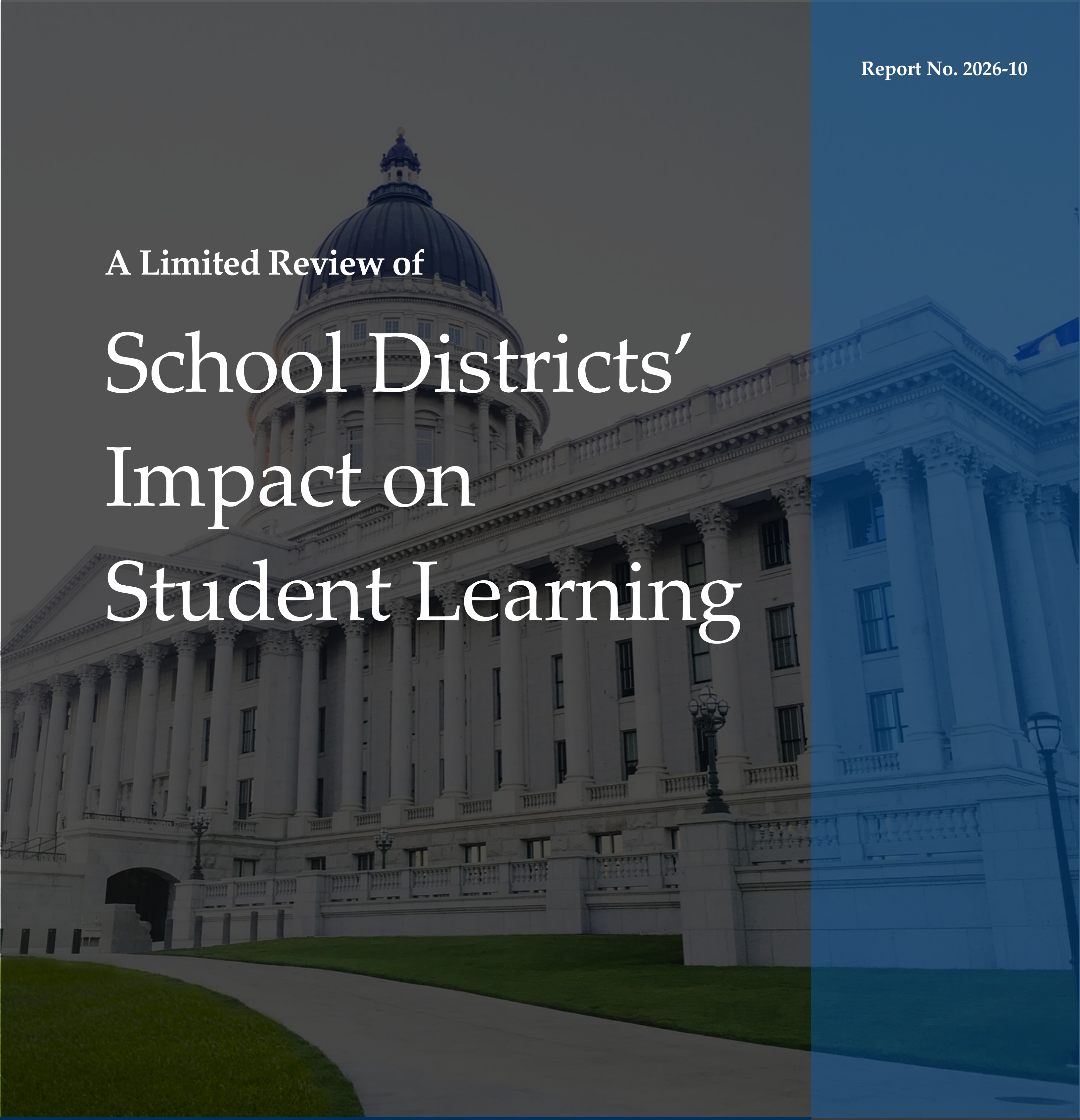


Report No. 2026-10

A Limited Review of

School Districts' Impact on Student Learning

Office of the Legislative Auditor General

Report to the **UTAH LEGISLATURE**

LEGISLATIVE AUDITOR GENERAL

LEGISLATIVE
AUDITOR GENERAL

Audit Subcommittee

President J. Stuart Adams, Co-Chair
President of the Senate

Senator Kirk Cullimore
Senate Majority Leader

Senator Luz Escamilla
Senate Minority Leader

Speaker Mike Schultz, Co-Chair
Speaker of the House

Representative Casey Snider
House Majority Leader

Representative Angela Romero
House Minority Leader

Audit Staff

Kade R. Minchey, Auditor General, CIA, CFE

Darin Underwood, Manager, CIA

Nick Varney, Audit Supervisor

Christopher McClelland, Audit Supervisor, CIA, CFE

Tyson Cabulagan, Methodologist, CFE

Office of the Legislative Auditor General

Kade R. Minchey, Legislative Auditor General

W315 House Building State Capitol Complex | Salt Lake City, UT 84114 | Phone: 801.538.1033

June 16, 2026

TO: THE UTAH STATE LEGISLATURE

Transmitted herewith is our report:

"A Limited Review of School Districts' Impact on Student Learning" [Report #2026-10].

A summary is found at the front of the report to introduce the consultant report.

We will be happy to meet with appropriate legislative committees, individual legislators, and other state officials to discuss any item contained in the report.

Sincerely,

Kade R. Minchey, CIA, CFE

Auditor General

kminchey@le.utah.gov

LIMITED REVIEW

SCHOOL DISTRICTS' IMPACT ON STUDENT LEARNING

CONTEXT

We hope this report can be used as a part of the statewide discussion and strategy to help districts continuously improve student growth and achievement.

We caution against too strict interpretation of this analysis, as student achievement is influenced by a wide range of factors. Nevertheless, this information offers a more meaningful starting point for understanding district performance than relying on raw test scores alone. We also intend to use this analysis to inform risk assessment in our LEA performance audits.

Student demographics and attributes play the largest role in explaining the differences in student achievement. While district practices have an effect on student performance, it may be smaller than people believe. We contracted with researchers to help determine the size of the effect of those practices. This analysis helps identify which districts have recently produced learning gains and which may need additional support.

After controlling for student demographics and attributes, some districts who have historically had higher average test scores drop in the relative performance rankings found in this report. Similarly, some districts who have historically had lower average test scores rise in relative performance rankings. The figure on page two shows how district performance changes when adjusting for student demographics and attributes.

Average student performance becomes more similar across districts after adjusting for student demographics and attributes

Figure 1.1 displays average student scores by district. Values to the left of zero are below mean district performance, and values to the right of zero are above. After adjusting for student demographics and attributes, the gaps in district performance shrink considerably.

Figure 1.1. Differences in School District Performance

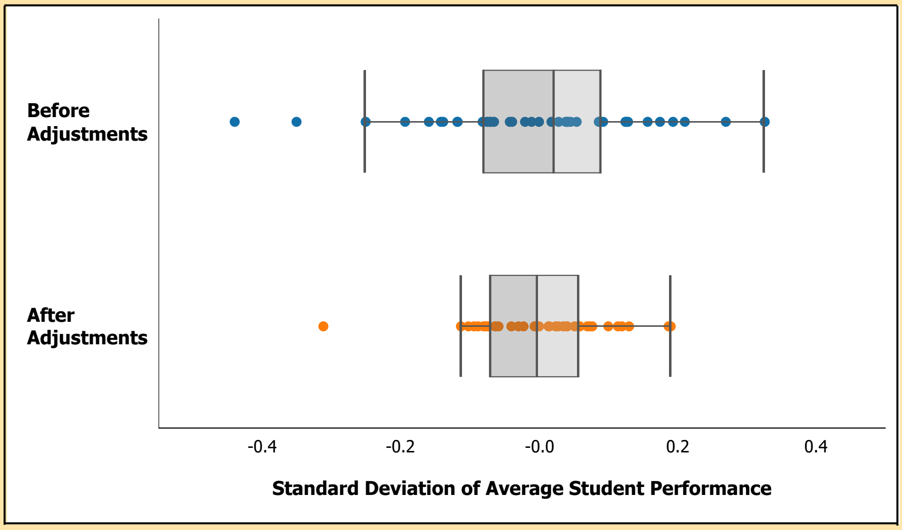
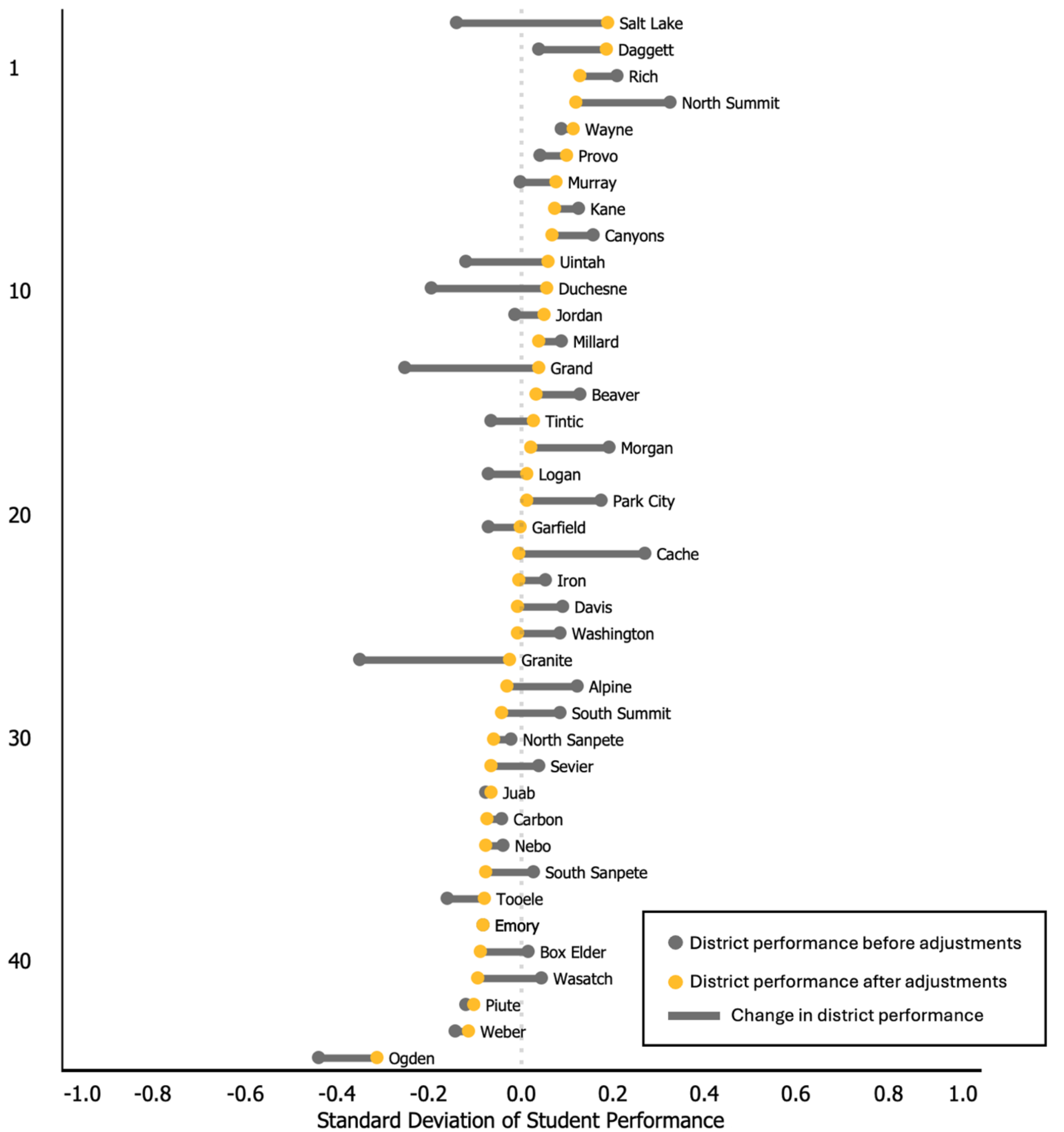


Figure 1.2. Some School Districts Demonstrate Stronger Learning Gains than Others. This analysis used data from Acadience, SAGE, RISE, and Utah Aspire Plus. Subjects included Language Arts, Math, and Science.



Source: Auditor generated from Notalys analysis. Because of data complications, San Juan School District was omitted from the list ranking.

To better understand standard deviations, it is useful to reference actual test results. In 2025, the range of 8th grade Math RISE scores was 279 – 667, and the standard deviation was 83 points. This means that a 0.1 standard deviation change between average district scores would be around 8 points for that test in that year.

District Variation in the Test Scores of Utah Students

Jay Goodliffe, PhD

Sven Wilson, PhD

NOTALYS

Notalys-Utah.com

A Report Submitted to the Office of Legislative Auditor General

6/26/2025

Executive Summary

This analysis standardizes and combines the comprehensive data on testing in Utah from grades 1-10 for the years 2017-2019 and 2021-2023 in order to assess the effectiveness of the state's 41 school districts. We use multi-level value-added assessment to obtain our district effectiveness measures.

Our principal findings are:

- Student performance on tests is primarily a function of student capabilities, family contributions, neighborhood environments, and random factors related to testing. Only 11% of the variance in student performance can be attributed to factors that are related to districts and schools, including their income and racial composition. These statistics are low, but they generally align with the scientific literature.
- Relatively small differences in average test scores exist across districts, but those differences are largely due to student characteristics such as income and race.
- After adjusting for student characteristics, the year-to-year value added performance varies only modestly across districts, with all but one district scoring within 0.2 standard deviations of the state mean.
- District spending decisions, including spending on teachers, are not associated with higher performance.

In this analysis, we provide a valued-added performance ranking of district effectiveness that incorporates adjustments for student characteristics. These adjustment factors have a significant impact on how districts are ranked (Figures 2a-2c). In summary:

- The 10 districts with the highest *unadjusted* scores are, in order: North Summit, Cache, Rich, Morgan, Park City, Canyons, Beaver, Kane, Alpine and Davis. *However, these scores do not measure value added or adjust for student characteristics.*
- The 10 districts with the *highest value-added* performance measures are, in order: Salt Lake, Daggett, Rich, North Summit, Wayne, Provo, Murray, Kane, Canyons, and Uintah.
- The 10 districts with the *lowest value-added* performance measures, starting from the bottom: Ogden, Weber, Piute, Wasatch, Box Elder, Emery, Tooele, South Sanpete, Nebo, and Carbon.
- Some districts, especially Salt Lake, Duchesne, Grand, and Granite, that appear to be low-performing, improve considerably after adjusting for prior performance and student characteristics.
- Others, especially Cache, North Summit, Morgan, Park City, and Alpine, fall considerably after the adjustment.
- Even after adjustment, Ogden's performance is considerably below other districts.

Introduction

The state of Utah invests considerable resources into educating its children. Though many important decision makers affect how those resources are used, the state school districts make a variety of crucial decisions that influence the learning of students in our public schools. The analysis that follows seeks to assess and rank the 41 school districts in terms of student performance as measured by student test scores taken from 2017-2023.

Districts are charged with educating Utah's children subject to state law and are constrained by the resources granted to them. Most important, they must provide education to all families within their boundaries, regardless of the capabilities and resources students bring with them to school. Our analysis shows that even though most of the variation in student performance is driven by the students themselves (and their families), small but meaningful differences exist that can be statistically attributed to districts. This report documents and seeks to (partially) explain those differences.

A multi-level approach to assessing district effectiveness

Student performance depends on numerous factors. For instance, socioeconomic status (SES) of the students remains the single biggest factor (Sirin, 2005; Song, et al., 2025). Race/ethnicity disparities are profound and often interact with SES (Benner, Boyle, and Sadler, 2016). Parental education and involvement play an important role (Desimone, 1999), as does neighborhood and environmental stress (Eamon, 2005). Berkowitz et al, (2017) show that a positive school climate buffers the negative impact of low SES on academic achievement. The impact of state expenditures is characterized by a long and controversial scholarly literature (discussed later).

This partial list includes some things we can measure with state administrative data (race, income) as well as many we cannot, such as parental support and the neighborhood and school environment. In this analysis we adopt a multi-level modeling approach which allows us to best exploit the information we have as well as accounting for important unobservable variables with the multi-level structure of the data, as we discuss below.

Our analysis subdivides these factors into the following four levels: district-level, school-level, student-level, and test-level. Each of these levels encompasses numerous relevant factors. For instance, a few of these are:

1) *Districts* create school boundaries, make policies, hire school principals, set curricula, and make decisions on how district funds are allocated.
2) *Schools* hire and supervise teachers, determine classroom composition, develop lesson plans, and hire and supervise school staff.

3) *Students* and their families are the primary drivers of performance. Students have innate and acquired characteristics that affect their learning and school performance. The family and neighborhood environment in which students live also influences their performance.
4) *Tests* differ by subject, by year, and by the testing experience. A given student will have varying performance from test-to-test for a variety of random reasons.

The better we can identify and control for the influence of each of the levels above, the better we can isolate the contributions that school districts are making to student success. Fortunately, a well-developed tool of statistical analysis, called multi-level modeling, is tailor-made for estimating district effectiveness.

Rationales for using multi-level analysis

An important feature of the data is that we observe the same student taking multiple exams across time and in different subjects. We also know the schools and districts in which the students are taking the exams. This complex data structure allows us to perform a sophisticated type of statistical analysis referred to as multi-level modelling. Reynolds and colleagues (2014, 205) notes that "school effects generally have been measured through the application of cross-sectional multi-level models." Indeed, when multi-level models are taught in textbooks, the standard application is assessing schools (Raudenbush and Bryk 2002; Goldstein 2007, 2011; Snijders and Bosker 2012; Hox, Moerbeck, and van de Schoot 2018; Leckie 2019; Finch and Bolin 2024).

Much of the existing literature examines the performance of schools, but the same method can also be used to assess districts (or districts and schools together). The work that includes both schools and districts together finds the larger effects at schools rather than districts (Yang and Woodhouse 2001, Fielding et al. 2006, Tymms et al. 2008, Rasbash et al. 2010, Leckie et al. 2010)

Rasbash et al. (2010) note that "Models of school differences in educational achievement typically assess the progress that pupils make between two test occasions and attempt to assess the extent to which unexplained variation between pupils is attributable to differences between schools. These models are commonly referred to as 'value-added' or 'school effectiveness' models" (658). Multi-level modelling is the standard approach in the educational literature to estimate the "value added" by schools and districts (Leckie 2019, 13).

One of the most significant shortcomings of empirical models is that every analysis has variables that are not available to the analyst. For each of the four levels discussed above we lack crucial data that is an important part of the story. Our data contains no information, for instance, on teacher training and quality, the effectiveness of curricula, student expectations, the economic resources of the neighborhoods or communities in the district, parental involvement in the schools, the home environment of students, and a host of other factors.

But an advantage of multi-level models is that we can overcome some of the bias inherent in having missing variables. This is because we observe many repeated observations at each level of analysis: many schools within each district, many students within each school, and, most important, many tests over many years for each student. This allows observations to essentially function as their own controls. For example, when we observe students many times, we can isolate the ability of that student and winnow out the noise associated with testing.

What variables to include in multi-level models

When applying multi-level models, an essential component of assessing value added is including a measure of incoming or prior performance (Hanushek and Hoxby 2005, Prior et al. 2021, Leckie and Prior 2022). Prior (2021, 9) notes that "In the academic literature, adjusting for pupil prior attainment is widely considered the minimum requirement for school performance comparisons." Including student demographics alone is a poor proxy for incoming attainment (Lieckie and Prior, 2022).

Most researchers also argue for the importance of including student characteristics. Leckie and Goldstein (2017) argue student characteristics are important, since they are beyond the control of the schools. When assessing schools, researchers often include school characteristics to capture contextual and peer effects as well (Goldstein and Leckie 2008, Lauder et al. 2010). For the same reasons, district characteristics are important in assessing districts. In their study of 13 US states (including Utah), Pivovarova, Amrein-Beardsley, and Geiger (2024) argue that most of the variance in school performance is driven by demographics. There are some disagreements, however. Marks (2021) argues against including student characteristics and school characteristics.

We will show, consistent with prior research, that most of the variation in student scores occurs at the student level. However, some of that variation is due to family factors (Rasbash et al. 2010, Leckie et al. 2010, Myhr et al. 2017), but we do not have information on family background, nor do we know the family connections between students in the state.

An additional "missing level" in our analysis is that we do not have any information on classrooms and teachers. We anticipate that the classroom contribution to a student's performance is significant. Since we do not know the classrooms students are in or anything about the teachers they have, we cannot determine the effects of teachers. Our anticipation would be that "teacher effects" are high. Some, perhaps a significant amount, of the student-level variance in outcomes could be attributed to teachers if we had such data.

In sum, multi-level models account for what portion of the performance is due to districts, schools, and students and what portion is due to simple random variation in testing across time. This analysis allows us to draw conclusions regarding the importance of district decision-making in general and, to a more limited extent, the impact of specific spending decisions made by districts.

How do we use Utah test score data to assess district effectiveness?

Summary of Utah test data used in this analysis

In this analysis we use six different tests. We constrain the sample to include those years and grade levels for which a reasonably comprehensive set of students took the exam. The six tests are as follows:

1) Acadiance literary: grades 1-3,
2) Acadiance math: grades 1-3,
3) SAGE/Rise/UA language: grades 4-10
4) SAGE/Rise/UA math: grades 4-10
5) SAGE/Rise/UA science: grades 4-10
6) ACT, composite: grade 11

Some of these tests show up in other grade levels, but not for the full sample, so we restrict the sample to the groups above. For most of the analysis that follows we exclude ACT tests because students usually take it only once, which means we don't have a measure of prior year performance to include in the model.

Analytical goal

We desire a method that allows us to use as much of the test score information as possible. In particular, we want to incorporate different academic subjects (language, math, and science) across the full range of grade levels. But since the tests measure different things and change over time, we need an approach that consolidates those different measures into a unified assessment.

Our basic analytical assumption is that the most important information contained in a test is how a student performs relative to other students taking the same test at the same time. Using this assumption as our foundation, we can then assess how school districts compare to one another based on a comparison of how the students in a particular district relate to students in the state as whole.

Score re-standardization process

Although the tests we are using are often referred to as "standardized tests," they are on different scales and the variation of scores is different across tests. Thus, our first task to re-standardize the tests so that they are directly comparable to each other and can be combined in the same analysis. For instance, we want to create scores that capture how eighth graders did on the state math test in 2019 that is directly comparable to how fifth graders did on the state language test in 2022. To do this, we create new standarized scores for each of the test-grade-year combinations in the data. For a given student score on a test, we calculate the standardized score as:

Standardized score = (raw score – average score) / (test standard deviation)

Another name for this type of measure is a z-score. We calculate these for each test-grade-year combination in the data. This normalization process puts all the test scores in the data on the same scale: each test-grade-year distribution has a mean of zero and a standard deviation of one.

Interpretation

The value of the standardized score is that it makes interpretation relatively simple. Each score refers to how far a particular student's performance is from the average student. For example, consider a student who takes the seventh-grade math test in 2019. If this student has a standardized score of 0.5, this means that she has scored 0.5 standard deviations above the average student taking that test. Scores above zero indicate better than average performance, while negative scores indicate performance below average.

What does this mean in practical terms? What does it mean, to continue the example, to be 0.5 standard deviations above the mean? A rough rule-of-thumb is as follows: about two-thirds of students will score within one standard deviation of the mean. In the data we are using (which is slightly skewed, in that the mean of 0 is at the 47th percentile), an increase in the score from the average of 0 to 0.1 is associated with moving from the 47th percentile to the 51st percentile. A score of 0.25 is at the 57th percentile, 0.5 is at the 67th percentile and 1.0 is at the 84th percentile.

We stress here that in this method all we capture is *relative* performance. The score does not capture what percentage of the content is mastered by the student. The score of 0.5 places the student at around the 69th percentile, but we do not know what that means in terms of the actual knowledge gained by the student.

The standardized scores also dissolve differences across time so we do not have to speculate whether a change from year to year represents a change in performance or a change in, for instance, the difficulty of the test. The standardized scores always have a mean of zero in each year, so the average cannot improve over time. When an individual student improves over time, say from .5 to .7 between two years, what that means is that the student's performance *relative to other test-takers* has improved from .5 to .7 standard deviations above the average.

Our main analysis involves estimating how much districts are responsible for the variation in student scores. Most of the districts play a very small role. A relatively influential district would be one that moves average scores by 0.25 standard deviations. As noted above, this would mean going from the 47th percentile to the 57th percentile—a change of 10 percentiles. We will see that most of the district effects are much smaller than that.

The standardized score is ideal for this analysis because comparisons can be made across all tests and all years (or across any subset of them). For instance, we can combine all the standardized scores from 2017-2023 in all five tests for all grade levels to obtain a broad-based measure of how a district is doing over the whole period. A district whose students have an average score of 0.2 are, on average, performing 0.2 standard deviations above the mean for the state.

How much do test scores vary across districts?

A first comparison of district performance reveals differences in performance across districts without adjusting for any of the factors that are known to affect test scores, such as income or race/ethnicity. Table 1 below provides the unadjusted rankings of districts based on the average standardized test scores for all test-grade-year combinations in the data. Table 1 also includes some key district features that will be discussed later.

The first thing to notice from this table is that most district averages are relatively close to zero. Only a handful of districts have average standardized scores more or less than 0.2 standard deviations from the mean. The differences range from 0.44 standard deviations below the average (Ogden) to 0.34 standard deviations above the average (North Summit).

Though these rankings are based on millions of test score observations and are, therefore, highly reliable, we should remember that the differences reflected in the rankings are, in general, quite small. We show below that the percentage of the test score variance attributable to districts is very small when compared to the variance attributable to schools and, especially, to student characteristics.

The data in Table 1 also suggest the importance of the racial composition and income status of students in accounting for student test scores—and, thus, the importance of controlling for these factors in our analysis.

Table 1: Average Re-Standardized Scores (Unadjusted) by District

District	Average Score	Percent Non-White	Percent Low-Income	Student/ Teacher Ratio	Instruction Spending
North Summit	0.336	17.0%	29.2%	15.8	$ 8,070
Cache	0.278	13.6%	29.3%	23.0	$ 5,317
Rich	0.219	6.6%	38.0%	12.8	$ 10,183
Morgan	0.209	5.2%	11.1%	20.5	$ 4,744
Park City	0.200	25.8%	20.1%	16.6	$ 8,048
Canyons	0.172	27.0%	29.4%	22.5	$ 5,094
Alpine	0.138	19.4%	19.0%	24.4	$ 5,119
Kane	0.133	10.3%	38.8%	17.8	$ 6,996
Beaver	0.126	20.1%	47.1%	17.2	$ 6,573
Davis	0.109	17.3%	18.6%	23.4	$ 5,324
Washington	0.097	21.1%	37.8%	22.1	$ 4,843
South Summit	0.095	13.9%	20.6%	17.5	$ 6,581
Millard	0.095	20.2%	48.3%	19.5	$ 7,373
Wayne	0.091	12.8%	45.3%	13.3	$ 7,886
Iron	0.064	17.1%	41.9%	21.3	$ 4,992
Wasatch	0.062	22.1%	27.9%	21.6	$ 6,908
Provo	0.058	38.4%	42.0%	21.5	$ 5,211
Daggett	0.053	8.6%	19.8%	11.3	$ 10,906
Sevier	0.043	9.3%	48.0%	21.0	$ 5,477
South Sanpete	0.036	16.2%	47.4%	20.9	$ 7,050
Box Elder	0.028	14.5%	32.6%	22.3	$ 5,588
Murray	0.012	31.8%	33.6%	21.5	$ 5,714
Jordan	0.004	24.4%	20.0%	22.9	$ 4,889
North Sanpete	-0.010	19.0%	52.1%	21.2	$ 5,591
Nebo	-0.024	17.5%	25.4%	23.2	$ 4,874
Carbon	-0.038	15.7%	44.2%	20.2	$ 6,393
Tintic	-0.055	11.2%	32.3%	11.8	$ 11,190
Logan	-0.059	41.6%	55.6%	22.5	$ 5,860
Garfield	-0.061	12.8%	45.2%	13.7	$ 6,311
Juab	-0.067	7.7%	33.9%	21.7	$ 4,457
Emery	-0.075	10.2%	50.7%	18.0	$ 7,066
Uintah	-0.106	21.0%	44.6%	22.6	$ 5,185
Piute	-0.113	16.1%	63.0%	8.9	$ 10,580
Salt Lake	-0.126	56.9%	53.8%	20.1	$ 6,222
Weber	-0.129	18.3%	28.8%	20.7	$ 5,483
Tooele	-0.145	20.7%	32.9%	22.0	$ 4,779
Duchesne	-0.184	18.2%	40.0%	19.9	$ 5,535
Grand	-0.236	26.9%	46.8%	16.4	$ 6,810
Granite	-0.334	50.7%	55.2%	23.6	$ 5,176
San Juan	-0.350	61.4%	100.0%	17.8	$ 7,287
Ogden	-0.432	57.5%	74.7%	21.0	$ 5,074
Total	0.000	25.8%	32.5%	22.5	$ 5,317

Figure 1a shows district averages by percent of students categorized as low-income. Figure 1b shows the district averages across different levels of racial composition (measured here as the percent non-white in the district). In both cases, we see a strong negative relationship: test scores are lower for districts with higher minority or low-income percentages.

Table 1 also includes measures of two variables that often receive significant public attention: student-teacher ratio and spending on instruction. The relationship between these variables and test scores are shown in Figure 1c and 1d. These variables show no evident relationship between either variable and test scores. This non-relationship is confirmed in the statistical analysis on district spending patterns at the end of this report.

To what extent can student performance be attributed to districts?

A primary advantage of using multi-level models is that we can attribute the sources of variation in scores across the four levels of analysis. To understand how this works, consider a hypothetical scenario where two students take a series of exams. In this scenario, Student A scores consistently higher than Student B and, furthermore, both students achieve roughly the same score every time they take the test (for instance, suppose Student A regularly comes in around the 90th percentile and Student B at the 70th percentile). In this case we could conclude that the difference in average scores between Students A and B is due to differences between the students, not differences between the test—even though they may be taking different tests across different time periods.

Now, if we had many students similar to Student A and Student B repeatedly taking tests and their performance test-to-test stays about the same, it would be logical to conclude that most of the calculated variance in test scores is due to differences in students, not differences in tests.

On the other hand, consider the case where Student C and Student D take a series of tests but the results are not as consistent as they were for A and B. In this case, the scores are "all over the place." Even if Student C has a higher average than Student D, we wouldn't be confident attributing that higher average to student performance because of how much performance varies test-to-test (e.g., 40th percentile one test, 80th the next). These differences could be due to differences in tests or differences in testing experiences. And in a dataset with many students like C and D, we would attribute most of the calculated variance of the scores in the dataset to variance across the tests, not variance across students.

Multi-level modeling extends this analysis to all the levels in the model. It allows us to exploit the fact that we observe students A, B, C, and D many times—across tests and across years. We also have many students within a given school and many schools within a given

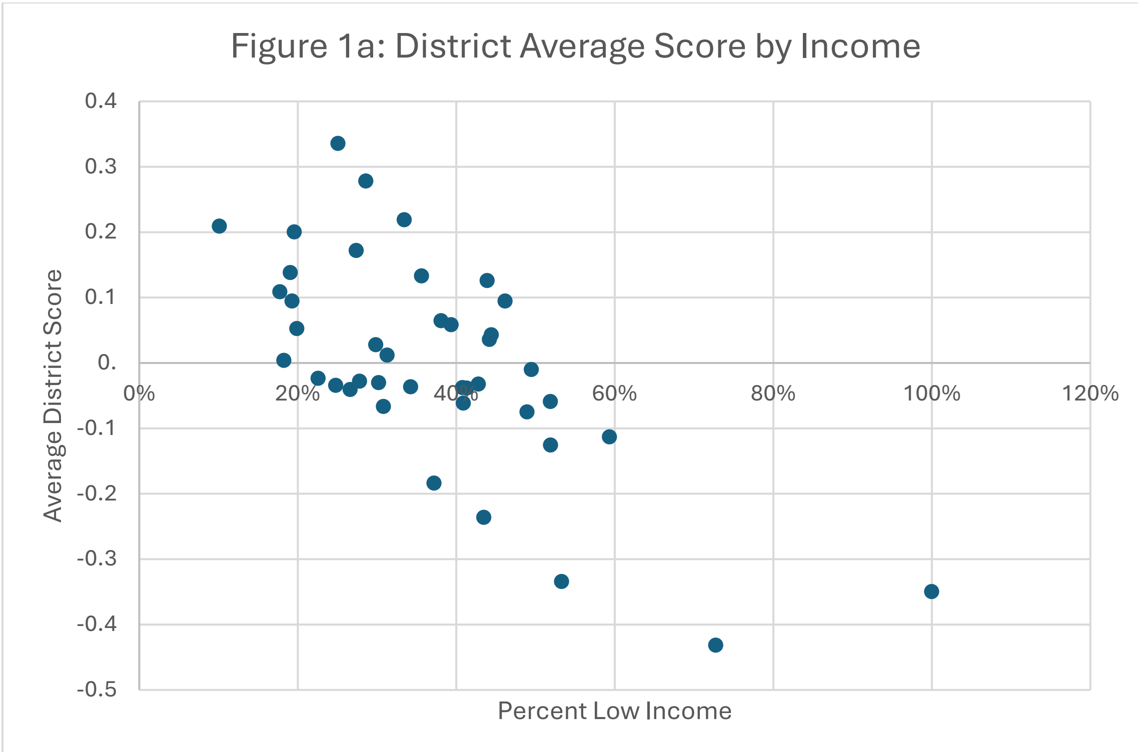


Figure 1a: District Average Score by Income
0.4
0.3
0.2
0.1
0.
-0.1
-0.2
-0.3
-0.4
-0.5
0%
20%
40%
60%
80%
100%
120%
Average District Score
Percent Low Income

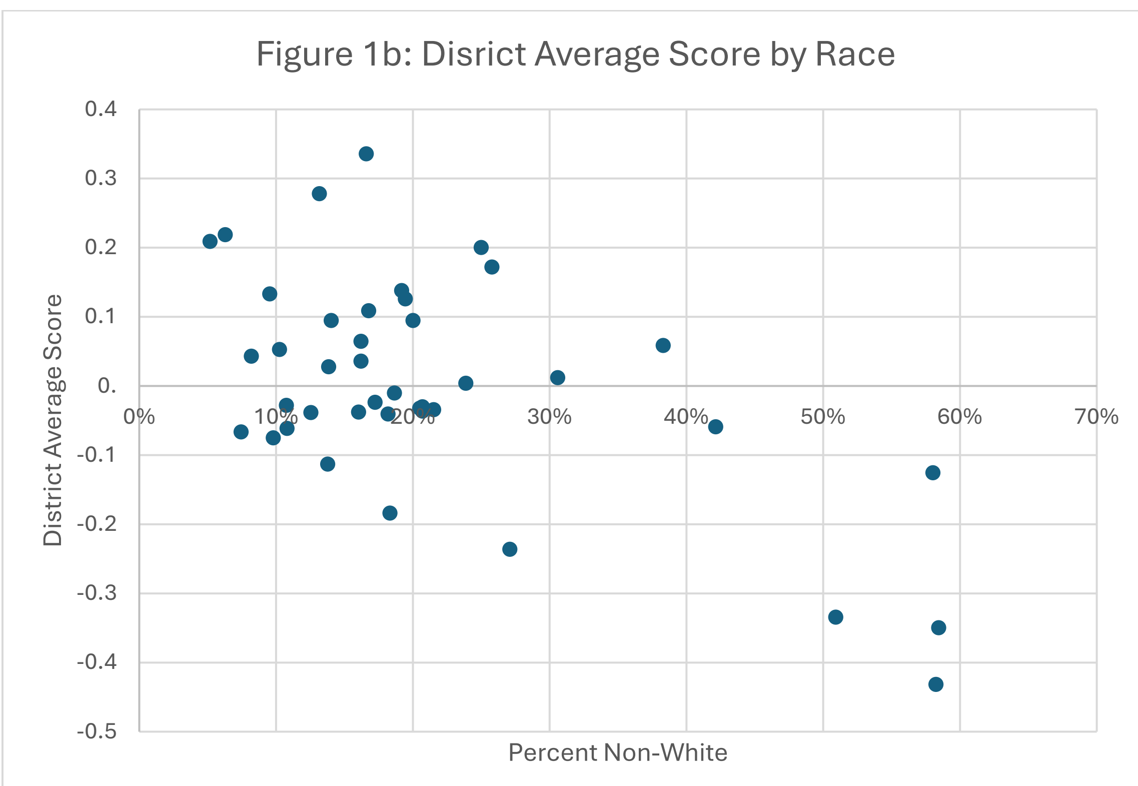


Figure 1b: Disrict Average Score by Race
0.4
0.3
0.2
0.1
0.
-0.1
-0.2
-0.3
-0.4
-0.5
0%
10%
20%
30%
40%
50%
60%
70%
District Average Score
Percent Non-White

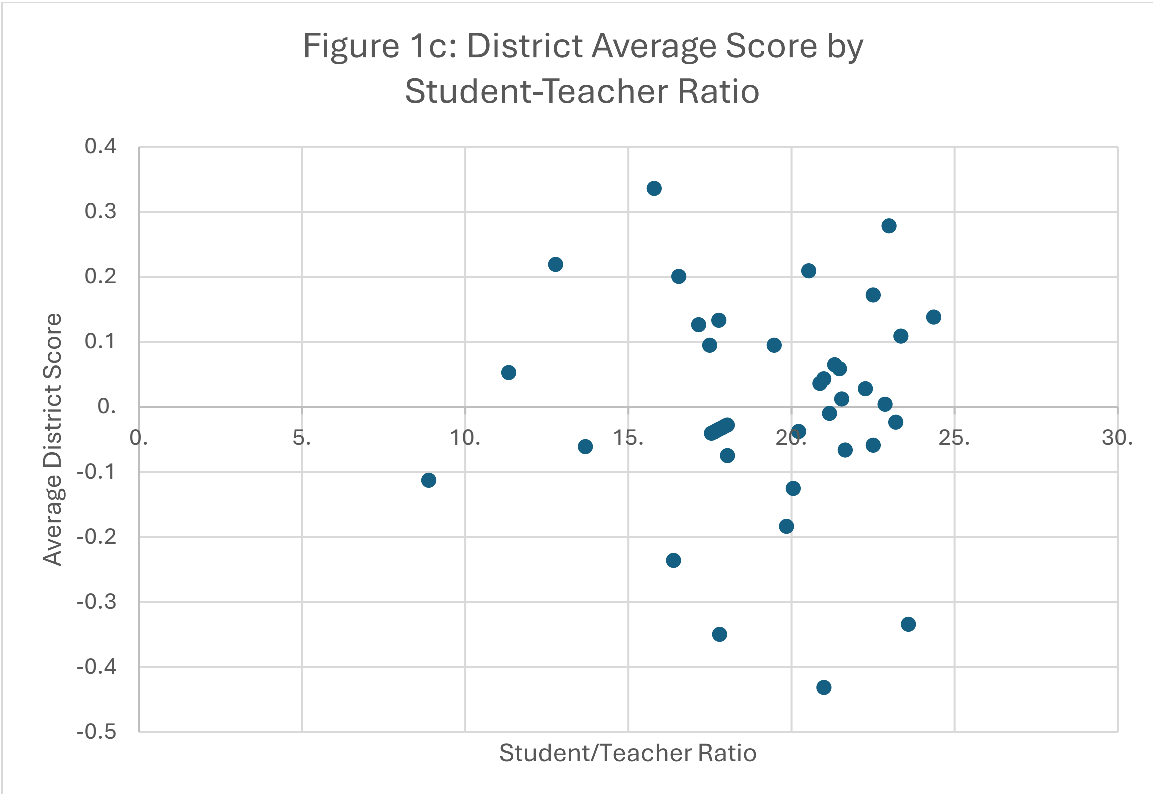


Figure 1c: District Average Score by Student-Teacher Ratio
Average District Score
0.4
0.3
0.2
0.1
0.
-0.1
-0.2
-0.3
-0.4
-0.5
0.
5.
10.
15.
20.
25.
30.
Student/Teacher Ratio

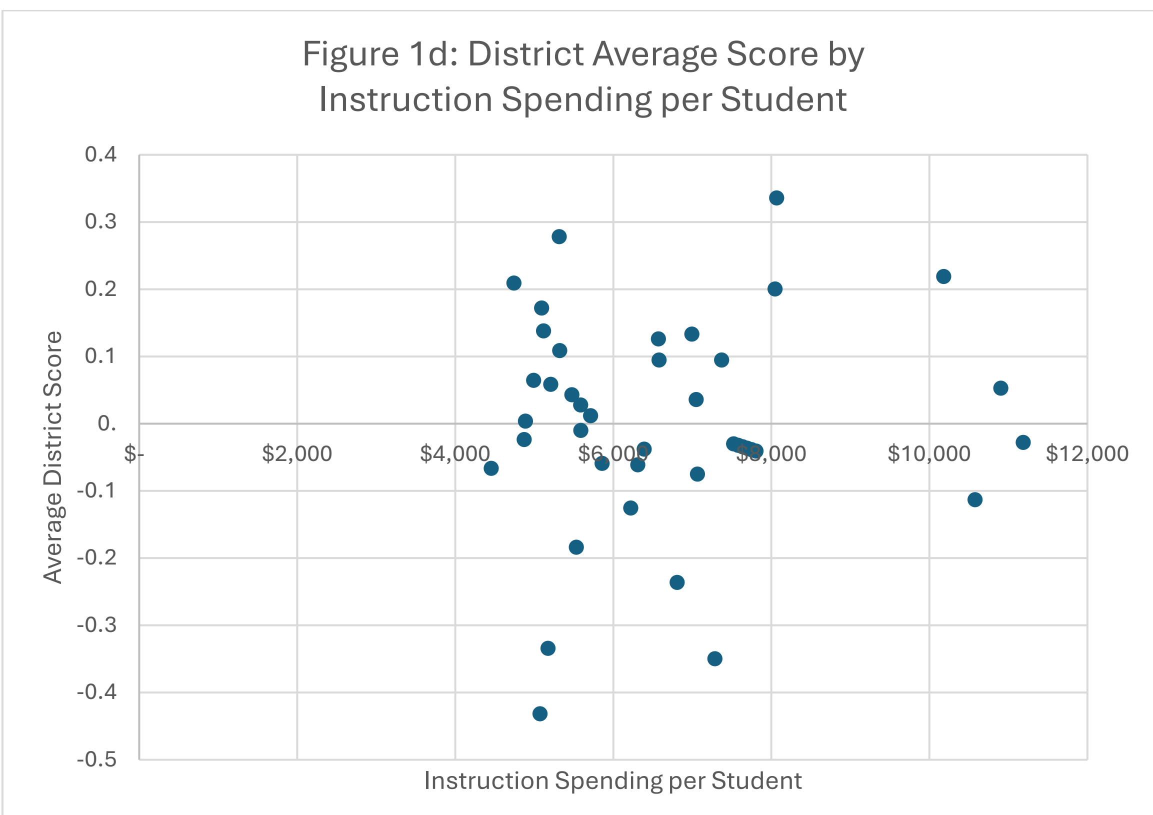


Figure 1d: District Average Score by Instruction Spending per Student
Average District Score
0.4
0.3
0.2
0.1
0.
-0.1
-0.2
-0.3
-0.4
-0.5
$-
$2,000
$4,000
$6,000
$8,000
$10,000
$12,000
Instruction Spending per Student

district; in other words, we have many observations at each level to help us understand how much schools and districts contribute to performance. In much the same way that we know Student A has higher performance than Student B, we can differentiate between districts by observing many observations of students across districts. Furthermore, including data on student demographic characteristics allow us to control for other factors that affect test scores.

Though the estimation methods for multi-level models are quite complex and take a lot of computing power, the heart of the method consists of identifying patterns at each level (just as we looked for patterns when comparing students A, B, C, and D). Such an exercise is commonly called "decomposition of variance."

In Table 2 below we decompose the variance in scores across the four levels using what is called the Random Effects ANOVA model. This model does not include any other variables such as prior performance, student characteristics or district spending patterns; those will come later. We do this decomposition for the full set of re-standardized test scores and then sub-divide the analysis into three categories: Lower grades (1-3), Middle Grades (4-10) and High School (ACT, grade 11).

Table 2: What Accounts for Variation in Test Scores?

	Districts	Schools	Students	Tests	
Combined	2.2%	8.9%	65.5%	23.4%	100.0%
Lower Grades (1-3)	2.0%	7.5%	64.9%	25.6%	100.0%
Middle Grades (4-8)	2.3%	9.7%	66.5%	21.5%	100.0%
High School (ACT)	1.0%	15.0%	72.8%	11.2%	100.0%

The results of this table tell an unsurprising story: the majority of the variance in test scores is due to differences across students. This speaks, fundamentally, to the vast variability in human capacity. Students come to the schools with very different capacities and are from very different families and neighborhoods. 65.5% of the total variance in tests scores is attributable to these differences across students.

Another 23.4% is due to variation across testing events. This includes different subject matter areas, differences across grade levels and year to year variation in how well a given student will do in a given testing environment on a given day. Some of this variance is just noise—the inherent randomness present in testing.

Together, districts and schools account for 11.1% (2.2% from districts + 8.9% from schools) of the variation in test scores. Since districts are responsible for how schools are governed, the contribution of schools to student performance is due in part to decisions made by districts. Teacher hiring, for instance, is shaped by both district policies, practices, and resources and by the decisions of school principals. The 2.2% number attributable to districts would be the factors that are shared in common by all the schools in the district. All schools, for instance, have the same district administrators.

In the scientific literature, estimates of school effects vary widely, depending on location, grade, subject, and other factors. Estimates of the proportion of total variance at the school-level typically range from around 5-25% of the total variance (Lee & Bryk, 1989; Scheerens & Bosker, 1997; Rumberger & Palardy, 2004; Hedges & Hedberg, 2007,2014; Munoz & Prather, 2011, Scheerens, 2016)). Thus, our estimate of 8.9% is in the low end of the range. Our conjecture is that this is partly because we are using comprehensive data for the whole state population, rather than the random samples used in most studies. It may be that the sampling approach systematically misses many lower performers, thereby lowering the variance in total performance.

A significant limitation of our analysis is that we were not provided data on teachers and classes. In the literature, the effect of teachers is generally thought to be higher than the effect of schools, since it is teachers that are having the most direct interaction with students, and students are also influenced by fellow classmates. Scheerens (2016) reports that the school-level component of variance in the U.S. is 10% for schools and 46% for teachers/classes. A recent international study (Zhang, 2025) finds that in developed countries, including the U.S., the school-level variance component is about 10% and the teacher/class-level component is about 15%. In our data, the effect of teachers is attributed to student-level and test-level factors.

Very few studies look at the effect of school districts, largely because of the sampling design. The works by Hedges and Hedberg (2013) identifies both school and district effects across several states (not Utah). They find that school-level factors account for 10-25% of the variance, while districts account for only about 2-10%. Thus, our results for minimal district influence are within that range.

How do districts differ in their contribution to student performance?

Method

The estimates in Table 2 allocate the variance in test scores across the four different levels of analysis. But those estimates have no control variables included. In this section, we expand the analysis to include the control variables that yield our “adjusted” district effects.

The most important control variable needed to obtain a true value-added measure is prior performance. Ideally, we would have a test at the beginning of the school year and one at the end, but that testing method is not used in the state. Instead, we control for prior performance by using the test score from the previous year. For example, we use Acadiance language and math tests from grades 1-3. For this analysis, our measure of prior performance is the student's score from the previous year within the given subject matter area (this means that only grades 2 and 3 actually enter the model). The same procedure is applied for the state tests for grades 4-10.

We also include student level characteristics corresponding with race and socioeconomic status. These include low-income, unhoused status, full year attendance, s504 disability participation, ELL (English language learner) participation , absenteeism and mobility (moved within the school year). We also include district level aggregates for all of these variables. The individual-level variables are powerful predictors of performance, but the district-level variables also have some explanatory power, though the effects are much smaller.

We do not include school-level characteristics because those are, to a degree, a function of district decisions. The modelling strategy is to include only those variables that are not chosen by the district. In other words, how well does the district do with what they can choose while controlling statistically for the factors they cannot affect, such as the racial composition of the students in the district?

An important feature of multi-level models is that they include what are called "random effects" at each level of analysis, which captures the unobserved variables at each level that we discussed. We can estimate these effects because we have multiple observations for each student and multiple schools within each district. In the models reported below we do not include a school-level random effect, but we do have a student-level and a district-level random effect (thus the district effect includes contributions of schools within the district and the district itself).

Interpretation

The estimates below are the estimated random effect at the district level for each of the districts in the state. We interpret this as a measure of district effectiveness that captures the value added (above or below the prior performance) and that adjusts student and district level characteristics, both observed and unobserved. In simpler terms, what value does the district add relative to the hand they were dealt?

In interpreting these measures it is important that we keep in mind two important things. First, these scores have been standardized to have a mean zero. Thus, even when all districts are adding value, some will have a positive score and some a negative score. A more nuanced interpretation is, therefore, what value is a district adding *relative to other districts?* The important thing that comes out of this measure is a ranking of districts. Our assessment method is agnostic to the question of how much absolute value districts are adding to student learning.

Second, the district effects we estimate are more accurately district-level effects in that they capture all the things going on within the district, *whether or not they have anything to do with actions taken by the school district itself*. For example, we know how many students in a district are low-income, but we don't know anything about the total economic resources of the district. Imagine two districts that each have 15% low-income students, but one district is much wealthier than the other. For a variety of reasons, we would expect that the wealthier district would have more advantages for student learning, such as parent volunteers in schools or access to private tutoring. But because the advantages are unobserved, they are lumped in with the district-level effect. For these reasons, we should interpret the district effectiveness rankings cautiously.

Results

We include the main results of the analysis in Figure 2a, 2b, and 2c. Districts are ranked according to the estimated "district effect," which we estimate here in two ways. The first rankings are in Figure 2a and are the *unadjusted* district effects. These range from -0.44 (Ogden) at the bottom to +0.33 (North Summit) at the top. Blue bars represent confidence intervals, but since we have so many test scores, the confidence intervals are narrow. The top five unadjusted district effects are North Summit, Cache, Rich, Morgan, Park City.

Figure 2b represents the rankings of the districts after adding the control variables to the model. Controls include, most importantly, the previous year's standardized score. We also control for individual student characteristics such as race and the district averages for those same variables. The models used to obtain the adjusted estimates do not include any choices made by districts, such as how they are spending their funds or student-teacher ratios, since student-teacher ratios are a function of district decisions. In short, the adjusted district estimates answer this question: how well is the district doing given the hand it has been dealt? The top five districts after making for adjustments are Salt Lake, Daggett, Rich, North Summit, and Wayne. Because of problems with the income data. Because of problems with the income data, the models used to estimate Figures 2b and 2c do not include the San Juan District (see the Technical Appendix).

Figure 2c combines the data from Figure 2a and 2b to illustrate the impact of including the control variables, which is illustrated by the blue arrows. In some cases these adjustments are sizable. Districts like Salt Lake, Grand, and Granite make particularly noticeable increases in their ranking. Other districts like Cache and North Summit make noticeable declines once adjustments are made. Of course, many districts do not move with the addition of the adjustment factors. These include districts at the upper end of the rankings, such as Wayne and Provo, and districts at the bottom, such as Piute and Weber.

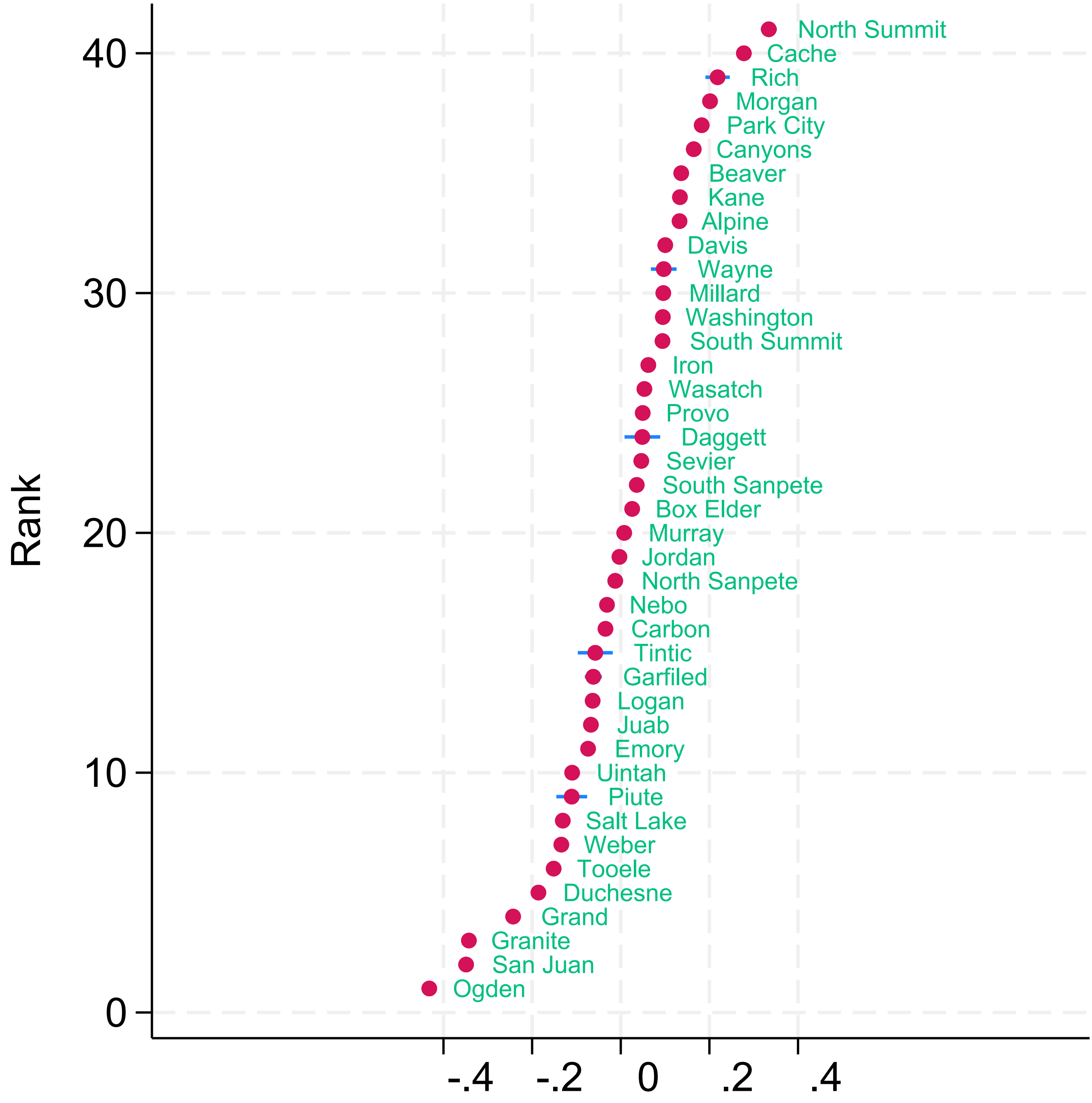


Figure 2a: Unadjusted District Scores

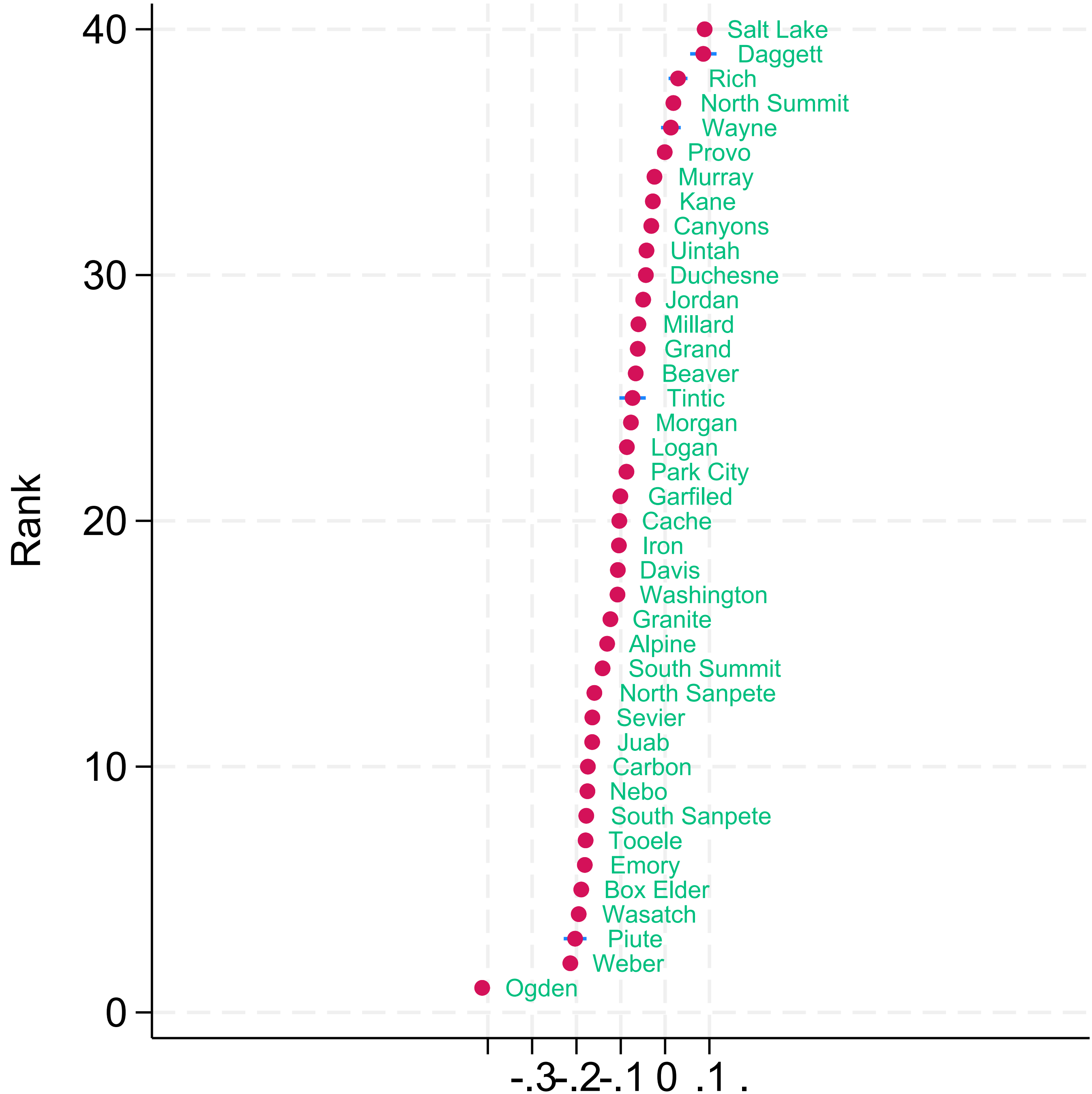


Figure 2b: Adjusted District Scores
(Value Added, with controls)

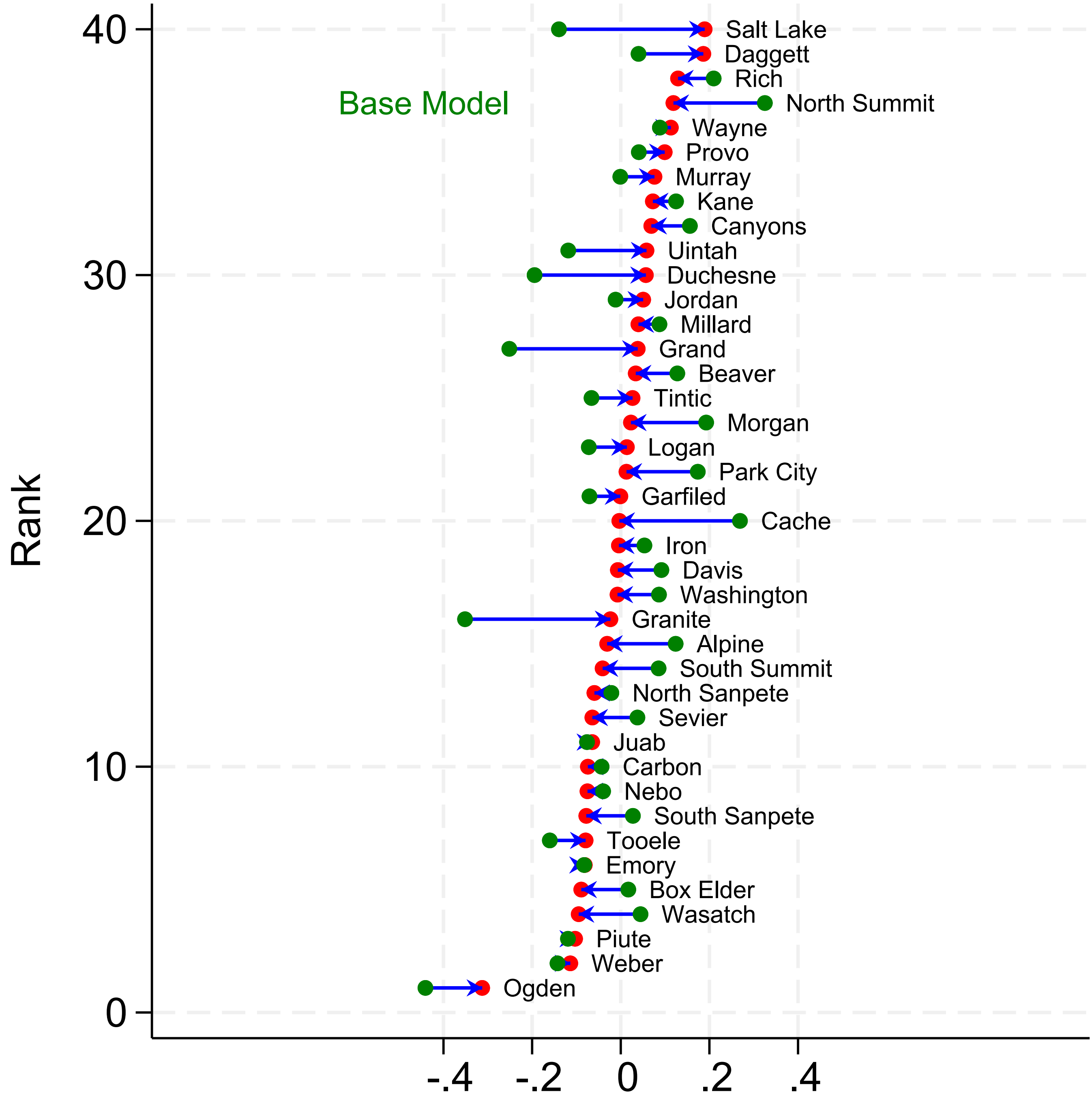


Figure 2c: Impact of model adjustment on District Performance

These rankings have two important features worth remembering. First, including controls for prior performance and student characteristics has a narrowing effect on the distribution of district effects. Note in Figure 2b that only the Ogden district has an adjusted score more than 0.2 standard deviations from the mean. The second feature of this analysis comes directly from the first: the magnitude of district effects in general are very small. Most districts are not very far from the state average.

How are spending patterns associated with student performance?

Prior research related to school finance

In a recent study, Jackson & Mackevicius (2024) do a meta-analysis of US school spending studies. Their central conclusion is that increases in per-pupil spending result in higher tests scores. They find that a $1,000 per-pupil spending increase over four years raises test scores by 0.03 standard deviations. They also find that these results are consistent across geography and level of baseline spending and that the results are even stronger for low-income populations. The authors also find that both operational and capital expenditures have similar effects but operational expenses are slightly more effective.

The studies incorporated in this meta-analysis have results that contrast to the previous prevailing wisdom in economics. Based on an extensive review of the existing literature of the time, Hanushek (2003) argued that input-based policies, which focus on increasing educational resources have largely failed to yield significant improvements in student achievement. He argues, instead, for reforms that incentivize teachers to perform better and that base funding for schools on performance.

The Hanushek review was based on many studies that relied on cross-sectional, OLS regression analysis, rather than the quasi-experimental work done since 2010 that allowed for more robust causal inference. For instance, Brunner and colleagues (2020) use a natural experiment from changes in the school funding formula in California to show an increase in academic performance. Baron (2022) examines the causal effects of school spending on student achievement using a natural experiment created by Wisconsin's revenue limit law. He finds that increased operational spending significantly improves student test scores, particularly in lower grades. Using data from Ohio's school finance reform, Carlson and Lavertu (2018) use regression discontinuity design to obtain causal evidence that increased funding improves test scores.

Methods

To estimate the effects of spending in the state we employ the same multi-level model as employed above, but we add additional control variables. In the district performance model above we included only variables that were out of the control of the district. In the spending model estimated here we want to include choices made by districts, particularly the spending choices.

Spending falls into a long list of categories and is reported at the district level. Consistent with the literature, we divide expenditures into two categories: operational (which includes instruction, school maintenance, support services, food services, etc.) and capital (land acquisition, new buildings, debt management, etc). All of these are measured on a per-student basis. We estimate the effect of spending with two models. The first includes variables for the sum of operational and capital expenditures. The second excludes capital expenditures, which we don't think are highly relevant to year-to-year value added performance, and divides operational expenditure into sub-categories such as instruction, transportation, student support, etc.

We also include school-level variables in the spending models, such as the racial and income composition of the school. Districts determine the number of schools and the boundaries for those schools, which, in turn, determine things like racial and income composition of the different schools in the district. These school-level compositional variables are more important in explaining the performance of students within a given school than are the district-level compositional variables. One important school-level variable is student teacher ratio, which we obtain by taking the number of students taking a test in a given year in the school divided by the number of teachers.

Results

Previous research in other states has shown that both operational and capital expenditures can be important. In Utah, averaging across the years 2017-2023 and adjusting for inflation gives an average operational expenditure of $9,795 per student; average capital expenditure is $1,995 per student (these estimates are represented in constant 2023 dollars). We report the complete multi-level regression results in the technical appendix. That analysis shows that increasing operational expenses by $1,000 per student is actually associated with a decline in the re-standardized score of 0.002 standard deviations. An increase of $1,000 in capital expenditure is associated with an increase of 0.006 standard deviations. Though statistically significant because of the large sample size, these effects are trivial in size.

Our second estimated model breaks down the operational expenditure to see if some categories are more influential than others. Since capital expenditure is typically devoted to projects spread across multiple years, we do not include them here because we are looking for factors that affect the year-to-year value added in terms of test scores. The spending categories used in the analysis come from the Utah State Board of Education's chart of accounts. The results are given in Table 3 below.

Table 3: Student Performance and Spending

	Mean	Marginal Effect
Student Transportation (2700)	$ 318	***0.023***
Support Services-General Admin (2300)	$ 99	***0.021***
Support Services-School Admin (2400)	$ 590	***0.016***
Food Services (3100)	$ 439	***0.013***
Support Services-Instruction (2200)	$ 448	***0.004***
Central Services (2500)	$ 286	***0.004***
Community Services (3300)	$ 316	***0.000***
Operation and Maintenance (2600)	$ 849	***-0.001***
Support Services-Students (2100)	$ 451	***-0.002***
Instruction (1000)	$ 5,997	***-0.004***

Note: Marginal Effect represents change in average test performance from a $100 increase in spending per student

Note: Spending types exclude capital expenditures (codes 4000 and above)

Because expenditures per student are less than $1,000 for all categories except instruction, we estimate here the marginal impact of an increase of $100. Somewhat surprisingly, the categories with the largest impact are not instruction related but are related to transportation, food, and district and school administration. Additional spending on instruction, by far the biggest category, is actually negatively correlated with test scores.

These negative effects for instruction related spending are surprising, though it is necessary to keep in mind that they are small effects, and we cannot interpret them in a causal fashion. They tell us only that, on average, districts that spend more on instruction per student are not getting better outcomes, even after controlling for the demographic and economic composition of the schools. It may also be the case that we are observing reverse causality here: schools that are not performing as well as hoped are seeking better results by spending more on teaching.

A related finding from both of the models discussed in this section is that the effect of student-teacher ratio is small but positive, meaning the larger the class size, the higher performance. A one unit increase in the student-teacher ratio raises school performance by .001 standard deviations (a small but statistically significant increase). School reformers frequently advocate higher salaries and reduced class sizes, but the evidence found here does not suggest such reforms will be effective.

As noted above, quasi-experimental research in the last two decades has found that increases in spending does have modest positive effects on test scores. This may well be the case in Utah as well. Our data cannot be exploited to do the causal analysis that this recent research has done. We only know that in comparing test outcomes across districts that higher spending districts are not doing better than lower spending districts in terms of adding value to test scores.

Concluding Notes

Our analysis in this report is able to provide a comprehensive view of district variation in test scores during the 2017-2023 period. The central finding is that, once we control for student characteristics, we do not find much variation. Many of the districts that appear to be doing poorly when looking at raw test scores (Fig. 2a) are actually doing quite well. Salt Lake is the prime example of this: near the bottom in raw scores but at the top when adjustments are made for student characteristics.

Because of the very large data set, we can be quite confident in these findings. But we warn the reader that the differences between most districts is very small. Because we have only a limited set of information on students—and no information on the neighborhoods and communities in which those students live—we should be very cautious in interpreting the findings. The performance measures provided here provide a good starting point for understanding variation across districts, but more information, especially about the variation in financial assets across different communities, would likely lead to a different ranking.

Finally, nothing in this report should be construed as an assessment of the state's performance overall or of the absolute performance of any particular district. The estimates here are all in relative terms. They are useful only to rank districts against each other using the metric of value-added testing (improvement from one year to the next, controlling for student characteristics). Our methods have nothing to say about whether the districts at the bottom are doing poorly or even if the districts at the top are doing well.

References

Baron, E. J. (2022). School spending and student outcomes: Evidence from revenue limits. *American Economic Journal: Economic Policy*, 14(2), 1–33. https://www.aeaweb.org/articles?id=10.1257/pol.20200226

Benner, A. D., Boyle, A. E., & Sadler, S. (2016). Parental involvement and adolescents' educational success: The roles of prior achievement and socioeconomic status." *Journal of*

Youth and Adolescence, 45, 1053-1064. *https://link.springer.com/article/10.1007/s10964-016-0431-4*

Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2017). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement: *Review of Educational Research*, 87(2), 425–469. https://doi.org/10.3102/0034654316669821

Brunner, E. J., Hyman, J. M., & Ju, A. (2020). School finance reforms, teachers' unions, and the allocation of school resources. *Review of Economics and Statistics*, 102(3), 473–489. https://doi.org/10.1162/rest_a_00828

Caldas, S.J., and Bankston, C.L III. (1999). Multi-level examination of student, school, and district-level effects on academic achievement. *The Journal of Educational Research,* 93(2), 91–100. https://doi.org/10.1080/00220679909597633

Carlson, D., & Lavertu, S. (2018). School improvement grants in Ohio: Effects on student achievement and school administration. Educational Evaluation and Policy Analysis, 40(3), 287-315. https://doi.org/10.3102/0162373718760218

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2013). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679. https://doi.org/10.1257/aer.104.9.2633

Desimone, L. (1999). Linking parent involvement with student achievement: Do race and income matter? *The Journal of Educational Research*, 93(1), 11–30. https://doi.org/10.1080/00220679909597625

Duncan, G. J., & Magnuson, K. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children*, 15(1), 35–54. https://files.eric.ed.gov/fulltext/EJ795842.pdf

Eamon, M. K. (2005). Social-demographic, school, neighborhood, and parenting influences on the academic achievement of Latino adolescents. *Journal of Youth and Adolescence*, 34(2), 163–174. https://doi.org/10.1007/s10964-005-3214-x

Fahle, E.M., and Reardon, S.F. (2018). How much do test scores vary among school districts? New estimates using population data, 2009–2015. *Educational Researcher,* 47(4), 221–234. https://doi.org/10.3102/0013189X18759524

Fielding, A. et al. (2006). Using cross-classified multi-level models to improve estimates of the determination of pupil attainment: A scoping study. *Research Report to Department for Education and Skills, University of Birmingham.* https://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/dfes-scoping-report.pdf

Finch, W. H., and Bolin, J.E.. (2024). *Multi-level Modeling Using R*, 3rd ed. CRC Press.

Goldstein, H. (2007). "Becoming familiar with multi-level modeling.." *Significance*, 4(3),133–135. https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1740-9713.2007.00249.x

Goldstein, H. (2011). *Multi-level Statistical Models*, 4th ed. Wiley.

Goldstein, H., Burgess, S. & McConnell, B. (2007). Modelling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 170(4), 941–954. https://doi.org/10.1111/j.1467-985X.2007.00491.x

Goldstein, H., & Leckie, G. (2008). School league tables: what can they really tell us? *Significance*, 5(2), 67–69. https://doi.org/10.1111/j.1740-9713.2008.00289.x

Hanusheck, E.A., (2003). The failure of input-based schooling policies. *The Economic Journal, 113, F4-F98.* https://doi.org/10.1111/1468-0297.00099

Hanushek, E.A., and Hoxby, C. (2005). Developing value-added measures for teachers and schools. In *Reforming Education in Arkansas: Recommendations from the Koret Task Force, 2005,* 99–104. Hoover Institution Press. https://doi.org/10.1111/1468-0297.00099

Hedges, L.V. & Hedberg, E.C. (2007). Intraclass Correlation Values for Planning Group-Randomized Trials in Education. *Educational Evaluation and Policy Analysis, 29, 60-87. https://doi.org/10.3102/0162373707299706*

Hedges, L.V. & Hedberg, E.C. (2013). Intraclass correlations and covariate outcome correlations for planning two- and three-level cluster-randomized experiments in Education. *Education Review 37, 445-489. DOI: 10.1177/0193841X14529126*

Hox, J.J., Moerbeek, M., & van de Schoot, R. (2018). *Multi-level Analysis: Techniques and Applications*, 3rd ed. Routledge.

Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *Quarterly Journal of Economics*, 131(1), 157–218. https://doi.org/10.1093/qje/qjv036

Jackson, C. K., & Mackevicius, C. (2024). The distribution of school spending impacts. *American Economic Journal: Applied Economics*, Forthcoming. https://doi.org/10.3386/w28517

Lauder, H., Kounali, D., Robinson, T. & Goldstein, H. (2010). Pupil composition and accountability: An analysis in English primary schools. *International Journal of Educational Research,* 49(2–3), 49–68. https://doi.org/10.1016/j.ijer.2010.08.001.

Leckie, G. (2018). Avoiding bias when estimating the consistency and stability of value-added school effects." *Journal of Educational and Behavioral Statistics* 43(4), 440–468. https://doi.org/10.3102/1076998618755351

Leckie, George. (2019). Multi-level models for continuous outcomes." *arXiv preprint* arXiv:1907.05941. https://arxiv.org/abs/1907.05941

Leckie, G., & Goldstein, H. (2017). The evolution of school league tables in England 1992–2016: 'Contextual value-added', 'expected progress' and 'progress 8'." *British Educational Research Journal,* 43(2), 193–212. https://doi.org/10.1002/berj.3264

Leckie, G., and Goldstein, H. (2019). The importance of adjusting for pupil background in school value-added models: A study of Progress 8 and school accountability in England. *British Educational Research Journal*,45(3), 518–537. https://doi.org/10.1002/berj.3511

Leckie, G., Pillinger, R., Jenkins, J. & Rasbash, J. (2010). School, family, neighborhood: Which is most important to a child's education? *Significance,* 7(2), 67–70. https://doi.org/10.1111/j.1740-9713.2010.00421.x

Leckie, G., & Prior, L. (2022). A comparison of value-added models for school accountability. *School Effectiveness and School Improvement,* 33(3), 431–455. https://doi.org/10.1080/09243453.2022.2032763

Lee, V. & Bryk, A. (1989). A multi-level model of the social distribution of education achievement. Sociology of Education. 62, 172-192. https://www.jstor.org/stable/2112866

Marks, G.N. (2021). Should value-added school effects models include student- and school-level covariates? Evidence from Australian population assessment data. *British Educational Research Journal,* 47(1), 181–204. https://doi.org/10.1002/berj.3684

Munoz, M.A., & Prather, J.R. (2011). Exploring teacher effectiveness using hierarchical linear models: student- and classroom-level predictors and cross-year stability in elementary school reading. *Planning and Changing* 42, 241-273.

Myhr, A., Lillefjell, M., Espnes, G.A.. & Halvorsen, T. (2017). Do family and neighbourhood matter in secondary school completion? A multi-level study of determinants and their interactions in a life-course perspective. *PLoS ONE,* 12(2), e0172281. https://doi.org/10.1371/journal.pone.0172281

Pivovarova, M., Amrein-Beardsley, A., & Geiger, T. (2024). What's in a school grade? Examining how school demographics predict A-F letter grades. Practical Assessment, Research & Evaluation, 29(12). http://files.eric.ed.gov/fulltext/EJ1440455.pdf

Prior, L., Jerrim,J., Thomson, D., & Leckie, G. (2021). A review and evaluation of secondary school accountability in England: Statistical strengths, weaknesses and challenges for

"Progress 8" raised by COVID-19. *Review of Education,* 9(3), e3299. https://doi.org/10.1002/rev3.3299

Rasbash, J. Leckie, G., Pillinger, R. & Jenkins, J. (2010). Children's educational progress: Partitioning family, school, and area effects. *Journal of the Royal Statistical Society Series A: Statistics in Society,* 173(3), 657–682. https://doi.org/10.1111/j.1467-985X.2010.00642.x

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed. Sage Publications.

Reynolds, D., Sammons, P., Fraine, B.D., Van Damme, J., Townsend, T., Teddlie, C. & Stringfield. S. (2014). Educational effectiveness research (EER): a state-of-the-art review. *School Effectiveness and School Improvement,* 25(2), 197–230. https://doi.org/10.1080/09243453.2014.885450

Reardon, S. F. (2016). The widening income achievement gap. *Educational Leadership,* 70, 10-16. https://stonecenter.gc.cuny.edu/files/2022/09/Conwell-2.pdf

Rumberger, R. & Palardy, G.J. (2004) Multilevel models for school effectiveness research. *The SAGE Handbook of Quantitative Methodology for the Social Sciences.*

Scheerens, J. (2016). Educational effectiveness and ineffectiveness. *A critical review of the knowledge base*. Springer

Scheerens J, Bosker RJ. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.

Scheerens J, Vermeulen CJAJ, Pelgrum WJ. Generalizability of instructional and school effectiveness indicators across nations. *Int J Educ Res*, 3(7),789–99. https://doi.org/10.1016/0883-0355(89)90029-3

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453. https://doi.org/10.3102/00346543075003417

Snijders, T.A.B., & Bosker, R.J. (2012). *Multi-level Analysis: An Introduction to Basic and Advanced Multi-level Modeling*, 2nd ed. Sage Publications.

Song, Q., Liu, Y., & Tan, C. Y. (2025). Effects of family socioeconomic status on educational outcomes in primary and secondary education: A systematic review of the causal evidence. *Educational Psychology Review*, 37(29). https://link.springer.com/article/10.1007/s10648-025-10004-8

Tymms, P., Merrell, C., Heron, T., Jones, P., Albone, S. & Henderson., B. (2008). The importance of districts. *School Effectiveness and School Improvement,* 19(3), 261–274. https://doi.org/10.1080/09243450802332069

Yang, M., & Woodhouse, G. (2001). Progress from GCSE to A and AS Level: Institutional and gender differences, and trends over time. *British Educational Research Journal,* 27(3), 245–267. https://doi.org/10.1080/01411920120048296

Zhang, L. (2025). Estimating school and teacher effects on students' academic performance using multilevel models with three or more levels: a literature review. Discover Education, 4:4. https://doi.org/10.1007/s44217-024-00392-4

THE MISSION OF THE LEGISLATIVE AUDITOR GENERAL IS TO

AUDIT · LEAD · ACHIEVE

WE HELP ORGANIZATIONS IMPROVE