Research Report
2018-1

Differences in Improvement Trends across Arkansas School Districts

CHRYS DOUGHERTY, PHD

ACT.org

Acknowledgements

The author would like to thank Joann Moore, Paul Westrick, and Krista Mattern for their helpful comments on drafts of the report. In addition, I would like to thank the staff of the Arkansas Department of Education for making available the data that made this report possible.

Author

Chrys Dougherty, PhD

Chrys Dougherty was a principal research scientist in Statistical and Applied Research at ACT.

Table of Contents

Table of Figures

Table of Tables

Abstract

This report looks at differences in test score improvement in Arkansas school districts between above- and below-average-improving districts. Test score improvements were measured relative to what would be statistically predicted given a district's student demographics, so that above-average-improving districts improved faster than a typical district would with the same student population. The report looks at trend improvement in performance relative to predicted over a four- or five-year period (depending on availability of consistent data), because long-term trend data are less subject to the influence of year-to-year random fluctuations and are more likely to show statistically and practically significant change over time. The report focuses on school districts as these are typically the smallest administrative unit in charge of students' educations from preschool or kindergarten through 12th grade. The report also examines the issue of consistency: whether districts that were above-average-improving in one subject or grade were also above-average-improving in others.

The data showed that differences in improvement statistics between above- and below-average-improving districts were substantial: Over time, the difference between those two groups of districts widened by an amount equivalent to between one and two years' typical student growth. However, few districts were above-average-improving in more than one subject in the same grade, and even fewer districts were above-average-improving in more than one grade in the same subject. Improvement statistics in the same grade were often moderately correlated, particularly across "similar" subjects such as English and reading, but improvement in the same subject in different grade levels (e.g. 4th, 8th, and 12th grade) was essentially uncorrelated. Thus, educators and policymakers trying to understand district improvement trends need to look at improvement in specific subjects and grade levels.

Differences in Improvement Trends across Arkansas School Districts

Chrys Dougherty, PhD

Introduction

Discussions of school district improvement among school board members and in the news media often focus on year-to-year improvement, i.e., whether this year's test scores were better than last year's. However, these year-over-year improvement numbers are often not statistically reliable, particularly for small districts or schools (Sawyer, 2013). This report looks at trend improvement in test scores relative to predicted over a four- or five-year period, depending on the availability of data from the same test. Each year's test score improvements were measured relative to what would be statistically predicted given a district's student demographics, so that above-average-improving districts improved faster than what would be typical for a district with the same student population. The report focuses on school districts, as these are typically the smallest administrative unit in charge of students' educations from preschool or kindergarten through 12th grade.

Specifically, the report looks at whether differences in improvement between above- and below-average-improving districts in specific grades (4, 8, and 12) and subject areas (English, mathematics, reading, and science) were large enough to be of practical importance. The report also examines the issue of consistency: whether districts that were above-average-improving in one subject or grade were also above-average-improving in others. If improvement is caused by subject- and grade-specific changes in curriculum and instruction– for example, changes in the mathematics program in the upper elementary grades– then the improvement is likely to be grade- and subject-specific and not necessarily correlated with improvement in other subjects or in the same subject in other grades (e.g., in high school mathematics). On the other hand, district-wide process improvement– for example, better aligned K-12 curriculum, or expanded use of coaching or professional learning communities in every school– might result in simultaneous improvement across multiple grade levels and subject areas.[1]

The report addresses the following four questions:

1. How did the change in student scores over time and the number of districts with improving or declining scores vary by grade and subject?

2. What was the difference in improvement between above- and below-average-improving districts in each grade and subject?

3. How consistent were district improvement statistics across subjects at the same grade level?

4. How consistent were district improvement statistics across grade levels in the same subject?

In this report, the phrase "district improvement statistics" is used to refer to model-generated estimates of each district's improvement in student test scores over time relative to the scores that would be predicted based on the district's student characteristics. This may be thought of as a measure of whether a district is above-average-improving based on the district's student characteristics. These trends may differ from trends based on non-statistically adjusted student scores. For example, unadjusted scores might improve in every district over time, but not every district can be above average in improvement. In general, trends in unadjusted student scores are better at addressing the question, “How much did student achievement improve?” District improvement statistics based on student characteristics are better at addressing the question, “Which districts improved student achievement more than others, after adjusting for student demographics?” In all cases, districts were classified as above-average-, average-, or below-average-improving based on these improvement statistics, not their unadjusted changes in student scores.[2]

Data

This report used longitudinal cohorts created from student-level enrollment and test data supplied by the Arkansas Department of Education for the 2006-07 through the 2014-15 school years. The test data included student scores in literacy and mathematics on the Arkansas Benchmark Exams (ABE) in grade 4, and English, mathematics, reading, and science on the ACT Explore® tests in grade 8 and the ACT® test in grades 11 and 12.[3] All enrollment and test datasets contained state-encrypted student IDs so that records for the same students could be linked anonymously across enrollment and test datasets.

The statistical analysis used data both on students' individual demographic characteristics and the demographics of their districts. This required three steps to create the necessary datasets: (1) create longitudinal cohorts of students to be included in the analysis; (2) calculate district-level statistics and apply rules for including districts in the analysis; and (3) merge student- and district-level datasets together based on the district in which each student was enrolled in the initial cohort year.

The analysis used longitudinal cohort data on students who had been enrolled in the same district for multiple years— as opposed to using snapshot data on all students enrolled in the final year— for three reasons. The first was to focus on those students whose test results would be more likely to reflect the instructional program in the district where they were tested. The second reason was to make the data comparable to that used in other studies (Dougherty & Shaw, 2016; Dougherty & Shaw, 2017a; Dougherty & Shaw, 2017b). Third, the use of longitudinal cohort data following students forward from an initial grade (8th grade in the case of the grades 11-12 analysis) made it possible to account for higher student attrition in some districts in the high school grades.

1. Creation of Student Cohorts and Identification of Student Characteristics

This section describes how longitudinal cohorts of students were created for the analysis and how their demographic characteristics were derived from the Arkansas state enrollment and test data.

Creation of student cohorts. The analysis of 4th-grade test scores began with students enrolled in kindergarten in the initial cohort year (for example, the 2006-07 school year) and followed them forward for four subsequent school years, keeping students who took both 4th-grade ABE tests four years later and who were continuously enrolled in the same district during the entire period. Using this process, four kindergarten through 4th-grade (K-4) cohorts were created, referred to as the 2007-2011, 2008-2012, 2009-2013, and 2010-2014 cohorts.[4]

Likewise, five grades 4-8 cohorts were created for 2007-2011 through 2011-2015 by following 4th-grade students forward for four subsequent years, keeping students who were enrolled in the district the entire time and who took both 4th-grade ABE tests in the initial cohort year and all four ACT Explore tests in 8th grade in the final cohort year. Likewise, students in five grades 8-12 cohorts were followed from 8th grade forward for four subsequent years, keeping students who were 12th-graders four years later, who had been enrolled in the district the entire time, and who took all four ACT Explore tests in the initial cohort year and all four ACT tests in grade 12 in the final cohort year or in grade 11 in the next to final year. The most recent score was used for students who took the ACT more than once in grades 11 and 12. These rules for selecting students make the cohorts in this analysis similar to those analyzed in previous studies (Dougherty & Shaw, 2016, 2017a, 2017b).

Thus, the process created four longitudinal K-4 cohorts and five cohorts each in grades 4-8 and 8-12, or 14 total student cohorts. At each grade level (K-4, 4-8, and 8-12), the student-level cohorts were concatenated into a single dataset, and students who were retained in the initial grade were counted in the cohort beginning in the year they were retained. We also created an indicator variable for those retained students. In all, the 146,048 students in the four K-4 cohorts, 182,021 students in five 4-8 cohorts, and 179,409 students in five 8-12 cohorts comprised respectively 49%, 51%, and 24% of the students enrolled in the initial cohort years (Appendix A, Tables A1 – A3).

Identification of students' demographic and program participation status. Student characteristics may vary over time: For example, a student's family may qualify for the free and reduced price lunch program when the student is in 4th grade but not when the same student is in 8th grade. Likewise, a student's special education or English language learner status or a student's self-identified ethnicity may change over time. Because an indicator of low-income, English language learner, or special education status may signal a level of disadvantage even if the status is not consistent every year, students were identified as low-income, English language learners, or special education if they had that status in either the initial or final cohort grade level (e.g. either kindergarten or 4th grade for the students in the grade 4 analysis). Because no such logic applies to inconsistent reporting of student ethnic status, the student's reported ethnicity in the earliest cohort year was used as the determining factor for the student's overall ethnic status.[5] Students whose ethnicity, low-income status, special education status, or English language learner status could not be ascertained using these criteria were dropped from the analysis.[6]

2. District-Level Statistics and Inclusion of Districts in the Analysis

The calculation of district-level statistics began with 236 K-12 school districts that were in existence continuously from the 2006-07 through the 2014-15 school years.[7] Since the focus was on traditional K-12 school districts, charter schools that were not part of such a district were omitted from the analysis.[8] For these 236 districts, statistics were calculated on district-wide demographics and the district's number and percentage of students included in the analysis. Next, districts were classified as rural or non-rural, and districts that met criteria for eligibility for the analysis were divided into the three poverty categories used in this report.

District-wide demographics. Each district's fall student-level enrollment data for kindergarten through 12th grade for each year from 2006-07 through 2010-11 were used to derive annual statistics on the overall district-wide percentage of low-income, African American, Hispanic, Asian, White, Native American, English language learner, and special education students. These statistics were used as district-level predictors in the statistical models.[9]

District percentage of students in the analysis. For each cohort, the size of the cohort in each district was calculated as a percentage of the total number of students enrolled in the district in the initial cohort grade and year. A low district percentage of students in the analysis– reflecting a high rate of student mobility or a low percentage of students taking the test– might either raise or lower the estimated relative performance of the district. For example, if students whose families face the most difficulties leave the sample in disproportionate numbers, that could bias the results in favor of districts with high attrition.[10] High attrition could also result from the presence of a nearby military base or from the district's being less effective at retaining and educating students. In the last of these cases, controlling for attrition rates in the statistical model picks up some of the district performance the analysis is trying to measure. Further research may explore the variables that are associated with student attrition to identify when attrition should be treated as a district performance indicator (e.g., as in the case of high school dropout rates) and when it is simply an aspect of the environment in which the district operates.

Rural district status. Using school-level information from the 2013 Common Core of Data,[11] rural districts were defined as those in which all schools had a two-digit NCES locale code beginning with 3 (small town) or 4 (rural).

Selection of districts for inclusion in the analysis. Two additional criteria were used to identify which of the 236 continuously existing regular K-12 districts should be included in the analysis:

1. *Accuracy of low-income statistics.* The use of students' low-income status as an important control in the statistical models made the accuracy of this classification an important consideration. To assess the accuracy of each district's low-income statistics in a given year, districts' overall percentages of low-income students in that year were regressed on Census estimates of poverty rates of individuals age 5-17 in the district to get a statewide relationship between the two variables, which in turn yielded a Census-predicted district low-income percentage for each year.[12] To have its students included in the analysis, a district's percentage of low-income students in kindergarten through grade 12 had to fall within 20 percentage points of its Census-predicted value in each school year from 2006-07 through 2009-10, the starting years for the first four cohorts in this report. 201 out of 236 Arkansas K-12 districts met this requirement.

2. *Number of students in the analysis.* To be included in the analysis for a given grade level (4, 8, or 11-12), districts were required to have at least 20 total students in the first two longitudinal cohorts combined and at least 20 in the last two (or three) longitudinal cohorts for that grade level. In grades 4 and 8, 200 districts that met the income data requirement also met this criterion, despite the fact that Arkansas has many small districts (Appendix B, Figure B1 and Table B1).[13] Because of low ACT Explore and ACT participation rates, 67 of the 201 districts meeting the low-income data criterion did not have enough students for the grades 11-12 analysis, leaving 134 eligible districts at that grade level.[14]

3. Combining Student- and District-Level Data

At each grade level (K-4, 4-8, and 8-12), the file of district-level data created in the previous step was merged into the single concatenated student-level file for that grade level, based on the district in which each student was continuously enrolled. This process created a single dataset at each grade level with matched student- and district-level data.

Methods

Once the datasets for the study were built, the analysis had five steps: (1) fit simple linear trends at the statewide and district level to estimate statewide and district-level improvement trends in scores for each subject and grade level; (2) use statistical models to estimate district improvement statistics for each subject and grade level; (3) use these improvement statistics to classify districts into above-average, average, or below-average improvement categories by subject and grade level; (4) calculate additional district-level descriptive statistics based on the percentage of students On Track in earlier and later cohorts; and (5) aggregate the district performance statistics and student achievement statistics by district performance category to address the research questions in the study. These steps are described here.

1. Estimation of Score Trends across Cohorts

To estimate the average yearly change in test scores across the four or five cohorts at a given grade level, ordinary least squares (OLS) regression was used to regress student scores in each grade and subject on a linear trend variable. The trend variable was mean-centered so that its values were -1.5, -.0.5, 0.5, and 1.5, respectively, for the four consecutive K-4 cohorts and -2, -1, 0, 1, and 2 for the five consecutive 4-8 and 8-12 cohorts. With mean-centering, the intercept represents the average score across the five (or four) years. With two subjects in K-4 and four each in 4-8 and 8-12, this process used ten OLS regression models. In addition to statewide trend regressions, separate trend regressions were run by district (one per district for each of the 10 subject-grade combinations) to see if each district had statistically significantly improving or declining scores in the grade and subject in question.

2. Use of Statistical Models to Create District Improvement Statistics

The student- and district-level predictors shown in Table 1 were used to predict student-level scores on each of the two 4th-grade Arkansas Benchmark Exams (ABE), four 8th-grade ACT Explore tests, and four ACT tests for students in grades 11-12.[15] The models (one per subject and grade level) contained student-level predictors on students' low-income, ethnic, English language learner (ELL), and special education status, and district-level averages of these predictors. The district-level averages might be related to a school district's academic culture, funding, and priorities; these influences might in turn affect students' test scores. The models also contained predictors on the district's number and percentage of students included in the analysis, in order to explore the effects of district size and cohort attrition, respectively. The models also included a dummy variable for whether the district was located in a rural area, on the theory that that might affect teacher recruitment and thus, indirectly, student performance.

Because the models used in the study contained both student- and district-level predictors, they were estimated as hierarchical linear models.[16] Average district performance in each grade and subject was represented by a district random intercept, and district improvement was represented by a district random linear trend. In 4th grade, this process generated two sets of district improvement statistics, one for each ABE subject. In 8th and 12th grade, this process generated four sets of district improvement statistics at each grade level, one for each ACT Explore or ACT subject.

Table 1. Predictors in the Models for Grades 4, 8, and 11-12

Type of Data	Predictor
Student-Level	Intercept
	Low-income status
	African American status
	Hispanic status
	Asian status
	Native American status
	ELL status
	Special education status
	Flag for retained student*
	Linear trend (as in the OLS models)**
District-Level	% low-income students
	% African American students
	% Hispanic students
	% Asian students
	% Native American students
	% ELL students
	% special education students
	Number of students in model
	% of students in model
	Flag for rural district
	Random intercept
	Random linear trend (district improvement statistic)

* Retained students had enrollment records in consecutive initial cohort years. If the student met the other criteria for inclusion in the analysis based on the second initial cohort year, the student was included in that year's data and assigned a flag as a retained student.

** This is the mean-centered linear trend variable discussed above.

3. Classification of Districts into Improvement Categories

For a given subject and grade level, "above-average-improving" districts were defined as those whose improvement statistics fell in the top quintile for the grade and subject in question and also were statistically different from average at the .05 confidence level. Similarly, "below-average-improving" districts were those with improvement statistics in the bottom quintile and also different from average at the .05 confidence level. Districts not meeting these requirements– i.e., in the middle three quintiles, or in the top or bottom quintile but not statistically different from average at the .05 level– were classified as average-improving.[17]

The same district might fall into different improvement categories in different subjects and grade levels. For example, a district could be above-average-improving in grade 4 mathematics and below-average-improving in grades 11-12 reading.[18] Thus, a performance category such as "above-average districts" comprised different districts depending on the grade and subject. Questions 3 and 4 in this analysis examine how often these sorts of inconsistencies occur.

4. Aggregation of Statistics by District Improvement Category

For each grade and subject, weighted averages of the district improvement statistics were calculated for districts in each improvement category. These statistics were used to calculate the difference in improvement between above- and below-average-improving districts in each grade and subject.

Limitations

Though this report looked at district improvement statistics, it was not possible in most cases to differentiate "district effects" from "school effects." Thus, no distinction is made between "improvement of grade X student achievement due to district-level effects" and "improvement in grade X student achievement due to school-level effects." The majority of Arkansas school districts are small and rural, and many districts have only one school at a given level. For example, in 2014-15, 116 (87%) of the 134 districts in the grades 11-12 analysis had only one high school serving grades 11 and 12. Likewise, 171 of the 200 districts in the grade 8 analysis had only one school serving 8th grade, and 149 of the 200 districts in the grade 4 analysis had only one school serving 4th grade. Thus, for the great majority of Arkansas districts, improvement statistics in grades 11-12 could also be thought of as an indicator of improvement of the district's single high school, likely contributed to in part by the earlier improvement of its feeder elementary and middle school(s). The comparable statistic for 8th grade could be used as an indicator of the improvement of the district's single middle or junior high school and the earlier improvement of its feeder elementary school(s).[19] The value of treating the district as the unit of analysis is to examine consistency of improvement across grade levels within the same district and to draw attention to the fact that students' earlier schools are likely to have contributed to improvement in achievement levels in grades 8 and 11-12.

Second, the report does not compare the wide range of statistical models that could be used to generate district improvement statistics. The goal was to examine results from relatively straightforward models that control for generally available student- and district-level demographic statistics. These models were not refined to eliminate variables that did not add much explanatory power to the models.

Third, measured district improvement differences were studied with the understanding that these differences may reflect the effects both of educator practices and of unmeasured student, parent, and community influences that were not picked up as controls in the statistical analysis (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010). For example, some districts may operate in more favorable community environments than other districts with similar student demographics. Thus, the measured improvement differences such as those discussed in this report should be treated as the starting point for further inquiry into why these differences exist and what can be done to improve student outcomes in all school systems.

Fourth, because the data in this report are for Arkansas students and districts, further research in other states is needed to determine how the results generalize across states.

Finally, because of concerns about the statistical reliability of results from small student groups, results are not reported for groups of less than 20 students.

Results

Question 1: How did the statewide change in average scores and the number of districts with improving or declining scores vary by grade and subject?

Table 2 shows that statewide scores rose on average in 4th-grade literacy on the Arkansas Benchmark Exams and in grades 11-12 on the ACT in all subjects. Scores declined during the period in 4th-grade mathematics and in all subjects on the 8th-grade ACT Explore.[20] The average rate of change per year is shown to permit a comparison between the rates of change in grade 4, which spanned four cohorts, and in grades 8 and 11-12, spanning five cohorts. Score changes in Table 2 are measured in standard deviation units to facilitate comparisons across tests that have different score scales.[21]

Table 2. Statewide Trend in Unadjusted Student Scores

Grade	subject	average annual score change*	cumulative score change**	number of districts with improving scores***	number of districts with declining scores***
4	Literacy	0.04	0.11	42	5
	Math	-0.07	-0.20	6	77
8	English	-0.01	-0.04	9	28
	Math	-0.06	-0.23	2	81
	Reading	-0.02	-0.07	10	27
	Science	-0.01	-0.03	17	19
11-12	English	0.01	0.02	4	11
	Math	0.01	0.02	0	9
	Reading	0.02	0.08	10	4
	Science	0.02	0.07	6	4

N = 200 districts in grades 4 and 8 and 134 districts in grades 11-12.

* Statewide score changes were statistically significant at the .01 level except for high school English and mathematics, which were significant at the .10 level.

** Cumulative score change = three times average annual score change in grade 4 and four times average annual score change in grades 8 and 11-12. Apparent discrepancies in the table are due to rounding: unrounded annual changes are multiplied by three or four and the results were rounded off.

*** The .05 significance level was used to identify districts with improving or declining score trends.

To get an idea of the magnitude of the score changes shown in Table 2 cumulated over multiple cohorts, the 4th-grade average annual change can be multiplied by three and the 8th- and 12th-grade average annual changes by four to encompass the cumulative change between the first and last cohorts in the analysis. In turn, the 4th-grade cumulative changes in Table 2 of approximately 0.11 of a standard deviation in literacy and -0.20 of a standard deviation in mathematics may be compared with average growth per year between grades 4-8 on the ABE exam ranging from 0.24 to 0.30 standard deviations in literacy and from 0.27 to 0.35 standard deviations in mathematics (depending on student cohort), calculated for the three study cohorts in Dougherty and Shaw (2016). Thus, the cumulative change in 4th-grade literacy amounts to a gain in a range of roughly a third to just under a half of a year's typical growth in literacy; likewise, the cumulative change in 4th-grade mathematics amounts to a decline in a range of just under three-fifths to about three-quarters of a year's typical growth in mathematics.

Likewise, the cumulative 8th- and 12th-grade annual changes may be compared with average growth per year between the ACT Explore and ACT exams of 0.26 standard deviations in English, 0.28 in mathematics, 0.31 in reading, and 0.24 in science (ACT, 2012), using the average of the ACT Explore and ACT standard deviations to convert typical growth in score points to standard deviations.[22] Thus, the cumulative 8th-grade score declines amount to around four-fifths of a year in mathematics, about a fifth of a year in English and reading, and about a tenth of a year in science. 12th-grade cumulative score increases amounted to about a tenth of a year in English and mathematics to around a quarter of a year in reading and science.

Score changes in individual districts did not always follow the statewide trend: for example, in 8th-grade mathematics where average scores statewide declined, scores increased in two districts (Table 2). The majority of districts did not have statistically significant score changes at the .05 level, likely due in large part to smaller districts' relatively low numbers of students in the analysis. For example, 117 (= 200 – 81 – 2) districts did not have a statistically significant score trend in 8th-grade mathematics.

Question 2: What was the difference in improvement between above- and below-average-improving districts in each grade and subject?

Table 3 shows the difference in the improvement statistics of above- and below-average improving districts. These statistics represent the estimated difference between annual improvement in those districts and the improvement that would be predicted for a district with the same student characteristics.

Since the differences shown in the last column of Table 3 represent differences in yearly improvement, they should be multiplied by three (in the case of 4th grade) and four (in the case of 8th or 11th and 12th grades) to represent cumulative improvement between the first and last cohorts. Rounded off, this gives estimated improvement differences of 0.46 and 0.59 of a standard deviation in 4th-grade literacy and mathematics, respectively, and differences of 0.32, 0.49, 0.38, and 0.51 standard deviations in 8th-grade English, mathematics, reading, and science (Table 4).

In turn, the cumulative 4th-grade differences may be compared with average growth per year between grades 4-8 on the ABE exam ranging from 0.24 to 0.30 of a standard deviation in literacy and from 0.27 to 0.35 of a standard deviation in mathematics (Table 4). Using this metric, the cumulative differences in 4th-grade literacy and mathematics improvement represent from 1.5 to 1.9 years' growth in literacy and 1.7 to 2.2 years' growth in mathematics.

Likewise, the cumulative 8th-grade differences may be compared with average student growth between ACT Explore and the ACT of 0.26, 0.28, 0.31, and 0.24 standard deviations in English, mathematics, reading, and science, respectively (Table 4). Thus, the cumulative differences in 8th grade represent approximately 1.2, 1.8, 1.1, and 2.1 years' growth in those four respective subjects.

Table 3. Difference in Relative Improvement Statistics between Above- and Below-Average-Improving Districts

		Above-average-improving		Below-average-improving		
Grade	subject	number of districts	average of improvement statistics	number of districts	average of improvement statistics	difference in improvement statistics
4	Literacy	12	0.08	14	-0.07	0.15
	Math	20	0.10	17	-0.10	0.20
8	English	3	0.04	4	-0.04	0.08
	Math	12	0.06	15	-0.06	0.12
	Reading	6	0.05	9	-0.04	0.09
	Science	9	0.06	7	-0.07	0.13
11-12	English	1	0.04	0	--	--
	Math	0	--	2	-0.03	--
	Reading	1	0.04	0	--	--
	Science	1	0.03	0	--	--

N = 200 districts in grades 4 and 8 and 134 districts in grades 11-12.

Table 4. Comparison of Cumulative Relative Improvement Differences with Typical Yearly Student Growth

grade	subject	number of districts	difference in improvement per year	number of years	cumulative improvement difference	comparison: one year's typical student growth
4	Literacy	12	0.15	3	0.46	0.24-0.30
	Math	20	0.20	3	0.59	0.27-0.35
8	English	3	0.08	4	0.32	0.26
	Math	12	0.12	4	0.49	0.28
	Reading	6	0.09	4	0.38	0.31
	Science	9	0.13	4	0.51	0.24

Question 3: How consistent were district improvement statistics across subjects at the same grade level?

Table 5 shows moderate to high correlations between district improvement in different subjects in the same grade level, ranging from .38 between improvement in high school reading and mathematics to .81 between improvement in high school English and reading. Not surprisingly, more closely related subjects (e.g., English and reading) had the highest improvement correlations. Improvement correlations between reading and science were higher than those between mathematics and science, which may be less surprising given the heavy reading component of the ACT Explore and ACT science tests.

A substantial number of districts showed above- or below-average improvement in individual 4th- and 8th-grade subjects (Table 5); for example, 29 districts had above-average and 26 districts below-average improvement in either 4th-grade literacy or mathematics. However, despite the cross-subject correlations in improvement statistics, few districts showed statistically reliable above- or below-average improvement in more than one subject; in 4th grade, for example, only three districts showed above-average and five districts below-average improvement in both literacy and mathematics. In high school, few districts showed statistically significant improvement in any subject. However, one district was above-average-improving in English, reading, and science.

Table 5. Consistency of District Improvement Statistics across Subjects

			number of districts above average		number of districts below average	
Grade	Subject Comparison	Correlation of Improvement Statistics*	in either subject	in both subjects	in either subject	in both subjects
4	Literacy, Math	.51	29	3	26	5
8	English, Math	.44	13	2	18	1
	English, Reading	.76	9	0	10	3
	English, Science	.62	10	2	11	0
	Math, Reading	.39	18	0	23	1
	Math, Science	.55	18	3	18	4
	Reading, Science	.75	12	3	14	2
11-12	English, Math	.42	1	0	2	0
	English, Reading	.81	1	1	0	0
	English, Science	.74	1	1	0	0
	Math, Reading	.38	1	0	2	0
	Math, Science	.58	1	0	2	0
	Reading, Science	.77	1	1	0	0

* All correlations were significant at the .01 level. N = 200 districts in grades 4 and 8 and 134 districts in grades 11-12.

Tables 6 and 7 examine how many districts were above- or below-average-improving in different numbers of subjects. In 4th grade, out of 29 districts that were above-average-improving in at least one subject, three had above-average improvement in both subjects (Table 6). Of 26 districts with below-average improvement in at least one subject, five had below-average improvement in both subjects (Table 7).[23] In 8th grade, out of 22 districts that were above-average-improving in at least one subject, four had above-average improvement in two subjects, two districts in three subjects, and none had better-than-average improvement in all four subjects (Table 6). Of the 25 districts that were below-average-improving in one or more subjects, eight had below-average improvement in two subjects, one in three subjects, and none in all four subjects (Table 7).

Table 6. Number of Districts by Number of Subjects with Above-Average Improvement

	Number of subjects with above-average improvement				
Grade	0	1	2	3	4
4	171	26	3	N/A	N/A
8	178	16	4	2	0
11-12	133	0	0	1	0

Table 7. Number of Districts by Number of Subjects with Below-Average Improvement

	Number of subjects with below-average improvement				
Grade	0	1	2	3	4
4	174	21	5	N/A	N/A
8	175	16	8	1	0
11-12	132	2	0	0	0

Question 4: How consistent were district improvement statistics across grade levels in the same subject?

Table 8 shows low correlations between district improvement statistics in the same subject in different grade levels; for example, there was little relationship between districts' improvement statistics in mathematics on the grade 4 ABE, grade 8 ACT Explore, and grades 11-12 ACT. Thus, few districts that were above-average-improving in a subject at one grade level were above-average-improving in the subject at another grade level, and the same was true for below-average-improving districts.

Table 8. Consistency of District Improvement Statistics across Grade Levels

Grade Comparison	Subject	Correlation of Improvement Statistics*	number of districts above average: in either grade	number of districts above average: in both grades	number of districts below average: in either grade	number of districts below average: in both grades
4 vs. 8	Literacy/Reading	0.00	17	0	20	3
	Mathematics	0.04	30	2	30	2
4 vs. 11-12	Literacy/Reading	0.01	11	0	12	0
	Mathematics	-0.13	13	0	12	1
8 vs. 11-12	English	0.04	3	0	2	0
	Mathematics	-0.06	7	0	13	0
	Reading	-0.13	6	0	6	0
	Science	0.07	6	1	6	0

* No correlation was significant at the .10 level or better.

N = 199 districts in the grades 4 vs. 8 comparisons and 133 districts in the other comparisons.

Tables 9 and 10 examine the number of grades levels (0, 1, 2, or 3) in which districts were above- or below-average-improving by subject area. From Table 9, of 19 above-average-improving districts in mathematics and 15 in literacy/reading at least one level, only one was above-average-improving in mathematics and none was above-average in literacy/reading in at least two grade levels. Of 21 below-average-improving districts in mathematics and 15 in literacy/reading, three in each subject were below-average-improving in two grade levels (Table 10). These low counts of districts that were above or below average in multiple grades are what would be expected if improvement in the same subject is uncorrelated across grade levels, as shown in Table 8.

Table 9. Number of Districts by Number of Grade Levels with Above-Average Improvement

	Number of grade levels with above-average improvement			
Subject	0	1	2	3
English	130	3	0	N/A
Mathematics	113	18	1	0
Reading/Literacy	117	15	0	0
Science	127	5	1	N/A

N = 132 districts for mathematics and reading/literacy (grades 4, 8, and 11-12); 133 districts for English and science (grades 8 and 11-12)

Table 10. Number of Districts by Number of Grade Levels with Below-Average Improvement

	Number of grade levels with below-average improvement			
Subject	0	1	2	3
English	131	2	0	N/A
Mathematics	111	18	3	0
Reading/Literacy	117	12	3	0
Science	127	6	0	N/A

N = 132 districts for mathematics and reading/literacy (grades 4, 8, and 11-12); 133 districts for English and science (grades 8 and 11-12)

Conclusion

After adjusting for predicted improvement differences due to student characteristics, the analysis found substantial differences in improvement between above- and below-average-improving districts. Not surprisingly, consistency of improvement across subjects in the same grade level was higher in similar subjects such as English and reading, and lower for less similar subjects such as reading and mathematics. However, relatively few districts that were above-average-improving in one subject in a grade level were above-average-improving in other subjects in the same grade. In addition, improvement was highly inconsistent across grade levels in the same subject; few districts that were above-average-improving in one grade level in a subject were above-average-improving in a different grade level in the same subject.

To understand why improvement occurs in one area more than another, district leaders can improve how well they keep track of changes in educational practices (Dougherty, 2016), and look at a variety of data indicators to determine when changes in practice are closely associated with changes in student outcomes (Dougherty, 2014). Keeping track of practices means collecting the information needed to understand what is going on in classrooms– how teachers respond to district initiatives and what practices they choose to implement. Gathering good information on implementation of practices in schools and classrooms requires a high level of trust between teachers and school and district leaders (Knight, 2007).

References

ACT. (2012). *Principles for measuring growth towards college and career readiness*. Iowa City, IA: ACT. http://www.act.org/content/dam/act/unsecured/documents/GrowthModelingReport.pdf.

ACT. (2013). *ACT Explore technical manual*. Iowa City, IA: ACT. http://www.act.org/content/dam/act/unsecured/documents/Explore-TechManual.pdf.

ACT. (2014). *The ACT technical manual*. Iowa City, IA: ACT. http://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf.

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

Dougherty, C. (2014). *Catching up to college and career readiness: The challenge is greater for at-risk students*. Iowa City, IA: ACT. http://www.act.org/content/dam/act/unsecured/documents/CatchingUp-Part3.pdf.

Dougherty, C. (2016). *Keeping track of improvement in educational practices*. Iowa City, IA: ACT. http://www.act.org/content/dam/act/unsecured/documents/5516-insights-keeping-track-of-improvement-in-educational-practices.pdf.

Dougherty, C. & Shaw, T. (2016). *Size and consistency of school district performance measures across models, subjects, and grade levels*. Iowa City, IA: ACT. http://www.act.org/content/dam/act/unsecured/documents/5293-Research-Report-2016-2-Size-and-Consistency-of-Performance-Measures.pdf.

Dougherty, C. & Shaw, T. (2017a). *Comparisons of student achievement levels by district performance and poverty*. Iowa City, IA: ACT. http://www.act.org/content/dam/act/unsecured/documents/R1612_Comparisons_of_Student_Achievement_Levels.pdf.

Dougherty, C. & Shaw, T. (2017b). *Comparisons of student growth by district performance and poverty*. Iowa City, IA: ACT. http://www.act.org/content/dam/act/unsecured/documents/R1615-growth-by-performance-poverty-2017-05.pdf

Knight, J. (2007). *Instructional coaching: A partnership approach to improving instruction.* Thousand Oaks, CA: Corwin Press.

Sawyer, R. (2013). *Interpreting changes over time in high school average ACT® college readiness assessment composite scores and ACT College Readiness Benchmark attainment rates*. Iowa City, IA: ACT. http://www.act.org/content/dam/act/unsecured/documents/ACT_RR2013-9.pdf.

Appendix A: Descriptive Statistics on Students in the Analysis

Tables A1-A3 show the number and percentage of students from each cohort who were included in the statistical analysis. In 4th grade, there was a modest decrease in the percent of cohort students in the analysis (from 53% to 46%); little change in 8th grade; and a modest increase (from 20% to 26%) in the percentage of students in the analysis in grades 11-12. However, it is unlikely that these changes in percentages were large enough to have had much impact on the score trends described in the paper.

Tables A4 and A5 compare student attrition in the grades 4 and 8 and grade 8 and 11-12 analyses, respectively. Comparing grades 4 and 8 (Table A4), enrollment attrition, defined as students not enrolled in the expected grade four years later, was higher between kindergarten and grade 4 than between grades 4 and 8. Higher enrollment attrition in the grade 4 analysis was likely due to a larger number of retained students in kindergarten and 1st grade who were not picked up in a later cohort. However, this effect was offset by a higher percentage of students in the 8th-grade analysis not taking the ACT Explore test than students in the 4th-grade analysis not taking the ABE, so the overall percentages of students in the analysis were similar in grades 4 and 8.

Comparing grades 8 and 11-12 (Table A5), higher enrollment attrition in high school was likely a result of students dropping out. In addition, fewer students in the grades 11-12 analysis took ACT Explore in grade 8 than the percentage of students in the grade 8 analysis who took the ABE in grade 4, resulting in attrition of 24% from the grades 11-12 analysis versus only 3% for the grade 8 analysis (Table A5). An additional 13% attrition in the grades 11-12 analysis was accounted for by students not taking the ACT.

Table A1. Percentage of Arkansas Kindergarten Students in Grade 4 Analysis

Student cohort*	Total kindergarten enrollment	Students tested in 4th grade	Students eligible for statistical analysis	Eligible students in eligible districts	Percent of students in statistical analysis
2007-2011	33,072	25,783	19,810	17,563	53%
2008-2012	35,950	27,499	21,034	18,045	50%
2009-2013	37,354	28,183	21,211	18,154	49%
2010-2014	39,672	27,278	21,026	18,203	46%
Total	146,048	108,743	83,081	71,965	49%

* For example, the 2007-2011 cohort consists of students who were enrolled in kindergarten in the 2006-07 school year and who took the Arkansas Benchmark Exams in 4th grade in the 2010-11 school year.

Table A2. Percentage of Arkansas 4th-Grade Students in Grade 8 Analysis

Student cohort	Total 4th-grade enrollment	Students tested in 4th and 8th grade	Students eligible for statistical analysis	Eligible students in eligible districts	Percent of students in statistical analysis
2007-2011	34,570	25,512	20,284	17,602	51%
2008-2012	35,418	26,499	21,136	18,332	52%
2009-2013	37,954	28,018	21,887	18,864	50%
2010-2014	37,556	26,478	21,340	18,812	50%
2011-2015	36,523	28,745	22,204	19,242	53%
Total	182,021	135,252	106,851	92,852	51%

Table A3. Percentage of Arkansas 8th-Grade Students in Grades 11-12 Analysis

Student cohort	Total 8th-grade enrollment	Students tested in 8th and 11th or 12th grade	Students eligible for statistical analysis	Eligible students in eligible districts	Percent of students in statistical analysis
2007-2011	34,810	10,020	8,470	6,928	20%
2008-2012	35,421	10,768	9,181	8,023	23%
2009-2013	36,767	14,078	11,463	9,148	25%
2010-2014	36,536	15,348	13,241	10,033	27%
2011-2015	35,875	17,370	14,954	9,368	26%
Total	179,409	67,584	57,309	43,500	24%

Table A4. Comparing Attrition in the Grade 4 and Grade 8 Analysis

	Grade 4		Grade 8	
Student population	Number of students*	% of students in initial grade	Number of students	% of students in initial grade
all enrolled students in the initial grade	146,048	100%	182,021	100%
...with complete demographic information	145,611	99.7%	181,908	99.9%
...and enrolled in final grade four years later	112,032	77%	153,138	84%
...and taking all tests in initial grade	N/A	N/A	147,788	81%
...and taking all tests in final grade	108,743	74%	135,252	74%
...and continuously enrolled in the district	83,081	57%	106,851	59%
...and in an eligible district	71,965	49%	92,852	51%

* Four K-4 cohorts and five 4-8 cohorts were included in the analysis.

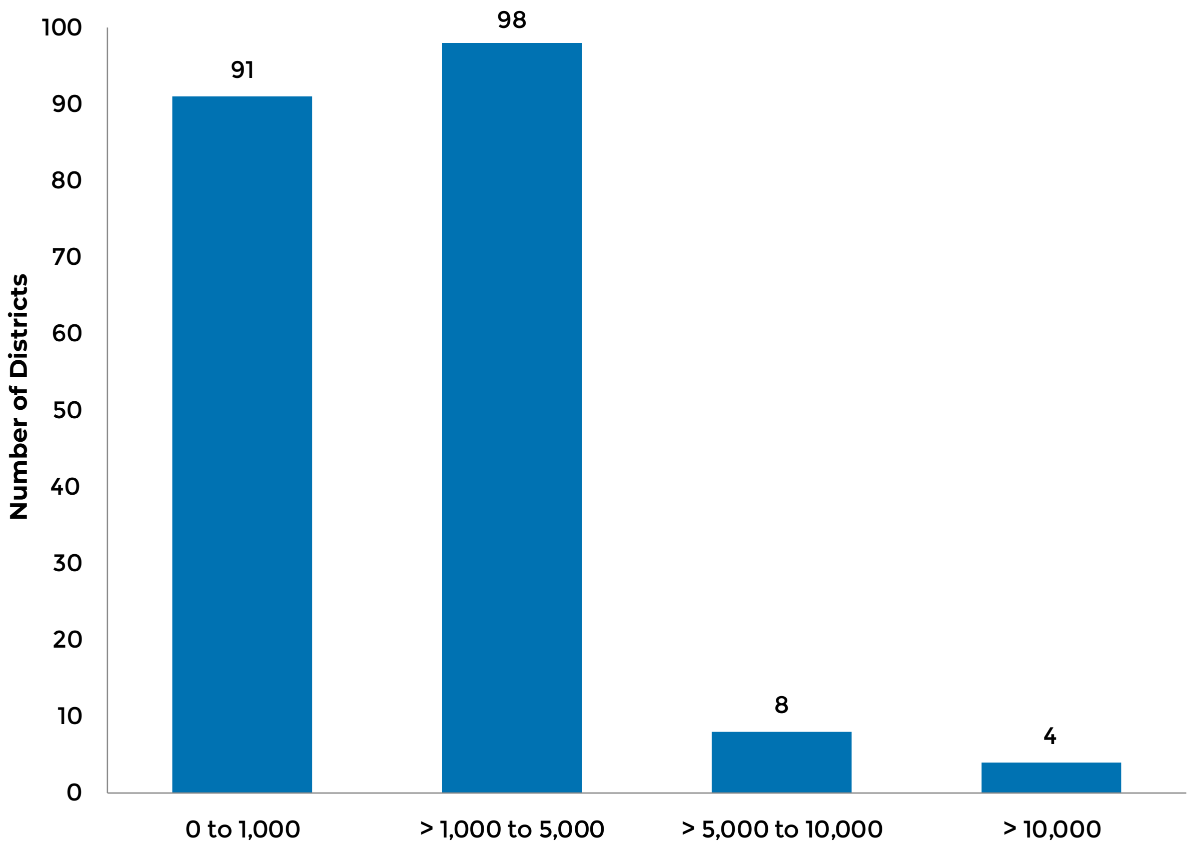
Table A5. Comparing Attrition in the Grade 8 and Grade 11-12 Analysis

	Grade 8		Grades 11-12	
Student population	Number of students*	% of students in initial grade	Number of students	% of students in initial grade
all enrolled students in the initial grade	182,021	100%	179,409	100%
...with complete demographic information	181,908	99.9%	179,256	99.9%
...and enrolled in final grade four years later	153,138	84%	133,804	75%
...and taking all tests in initial grade	147,788	81%	91,601	51%
...and taking all tests in final grade	135,252	74%	67,584	38%
...and continuously enrolled in the district	106,851	59%	57,309	32%
...and in an eligible district	92,852	51%	43,500	24%

* Five cohorts were included in the analysis at each grade level.

Appendix B: Descriptive Statistics on Districts in the Analysis

Arkansas is a largely rural state whose largest district, the Little Rock School District, had approximately 25,000 K–12 students, averaged across the five initial cohort years.[24] Overall, the majority of Arkansas school districts were small; only 12 (6%) of the 201 eligible districts in the analysis had more than 5,000 students, while 91 (45%) had fewer than 1,000 students (Figure B1). 115 districts were in the medium-poverty category with 50–70% low-income students, while 40 and 44 districts respectively were in the lower- and high-poverty categories (Figure B2).



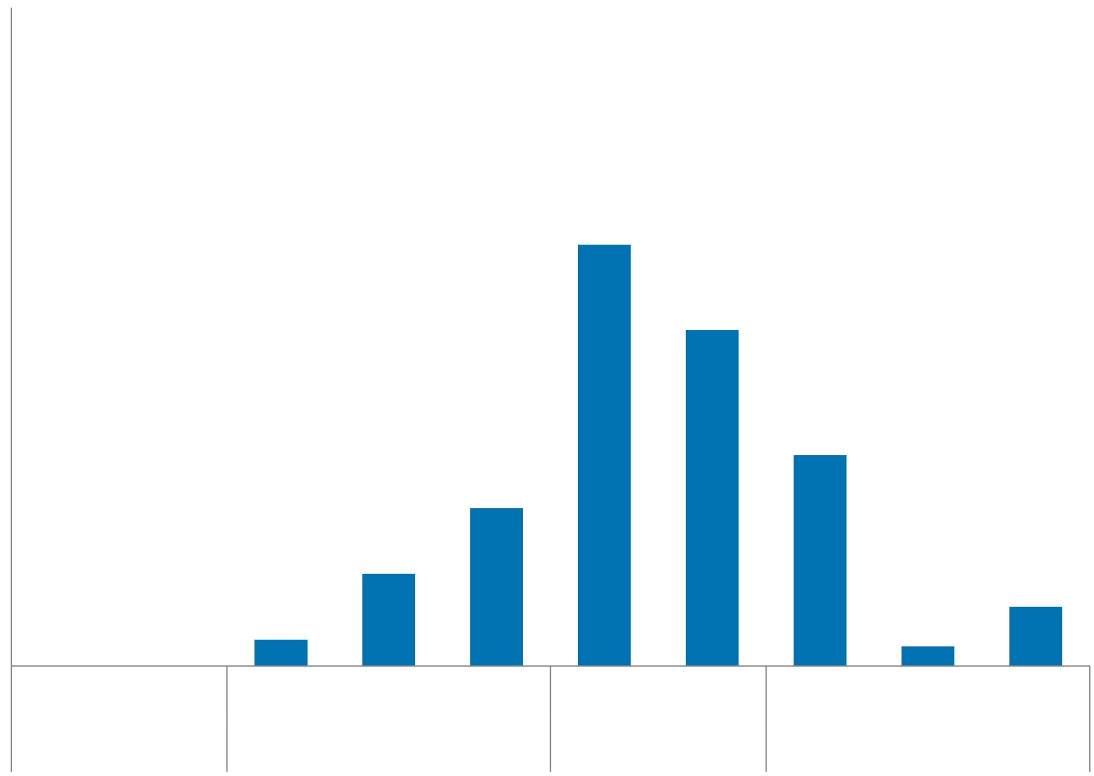


Figure B2. Distribution of eligible districts by their percentage of low-income students (N = 201 districts)

Disaggregating districts by size, poverty, and rural status, 175 (85%) of the eligible districts were rural, and rural districts made up 90% and 93% of the medium and high-poverty districts in the state (Table B1). Minority students were generally concentrated in high-poverty rural districts and medium- and high-poverty nonrural districts (Table B2).

Table B1. District Size by Poverty and Rural Status (N=201 districts)

Rural status	Poverty category	Number of districts	Median size	Size of largest district
Nonrural	Lower	15	3,084	12,586
	Medium	11	5,223	25,379
	High	3	3,737	4,994
Rural	Lower	27	1,080	4,238
	Medium	104	1,014	5,185
	High	41	784	3,537

Table B2. District Demographics by Poverty and Rural Status (N=201 districts)

rural status	poverty category	number of districts	% African American		% Hispanic		% White	
			median	highest	median	highest	median	highest
Nonrural	Lower	15	2	25	5	10	89	96
	Medium	11	41	68	6	41	50	91
	High	3	93	96	1	11	6	46
Rural	Lower	27	1	22	2	7	94	99
	Medium	104	1	55	2	26	93	99
	High	41	30	95	2	53	57	99

Appendix C: Fixed-Effect Coefficients from Statistical Models

Tables C1–C3 show the fixed-effect coefficients from the hierarchical models, measured in standard deviation units of student scores on the test used as the dependent variable. These are partial effects: For example, in Table C1, the fixed-effect coefficient for "low-income status" of -0.335 in 4th-grade literacy indicates that the predicted score of a low-income student is about 0.335 standard deviations (or about 62.4 scale score points) lower on the grade 4 literacy ABE than the predicted score of a non-low-income student who has the same values of the other variables in the model. No interaction effects were modeled. For example, the analysis did not model how a student's ethnicity might affect the differences in predicted scores between low- and non-low-income students.

The tables also show the standard deviation of the district relative performance statistics (random effects) in each model, labeled as "SD of district effect," and the standard deviation of the district improvement statistics (random trend effect), labeled as "SD of change in district effects." These statistics estimate the variation across districts in the true random effects. For example, in Table C1, the "SD of change in district effects" is 0.05 for 4th-grade literacy. Given the assumed normal distribution of the random effect in the model, for approximately two-thirds of the districts, the absolute value of the district's improvement statistic is 0.05 score standard deviations or less; for approximately 95% of the districts, the absolute value of the improvement statistic is 0.10 score standard deviations or less.

Table C1. Fixed-Effect Coefficients in Regressions Predicting Grade 4 Scores

Variable	Literacy		Mathematics	
Intercept	0.288	***	-0.058	
Low-income status	-0.335	***	-0.366	***
African American status	-0.323	***	-0.501	***
Hispanic status	-0.002		-0.063	***
Asian status	0.160	***	0.125	***
Native American status	-0.013		-0.045	
ELL status	-0.239	***	-0.183	***
Special education status	-0.920	***	-0.713	***
Retained student flag	-0.479	***	-0.455	***
District % low-income	-0.005	***	-0.005	***
District % African American	0.000		0.002	**
District % Hispanic	-0.010	**	-0.010	**
District % Asian	-0.023	**	-0.011	
District % Native American	-0.014		0.006	
District % ELL	0.021	***	0.021	***
District % special education	0.014	***	0.025	***
District # students in model†	-0.007		0.000	
District % students in model	0.156	**	0.195	**
Rural district	0.027		0.021	
Linear trend	0.079	***	-0.028	***
SD of district effect	0.16		0.18	
SD of change in district effect	0.05		0.07	

***Significant at the .01 level. **Significant at the .05 level. * Significant at the .10 level.

†100s of students

Table C2. Fixed-Effect Coefficients in Regressions Predicting Grade 8 Scores

Variable	English		Mathematics		Reading		Science	
Intercept	0.971	***	0.242	***	0.181	***	0.101	
Low-income status	-0.396	***	-0.377	***	-0.352	***	-0.344	***
African American status	-0.437	***	-0.427	***	-0.410	***	-0.371	***
Hispanic status	-0.032	**	0.010		-0.004		0.030	*
Asian status	0.208	***	0.248	***	0.217	***	0.238	***
Native American status	-0.073	**	0.021		-0.017		-0.015	
ELL status	-0.450	***	-0.355	***	-0.405	***	-0.357	***
Special education status	-0.829	***	-0.874	***	-0.676	***	-0.699	***
Retained student flag	-0.451	***	-0.445	***	-0.393	***	-0.400	***
District % low-income	-0.002	**	-0.004	***	-0.002	***	-0.003	***
District % African American	-0.001	**	0.000		-0.001	***	-0.001	
District % Hispanic	0.001		0.003		0.003		0.006	*
District % Asian	0.010		0.008		0.008		0.020	**
District % Native American	-0.009		-0.001		0.001		-0.003	
District % ELL	0.003		0.004		0.000		-0.003	
District % special education	0.006	**	0.008	***	0.007	**	0.006	**
District # students in model†	0.008		-0.003		0.023	***	0.014	
District % students in model	-0.299	***	-0.400	***	-0.374	***	-0.381	***
Rural district	-0.035		0.016		0.022		0.008	
Linear trend	0.000		-0.041	***	-0.004		0.003	
SD of district effect	0.11		0.14		0.12		0.13	
SD of change in district effect	0.03		0.04		0.03		0.04	

***Significant at the .01 level. **Significant at the .05 level. * Significant at the .10 level.

†100s of students

Table C3. Fixed-Effect Coefficients in Regressions Predicting Grades 11-12 Scores

Variable	English		Mathematics		Reading		Science	
Intercept	1.051	***	0.222	***	0.471	***	0.151	**
Low-income status	-0.372	***	-0.301	***	-0.272	***	-0.278	***
African American status	-0.636	***	-0.524	***	-0.624	***	-0.593	***
Hispanic status	-0.213	***	-0.139	***	-0.222	***	-0.205	***
Asian status	0.209	***	0.381	***	0.137	***	0.220	***
Native American status	-0.145	***	-0.147	***	-0.102	**	-0.150	***
ELL status	-0.625	***	-0.402	***	-0.598	***	-0.457	***
Special education status	-0.967	***	-0.701	***	-0.781	***	-0.727	***
Retained student flag	-0.482	***	-0.361	***	-0.391	***	-0.468	***
District % low-income	-0.003	**	-0.004	***	-0.003	***	-0.003	***
District % African American	0.001	*	0.002	**	0.000		0.001	
District % Hispanic	0.000		0.004		-0.002		0.000	
District % Asian	0.013		0.030	***	0.012		0.028	***
District % Native American	-0.001		0.002		0.011		-0.004	
District % ELL	0.007		0.002		0.008		0.005	
District % special education	-0.003		-0.002		-0.002		-0.003	
District # students in model†	0.004		0.001		0.002		0.003	
District % students in model	-0.384	***	-0.297	***	-0.395	***	-0.386	***
Rural district	-0.114	***	-0.062		-0.104	***	-0.089	**
Linear trend	0.013	**	0.012	**	0.026	***	0.021	***
SD of district effect	0.13		0.14		0.12		0.12	
SD of change in district effect	0.03		0.02		0.03		0.02	

***Significant at the .01 level. **Significant at the .05 level. * Significant at the .10 level.

†100s of students

Notes

1. In addition, due to the effect of cumulative learning in a cohort of students, high school improvement might be correlated with elementary school improvement several years earlier. Additional years of data would have been required to search for these lagged effects based on a multi-year trend several years earlier.
2. Cross-district comparisons may look different depending on whether district improvement statistics or unadjusted changes in student achievement levels are used. For example, a district with above-average improvement statistics, but more disadvantaged students, may improve by less in absolute terms than a district with average improvement statistics but more advantaged students.
3. The 4th-grade ABE data only extended through the 2013-14 school year. In 2014-15, Arkansas replaced ABE with tests provided by PARCC, the Partnership for the Assessment of Readiness for College and Careers. Results from these tests were not used in this study because the change in tests might make improvement trends difficult to interpret. Information on the PARCC tests may be found at http://parcc-assessment.org/ and https://parcc.pearson.com/.
4. In this nomenclature, school years are named after their spring semesters, so that students in the 2007-11 cohort were present in the district from the collection of enrollment data in the fall of the 2006-07 school year to the collection of test data in the spring of the 2010-11 school year.
5. The only exception was for a student with missing ethnic data for the earliest cohort grade (e.g., kindergarten in the 4th-grade analysis) but ethnic data present for the final cohort grade level (e.g., grade 4 in the 4th-grade analysis), in which case the data from the final grade level were used.
6. The percentage of records dropped due to incomplete data was around 0.3% in K-4, 0.1% in 4-8, and 0.1% in 8-12 (Appendix A, Tables A1 – A3).
7. If District A consolidated into District B at any time between the 2006-07 and 2014-15 school years, then A's students were combined with B's for the years prior to the consolidation and everyone was treated as part of District B. Thus, basing the analysis on the 236 districts that existed after consolidation did not, in itself, reduce the number of students in the analysis. The 236 districts did not include two statewide institutions, the Arkansas School for the Blind and Arkansas School for the Deaf.

8. Omitting students in charter schools that were not part of a K-12 district reduced the number of students in the analysis– after the other rules for inclusion were applied– by 336 students in grades K-4, 199 students in 4-8, and 86 students in 8-12.
9. The district-wide demographic statistics calculated this way differ from ones that would be calculated by aggregating the cohort data, which do not cover all grades. As was the case in the student cohorts, students with missing demographic data were dropped when calculating the district-wide statistics.
10. The poverty measure based on students' free and reduced price lunch status is an imperfect measure of those challenges, so using this measure as a predictor in the statistical models does not completely adjust for this possible bias.
11. School- and district-level datasets from the Common Core of Data may be downloaded from the NCES website at https://nces.ed.gov/ccd/ccddata.asp.
12. Census-defined poverty uses a lower income threshold than the state definition of low-income status, which is based on federal eligibility requirements for the free and reduced price school lunch program. Thus, it was necessary to derive a predicted low-income percentage from the Census data rather than just using the Census percentage. A district with accurate low-income data may be hypothesized to have a relationship between the two poverty measures that is not too different from the state average relationship between the two measures.
13. One district missed the criterion for the number of students in the analysis in 4th but not 8th grade, and a different district missed the criterion in 8th but not 4th grade.
14. The differences between the third and fourth columns of Tables A1 – A3 in Appendix A show the effect that removing ineligible districts (and charter schools) had on the number of students in the analysis. The percentages in the final column of Tables A1- A3 (and the bottom row of Tables A4 and A5) are based on the number of students in eligible districts.
15. The predictors used in these models are similar to those used in Dougherty and Shaw (2017a; 2017b).
16. Appendix C shows the fixed effects of the predictors in Table 2 estimated in each of these statistical models. SAS Proc Mixed was used for all of the statistical models in this report. Information on the SAS code used for the models is available on request.

17. In general, districts with smaller numbers of students in the analysis were more likely to be classified as average-improving. With fewer students, a larger change is required to reduce the suspicion that the change resulted from chance differences in student cohorts. For similar reasons, longer-term trends across multiple student cohorts provide stronger evidence of improvement than does a single year-over-year change between two consecutive cohorts (Sawyer, 2013).
18. Consistency in district performance levels across grades and subjects was explored in Dougherty and Shaw (2016). This report addresses consistency in rates of improvement.
19. In a state with a number of larger districts, one could partition the variance in performance across schools in those districts into the variance across districts and the variance across schools within districts.
20. The 8th-grade ACT Explore score trends in Table 2 were influenced by declines in average scores in the 2014-15 school year, when the state changed its main statewide accountability test from ABE to PARCC. ACT Explore scores may have been indirectly affected by changes in instruction resulting from this testing change. Excluding that school year, the average annual ACT Explore score change was 0.01, -0.02, 0.02, and 0.02 of a standard deviation in English, mathematics, reading, and science, respectively. All of these score changes were significant at the .01 level.
21. Standard deviations on the ACT Explore tests were 4.2 points in English, 3.5 in mathematics, 3.9 in reading, and 3.3 in science (ACT 2013, Table 4.11). Standard deviations on the ACT were 6.5 in English, 5.3 in mathematics, 6.3 in reading, and 5.3 in science (ACT 2014, Table 5.4). Standard deviations on the grade 4 ABE, calculated for all students tested in the 2006-07 and 2007-08 school years, were 186.37 scale score points in literacy and 100.93 in mathematics.
22. For example, average growth per year in mathematics = (4.7/3.77)/[(3.5+5.3)/2], where 4.7 is average growth across the average period of 45.2 months (= 3.77 years) between the ACT Explore and ACT tests, and 3.5 and 5.3 are the standard deviations of student scores on the two tests.
23. Since there were only two tested subjects in 4th grade, the same information also appears in Table 5.
24. District size and demographic percentages reported in this appendix are based on K–12 statistics averaged across the 2006–07 through the 2010–11 school years.