Title: Analysis of Kentucky School Performance on Grade 3 Mathematics and Reading State Assessments

Authors: Daniel Princiotta, Patrick Thornton, Neil Seftor, Sarah Dec, Victoria Schaefer, Deborah Jonas

Abstract: Regional Educational Laboratory Appalachia (REL AP) supported Kentucky Department of Education (KDE) staff with training, coaching, and technical support to execute quantitative analyses aimed at two research questions: (1) which schools performed better, worse, or about the same as predicted with respect to grade 3 students' mathematics performance and reading performance in 2017 and 2018, given student and school demographic characteristics; and (2) which schools have shown larger, smaller, or about the same as predicted average annual growth in grade 3 student mathematics performance and reading performance during the five years from 2014 to 2018, given student and school demographic characteristics and their changes over time? Analyses used deidentified student-level administrative data supplied by the Kentucky Center for Statistics (KYSTATS). The partners fit multilevel hierarchical linear models to predict student scale scores, average annual growth over time in schools' average scale scores, and school-level effects. Results identified high-performing schools whose students were doing better than statistically predicted in grade 3 mathematics and reading in 2017 and 2018 and high-growth schools showing above averages gains from 2014 to 2018 in grade 3 mathematics and reading. This document includes a methodological summary of quantitative analyses performed by REL AP and KDE analysts coupled with a PowerPoint slide deck describing results completed as of winter 2020.

Date: 9/14/2020

Peer reviewed: These materials underwent rigorous external peer review per Regional Educational Laboratory (REL) protocols to ensure that they met Institute of Education Sciences (IES) standards for scientifically valid research. For more information on the REL peer review process, see https://ies.ed.gov/ncee/edlabs/relwork/index.asp.

Methodological Summary:
Analysis of Kentucky School Performance on Grade 3 Mathematics and Reading State Assessments

Background

In 2018, the Kentucky Department of Education (KDE) released a new strategic plan (KDE, 2018) prioritizing improved outcomes for students in mathematics and reading. As described in the plan, KDE's retrospective analyses of Kentucky students' data demonstrated that a majority of students in the 2018/19 grade 9 cohort who scored proficient in mathematics did so initially in grade 3—the first year they were tested; the same was true for reading. Of those grade 9 students who had ever scored proficient in math, 63 percent did so initially in grade 3; the corresponding statistic for reading was 61 percent. Given these results, KDE concluded that having strong foundational literacy and numeracy skills set these students up for success. As a result, KDE is pursuing efforts to get more students on track academically in their early years so that by grade 3 they are scoring at or above proficient in mathematics and reading.

To further this objective, Regional Educational Laboratory Appalachia (REL AP) supports KDE staff with training, coaching, and technical support to execute quantitative analyses aimed at identifying schools that are doing better, worse, or about the same as statistically predicted on outcomes of interest, given certain non-malleable factors, such as demographic characteristics.[1] Two research analysts in the Kentucky Commissioner's Office codesigned the analysis and are in the process of replicating the quantitative analyses with REL coaching support. REL AP staff have worked with these analysts to enhance their capacity to design and execute relevant quantitative analyses and share results with their leadership and other stakeholders. REL AP plans to support KDE's continued learning about schools performing better, worse, or about the same as predicted through ongoing coaching with the two research analysts. Specifically, we will provide coaching on their analysis of extant survey data and their collection of qualitative data to identify practices associated with success in schools that outperform predictions. Part of this endeavor may be to identify whether practices identified as evidence-based in federal clearinghouses, including the What Works Clearinghouse, are more prevalent

[1] The REL program has several publications using similar analyses: see Abe and colleagues (2015); Culbertson and Billig (2016); Koon, Petscher, and Foorman (2014); Meyers and Wan (2016); Partridge and Koon (2017); and Partridge, Rudo, and Herrera (2017).

in schools that outperform predictions than in other schools.

Several individuals are currently involved in this project from KDE. Two research analysts who work in the Commissioner's office codeveloped this project with REL AP staff, with one taking the lead role and the other collaborating substantively throughout the project. With coaching and technical support from REL AP staff, these analysts make all final design decisions, replicate quantitative analysis, and will conduct the extant data analysis and any additional data collection in the follow-on activities. The project also involves the state's chief performance officer and the associate commissioner, Office of Teaching and Learning, who provide strategic guidance and oversight. KDE invites additional staff to meetings with REL AP as needed. For example, the director of the division of program standards and an academy program consultant guide and advise REL AP and the core KDE staff on the development of follow-on activities to ensure the results can inform KDE-supported professional development efforts.

This document is a *methodological summary* of quantitative analyses performed by REL AP and KDE analysts. It is coupled with a PowerPoint slide deck describing results from a subset of quantitative analyses completed as of winter 2020.

- The primary audience for the methodological summary is the KDE analysts who have worked with REL AP to design and execute the analyses. The summary will provide a reference for the KDE analysts moving forward as they perform similar work in the future. The summary will also provide reference information to any broader research audiences that REL AP may engage with in cooperation with KDE.
- The primary intended audience for the PowerPoint presentation is KDE leadership. As such, the presentation has a sharper focus. Per KDE analysts' request, after providing background information on the full set of quantitative analyses, the presentation focuses on results for the second of two research questions described below. REL AP may also repurpose slides for additional presentations delivered with KDE staff to broader audiences (for example, REL AP webinar, National Center for Education Statistics STATS-DC conference).

The *methodological summary* serves two purposes. First, it describes how REL AP and KDE staff generated statistical models to predict school performance and changes in school performance over time based on the demographic makeup of schools and shifts in these student populations over time. Second, it describes how REL AP and KDE staff compared these predictions with actual school performance and change over time. The quantitative analyses addressed four school-level outcomes of interest: grade 3 mathematics scale scores (math status), grade 3 reading scale scores (reading status), growth in grade 3 mathematics scale scores over time (math growth), and growth in grade 3 reading scale scores over time (reading growth). Schools in which observed status or growth was greater than predicted were classified as outperforming predictions with respect to status or growth, respectively.

Primary research questions

This investigation is based on two primary research questions that jointly address the status and growth over time of school performance in grade 3 students' mathematics and reading achievement. The status research question (RQ1) investigates schools' grade 3 mathematics and reading performance in the most recent two school years after accounting for student and school demographic characteristics. The growth research question (RQ2) examines schools' adjusted school-level gains in grade 3 mathematics and reading performance over five school years regardless of their starting point with respect to student performance.[2]

The status research question (RQ1) focuses on identifying high-performing schools. Some of these schools may not have shown substantial school-level gains in recent years, but they may have been consistently high-performing, with long-standing, well-developed strategies for supporting students' performance in early-grade mathematics and reading. RQ2 involves the identification of high-growth schools, which may have adopted new interventions, policies, or practices in recent years to boost student performance. Staff at low-performing schools may be more amenable to drawing lessons from high-growth schools that were similarly situated just five years ago than they would be from persistently high-performing schools. Over time, KDE can investigate both high-performing and high-growth schools in comparison to other schools to determine what is driving their success and, ultimately, to inform school improvement efforts in Kentucky.

The two research questions are as follows:

1. Status: Which schools performed better, worse, or about the same as predicted with respect to grade 3 students' (a) mathematics performance and (b) reading performance in 2017 and 2018, given student and school demographic characteristics?

[2] KDE data analysts decided to focus on the growth research question (RQ2) in their presentation of findings to KDE leadership. As a result, the accompanying PowerPoint slide deck focuses on RQ2 results. Although not prioritized in their presentation of findings to KDE leadership, KDE data analysts remain interested in the status research question (RQ1) results, as well.

2. Growth: Which schools have shown larger, smaller, or about the same as predicted average annual growth in grade 3 student (a) mathematics performance and (b) reading performance during the five years from 2014 to 2018, given student and school demographic characteristics and their changes over time?

Data

The quantitative analyses used deidentified student-level administrative data supplied by the Kentucky Center for Statistics (KSTATS), which collects and links data from KDE and other sources to evaluate education and workforce efforts in the commonwealth.

Analytic sample

The analytic sample comprised all first-time grade 3 students who had grade 3 mathematics and reading scale scores on the Kentucky Performance Rating for Educational Progress (K-PREP) assessment and who attended A1 schools. A1 schools, which serve 99.9 percent of public elementary students in the commonwealth,[3] are traditional public schools "under administrative control of a principal and eligible to establish a school-based decisionmaking council" and "not a program operated by, or as a part of, another school" (KDE, 2019). A1 schools serve the vast majority of Kentucky's students who receive special education services (more than 9 in 10) and all students in magnet schools. Education programs not included in the analysis, which jointly serve 0.1 percent of public elementary students in Kentucky, are district-operated alternative programs, special education programs where all enrollments are students in special education (for example, schools for the blind and schools for the deaf), and programs for children committed to or in the custody of Kentucky funded by the Kentucky Educational Collaborative for State Agency Children. The primary status analyses included student observations from the two most recent years available: the 2016/17 and 2017/18 school years.[4]

The growth analyses included observations from each school year from 2013/14 through 2017/18. The two-year analytic sample included 91,337 first-time grade 3 students enrolled in 700 elementary schools, and the five-year analytic sample included 233,343 first-time grade 3 students enrolled in 727 elementary schools.[5] Because only first-time grade 3 students were included in the sample, each

[3] Personal communication with A. Butler (July 11, 2019) from the Office of the Commissioner in the Kentucky Department of Education.

[4] As described in the supplemental analyses section, we also performed status analyses using five years of data.

[5] The discrepancy in the number of schools is because of schools opening and closing over time.

student contributes only a single record to the analyses.

Sample exclusions

In addition to excluding students enrolled in non-A1 schools, we excluded first-time grade 3 students enrolled in their school for less than 100 days because of the limited time the schools had to affect these students' academic performance.

Method

To identify which schools are performing better, worse, or about the same as predicted in mathematics and reading status based on student and school demographics, we fit two-level multilevel models to predict student scale scores and school-level effects on those scale scores.[6] As described below, we captured school effects by allowing the level-1 intercepts to vary randomly at the school level. The level-2 residuals associated with these parameters represent "school effects" after accounting for individual- and school-level demographics. As recommended in the literature (for example, Bowers, 2010; Trujillo, 2013), to reduce the possibility that findings from these status analyses are driven by chance differences across schools in student cohorts, we used the two most recent years of student data available (2016/17 and 2017/18) as opposed to basing status estimates off of a single year of data.

Building upon the status analyses, we investigated average annual growth over time in schools' average mathematics and reading scale scores. The growth analyses incorporated five years of data (2013/14, 2014/15, 2015/16, 2016/17, and 2017/18) so we could identify the schools that made the greatest improvements in grade 3 student mathematics and reading performance over the five school years.[7] As shown below, incorporating a year count variable in the first level of the model and allowing the coefficient on this variable to vary randomly at the school level enabled us to estimate the average annual growth in the outcomes of interest from 2014 to 2018 by school, accounting for the influence of changes in school demographics over time.

[6] Historically, multilevel modeling has been a relatively rare approach in the school and district effectiveness literature (Trujillo, 2013). Recent REL and other studies have used the approach (for example, Bowers, 2015; Partridge, Rudo, & Herrera 2017).

[7] KDE and REL AP chose to examine five years of growth data because it is a reasonable time frame for identifying schools that show sustained growth in student outcomes over time and allows KDE and REL AP to focus on relatively recent school performance.

Benefits of a multilevel model using student-level data

Multilevel models, like the hierarchical linear models (HLMs) used in the present study, are preferable to a more traditional approach, such as ordinary least squares, for several reasons. First, they generate standard errors that account for the nesting of data (in our case, observations of first-time grade 3 students and observations of schools from different years are nested within schools). Second, they allow investigations into the extent of variation in outcomes (and in changes over time in outcomes) at the student and school levels.[8] This provides a sense of the extent of variation in the overall outcomes that student- and school-level variables may be able to predict, along with information researchers can use when planning future studies. Third, multilevel modeling enables us to use the same analytical framework to investigate which schools have shown the most improvement in grade 3 student mathematics and reading performance (growth) and which schools have demonstrated the best relative performance in recent years (status), conditional on student and school demographics.

Potential benefits to using student-level data to estimate a multilevel model, as opposed to aggregating data to the school level and running a single-level model, also exist. Aggregating to a group level suppresses within-group variation, and this can lead to misleading results (for example, Aitkin & Longford, 1986). In contrast, multilevel models based on individual data nested within groups with individual- and group-level predictor variables can increase efficiency, reduce aggregation bias, and enable investigations into the extent of variation that lies at the student and school levels (Raudenbush & Bryk, 2002). Including student-level data in the multilevel model allows the researcher to account for both individual- and school-level influences on outcomes. For example, we know that there is both an individual effect on student achievement of living in a poor family and an effect of attending a school serving a high concentration of poor students (for example, Caldas & Bankston, 1999). Models based on student-level data can help disentangle individual-level and contextual effects in a way that aggregate school-level models cannot.

Variables

The analyses drew on an array of variables from KDE administrative data. Table 1 describes each variable included in the analyses: outcomes of interest; student-level covariates; school-level covariates; time variables; sample inclusion and exclusion variables; and reporting variables, such as school name or

[8] We report intraclass correlation coefficients when presenting findings to describe the extent of variation that exists at different levels of the analyses.

magnet status, which identify schools and provide context when presenting results. The outcomes of interest are grade 3 mathematics and reading scale scores. The student-level covariates are student age (in years), as well as indicator (dummy) variables for English learner status, free and reduced-price lunch eligibility, individualized education program (IEP) status, male, and race and Hispanic origin (variables for Black alone, non-Hispanic; Hispanic; and Other race, non-Hispanic; with White alone, non-Hispanic as the reference category). The school-level covariates are school means of the student-level covariates, such as mean student age. Note that taking the mean of a student-level indicator variable at the school level generates a proportion ranging from 0 to 1. Time variables include an indicator variable for the 2017/18 school year in the status analyses and a year count variable in the school-level growth analyses. The sample inclusion and exclusion variables align with the concepts discussed above in the analytic sample and sample exclusion sections. The reporting variables are school and district name, magnet status, and variables describing receipt of support under the Every Student Succeeds Act (ESSA) via Comprehensive Support and Improvement (CSI) or Targeted Support and Improvement (TSI) efforts.

Magnet schools. These are public schools with specialized schoolwide curricula that typically draw students from across a school district via an application process. The school district may provide transportation to magnet schools for participating students.

CSI schools. Identified by Kentucky for the first time in the 2018/19 school year, these schools are the lowest-performing 5 percent of schools in the commonwealth, according to its accountability system.

TSI schools. Any school with at least one ESSA student subgroup (such as economically disadvantaged students) whose performance was at or below that of all students in any of the lowest 5 percent of all schools (Kentucky Revised Statutes Title XIII. Education § 160.346).

KDE works with local education agencies to help improve CSI and TSI schools by providing interventions, allocating resources, and delivering technical assistance.

Table 1. Variables in the analyses

Variable	Description
Outcomes of interest	
Grade 3 mathematics scale score	Student scale score on the grade 3 Kentucky Performance Rating for Educational Progress (K-PREP) mathematics assessment, a mandatory criterion-referenced test to measure student performance on Kentucky's mathematics standards and to provide data for the state accountability system.
Grade 3 reading scale score	Student scale score on the grade 3 K-PREP reading assessment, a mandatory criterion-referenced test to measure student performance on Kentucky's reading standards and to provide data for the state accountability system.
Student-level covariates	
Age	Student age estimated by subtracting the student's year of birth from the year of the spring when the student first participated in the grade 3 K-PREP in mathematics or reading.
English learner status	Indicator variable for whether the student was identified as an English learner in the current school year. English learners are students whose primary language is a language other than English whose difficulties in English may undermine their ability to meet state proficiency standards, achieve in classes taught in English, or participate fully in society.[a] Kentucky is part of the World-Class Instructional Design and Assessment Consortium.[b] As such, students are identified as English learners if they score below a cut point on a placement test or screener and if they have not later scored above a cut point on an annual assessment of English proficiency.[a]
Free and reduced-price lunch eligibility	Indicator variable for whether a student is eligible to participate in the National School Lunch Program.
Individualized education program (IEP) status	Indicator variable for whether a student is receiving special education services via an IEP.
Male	Indicator variable for whether a student reported gender as male (female is the reference category). Students not reporting gender as male or female are counted as missing for this variable.
Black	Student is Black alone, non-Hispanic.
Hispanic	Indicator variable for whether the student traces his or her origin or descent to Mexico, Puerto Rico, Cuba, Central and South America, or other Spanish cultures, regardless of race.
Other race	Student is non-Hispanic and either American Indian or Alaska Native, Asian, Hawaiian or Other Pacific Islander, two or more races, or of unknown race and ethnicity.
School-level covariates	
Mean age	School average student age among students in the analytic sample by year.
Proportion English learners	School proportion of English learners among students in the analytic sample by year.
Proportion eligible for free and reduced-price lunch	School proportion eligible for free and reduced-price lunch among students in the analytic sample by year.
Proportion with an IEP	School proportion with an IEP among students in the analytic sample by year.
Proportion male	School proportion male among students in the analytic sample by year.

Variable	Description
Proportion Black	School proportion Black alone, non-Hispanic among students in the analytic sample by year.
Proportion Hispanic	School proportion Hispanic among students in the analytic sample by year.
Proportion Other race	School proportion Other race (not White or Black only or Hispanic) among students in the analytic sample by year
Time variables	
Year 2018	Indicator variable in the status analyses identifying observations from the 2017/18 school year.
Year count	School year count, centered at the 2013/14 school year, so that 2013/14 is 0, 2014/15 is 1, 2015/16 is 2, 2016/17 is 3, and 2017/18 is 4. This variable is used in the growth analyses.
Sample inclusion and exclusion variables	
First-time grade 3 student status	Using data from student enrollment over time, we include students who are first-time grade 3 enrollees in the school district. Students enrolled in grade 3 in the school district for the second time (or beyond) will be excluded from the analyses.
A1 school	Indicator variable for traditional public school, including magnet schools. Excludes district-operated special education programs, alternative programs, and programs for children committed to or in the custody of Kentucky funded by the Kentucky Educational Collaborative for State Agency Children. No charter schools exist in Kentucky.
Enrolled 100 days or more	Indicator variable for whether students were enrolled in their school for at least 100 days in their first-time grade 3 school year. We excluded from the analyses students who did not meet this criterion.
Reporting variables	
Comprehensive Support and Improvement (CSI) school	Indicator variable showing whether the school is receiving CSI under the Every Student Succeeds Act (ESSA).
Targeted Support and Improvement (TSI) school	Indicator variable showing whether the school is receiving TSI under ESSA.
District name	Name of the school district.
Magnet status	Indicator variable for whether the school is a magnet school.
School name	Name of the school.

[a] **https://education.ky.gov/districts/tech/sis/Documents/Standard-LEP.pdf**
[b] **https://education.ky.gov/AA/Assessments/Pages/EL-Testing.aspx**

Approach to missing data

In accord with KDE's typical approach to missing data, we used complete case analysis. Any individual students with data missing on any of the outcomes of interest or covariates were excluded from the analyses. Because the analyses relied on variables that typically have little missing data, such as student assessment scores or demographic characteristics, the level of missingness in the data was limited. Just 5.57 percent of first-time grade 3 students were excluded from the analyses, mainly due to missing assessment data. As a result of low levels of missingness, complete case analysis was warranted.

That being said, it is important to note that results of the present analysis only pertain to students who participated in state assessments, and some students are less likely to participate in state assessments than others (table 2). For example, compared with those students who participated in assessments, more non-participants received special education services via an IEP (34 versus 15 percent), were English learners (7 versus 4 percent), and were eligible for free or reduced-price lunch (76 versus 63 percent).

Table 2. Descriptive statistics of analytic sample students and those excluded due to missing assessment or other data.

Student characteristics	Analytic sample average	Analytic sample SD	Excluded student average	Excluded student SD	Effect size of average difference
Age	9.41	0.536	9.65	0.654	-0.44
English learner	0.04	0.189	0.07	0.261	-0.43
Free or reduced-price lunch eligible	0.63	0.484	0.76	0.428	-0.38
Male	0.51	0.500	0.55	0.498	-0.09
Race and Hispanic origin (reference category is white, non-Hispanic)					
Black	0.11	0.313	0.15	0.355	-0.21
Hispanic	0.07	0.260	0.08	0.272	-0.06
Other race	0.04	0.189	0.04	0.207	-0.12
Receiving special education services via IEP	0.15	0.360	0.34	0.474	-0.64

NOTE: There were 233,341 cases in the analytic sample, and 13,764 cases were excluded due to missing data. All excluded cases had information on English learner status, gender, and race and Hispanic origin, 13,762 had information on eligibility for free or reduced-price lunch and receipt of special education services via an IEP, and 2,458 had age data. Effect size of average difference is Hedges' *g* for continuous variables and Cox index for dichotomous variables.

Status models

For the status models, using data from 2016/17 and 2017/18, we fitted two-level models separately for each of two different student outcomes of interest: grade 3 mathematics scale score and grade 3 reading scale score. These two outcomes are represented by the subscript *k* in the following two-level model:

Level 1

$$Y_{ijk} = \beta_{0j} + \beta_{1j}AGE_{ij} + \beta_{2j}ELL_{ij} + \beta_{3j}FRPL_{ij} + \beta_{4j}IEP_{ij} + \beta_{5j}MALE_{ij} + \beta_{6j}BLACK_{ij} + \beta_{7j}HISP_{ij} + \beta_{8j}OTHRACE_{ij} + \beta_{9j}\overline{AGE}_{jt} + \beta_{10j}\overline{ELL}_{jt} + \beta_{11j}\overline{FRPL}_{jt} + \beta_{12j}\overline{IEP}_{jt} + \beta_{13j}\overline{MALE}_{jt} + \beta_{14j}\overline{BLACK}_{jt} + \beta_{15j}\overline{HISP}_{jt} + \beta_{16j}\overline{OTHRACE}_{jt} + \beta_{17j}Y2018_t + r_{ijt} \quad (1)$$

Level 2

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} \quad (3)$$

$$\vdots$$

$$\beta_{17j} = \gamma_{17,0} \quad (4)$$

where each outcome of interest for individual *i* in school *j* is a function of student demographic characteristics and school-level averages of the same demographic characteristics at time *t*, along with a year effect ($Y2018_t$) representing the effect of being in the 2017/18 school year as opposed to the 2016/17 school year. Student-level demographic variables include age in years (AGE_{ij}) and dummy variables (which take the value of 0 for no and 1 for yes) for whether in grade 3 the student was:

- An English learner (ELL_{ij}).
- Eligible for free and reduced-price lunch ($FRPL_{ij}$).
- An IEP holder (IEP_{ij}).
- Male ($MALE_{ij}$).
- Black ($BLACK_{ij}$).
- Hispanic ($HISP_{ij}$).
- Other race ($OTHRACE_{ij}$).

School-level means of these student demographic characteristics are represented by variable names with single bars over their tops, with subscripts *j* and *t*, as the variables vary across *j* schools and over *t* years. For example, the school mean age of first-time grade 3 students in school *j* at time *t* is represented by $\overline{AGE}_{jt}$. All school-level means of dummy variables are proportions that can range from 0 to 1. For example, if no students in a school in a given year were eligible for free and reduced-price lunch, the variable $\overline{FRPL}_{jt}$ would be 0; if 100 percent were eligible, the variable would be 1; and if 50 percent of students were eligible, $\overline{FRPL}_{jt}$ would take on the value 0.5. School-level means of demographic characteristics are included at level 1 of the model because they vary over time. Variable coefficients are represented by the vector β′, with β_{0j} representing the model intercept. For the status model, all coefficients are held fixed at level 2 (the school level), except for the level-1 intercept, which we allow to vary randomly around a cross-school mean (γ_{00}).

We assume that the level-1 error term (r_{ijt}) and the error term associated with the random intercept at level 2 (u_{0j}) are normally distributed with means of zero. The level-2 error term associated with the random intercept (u_{0j}) represents the deviation of school *j* from the cross-school mean (γ_{00}) (see equation 2). As such, it represents the extent to which a school is over- or underperforming predictions with respect to the outcome of interest after accounting for student and school demographic factors and a year fixed effect. Some of this deviation from predicted performance may be due to chance, and some may be due to systemic factors not accounted for in the model. Some of these systemic factors may be school-caused and others may be the result of non-school factors. To the extent that these systemic factors represent factors within the purview of the school (for example, school policies, practices, procedures, climate, curricula, instruction, staffing, and decisions and efforts of teachers and leaders), they jointly represent school influences on student performance. For each school, we reported the level-2 error term associated with the random intercept (u_{0j}) and tested whether the empirical Bayes residual was statistically significantly different from zero ($p < .05$) using a two-tailed *t*-test. We then categorized each school as:

- Overperforming relative to predictions based on its students' demographic characteristics (those schools with u_{0j}'s that are positive and statistically significant).
- Underperforming relative to predictions based on its students' demographic characteristics (u_{0j}'s that are negative and statistically significant).
- Performing in accordance with predictions based on its students' demographic characteristics (schools with u_{0j}'s that are not statistically significantly different from zero).

To facilitate interpretation, we presented the status school effects both on the assessment scale and a standard deviation scale (based on the standard deviation of the relevant assessment among the two-year status model analytic sample). At KDE's request, to ease interpretation, we also grouped schools with statistically significant effects according to the size of their effects on the assessment scale: less than 5 points, 5 to 9.99 points, or 10 points or higher than predicted. Five points is roughly a quarter, and 10 points is roughly one half, of a standard deviation for both tests.

Growth models

As with the status models, for the growth models we fit two-level models separately for each of two different student outcomes of interest: grade 3 mathematics scale score and grade 3 reading scale score. These two outcomes are represented by the subscript *k* in the following two-level model:

Level 1

$$Y_{ijk} = \beta_{0j} + \beta_{1j}AGE_{ij} + \beta_{2j}ELL_{ij} + \beta_{3j}FRPL_{ij} + \beta_{4j}IEP_{ij} + \beta_{5j}MALE_{ij} + \beta_{6j}BLACK_{ij} + \beta_{7j}HISP_{ij} + \beta_{8j}OTHRACE_{ij} + \beta_{9j}\overline{AGE}_{jt} + \beta_{10j}\overline{ELL}_{jt} + \beta_{11j}\overline{FRPL}_{jt} + \beta_{12j}\overline{IEP}_{jt} + \beta_{13j}\overline{MALE}_{jt} + \beta_{14j}\overline{BLACK}_{jt} + \beta_{15j}\overline{HISP}_{jt} + \beta_{16j}\overline{OTHRACE}_{jt} + \beta_{17j}YEAR_t + r_{ijt} \quad (5)$$

Level 2

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (6)$$

$$\beta_{1j} = \gamma_{10} \quad (7)$$

$$\vdots$$

$$\beta_{17j} = \gamma_{17,0} + u_{17j} \quad (8)$$

where, as in the status models described above, each outcome of interest for individual *i* in school *j* is a function of student demographic characteristics and school-level averages of the same demographic characteristics at time *t*. The only differences between the specification of the status and growth models are that time is no longer accounted for with a single year dummy. Rather, because the growth models are drawing on data from five years (2013/14 through 2017/18), we have replaced the year dummy with a year count variable ($YEAR_t$), centered at the 2017/18 school year so that it ranges from –4 in 2013/14 to 0 in 2017/18. By including this year count variable, we have specified a linear growth model, where the coefficient on year (β_{17j}) represents the average annual change in our outcomes of interest from 2013/14 to 2017/18, and the intercept (β_{0j}) represents the status of those outcomes in 2017/18.[9]

Furthermore, we have allowed the coefficient, or slope parameter, on the year count variable to vary randomly at the school level (equation 8). The error term for this slope parameter (u_{17j}), which we assume to have a normal distribution and mean of zero, represents the deviation of each school, *j*, from the cross-school average annual change in the outcome of interest over time ($\gamma_{17,0}$). For each school, we tested whether the error term (u_{17j}) is statistically significantly different from zero. We reported the magnitude of the empirical Bayes residuals for each school, and those schools with residuals that are positive and statistically significant at the $p < .05$ level are classified as overperforming statistical predictions based on their students' demographic characteristics with respect to change over time. We categorized those schools with u_{17j}'s that are negative and statistically significant as underperforming with respect to change over time in the outcome of interest. Finally, we categorized those schools with u_{17j}'s that are not statistically significantly different from zero as performing roughly as statistically predicted with respect to the average annual change in the outcome of interest over time.

In addition to testing the significance of these estimates, we set cut points to ease interpretation at KDE's request. Per KDE's guidance, we grouped schools into categories according to whether their cumulative average annual gains were less than 5 points, 5 to 9.99 points, or 10 points or higher than predicted over the five-year period. Ten points is roughly equal to a half a standard deviation, and the 5 points is about a quarter of a standard deviation of first-time grade 3 students' scale scores on the mathematics and reading assessments. Unlike the random intercept estimate results from the status model, few random slope estimates under 5 points were statistically significantly different from zero due to relatively larger confidence intervals associated with the slope estimates.

[9] This intercept varies randomly at level 2; thus, the empirical Bayes residuals associated with u_{0j} provide alternate status estimates of the extent to which schools are over- or underperforming predicted performance in 2017/18.

Supplemental analyses

School-readiness analyses

Quantitative analyses aimed at understanding whether schools are performing in ways that differ from statistical predictions often include students' prior achievement in their models to identify schools that are doing better than predicted in improving student performance, given baseline student performance. That is, to measure school performance more accurately, these analyses often model school effects on growth in individual student achievement over time. Because grade 3 is the first year in which students participate in mandatory state assessments, comparable baseline student performance data were not readily available statewide.

Kentucky collects school-readiness data on students from teacher observations during kindergarten using the BRIGANCE Early Childhood Kindergarten Screen III. These screener data, however, are not directly comparable to grade 3 state assessment data. Unlike the summative grade 3 state assessment data, kindergarten screener data are designed to help teachers identify students with potential delays, support referrals for special education services, and inform personalized instruction. Furthermore, comparable and appropriately lagged data on school readiness are available in Kentucky only for 2016/17 and 2017/18 grade 3 students (who received the kindergarten screener in 2013/14 and 2014/15, respectively), meaning that school-readiness data could not be used for the five-year school growth analyses. Finally, in any potential cases where large numbers of students transferred into a school district after kindergarten, any complete case analyses including measures of school readiness could substantially reduce the analytic sample size, potentially undermining generalizability of results.

To investigate how the inclusion of school-readiness data in the status analyses might affect results, REL AP and KDE investigated which schools were performing better, worse, or about the same as predicted on grade 3 students' mathematics and reading scale scores in 2017 and 2018, given student and school demographic characteristics and school readiness as measured in kindergarten for the subsample of students who had kindergarten screening data and grade 3 test scores. For the same subsample, we also ran our original status models without information on student school readiness as measured in kindergarten, as described in equations 1–4, and compared the school categorizations. When we ran our original status models on both the overall sample and the subsample, we found similar results, leading us to determine that estimating school effects based on the subsample (limited to students with kindergarten-readiness information) was a reasonable approach.

Drawing on additional years of data for status estimates

To investigate the stability of status estimates, REL AP and KDE ran the status models on five years of data using two approaches. The first generated status estimates by incorporating all five years of data in a modified version of the model that included dummy variables for four of the years in level 1, holding the year effects fixed at level 2. We then compared each school's estimated effects from the two-year and the five-year models. The second approach measured status using the level-2 empirical Bayes residuals associated with the randomly varying intercept of the growth model, providing alternate status estimates. These status estimates indicated the extent to which schools were over- or underperforming predictions in the 2017/18 school year. We compared these estimates with our previously described status model estimates to determine whether the growth models provided status estimates consistent with our preferred status models.

Summary of supplemental analysis results

Tables 3 and 4 offer Pearson correlation coefficients among school performance status model estimates for math and reading for the two-year status model, and the supplemental status models. These supplemental models include the:

- Five-year status model,
- Two-year status model based on the restricted sample,
- Two-year status model based on the restricted sample including school-readiness predictor variables, and
- Supplemental status estimates based on the intercept of the five-year growth model.

The two-year status model estimates were very highly positively correlated (0.97 or above) with all supplemental model estimates aside from those associated with the five-year status model, with which they had a correlation of 0.86 for both math and reading.

Table 3. Pearson correlation coefficients among school math performance status model estimates

	School math performance status model estimates				
			Two-year restricted sample[a]		
School math performance status model estimates	Two-year	Five-year	Without school readiness[b]	With school readiness[b]	Five-year growth intercept[c]
Two-year	1.00	0.86	0.99	0.97	0.97

School math performance status model estimates	School math performance status model estimates				
			Two-year restricted sample[a]		
	Two-year	Five-year	Without school readiness[b]	With school readiness[b]	Five-year growth intercept[c]
Five-year	0.86	1.00	0.85	0.82	0.86
Two-year restricted sample[a]					
Without school readiness[b]	0.99	0.85	1.00	0.98	0.96
With school readiness[b]	0.97	0.82	0.98	1.00	0.94
Five-year growth intercept[c]	0.97	0.86	0.96	0.94	1.00

[a]The restricted sample includes only those first-time grade 3 students who had school-readiness data collected in kindergarten.

[b]School-readiness variables included (1) whether the student scored "ready," (2) whether the student scored "ready with enrichments," (3) the proportion of sample students in the school who scored "ready," and (4) the proportion of students in the school who scored "ready with enrichments" on the BRIGANCE Early Childhood Kindergarten Screen III.

[c]This is a 2017/18 status estimate based on the intercept of the five-year growth model with random intercept and random slope on year, with year centered at 2017/18.

Table 4. Pearson correlation coefficients among school reading performance status model estimates

School reading performance status model estimates	School reading performance status model estimates				
			Two-year restricted sample[a]		
	Two-year	Five-year	Without school readiness[b]	With school readiness[b]	Five-year growth intercept[c]
Two-year	1.00	0.86	0.99	0.97	0.97
Five-year	0.86	1.00	0.84	0.82	0.90
Two-year restricted sample[a]					
Without school readiness[b]	0.99	0.84	1.00	0.98	0.95
With school readiness[b]	0.97	0.82	0.98	1.00	0.93
Five-year growth intercept[c]	0.97	0.90	0.95	0.93	1.00

[a]The restricted sample includes only those first-time grade 3 students who had school-readiness data collected in kindergarten.

[b]School-readiness variables included (1) whether the student scored "ready," (2) whether the student scored "ready with enrichments," (3) the proportion of sample students in the school who scored "ready," and (4) the proportion of students in the school who scored "ready with enrichments" on the BRIGANCE Early Childhood Kindergarten Screen III.

[c]This is a 2017/18 status estimate based on the intercept of the five-year growth model with random intercept and random slope on year, with year centered at 2017/18.

Limitations

The primary limitation of our analyses is that while they identified schools that were performing better or worse than statistically predicted or showing larger or smaller school-level gains than statistically predicted, they cannot, in and of themselves, explain why schools were doing so. Attributing school performance and changes in school performance solely to the effectiveness of the schools themselves or to changes in the effectiveness of schools would be naïve. In fact, any factors omitted from the initial models could be driving the school effects we estimated from these analyses, even factors outside the realm of a school's direct influence. For example, due solely to the luck of the draw, a school may have ended up with grade 3 cohorts that have, on average, greater cognitive abilities, more perseverance, or parents with higher educational expectations for their children than is the norm. Furthermore, some schools may be in communities with increasing levels of drug abuse, declining access to health care, or decreasing availability of social services.

This is not to say that factors within schools' purviews do not play a role in whether a school is over- or underperforming predictions. In fact, a wide array of literature on school effects suggests that numerous school factors, including principal and teacher effectiveness, educator expectations for student performance, data use, school climate, enacted curriculum, and instructional practices, can drive school performance (for example, Bryk, Sebring, Allensworth, Easton, & Luppescu, 2010; Edmunds, 1979; Teddlie & Reynolds, 2000). However, to successfully investigate the effect of malleable school-related factors on the results requires additional research. The results of the present analyses should be considered the launching point for a more thorough investigation.

A related limitation, unique to the present investigation, is the lack of baseline measures clearly aligned to the outcomes of interest. The absence of student mathematics and reading achievement measures prior to grade 3 may increase the likelihood that student cohort effects, and not school performance, are driving results. Incorporating demographic variables associated with the outcomes of interest helps mitigate this problem but does not eliminate it.[10]

[10] Similarly, using two cohorts of student data may mitigate this concern somewhat, but the results of the status models focused on the two most recent cohorts of student data are not necessarily generalizable to prior cohorts.

References

Abe, Y., Weinstock, P., Chan, V., Meyers, C., Gerdeman, R. D., & Brandt, W. C. (2015). *How methodology decisions affect the variability of schools identified as beating the odds* (Applied Research Methods No. REL 2015-071). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. https://eric.ed.gov/?id=ED558161

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society: Series A (General), 149*(1), 1–26.

Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: University of Chicago Press.

Bowers, A. J. (2010). Toward addressing the issues of site selection in district effectiveness research: A two-level hierarchical linear growth model. *Educational Administration Quarterly, 46*(3), 395–425. https://doi.org/10.1177/0013161X10375271

Bowers, A. J. (2015). Site selection in school district research: A measure of effectiveness using hierarchical longitudinal growth models of performance. *School Leadership & Management, 35*(1), 39–61. https://doi.org/10.1080/13632434.2014.962500

Caldas, S. J,. & Bankston, C. L. (1999). Multilevel examination of student, school, and district-level effects on academic achievement. *The Journal of Educational Research*, *93*(2), 91–100.

Culbertson, M. J., & Billig, S. H. (2016). *Decision points and considerations for identifying rural districts that have closed student achievement gaps* (Applied Research Methods No. REL 2016-130). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. https://eric.ed.gov/?id=ED565611

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15–24.

Kentucky Department of Education. (2018). *Our kids, our future: 2018–2023 strategic plan*. Frankfort, KY: Author. https://web.archive.org/web/20190701161652/https://education.ky.gov/CommOfEd/CDU/Documents/Strategic%20Plan%20Final%202018.pdf

Kentucky Department of Education (KDE). (2019). *School classification*. Frankfort, KY: Author. https://applications.education.ky.gov/sdci/Classification.aspx

Koon, S., Petscher, Y., & Foorman, B. R. (2014). *Beating the odds: Finding schools exceeding achievement expectations with high-risk students* (REL 2014-032). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. https://eric.ed.gov/?id=ED544802

Meyers, C. V., & Wan, Y. (2016). *A comparison of two methods of identifying beating-the-odds high schools in Puerto Rico* (Applied Research Methods No. REL 2017-167). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. https://eric.ed.gov/?id=ED570966

Partridge, M. A., & Koon, S. (2017). *Beating the odds in Mississippi: Identifying schools exceeding achievement expectations* (What's Happening No. REL 2017-213). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. https://eric.ed.gov/?id=ED572606

Partridge, M. A., Rudo, Z., & Herrera, S. (2017). *Identifying South Carolina charter schools that are "beating the odds"* (What's Happening No. REL 2017-236). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. https://eric.ed.gov/?id=ED572602

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Teddlie, C., & Reynolds, D. (Eds.). (2000). *The international handbook of school effectiveness research*. London, ENG.: Falmer Press.

Trujillo, T. (2013). The reincarnation of the effective schools research: Rethinking the literature on district effectiveness. *Journal of Educational Administration, 51*(4), 426–452. https://doi.org/10.1108/09578231311325640

Analysis of Kentucky School Performance on Grade 3 Mathematics and Reading State Assessments

Daniel Princiotta
Senior Education Researcher

Patrick Thornton
Data Analyst

Neil Seftor
Principal Education Researcher

Sarah Dec
Research Associate

Sarah Mehrotra
Research Associate

Victoria Schaefer
Principal Education Researcher

Deborah Jonas
REL Appalachia Director

[NOTE: The primary intended audience for the PowerPoint presentation is KDE leadership. As such, the presentation has a sharper focus than the accompanying methodological summary. Per KDE analyst request, after providing background information on the two research questions jointly addressed by KDE and REL AP data analysts, the presentation focuses on results from the second of the two research questions. REL AP may also repurpose slides for additional presentations delivered with KDE staff to broader audiences (for example, a REL AP webinar or professional conference).

The primary audience for the accompanying methodological summary is the KDE analysts who REL AP supported to design and execute the analyses. The summary will serve as a reference for the KDE analysts moving forward as they perform similar work in the future. The summary will also provide reference information to any broader research audiences that REL AP may engage with in cooperation with KDE.]

Overview of Kentucky Early Mathematics and Reading Study

- The Kentucky Department of Education's strategic plan aims to increase grade 3 student proficiency rates for mathematics and reading.
- One of the State Consolidated Plan Goals is to reduce the percentage of students scoring lower than proficient on mathematics and reading by 50 percent by 2030 for students and student subgroups in tested grades.
- As part of this effort, KDE is working in partnership with Regional Educational Laboratory Appalachia (REL AP) to identify schools with substantial gains in grade 3 mathematics and reading to inform educator development and school improvement efforts throughout Kentucky.

[CLICK]
The Kentucky Department of Education (KDE) released a strategic plan in 2018 that prioritizes improved outcomes for students in mathematics and reading.
It included a retrospective analysis of Kentucky students' data that demonstrated that most of the 2018/19 grade 9 cohort who scored proficient in mathematics did so initially in grade 3—the first year they were tested; the same was true for reading.
Given these results, KDE concluded that strong foundational mathematics and reading skills set these students up for success.
KDE is developing a comprehensive statewide early mathematics and reading plan.

[CLICK]
A key objective of this effort is to get more students on track academically in their early years, so that by grade 3 they are performing well in mathematics and reading.
One of the State Consolidated Plan Goals is to reduce the percentage of students scoring lower than proficient by 50 percent by 2030.

[CLICK]
KDE is working in partnership with REL Appalachia to identify the practices of high-growth schools to inform educator development and school improvement efforts throughout Kentucky.

Partnership with REL Appalachia

Support KDE staff to foster the adoption of evidence-based mathematics and reading practices in the early grades across Kentucky to improve student achievement.

[CLICK]
In this partnership, the role of REL Appalachia is to support KDE staff to foster the adoption of evidence-based mathematics and reading practices in the early grades across Kentucky to improve student achievement.

[CLICK]
Specifically, this project has three key elements:

- A quantitative analysis to identify high-performing and high-growth schools,
- Qualitative analysis of these schools to identify practices contributing to their success, and
- Application of the findings in Kentucky schools and districts to foster the adoption of evidence-based mathematics and reading practices in the early grades.

[CLICK]
In this presentation, we will focus on the findings from the quantitative analysis and how they can be used to inform the next part of the project.

Quantitative Analysis

Goal and research questions

Goal: Identify **high-performing and** high-growth schools **to inform school improvement efforts**

Status	Growth
Using data from 2017 and 2018, how did each school's actual grade 3 mathematics and reading performance compare to a set of predictions based on student and school demographic characteristics?	Using data from 2014–2018, how did each school's change in performance over time compare with the average school's change in performance over time, accounting for demographics?

Predicted reading score	205
Estimated true reading score	215
Status	**215 – 205 = +10**

Predicted school-level change	2.5
Estimated true school-level change	7.5
Growth	**7.5 – 2.5 = +5**

[CLICK]
As we began our work together, we identified two primary research questions aimed at identifying schools to inform school improvement efforts.

For our status research question, we wanted to identify high-performing schools – schools whose students were doing better than statistically predicted in grade 3 mathematics and reading in 2017 and 2018.

For our growth research question, we wanted to identify high-growth schools – schools showing above averages gains from 2014 to 2018 in grade 3 mathematics and reading.

For both questions, we used historical third-grade test data to create a model that would allow us to predict a school's performance based on student and school demographic characteristics.

This approach is called predictive modeling.

[CLICK]
For the status research question, we used data from 2017 and 2018 to investigate how each school's actual grade 3 mathematics and reading performance compared to a set of predictions based on student and school demographic characteristics.

[CLICK]
For example, suppose that a school was predicted to have a reading score of 205 based on the demographics of the students it served.
If it had an actual score of 215, we would say that this school performed better than predicted.
We called this difference the Status of the school.
In this case, the Status would be 10, since the school performed 10 points above the level predicted by the model.

[CLICK]
For the second research question, we used data over a longer period – from 2014 to 2018 – to investigate how the school's performance changed over time.
Specifically, we looked at average annual change in school mean grade 3 reading and grade 3 math scale scores over that five-year period, accounting for both the demographics of the students served by the school and how those may have changed over time.
This change was our estimate for growth of the school.

[CLICK]
Conditional on demographic characteristics, suppose that all schools improved by an average of 0.5 scale score points per year (or 2.5 points over the five-year period).
Now, suppose one study school improved an average of 1.5 scale score points per year (or 7.5 points over the five-year period).
Our estimate of growth for that school would be the difference between how much it actually changed and how much it was predicted to change, or 5 points.

The status and growth research questions are complementary.

Although some high-growth schools will be high-performing, not all will. That said, high-growth schools that are not yet high-performing may have recently adopted new interventions, policies or practices to boost student performance. If KDE can determine what changes have fueled school-level growth, it can help other schools adopt similar changes as appropriate.

Similarly, some high-performing schools may not have shown substantial school-level gains in recent years. This may be due to consistent high performance, which may be driven by long-standing, well-developed strategies for supporting students' performance in early-grade mathematics and reading.

KDE can ultimately investigate both high-performing and high-growth schools in comparison to other schools in order to help KDE generate and test hypotheses about what may be driving their success. This may help inform school improvement and research efforts in the future.

To focus their efforts, however, KDE data analysts have decided to begin with high-growth schools. As a result, the rest of this presentation focuses on results from research question 2, focused on school-level growth from 2014 to 2018 in students' grade 3 mathematics and reading performance.

Dataset and sample

- Dataset
  - Obtained data from Kentucky Center for Statistics (KSTATS)
  - Examined grade 3 student scale scores on Kentucky Performance Rating for Educational Progress (K-PREP) mathematics and reading tests
  - Included key demographic information
    - Age
    - Gender
    - Race
    - English learner status
    - Free and reduced-price lunch (FRPL) status indicating economic disadvantage
    - Individualized education program (IEP) status indicating students with disabilities
- Sample
  - First-time grade 3 students who attended a school for at least 100 days between 2014 and 2018
    - Created school-level measures for 727 schools from student averages

We want to briefly draw your attention to the contents of the dataset we used and who was included.

[CLICK]
First, for the data, we worked with KSTATS to obtain deidentified student-level administrative data.
We focused on the third-grade student scale scores on the K-PREP mathematics and reading assessments.
We also had key demographic information, such as age, gender, race, and indicators for English learner, free and reduced-price lunch, and individualized education program status.

[CLICK]
For the analyses, we included all students who were in grade 3 for the first time, had attended for at least 100 days, and had K-PREP scores.
For each school, we took averages of student-level data to create school-level measures of demographics.

Next, we will explain the way we analyzed the data.

<u>ADDITIONAL NOTES</u>
These were students at A1 schools, which serve 99.9 percent of students.
100 days was the threshold for inclusion in accountability measures.

First-time grade 3 so that each student has only one observation in the data.

Analysis

- Determined relationships between student and school demographics and outcomes
- Computed predicted outcomes for each school based on its demographic composition
- Compared the actual outcomes to the outcomes predicted by the model

Identified high-growth schools as those with five-year growth of 5 points or more for both mathematics and reading.

Scale Score to Performance Level						
Grade 3 - K-PREP						
Subject	**Novice Low**	**Novice High**	**Apprentice Low**	**Apprentice High**	**Proficient**	**Distinguished**
Reading	100-187	188-197	198-203	204-209	210-225	226-300
Mathematics	100-183	184-191	192-201	202-209	210-233	234-300

Our predictive modeling had three steps.

[CLICK]
First, we looked at the relationships between math and reading outcomes and the demographics of schools and their students.
For example, increases in FRPL proportion are associated with lower scores.

[CLICK]
Next, we used the demographics of each student and the school he or she attended to predict the level of outcomes.
Continuing the example, if two schools were exactly alike other than FRPL, we would predict students at the school with a higher FRPL to have lower scores.

[CLICK]
Finally, we compared the actual outcomes observed at the schools to the prediction from the model.
As we noted earlier, after some discussions of preliminary findings with KDE, we focused on the Growth measure.

[CLICK]
Specifically, we identified schools that demonstrated statistically significant positive growth of at least five points over five years for both subjects as high performing.

[CLICK]
To give you an idea of how much that is, here are the score ranges for the grade 3 K-PREP mathematics and reading assessments.

[CLICK]
For reading, the lower cutoff for Proficient is 210 and for Apprentice High is 204.
So a five-point gain would be enough to move a school up nearly a full category.
Additionally, a five-year estimate covers half of the time between now and the Department's goals for 2030.

Number of schools by type

Schools	High-growth	All other schools
Number of schools	41	654
Comprehensive support and improvement schools (CSI)	0	32
Targeted support and improvement schools (TSI)	6	165
Magnet schools	0	25

First, let's look at the type of schools that are in the group.

[CLICK]
We identified 41 schools that met our growth threshold for both math and reading. Overall, this group makes up about 6 percent of all schools.

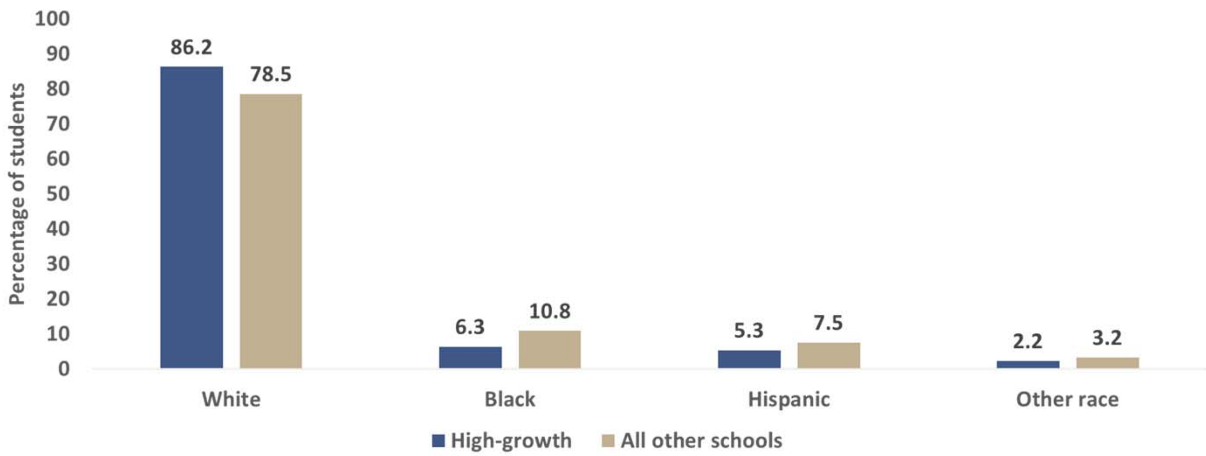
As required by the Every Student Succeeds Act, KDE identified CSI and TSI schools beginning the in 2018/19 school year.
CSI schools are those in the bottom 5 percent of the state, as measured by a combination of factors.
For elementary schools, the indicators are:

- Students' performance on math and reading on end-of-year K-PREP tests
- Students' performance on writing, social studies, and science K-PREP tests
- Students' growth on the math and reading tests, as well as growth demonstrated on a separate exam by students still learning English

TSI schools are those that have student subgroups performing significantly lower than their peers on the same set of indicators.

Of these schools, 6 were TSI schools, which have a lower representation in the high-growth group compared to other schools in Kentucky.
And there were no CSI or magnet schools.

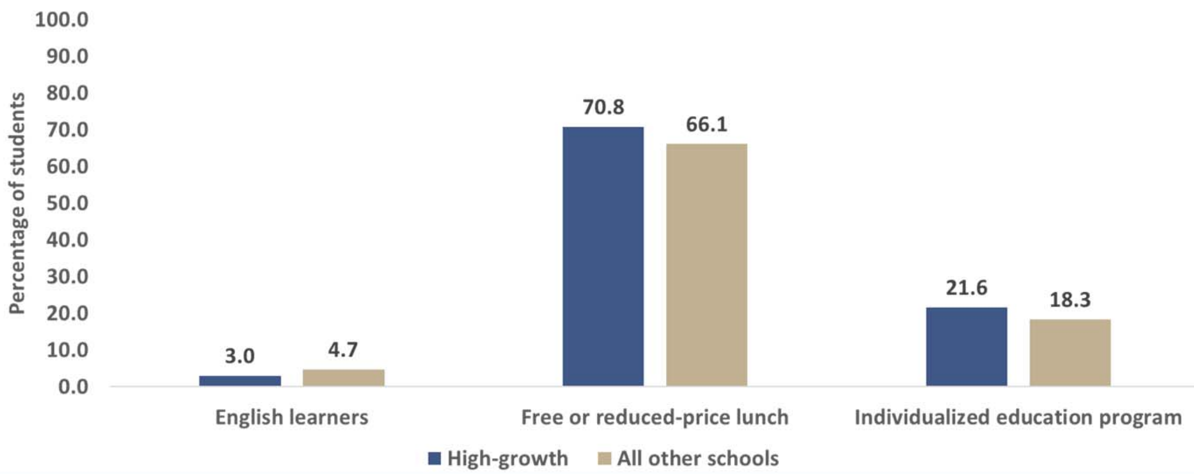
High-growth schools had a larger percentage of White students and a smaller percentage of other racial/ethnic group students than did other schools based on 2018 data.



Now, we can look at how the high-growth schools compare to other schools in Kentucky.
For each of these comparisons, we are looking at the averages of schools in each group with complete information in 2018.

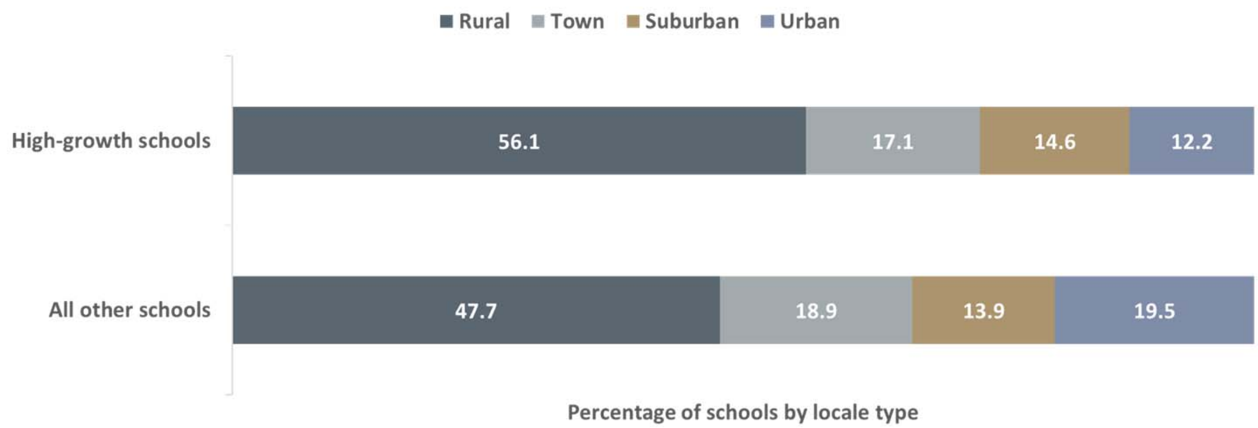
On average, high-growth schools served significantly higher percentages of White students than did other schools, offset by fewer students who were Black, Hispanic, or Other race.

High-growth schools had a larger percentage of students with disabilities or with economic disadvantages than did other schools based on 2018 data.



However, we see that high-growth schools also served higher percentages of students who were eligible for free or reduced-price lunch or had IEPs.
So while their students may have been less racially diverse, they were more frequently economically disadvantaged or students with disabilities.

Compared with other Kentucky schools, a greater percentage of high-growth schools were rural, and a lower percentage were urban based on 2018 data.

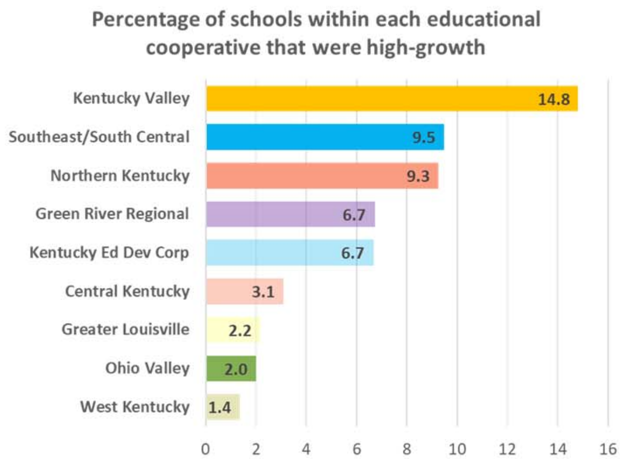


Next, we can look at where these schools are compared to other schools in Kentucky.

[CLICK]
First, let’s look at the type of location for both groups of schools.

[CLICK]
More of the high-growth schools were in rural areas, at 56 percent, than other schools in the state, at 48 percent.
This difference comes mainly from a smaller share of urban high-growth schools.
But generally, we see that the high-growth schools are distributed across the different types of locations in a way that is not too dissimilar from all other schools.

The percentage of schools within each educational cooperative that were high-growth varied across regions of Kentucky based on 2018 data.

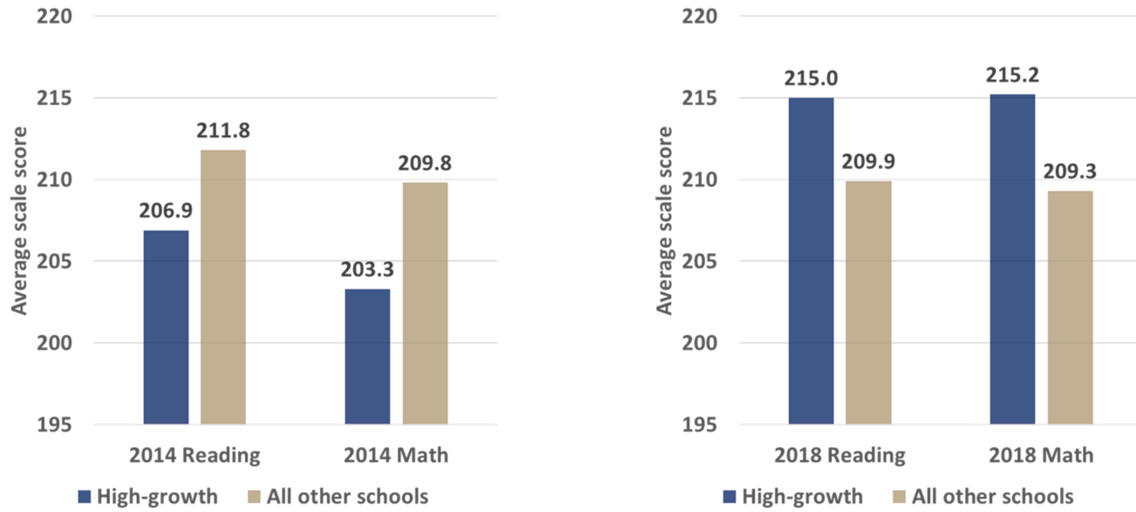


Educational cooperatives in Kentucky provide assistance and expertise for the benefit of their member school districts.
The cooperatives provide comprehensive educational services and programs that support the member districts and their schools in their school improvement efforts.
Member districts also work through the cooperatives to maximize their purchasing power to improve fiscal efficiency.

High-growth schools were not evenly distributed across the co-op regions.
Schools served by KVEC had the highest percentage of high-growth schools (nearly 15 percent of their schools) and more than 9 percent of the schools that Southeast/South Central and Northern Kentucky co-ops serve were high growth.
The remaining schools were in co-ops where less than 7 percent of schools served were identified as high growth.

The top two co-ops, Kentucky Valley and Southeast/South Central, are predominantly rural, which is consistent with the previous findings of higher percentages of high-growth schools in rural areas.
Similarly, consistent with the previous findings for urban areas, Central Kentucky and Greater Louisville have noticeably lower percentages of high-growth schools.

High-growth schools had lower math and reading scores in 2014 and higher scores in 2018 compared with all other schools.



Finally, it is useful to look at how the schools performed academically in 2014 and 2018.

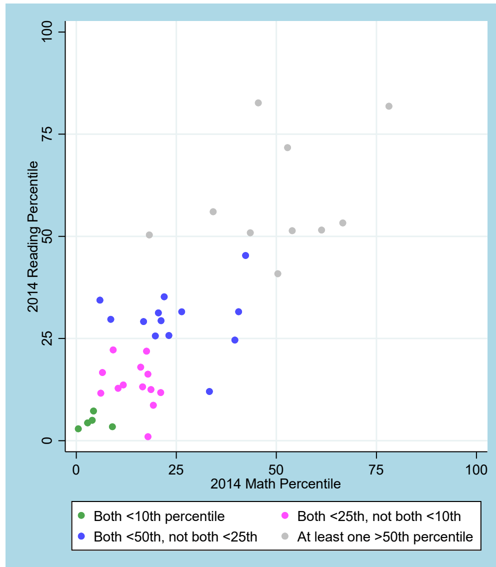
While we have identified a group of high-growth schools that had growth in both subjects over time, we may also want to know where they started.

First, let's look at the average scores for both groups of schools on the two tests in 2014 and 2018.

[CLICK]
In 2014, the schools we have identified as high-growth had average scores lower than all other schools, by 5 points in reading and 6 ½ points in math. In other words, they had more room to grow.

[CLICK]
By 2018, these schools had scores that were significantly higher than all other schools, by about 5 points in reading and 6 points in math. This suggests that room exists for other schools to grow, on average, as well.

Half of high-growth schools were in the bottom quartile for math and reading in 2014.



Within the group of high-growth schools, achievement in 2014 varied.

That is, while we just saw that these schools had lower than average math and reading scores in 2014, variation existed across schools.

This figure plots schools by their math and reading test scores in 2014, and the colors of the schools represent their percentiles on the distributions of all Kentucky schools with respect to math and reading test scores in 2014.

The green dots in the lower left are the five schools that were in the lowest 10 percent for both math and reading in 2014.

The pink dots represent 13 more schools that did not fall below the 10th percentile in both subjects but did score in the lowest quartile for both math and reading in 2014. Combined, those two groups make up almost half of the high-growth schools.

At the other end of the distribution, we see gray dots representing the 10 schools that scored above the 50th percentile for at least one subject in 2014. This group makes up 25 percent of the high-growth group.

High-growth schools spanned a wide distribution of academic starting points.

Limitations of the study

- Predictive analyses are not causal.
  - Identified schools that had larger school-level gains than statistically predicted, but no explanation for *why* it happened.
  - Attributing solely to school effectiveness would be inaccurate.
  - Factors omitted from the models or outside the school's control could affect estimates.
- The availability of baseline academic measures is limited.
  - Cannot account for student cohort effects.
  - Incorporating demographic variables associated with outcomes of interest helps but does not resolve.
- Results are not necessarily generalizable to years beyond those included in the analysis.

The primary limitation of the analyses is inherent to these types of predictive analyses.

The analyses identified schools that were performing better or worse than statistically predicted or showing larger or smaller school-level gains than statistically predicted, but they did not, in and of themselves, explain why schools were doing so. Attributing school performance and changes in school performance solely to the effectiveness of the schools themselves or to changes in the effectiveness of schools would be inaccurate. In fact, any factors omitted from the initial models could be driving the school effects we estimated from these analyses, even factors outside the realm of a school's direct influence, such as student cognitive abilities. To successfully investigate the effect of malleable school-related factors on the results requires additional research - the results of the present analyses should be considered the launching point for a more thorough investigation.

A related limitation, unique to the present investigation, is the lack of baseline measures clearly aligned to the outcomes of interest.
The absence of student mathematics and reading achievement measures prior to grade 3 may increase the likelihood that student cohort effects, and not school performance, are driving results. Incorporating demographic variables associated with the outcomes of interest helps mitigate this problem but does not eliminate it.

Finally, the results of the status models focused on the two most recent cohorts of student data are not necessarily generalizable to prior (or future) cohorts.

Summary and next steps

- The study identified 41 schools with statistically significant five-year growth of at least 5 points for both math and reading.
- On average, high-growth schools had math and reading scale scores that were 5–6 points below all other schools in 2014 and 5–6 points above all other schools in 2018.
- On average, high-growth schools had higher percentages of economically disadvantaged students, students with disabilities, and White students. More of the high-growth schools were in rural communities than all other schools.
- Next, KDE can investigate whether high-growth schools have adopted different practices or policies from other schools in recent years that could help generate and test hypotheses about possible reasons for their gains.
- If appropriate, this information could eventually help leaders and educators in *other Kentucky schools* adopt practices and policies to improve student outcomes.

Let's summarize some of the key takeaways from this analysis.

[CLICK]
We used predictive modeling to identify 41 schools with statistically significant five-year growth in K-PREP math and reading test scores of at least 5 points.

CLICK]
High-growth schools had lower average math and reading test scores in 2014, but they were spread across the distributions of scores, with five schools scoring in the lowest 10 percent on both and 10 schools scoring above average on both.

[CLICK]
These schools served more economically disadvantaged students, white students, and students with disabilities. And while spread across the state, there were more high-growth schools in rural areas and educational cooperatives and fewer high-growth schools in urban areas and educational cooperatives.

[CLICK]
What is driving these 41 schools to show substantial gains in mathematics and reading?

With some additional investigations, we can find out. We can identify what changes—around instruction, curriculum, professional development, leadership, student supports, or otherwise—were associated with gains for various schools. Some of these changes may have involved the adoption of evidence-based practices, and others may have been innovative approaches that deserve further study.

As appropriate, KDE can then seek to apply that knowledge to foster additional improvements in early mathematics and reading in similarly situated schools across Kentucky.

Supplemental Slides

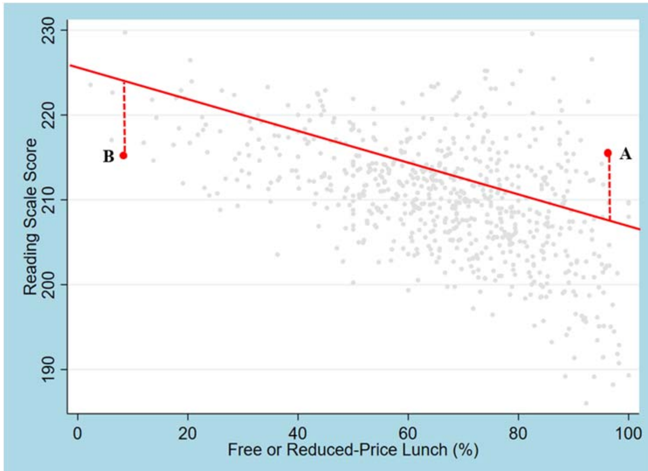
Motivation

Identifying high-growth schools is not simple.

- Many metrics are available to measure school performance, including quality of teaching; breadth, depth, or rigor of curricula; or level of student engagement (Trujillo, 2013). Most school effectiveness studies have focused on a narrow definition: student assessment performance in one or two core subjects (Bowers, 2010).
- School performance depends on a complex set of factors related to leadership, collaboration and professional learning, instructional quality, and family and community engagement, among a host of others (Beesley & Barley, 2005; Barr & Parrett, 2007; McREL, 2005).
- When focusing solely on *students*, research has shown a link between certain student characteristics and school performance (García & Weiss, 2017; Reardon, Weathers, Fahle, Jang, & Kalogrides, 2019).
- For example, research has shown connections between socioeconomic status and other demographic characteristics and academic achievement (American Psychological Association, n.d.; Duncan & Murnane, 2011).
- Some schools can demonstrate high performance when serving high concentrations of high-needs populations (Partridge, Rudo, & Herrera, 2017; Trujillo, 2013).
- Schools that have strong performance with different populations can inform strategies to maximize all students' learning and potential (Chenoweth, 2017).

- The present analyses draw on a relatively narrow definition of performance, examining student performance on state assessments in mathematics and reading in grade 3.
- Research has identified relationships between many student characteristics and achievement.
- However, schools that perform well typically enroll students with higher incomes and fewer special needs.
- Understanding more about schools that have strong performance under different circumstances, such as having a large percentage of students with risk factors beyond the schools' control (e.g., poverty), is helpful for learning how to maximize all students' learning and potential.

School performance is strongly linked to the percentage of economically disadvantaged students in Kentucky schools in 2018.



To illustrate this point, we can plot all Kentucky elementary schools by their reading score and the percentage of students eligible for free or reduced-price lunch.

There's a lot of variation, but it's clear that schools with lower percentages of students eligible for free- or reduced-price lunch tend to do better.

Now let's consider two schools, both of which had an average reading score of about 215.

- School A is on the right side of the figure, with a FRPL rate of 97 percent, and School B is on the left side, with a FRPL rate of 10 percent.
- Looking above and below these schools, we can see how other schools with similar FRPL rates performed.
- We see that School A has a score that is above many other schools with high FRPL levels.

On the other hand, School B has a score lower than nearly all schools with similarly low FRPL levels.

School A and School B have the same reading score, despite having significant differences in the number of students who qualify for FRPL.

One way to consider school performance is to compare a school's average student achievement to what might be predicted from an average school with a similar population.

- Looking at the data, we could estimate the relationship between reading score and FRPL with this line.

- We see that School A is well above the line, so it is performing better than predicted given its population.
- School B, however, is performing lower than we would predict given the population it serves.
- This is similar to the approach we used to classify schools in the present study, in which we used several student-level and school-level demographic measures to generate more accurate statistical predictions.

Empirical Models

Two-year status model

- **Level 1**

$$Y_{ijk} = \beta_{0j} + \beta_{1j}AGE_{ij} + \beta_{2j}ELL_{ij} + \beta_{3j}FRPL_{ij} + \beta_{4j}IEP_{ij} + \beta_{5j}MALE_{ij} + \beta_{6j}BLACK_{ij} + \beta_{7j}HISP_{ij} + \beta_{8j}OTHRACE_{ij} + \beta_{9j}\overline{AGE}_{jt} + \beta_{10j}\overline{ELL}_{jt} + \beta_{11j}\overline{FRPL}_{jt} + \beta_{12j}\overline{IEP}_{jt} + \beta_{13j}\overline{MALE}_{jt} + \beta_{14j}\overline{BLACK}_{jt} + \beta_{15j}\overline{HISP}_{jt} + \beta_{16j}\overline{OTHRACE}_{jt} + \beta_{17j}Y2018_t + r_{ijt} \quad (1)$$

- **Level 2**

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$
$$\beta_{1j} = \gamma_{10} \quad (3)$$
$$\vdots$$
$$\beta_{17j} = \gamma_{17,0} \quad (4)$$

The status research question investigates schools' grade 3 mathematics and reading performance in the most recent two school years after accounting for student and school demographic characteristics.
It focuses on identifying high-performing schools.
Some of these schools may not have shown substantial school-level gains in recent years, but they may have been consistently high-performing, with long-standing, well-developed strategies for supporting students' performance in early-grade mathematics and reading.

For the status models, each outcome of interest *k* for individual *i* in school *j* is a function of student demographic characteristics and school-level averages of the same demographic characteristics at time *t*, along with an indicator variable for the 2017/18 school year.
Student-level demographic variables include age in years and indicator variables for whether the student was an English learner, eligible for free or reduced-price lunch, had an IEP, male, Black, Hispanic, or another race.

School-level means of these student demographic characteristics are represented by variable names with bars over them and are subscripted with *j* and *t* as the variables vary across schools and time.
For example, the school mean age of first-time third graders in school *j* at time *t* is represented by $\overline{AGE}jt$.

All school-level means of dummy variables are proportions that can range from 0 to 1.
For example, if no students in a school in a year were eligible for free or reduced-price lunch, the variable $\overline{FRPL}jt$ would be 0; if 100 percent were eligible, the variable would be 1; and if 50 percent of students were eligible, $\overline{FRPL}jt$ would take on the value 0.5.
School-level means of demographic characteristics are included at level 1 of the model because they vary over time.

Variable coefficients are represented by the vector β’, with β0*j* representing the model intercept.
For the status model, all coefficients are held fixed at level 2 (the school level), except for the level-1 intercept, which we allow to vary randomly around a cross-school mean (γ00).

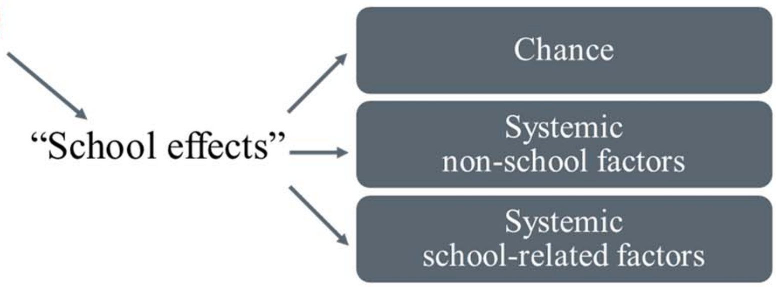
Two-year status model

- **Level 1**

$$Y_{ijk} = \beta_{0j} + \beta_{1j}AGE_{ij} + \beta_{2j}ELL_{ij} + \beta_{3j}FRPL_{ij} + \beta_{4j}IEP_{ij} + \beta_{5j}MALE_{ij} + \beta_{6j}BLACK_{ij} + \beta_{7j}HISP_{ij} + \beta_{8j}OTHRACE_{ij} + \beta_{9j}\overline{AGE}_{jt} + \beta_{10j}\overline{ELL}_{jt} + \beta_{11j}\overline{FRPL}_{jt} + \beta_{12j}\overline{IEP}_{jt} + \beta_{13j}\overline{MALE}_{jt} + \beta_{14j}\overline{BLACK}_{jt} + \beta_{15j}\overline{HISP}_{jt} + \beta_{16j}\overline{OTHRACE}_{jt} + \beta_{17j}Y2018_{t} + r_{ijt} \quad (1)$$

- **Level 2**

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$
$$\beta_{1j} = \gamma_{10} \quad (3)$$
$$\vdots$$
$$\beta_{17j} = \gamma_{17,0} \quad (4)$$



We assume that the level-1 error term (*rijt*) and the error term associated with the random intercept at level 2 (*u*0*j*) are normally distributed with means of zero.

The level-2 error term associated with the random intercept (*u*0*j*) represents the deviation of school *j* from the cross-school mean (γ00) (see equation 2). As such, it represents the extent to which a school is over- or underperforming predictions with respect to the outcome of interest after accounting for student and school demographic factors and a year fixed effect.

Some of this deviation from predicted performance may be due to chance and some may be due to systemic factors not accounted for in the model. Some of these systemic factors may be school-caused and others may be the result of non-school factors for which there are insufficient data to include in the model.

To the extent that these systemic factors represent factors within the purview of the school (for example, school policies, practices, procedures, climate, curricula, instruction, staffing, and decisions and efforts of teachers and leaders), they jointly represent school influences on student performance.

For each school, we reported the level-2 error term associated with the random intercept (*u*0*j*) and tested whether the empirical Bayes residual was statistically significantly different from zero ($p < .05$) using a two-tailed t-test.

We then categorized each school as:

- Overperforming relative to predictions based on its students' demographic characteristics (those schools with u_{0j}'s that are positive and statistically significant)
- Underperforming relative to predictions based on its students' demographic characteristics (u_{0j}'s that are negative and statistically significant)
- Performing in accordance with predictions based on its students' demographic characteristics (schools with u_{0j}'s that are not statistically significantly different from zero).

To facilitate interpretation, we presented the status school effects both on the assessment scale and a standard deviation scale (based on the standard deviation of the relevant assessment among the two-year status model analytic sample).

Five-year growth model

- **Level 1**

$$Y_{ijk} = \beta_{0j} + \beta_{1j}AGE_{ij} + \beta_{2j}ELL_{ij} + \beta_{3j}FRPL_{ij} + \beta_{4j}IEP_{ij} + \beta_{5j}MALE_{ij} + \beta_{6j}BLACK_{ij} + \beta_{7j}HISP_{ij} + \beta_{8j}OTHRACE_{ij} + \beta_{9j}\overline{AGE}_{jt} + \beta_{10j}\overline{ELL}_{jt} + \beta_{11j}\overline{FRPL}_{jt} + \beta_{12j}\overline{IEP}_{jt} + \beta_{13j}\overline{MALE}_{jt} + \beta_{14j}\overline{BLACK}_{jt} + \beta_{15j}\overline{HISP}_{jt} + \beta_{16j}\overline{OTHRACE}_{jt} + \beta_{17j}YEAR_t + r_{ijt} \quad (1)$$

- **Level 2**

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} \quad (3)$$

$$\vdots$$

$$\beta_{17j} = \gamma_{17,0} + u_{17j} \quad (4)$$

The growth research question examines schools' adjusted school-level gains in grade 3 mathematics and reading performance over five school years regardless of their starting point with respect to student performance.
It involves the identification of high-growth schools, which may have adopted new interventions, policies, or practices in recent years to boost student performance.
Staff at low-performing schools may be more amenable to drawing lessons from high-growth schools that were similarly situated just five years ago than they would be from persistently high-performing schools.

As with the status models, each outcome of interest *k* for individual *i* in school *j* is a function of student demographic characteristics and school-level averages of the same demographic characteristics at time *t.*
The only difference between the specification of the status and growth models is that time is no longer accounted for with a single year dummy.
Rather, because the growth models are drawing on data from five years (2013/14 through 2017/18), we have replaced the year dummy with a year count variable (*YEARt*), centered at the 2013/14 school year so that it ranges from 0 in 2013/14 to 4 in 2017/18.
By including this year count variable, we have specified a linear growth model where the coefficient on year (β17*j*) represents the average annual

change in our outcomes of interest from 2013/14 to 2017/18, and the intercept ($\beta 0j$) represents the initial status of those outcomes in 2013/14.

Furthermore, we have allowed the coefficient, or slope parameter, on the year count variable to vary randomly at the school level (equation 4).
The error term for this slope parameter ($u17j$), which we assume to have a normal distribution and mean of zero, represents the deviation of each school j from the cross-school average annual change in the outcome of interest over time ($\gamma 17,0$).
For each school, we tested whether the error term ($u17j$) was statistically significantly different from zero.
We reported the magnitude of the empirical Bayes residuals for each school.
Schools with residuals that are positive and statistically significant at the $p < .05$ level were classified as overperforming statistical predictions based on their students' demographic characteristics with respect to change over time.
We categorized those schools with $u17j$'s that are negative and statistically significant as underperforming with respect to change over time in the outcome of interest.
Finally, we categorized those schools with $u17j$'s that are not statistically significantly different from zero as performing roughly as statistically predicted with respect to the average annual change in the outcome of interest over time.

Two-year status model including school readiness

- **Level 1**

$$Y_{ijk} = \beta_{0j} + \beta_{1j}AGE_{ij} + \beta_{2j}ELL_{ij} + \beta_{3j}FRPL_{ij} + \beta_{4j}IEP_{ij} + \beta_{5j}MALE_{ij} + \beta_{6j}BLACK_{ij} + \beta_{7j}HISP_{ij} + \beta_{8j}OTHRACE_{ij} + \beta_{9j}KREADY_{ij} + \beta_{10j}KREADYE_{ij} + \beta_{11j}\overline{AGE}_{jt} + \beta_{12j}\overline{ELL}_{jt} + \beta_{13j}\overline{FRPL}_{jt} + \beta_{14j}\overline{IEP}_{jt} + \beta_{15j}\overline{MALE}_{jt} + \beta_{16j}\overline{BLACK}_{jt} + \beta_{17j}\overline{HISP}_{jt} + \beta_{18j}\overline{OTHRACE}_{jt} + \beta_{19j}\overline{KREADY}_{jt} + \beta_{20j}\overline{KREADYE}_{jt} + \beta_{21j}Y2018_{t} + r_{ijt} \quad (1)$$

- **Level 2**

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$
$$\beta_{1j} = \gamma_{10} \quad (3)$$
$$\vdots$$
$$\beta_{21j} = \gamma_{21,0} \quad (4)$$

When designing a school effects study, there are often different choices to be made, and *a priori,* it is not always clear how these choices might influence results.
We ran a series of supplemental models to investigate some of these alternatives, such as including a measure of school readiness and extending the status model to five years.
Ultimately, the results were quite similar across models.

Quantitative analyses aimed at understanding whether schools are performing in ways that differ from statistical predictions often include students' prior achievement in their models to identify schools that are doing better than predicted in improving student performance, given baseline student performance. That is, to measure school performance more accurately, these analyses often model school effects on growth in individual student achievement over time. Because grade 3 is the first year in which students participate in mandatory state assessments, comparable baseline student performance data were not readily available statewide.

Kentucky collects school-readiness data on students from teacher observations during kindergarten using the BRIGANCE Early Childhood Kindergarten Screen III. These screener data, however, are not directly comparable to grade 3 state assessment data.

Unlike the summative grade 3 state assessment data, kindergarten screener data are designed to help teachers identify students with potential delays, support referrals for special education services, and inform personalized instruction.
Furthermore, comparable and appropriately lagged data on school readiness are available in Kentucky only for 2016/17 and 2017/18 third-graders (who received the kindergarten screener in 2013/14 and 2014/15, respectively), meaning that school-readiness data could not be used for the five-year school growth analyses.
Finally, in any potential cases where large numbers of students transferred into a school district after kindergarten, any analyses based on complete cases including measures of school readiness could substantially reduce the analytic sample size, potentially undermining generalizability of results.

To investigate how the inclusion of school-readiness data in the status analyses might affect results, we investigated which schools were performing better, worse, or about the same as predicted on grade 3 students' mathematics and reading performance in 2017 and 2018 given student and school demographic characteristics and school readiness as measured in kindergarten for the subsample of students who had kindergarten screening data and grade 3 test scores.
For the same subsample, we also ran our original status models without information on student school readiness as measured in kindergarten, as described in equations 1–4, and compared the school categorizations.
By comparing the results from our original status models run on the overall sample to the subsample and finding similar results, we were able to determine that estimating school effects based on the subsample (limited to students with kindergarten readiness information) was a reasonable approach.

Five-year status model

- **Level 1**

$$Y_{ijk} = \beta_{0j} + \beta_{1j}AGE_{ij} + \beta_{2j}ELL_{ij} + \beta_{3j}FRPL_{ij} + \beta_{4j}IEP_{ij} + \beta_{5j}MALE_{ij} + \beta_{6j}BLACK_{ij} + \beta_{7j}HISP_{ij} + \beta_{8j}OTHRACE_{ij} + \beta_{9j}\overline{AGE}_{jt} + \beta_{10j}\overline{ELL}_{jt} + \beta_{11j}\overline{FRPL}_{jt} + \beta_{12j}\overline{IEP}_{jt} + \beta_{13j}\overline{MALE}_{jt} + \beta_{14j}\overline{BLACK}_{jt} + \beta_{15j}\overline{HISP}_{jt} + \beta_{16j}\overline{OTHRACE}_{jt} + \beta_{17j}Y2015_t + \beta_{18j}Y2016_t + \beta_{19j}Y2017_t + \beta_{20j}Y2018_t + r_{ijt} \quad (1)$$

- **Level 2**

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} \quad (3)$$

$$\vdots$$

$$\beta_{20j} = \gamma_{17,0} \quad (4)$$

To investigate the stability of status estimates for schools over time, we ran the status models on five years of data rather than just two.
We used two approaches.
The first generated status estimates by incorporating all five years of data in a modified version of the model that included dummy variables for four of the years in level 1, holding the year effects fixed at level 2.
We then compared each school's estimated performance from the two-year and the five-year models.

Five-year growth model: Supplemental status estimates

- **Level 1**

$$Y_{ijk} = \beta_{0j} + \beta_{1j}AGE_{ij} + \beta_{2j}ELL_{ij} + \beta_{3j}FRPL_{ij} + \beta_{4j}IEP_{ij} + \beta_{5j}MALE_{ij} + \beta_{6j}BLACK_{ij} + \beta_{7j}HISP_{ij} + \beta_{8j}OTHRACE_{ij} + \beta_{9j}\overline{AGE}_{jt} + \beta_{10j}\overline{ELL}_{jt} + \beta_{11j}\overline{FRPL}_{jt} + \beta_{12j}\overline{IEP}_{jt} + \beta_{13j}\overline{MALE}_{jt} + \beta_{14j}\overline{BLACK}_{jt} + \beta_{15j}\overline{HISP}_{jt} + \beta_{16j}\overline{OTHRACE}_{jt} + \beta_{17j}YEAR_t + r_{ijt} \quad (1)$$

- **Level 2**

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} \quad (3)$$

$$\vdots$$

$$\beta_{17j} = \gamma_{17,0} + u_{17j} \quad (4)$$

The second approach was used to see if we could be more parsimonious, using the same model to estimate both status and growth.

In this model, we measured status using the level-2 empirical Bayes residuals associated with the randomly varying intercept of the growth model (centered at 2013/14).

This approach provided alternate status estimates of the extent to which schools were over- or underperforming predictions in 2013/14.

We compared these estimates with our previously described status model estimates to determine whether the growth models provided status estimates consistent with our preferred status models.

Though the two- and five-year status models produced results that were highly correlated, we found that the status for a school did vary based on the amount of historical data used to estimate it (see accompanying methodological summary and slide 31).

Ultimately, KDE determined that for the status analyses, they wanted to focus on the most recent years only.

Analytical Estimates

Pearson correlation coefficients among model estimates

School performance status model estimates, reading / math	Two-year	Five-year	Two-year restricted, no readiness	Two-year restricted, readiness	Five-year growth intercept
Two-year	1.00 / 1.00				
Five-year	0.86 / 0.86	1.00 / 1.00			
Two-year restricted, no readiness	0.99 / 0.99	0.84 / 0.85	1.00 / 1.00		
Two-year restricted, readiness	0.97 / 0.97	0.82 / 0.82	0.98 / 0.98	1.00 / 1.00	
Five-year growth intercept	0.97 / 0.97	0.90 / 0.86	0.95 / 0.96	0.93 / 0.94	1.00 / 1.00

A correlation coefficient ranges from -1 to +1, with +1 representing a perfectly linear, positive relationship.

This table presents Pearson correlation coefficients among school performance status model estimates for math and reading for the two-year status model, the supplemental five-year status model, the supplemental two-year status model based on the restricted-use sample, the supplemental two-year status model based on the restricted-use sample and including school readiness predictor variables, and the supplemental status estimate based on the intercept of the five-year growth model.

The two-year status model estimates were very highly positively correlated (0.97 or above) with all supplemental model estimates aside from those associated with the five-year status model, with which they had a correlation of 0.86 for both math and reading.

A correlation coefficient ranges from -1 to +1, with +1 representing a perfectly linear, positive relationship.

In this context, a high correlation means that the results remained relatively consistent across the different sensitivity analyses and model specifications.

Model estimates for reading assessment

Student-level	Status	Growth
Intercept	218.98***	219.25***
Age of grade 3 students	0.01	0.10
Male	-1.95***	-1.90***
Black	-6.80***	-7.14***
Hispanic	0.13	-0.27
Other race	1.10***	0.90***
English learner	-9.16***	-10.47***
Free or reduced-price lunch (FRPL)	-7.07***	-7.90***
Individualized education program (IEP)	-7.24***	-7.88***

School-level	Status	Growth
Year	-0.74***	-0.07
Mean age of grade 3 students	3.88***	1.30*
Proportion Male	-3.93***	-3.05***
Proportion Black	-6.41***	-6.35***
Proportion Hispanic	-3.02	-2.49
Proportion Other race	12.62***	4.24*
Proportion English learner	-0.00	2.94
Proportion FRPL	-3.74***	-1.33*
Proportion IEP	5.03***	2.56**

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.
Status model explains 54 percent of between-school and 10 percent of within-school variance. Intraclass correlation = 0.118.
Growth model explains 38 percent of between-school and 12 percent of within-school variance. Intraclass correlation = 0.100.

At the student level, the coefficient estimates and significance were very close for the status and growth models.
Both models found that boys and Black students scored lower, while students in the Other race category scored higher.
In the schools included in the analysis, 46 percent of students in the “other race” category identified as Asian, 46 percent identified as two or more races, and about 3 percent each of American Indian, Hawaiian, and unknown.

Additionally, scores were lower for English learners, free and reduced-price lunch eligible students, and students with an IEP.

At the school level, there were some differences between the models, and with one exception, the coefficient estimates were smaller in the growth model.
Like the student-level findings, both analyses found that schools with higher proportions of boys and Black students scored lower, while schools with higher proportions of Other race students scored higher.
Both found that schools with older students in grade 3 or a higher proportion of Other race students had higher scores, and schools with a higher proportion of FRPL students had lower scores.

However, the estimates for English learners and students with IEPs are quite different for the school level.
At the student level, English learners had significantly lower reading scores, the

largest of the estimated coefficients. At the school level, however, the proportion of English learners in a school was unrelated to the school's score. As the proportion of English learners in a school increase, schools may be able to adapt their interventions (e.g., hire more ESL teachers, establish bilingual classes).

Even more striking is the finding for students with IEPs.
At the student level, having an IEP was associated with a significantly lower reading score. However, at the school level, scores increased significantly with the proportion of students with IEPs, perhaps due to the availability of additional or specialized resources.

Model estimates for mathematics assessment

Student-level	Status	Growth
Intercept	217.75***	218.38***
Age of grade 3 students	-0.14	-0.33***
Male	1.59***	1.31***
Black	-6.82***	-6.83***
Hispanic	0.20	0.01
Other race	4.02***	4.12***
English learner	-8.41***	-9.64***
Free or reduced-price lunch (FRPL)	-8.51***	-9.07***
Individualized education program (IEP)	-9.64***	-10.22***

School-level	Status	Growth
Year	-0.50***	0.34***
Mean age of grade 3 students	2.35	0.99
Proportion Male	-1.07	-4.26***
Proportion Black	-5.72***	-3.72***
Proportion Hispanic	-0.54	-2.40
Proportion Other race	18.07***	6.57***
Proportion English learner	-2.45	0.01
Proportion FRPL	-2.87***	-0.82
Proportion IEP	4.97***	4.37***

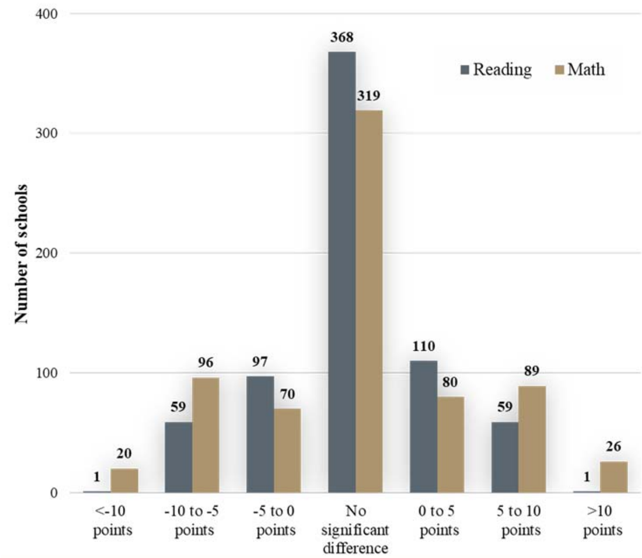
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.
Status model explains 40 percent of between-school and 11 percent of within-school variance. Intraclass correlation = 0.131.
Growth model explains 13 percent of between-school and 12 percent of within-school variance. Intraclass correlation = 0.106.

For math, the findings were generally similar in terms of direction, magnitude, and significance.
The only noticeable difference between the coefficients for math and reading were on the Other race indicator.
At the student level, the estimate was 1 point for reading and 4 points for math.
At the school level, the already large findings of 12 and 4 points for proportion of Other race students are about 50 percent larger here.

Findings

One in four schools outperformed predictions in the status model.

- Difference of 10 points or above
  - About ½ standard deviation
  - Example: 205 to 215 is a move from lower end of Apprentice High to Proficient for reading
- Difference of 5 to 10 points
  - About ¼ standard deviation
- Difference of less than 5 points
  - Less than ¼ standard deviation



This figure shows the distribution of schools for the status analyses of math and reading scores.
For both, there are 7 categories:

3 for positive differences, where the actual was higher than predicted and statistically significant
3 for negative differences, where the actual was lower than predicted and statistically significant
and 1 for schools for which the actual and predicted were not significantly different.

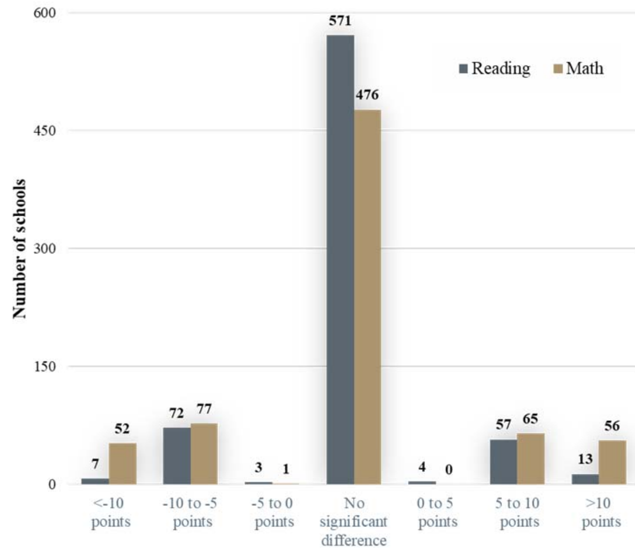
Working from right to left, the group furthest to the right reflects an actual score that is 10 points, or about one-half of a standard deviation, above what was predicted and statistically significant.
The next group difference is between 5 and 10 points and statistically significant.
And the difference in the third group is one that is fewer than 5 points, and statistically significant.

To give you an idea of the size of the difference, let’s go back to our earlier example.
School A was predicted to have a reading score of 205, which is near the bottom of Apprentice High, but had an actual reading score of 215, which is well into Proficient.

That difference is 10 points, enough to move the school up at least one K-PREP category for both math and reading.

One in eight schools outperformed predictions in the growth model.

- Change categories over five years on same scale as status
  - Difference of 10 points or above
  - Difference of 5 to 10 points
  - Difference of less than 5 points, but significant



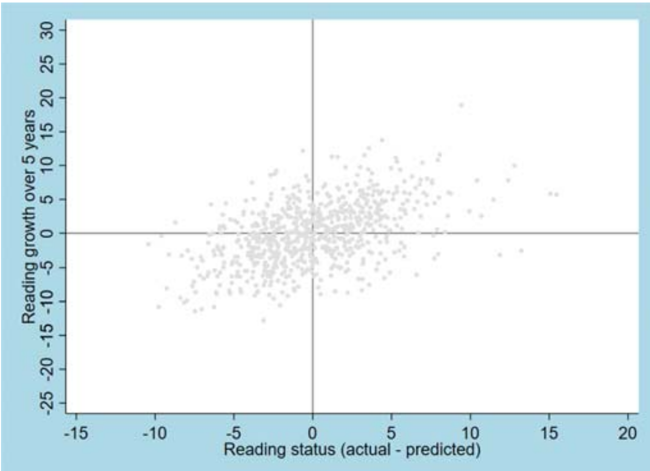
Here we examine our second research question, regarding how schools changed over time.
This figure shows the distribution of schools for the growth analyses of math and reading scores.
Because yearly changes for a school tend to be small, we estimated the change over five years, which is half the period between now and 2030, which is KDE's goal point.

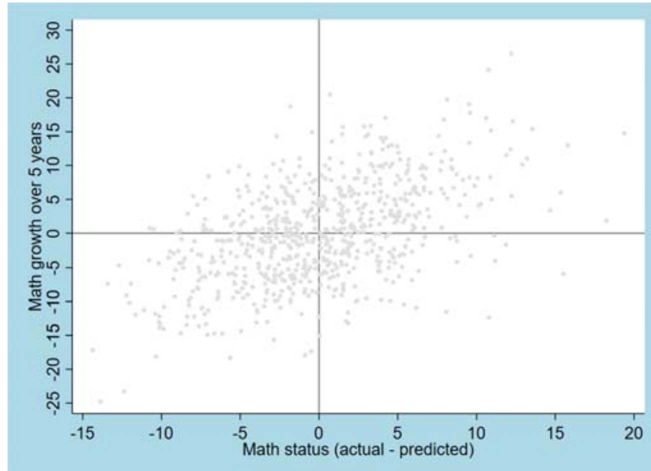
Positive change over time indicates that the actual outcome is rising over time relative to the predicted outcome.
Back to our earlier example, School A was predicted to have a reading score of 205 but had an actual reading score of 215, which was a difference of 10 points.
If the actual score grew at 2 points per year, it would be 10 points higher after five years, or 225.
In our analysis, we would say that School A had a change over five years of 10 points.

The seven categories are defined the same way in terms of size and significance.
In this case, the rightmost category reflects a change over five years of more than 10 points.
So again, for both math and reading, this would be enough to move a school's average student up at least one K-PREP category over five years.

There is again a distribution of schools across the categories, but fewer were statistically significant than in the other analysis.
The result is almost no schools in the ranges with the smallest values.
However, the number of schools in the top two categories is nearly identical to those in the previous analysis.

Combinations of status and growth are widely distributed.





Now we can put the findings together.

These figures show the distribution of schools by status and growth for both math and reading.
For both subjects, there is a positive correlation between status and growth.
However, there are a variety of combinations.
Some schools have positive status and negative growth, which suggests they are moving down over time to their predicted levels.
Some schools have negative status and positive growth, which suggests they are rising over time to their predicted levels.
Additionally, schools that have positive measures of performance and change over time for reading also often have positive measures for math.

Schools perform similarly relative to predictions from the mathematics and reading status models.

		Math Status (M)						
		M ≤ -10	-10 < M < -5	-5 < M ≤ 0	No diff.	0 < M < 5	5 ≤ M < 10	10 ≤ M
Reading Status (R)	10 ≤ R						1	8
	5 ≤ R < 10		1		10	5	28	12
	0 < R < 5				34	25	34	4
	No diff.	2	26	37	226	49	26	2
	-5 < R ≤ 0	2	39	27	41	1		
	-10 < R < -5	15	30	6	8			
	R ≤ -10	1	0	0				

Now we can put the findings together.
This table shows the distribution of schools across the categories for math and reading.
There is a strong correlation between the two, as schools tend to either exceed the predictions for both or fall below the predictions for both.

References

References

- American Psychological Society. (n.d.) Education & Socioeconomic Status. Washington, DC: Author. https://www.apa.org/pi/ses/resources/publications/education
- Barr, R. D., & Parrett, W. H. (2007). *The kids left behind: Catching up the underachieving children of poverty*. Bloomington, IN: Solution Tree.
- Beesley, A. D., & Barley, Z. A. (2005). *Rural schools that beat the odds: Four case studies*. Denver, CO: McREL.
- Chenoweth, K. (2017). *Schools that succeed: How educators marshal the power of systems for improvement.* Cambridge, MA: Harvard Education Press.
- Duncan, G. J., & Murnane, R. J. (Eds.). (2011). *Whither opportunity? Rising inequality, schools, and children's life chances.* New York: Russell Sage Foundation.
- García, E., & Weiss, E. (2017). *Education inequalities at the school starting gate: Gaps, trends, and strategies to Address them.* Washington, DC: Economic Policy Institute.
- *McREL insights: Schools that beat the odds*. (2005). Denver, CO: McREL.
- Partridge, M. A., Rudo, Z., & Herrera, S. (2017). *Identifying South Carolina charter schools that are "beating the odds"* (REL 2017-236). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. https://eric.ed.gov/?id=ED572602
- Reardon, S., Weathers, E., Fahle, E., Jang, H., & Kalogrides, D. (2019). *Is separate still unequal? New evidence on school segregation and racial academic achievement gaps* (CEPA Working Paper No. 19-09). Stanford, CA: Stanford Center for Education Policy Analysis. https://cepa.stanford.edu/content/separate-still-unequal-new-evidence-school-segregation-and-racial-academic-achievement-gaps
- Trujillo, T. (2013). The reincarnation of the effective schools research: Rethinking the literature on district effectiveness. *Journal of Educational Administration, 51*(4), 426–452. https://eric.ed.gov/?id=EJ1014244