056	0.046
High (>15%)	0.796	0.728	0.002	0.000	0.057	0.050	0.088	0.075
Enrollment per grade								
Very small (<40)	0.741	0.719	0.010	0.009	0.067	0.059	0.114	0.093
Small (40-99)	0.879	0.829	-0.002	0.000	0.041	0.035	0.059	0.050
Medium (100-199)	0.903	0.834	-0.002	-0.001	0.031	0.030	0.046	0.044
Large (200-399)	0.918	0.840	-0.004	-0.003	0.024	0.021	0.036	0.034
Very large (≥400)	0.975	0.912	-0.002	-0.004	0.017	0.018	0.021	0.024
School Type								
Elementary	0.879	0.830	-0.005	-0.003	0.042	0.037	0.064	0.055
Middle	0.848	0.757	0.003	0.005	0.035	0.033	0.061	0.058
Combined	0.664	0.703	0.014	0.010	0.052	0.042	0.086	0.064
Grade Span								
2-4 grades	0.871	0.811	-0.003	-0.001	0.039	0.036	0.063	0.056
5-6 grades	0.663	0.703	0.014	0.010	0.052	0.043	0.086	0.064
Number of Years								
1-3	0.838	0.782	0.018	0.021	0.085	0.076	0.127	0.107
4-6	0.759	0.801	0.013	0.009	0.056	0.048	0.087	0.070
All 7	0.868	0.807	-0.003	-0.002	0.035	0.031	0.055	0.048
Sector								
TPS	0.855	0.793	-0.002	0.000	0.040	0.036	0.063	0.056
Charter	0.818	0.816	0.024	0.013	0.065	0.049	0.102	0.072

Summary and Guidelines

Purpose of this Report:

- Stanford Education Data Archive (SEDA) estimates of cohort growth for a given school are intended to provide a measure of student learning between two time periods in cases where student-level data on achievement is unavailable.
- SEDA estimates are in theory identical to preferred longitudinal growth estimates of student-level growth in cases where the exact same students exist in the same unit (school or district) between two time periods.
- In practice, however, because students can transfer between schools, between district, or out of state, cohort and longitudinal estimates will diverge. This may be particularly true at the school level, where more students may transfer between schools within a district than between districts over time.
- Using student-level data from Massachusetts, Michigan and Tennessee, to create and compare SEDA-style cohort estimates and longitudinal growth estimates, this report provides guidance to both researchers and policymakers on the extent to which the implied differences between SEDA estimates and longitudinal growth estimates caused by student mobility render SEDA estimates inappropriate in certain cases.

Summary of Findings and Implications for SEDA Usage:

- In general, the SEDA-style cohort growth measures are strongly correlated with longitudinal growth: at the district-level, we find average correlations of r=0.87, and at the school-level of r=0.81. This suggests that in most cases, *researchers may use SEDA cohort growth measures to approximate longitudinal growth*.
- Correlations are weaker in districts or schools with high rates of mobility, which we define as more than 15% between districts or schools annually. In these cases, the district-level average correlations are r=0.84 and at the school-level r=0.74. This suggests that *researchers should use caution when approximating* longitudinal growth with SEDA data in districts or schools with more than 15% mobility.
- In this report's sample states of Massachusetts, Michigan and Tennessee, high mobility schools and districts tend to be small in size. Thus for the purposes of determining SEDA usage, *researchers may use the average number of students and grade-spans for schools and districts to approximate the conditions under which cohort growth and longitudinal growth diverge*.
- In schools or districts with more than 40 students per grade per year, cohort growth and longitudinal growth are highly correlated ($r>0.85$). This suggests that *researchers may use cohort growth to approximate longitudinal growth in districts or schools with more than 40 students per grade.*
- Schools with high grade-spans—5 or 6 tested grades, such as K-8 schools—have high rates of mobility and weaker correlations between cohort growth and longitudinal growth measures. Thus *researchers should not use cohort growth to approximate longitudinal growth for schools with more than 4 tested grades.*
- Correlations between cohort growth and longitudinal growth are similar in charter and traditional public schools and similar mobility or grade-span/size conditions. Thus researchers *may use cohort growth to approximate longitudinal growth to compare schools within the charter sector*.

- However, cohort growth measures are systematically larger than longitudinal growth in general, and that discrepancy is far larger for charter schools. Thus *researchers should not use cohort growth to compare longitudinal growth between charter and traditional public schools*.

Technically oriented readers may refer to the main body of this report and its appendices to further investigate these patterns and conditions.

Appendix

Table A-1. Effect of each mobility group on discrepancies and bias

		Discrepancy CG-LG	**CG bias** CG-target	**LG bias** LG-target
Leaver	t_1	$-r_l(\overline{y1}_l - \overline{y1}_s)$	$-r_l(\overline{y1}_l - \overline{y1}_s)$	0
Mover	t_1	$\frac{r_m}{1-r_n}(\overline{y1}_m - \overline{y1}_s)$	$r_m(\overline{y1}_m - \overline{y1}_s)$	$-\frac{r_m r_n}{1-r_n}(\overline{y1}_m - \overline{y1}_s)$
	t_2	$-\frac{r_m r_n}{1-r_n}(\overline{y2}_m - \overline{y2}_s)$	0	$\frac{r_m r_n}{1-r_n}(\overline{y2}_m - \overline{y2}_s)$
New	t_1	0	$r_n(\overline{y1}_n - \overline{y1}_s)$	$r_n(\overline{y1}_n - \overline{y1}_s)$
	t_2	$r_n(\overline{y2}_n - \overline{y2}_s)$	0	$-r_n(\overline{y2}_n - \overline{y2}_s)$

Table A-2. Mobility and bias terms for district-level growth measures

				Effect of group on discrepancy/bias		
		Ratio	**Gap relative to stayers**	**Discrepancy**	**CG bias**	**LG bias**
Leavers	t_1	0.101 (0.053)	-0.282 (0.143)	0.026 (0.017)	0.026 (0.017)	0
Movers	t_1	0.061 (0.039)	-0.274 (0.182)	-0.015 (0.014)	-0.015 (0.013)	0.001 (0.001)
	t_2	0.061 (0.039)	-0.286 (0.161)	0.001 (0.001)	0	-0.001 (0.001)
New	t_1	0.039 (0.018)	unobserved	0	unobserved	unobserved
	t_2	0.039 (0.018)	-0.351 (0.216)	-0.014 (0.010)	0	0.014 (0.010)
All mobility		0.102 (0.049)	----	-0.002 (0.018)	0.012 + *x* (0.017)	0.014 + *x* (0.010)

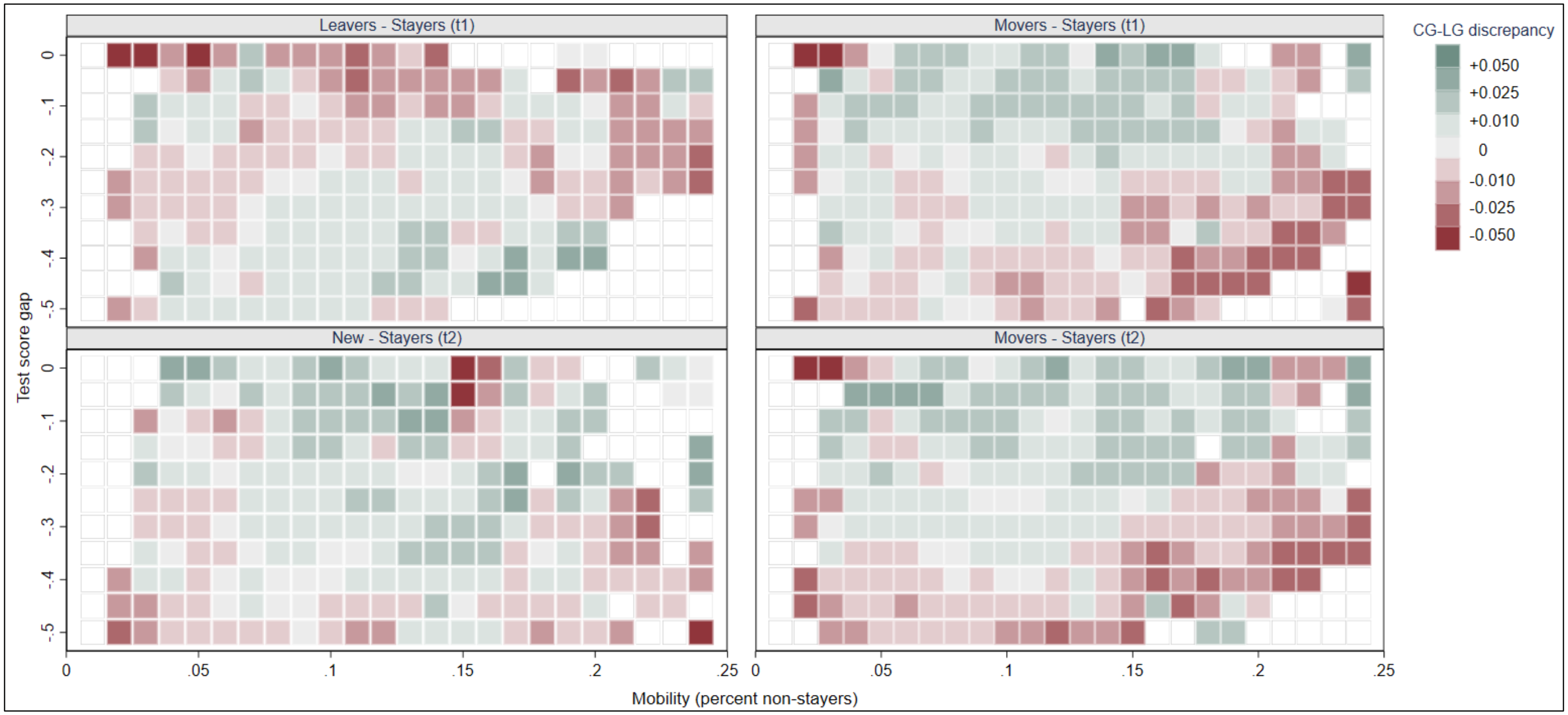


Figure A-1. Mean district-level CG-LG discrepancy by mobility and test score gap.

Appendix 2: Comparisons of data sources

As noted in the main text of this report, the data for this project focus on districts that appear in both the SEDA and state datasets, which is >99% of districts in these three states, and schools that appear in both the SEDA crosswalk and state datasets. **Table A-3** compares of the SEDA dataset with each of the source datasets from the three validation states**. Table A-4** outlines additional details of these datasets in including state-by-state differences in the source datasets, and decision rules applied while replicating the SEDA dataset and constructing each of the different growth measures.

Table A-3. Comparison of district-level data: SEDA vs state data correlations

	MA	MI	TN	Combined
Average grade size	0.999	0.999	0.999	0.999
Mean test scores: Math	0.993	0.993	0.957	0.989
Mean test scores: ELA	0.995	0.997	0.96	0.993
Cohort growth: Overall	0.934	0.974	0.975	0.953
Cohort growth: Math	0.961	0.975	0.973	0.959
Cohort growth: ELA	0.919	0.935	0.963	0.859

Table A-4. Details about the content and construction of each dataset used in the validation study.

State data (source files)	Grades/years in data	MI/TN: all grades (3-8) in all years (2009-2015) MA: missing 3rd grade 2009
	State assessment changes	MA 2015: Districts had choice of 2 assessment options. 47% of students took MCAS and 53% took PARCC. The state conducted an equating study to establish comparable scales for MCAS and PARCC scores. MI 2014: MEAP was replaced by M-STEP
Student-level datasets (constructed)	Observations dropped	Duplicate records (<1%); students repeating a grade and previous grade repeaters (<1%); TN students who took 8th grade EOC exam instead of state test (*all* records dropped for 3.8% of TN students)
	Uniqueness	Students with different state IDs but identical name & DOB (<1%) are treated as different students
	Standardization	MA 2015: standardized across equated MCAS and PARCC scores (not separately by assessment)
	Change relative to previous year	Change scores are calculated even if the assessment program is different for the two consecutive years (i.e. MA/MI transition years)
Grade-year-subject datasets (constructed)	Minimum cell size	TN cells dropped if n≤5 (<1% in school & district analyses) No minimum for MA/MI
Pooled datasets (constructed)	Cohort growth (CG) calculations	Grade-year-subject cells with missing/zero standard errors (<1%) are omitted from pooled CG calculations (due to precision weighting)

	Longitudinal growth (LG) calculations	Grade-year-subject cells for lowest grade and initial year a district/school appears in the data are omitted even if LG can be calculated using prior scores from a previous school/district (for consistency with CG measures)
	Validation analysis	A district/school must have both CG and LG estimates to be included