TXSmartSchools Methodology 2017

Technical Description

Contents

Apples to Apples ... 1
Academic Progress Measures ... 1
THE DATA ... 1
THE VALUE ADDED MODEL ... 2
THE CAMPUS-LEVEL AND DISTRICT-LEVEL MODELS: ... 3
ESTIMATION AND RANDOM EFFECTS ... 7
CONSTRUCTING THE ACADEMIC PROGRESS MEASURES ... 7
Spending Measures ... 7
THE DATA ... 8
IDENTIFYING FISCAL PEERS ... 12
District-Level Matches ... 13
Campus-Level matches ... 17
CONSTRUCTING THE SPENDING MEASURES ... 20
ENDNOTES ... 21
Appendix A: Estimating the Comparable Wage Indices ... 23
ENDNOTES ... 26

TXSmartSchools Methodology

Technical Description

Apples to Apples

Raw data seldom provide sufficient insight for effective decision-making. Because differences in educational context have to be taken into consideration to transform data into information, TXSmartSchools (TSS) uses recognized statistical methods to create our measures of academic progress and educational expenditures. Those methods are described below.

Academic Progress Measures

For the TXSmartSchools academic progress calculations, we follow the scholarly literature and use a value added model to generate our academic progress measures.[1] A value-added model measures the extent to which student performance in a school (or district) differs systematically from what would have been expected had the students attended school somewhere else. Schools where students perform better than expected, given their prior performance and demographic characteristics, have high academic progress. Schools where students perform worse than expected have low academic progress. For example, a school where none of the students are passing standardized tests would have high academic progress if the students are improving more rapidly than the norm elsewhere in the state. Similarly, a school where all of the students are passing standardized tests would have low academic progress if the students are failing to improve as much as their peers in other schools.

THE DATA

TEA provided all student-level data used in this analysis to the UT-Dallas Education Research Center. Students were included in the analysis of academic progress if they:

- were included in TEA's "Campus Accountability Subset";
- had valid current-year scores in required reading/language arts and/or mathematics exams;[2]
- had valid prior scores in reading/language arts and mathematics. It is not necessary that the prior score come from the previous school year;
- had valid indicators for grade level, race/ethnicity, eligibility for free or reduced-price lunches, Limited English Proficiency (LEP) status, Special Education status or Gifted and Talented status, and were sex-identified in the current year.

THE VALUE ADDED MODEL

The academic literature offers a variety of alternative value added models, some focused on estimating teacher effects instead of, or in addition to, campus effects. The TSS model is based on the Financial Allocation Study for Texas (FAST) model, which, in turn, was derived from a model developed by the Dallas Independent School District. The DISD and FAST models have been evaluated extensively over the years.[3]

The TSS model uses statistical methods based on a regression technique called hierarchical linear modeling (HLM) to predict the math and reading performance of individual students and separate out the contributions to growth for campuses and districts.[4] This approach measures academic growth by modeling current-year student achievement on STAAR/EOC reading or mathematics as depending on how the student performed in the prior year, and by other characteristics of students. These other factors, called “control” variables or “covariates,” were modeled to remove their influence on the Academic Progress Scores.

The TSS model uses, conceptually, a two-stage process, with the first stage adjusting for fair comparisons of all students and the second stage separating out the contributions of students, campuses and districts. This technique is known as a multi-level, random intercepts mixed model, with students, campuses, and districts each represented by a level. In the estimation, these stages are estimated simultaneously.

We use both a three-level campus model and a two-level district model. The first level in both models represents students; the next level(s) represent districts and/or campuses. Each level has its own equation and the components of each equation depend on the others. To produce estimates for each model, the levels were algebraically combined into a single equation called the mixed model. Estimates then were produced from statewide TEA data, with effects partitioned between districts, schools, and individual students.

The first level in both models has each student’s current year score (expressed as a standardized normal, or z-score) regressed on his or her same-subject prior-test score, the other-test (reading or math) prior-test score, and any characteristics important to maintaining fairness, including race/ethnicity, sex, whether the current and prior tests were taken in Spanish, free or reduced lunch status, and limited English proficiency (LEP) status—both of the latter two interacted across the three grade blocks of grades 4-5, 6-8, and 9-12 – the LEP status during prior-year test-taking, special education status, gifted status, and interactions with various combinations of the above. The full set of variables is included in the model section. The second and third levels only include random intercepts and do not include any covariates. This allows for the clustering of students within campuses, and campuses within districts, so that only the campus or district effect is measured.

The district model includes a second level that predicts the district effect as the residual over the level-one variables. The campus model includes second and third levels, which together provide value-added predictions at the campus level.

THE CAMPUS-LEVEL AND DISTRICT-LEVEL MODELS:

The campus model uses the notation of Raudenbush and Bryk (2002), where the student-level math or reading STAAR/EOC outcome is:

$$Y_{ijk} = \pi_{0jk} + \sum_{p=1}^{P} \pi_{pjk}\alpha_{pjk} + e_{ijk}$$

i = 1,…,m students (m varies by year)
j = 1,…,n campuses (n varies by year)
k = 1,…,o districts (o varies by year)
p = 1,…,v student-level variables (v varies as indicated in the variable list below)

Y_{ijk} = student STAAR/EOC reading or math score
π_{pjk} = student-level coefficients
α_{pjk} = student-level control variables
e_{ijk} = student-level random error, with $e_{ijk} \sim N(0;\sigma^2)$

Based on the FAST model, and with advice from other stakeholders, the following student-level control variables were included:

a_1 = Math prior-year test score
a_2 = Math prior-year test score squared
a_3 = Reading prior-year test score
a_4 = Reading prior-year test score squared
a_5 = African American (=1 if African American)
a_6 = Hispanic (=1 if Hispanic)
a_7 = Female (=1 if Female)
a_8 = African American x LEP
a_9 = Hispanic x LEP
a_{10} = African American x Female
a_{11} = Hispanic x Female
a_{12} = African American x Free or Reduced-Price Lunch
a_{13} = Hispanic x Free or Reduced-Price Lunch
a_{14} = LEP x Free or Reduced-Price Lunch
a_{15} = Female x Free or Reduced-Price Lunch
a_{16} = Free or Reduced-Price Lunch x Grades 4 or 5
a_{17} = Free or Reduced-Price Lunch x Grades 6 - 8
a_{18} = Free or Reduced-Price Lunch x Grades 9 - 12
a_{19} = LEP x Grades 4 or 5

a_{20} = LEP x Grades 6 - 8
a_{21} = LEP x Grades 9 - 12
a_{22} = Prior-year math x prior-year LEP
a_{23} = Prior-year reading x prior-year LEP
a_{24} = Spanish-language current-year test, grades 4-6 (=1 if Spanish test)
a_{25} = Spanish-language prior-year reading, grades 4-6 (=1 if Spanish test)
a_{26} = Spanish-language test prior-year math, grades 4-6 (=1 if Spanish test)
a_{27} = Spanish-language test prior-year reading, grades 4-6 x Reading prior-year test score
a_{28} = Spanish-language test prior-year math, grades 4-6 x Math prior-year test score
a_{29} = Gifted class (=1 if Gifted)
a_{30} = Special education class (=1 if Special Education)
a_{31}-a_q = Current grade x current test interactions (e.g. =1 if Algebra 1 and grade 8)(q varies for read/math)
a_{q+1}-a_v = Prior-year grade x prior-year test interactions, separate for both reading and math prior-year scores (v varies for read/math)

The campus-level is:

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$
$$\pi_{ljk} = \gamma_{l00}, \qquad l = 1,\dots,P$$

β_{00k} = campus-level coefficients
γ_{l00} = non-randomly varying intercepts
r_{0jk} = campus-level random effect, with $r_{0jk} \sim N(0;\tau_1^2)$

The district level allows for the clustering of campuses within school districts:

$\beta_{00k} = \gamma_{000} + \mu_{00k}$,
γ_{000} = non-randomly varying intercept
μ_{00k} = district-level random effect, with $\mu_{00k} \sim N(0;\tau_2^2)$.

The district model uses the same structure as the campus model for the student level, but without terms for campuses. Thus, student-level notation is the same as the campus model without the "j" terms:

The district level is:

$$Y_{ik} = \pi_{0k} + \sum_{p=1}^{P} \pi_{pk}\alpha_{pk} + e_{ik}$$

$$\pi_{0k} = \gamma_{00} + \mu_{00}$$
$$\pi_{lk} = \gamma_{l0}, \qquad l = 1,\dots,P$$

γ_{00} = non-randomly varying intercept

γ_{10} = non-randomly varying intercepts for student covariates
μ_{0k} = district-level random effect, with $\mu_{0k} \sim N(0; \tau_2^2)$

Table 1 presents selected regression coefficients and standard errors from the 2015-16 campus-level and district-level models for reading and mathematics.

Table 1

	Reading Campus-level Model	Reading District-level Model	Math Campus-level Model	Math District-level Model
Math prior-year test score	0.299 (0.001)***	0.301 (0.001)***	0.646 (0.001)***	0.651 (0.001)***
Math prior-year test score, squared	-0.025 (0.000)***	-0.025 (0.000)***	0.034 (0.000)***	0.035 (0.000)***
Reading prior-year test score	0.463 (0.001)***	0.468 (0.001)***	0.137 (0.001)***	0.140 (0.001)***
Reading prior-year test score, squared	-0.051 (0.000)***	-0.050 (0.000)***	-0.014 (0.000)***	-0.014 (0.000)***
African American	-0.056 (0.003)***	-0.067 (0.003)***	-0.093 (0.003)***	-0.109 (0.003)***
Hispanic	-0.017 (0.002)***	-0.025 (0.002)***	-0.048 (0.002)***	-0.058 (0.002)***
Female	0.159 (0.001)***	0.159 (0.001)***	0.002 (0.001)	0.003 (0.001)**
African American x LEP	0.140 (0.009)***	0.144 (0.009)***	-0.031 (0.009)***	-0.028 (0.009)***
Hispanic x LEP	-0.022 (0.004)***	-0.025 (0.004)***	-0.133 (0.004)***	-0.142 (0.004)***
African American x Female	-0.005 (0.003)**	-0.004 (0.003)	0.025 (0.003)*	0.025 (0.003)***
Hispanic x Female	-0.016 (0.002)***	-0.015 (0.002)***	0.004 (0.002)*	0.004 (0.002)*
African-American x Free or Reduced Price Lunch	-0.018 (0.003)***	-0.020 (0.003)***	0.007 (0.003)**	0.009 (0.003)***
Hispanic x Free or Reduced Price Lunch	0.005 (0.002)***	0.003 (0.002)	0.027 (0.002)***	0.034 (0.002)***
LEP x Free or Reduced Price Lunch	0.007 (0.003)**	0.008 (0.003)***	0.025 (0.003)***	0.026 (0.003)***
Female x Free or Reduced Price Lunch	-0.005 (0.002)***	-0.005 (0.002)***	0.004 (0.002)**	0.004 (0.002)**
Free or Reduced Price Lunch x Grades 4 or 5	-0.054 (0.002)***	-0.051 (0.002)***	-0.070 (0.002)***	-0.084 (0.002)***

Free or Reduced Price Lunch x Grades 6-8	-0.088 (0.002)***	-0.090 (0.002)***	-0.082 (0.002)***	-0.107 (0.002)***
Free or Reduced Price Lunch x Grades 9-12	-0.101 (0.002)***	-0.135 (0.002)***	-0.058 (0.003)***	-0.053 (0.003)***
LEP x Grades 4 or 5	-0.048 (0.004)***	-0.045 (0.004)***	0.092 (0.004)***	0.097 (0.005)***
LEP x Grades 6-8	-0.144 (0.004)***	-0.144 (0.004)***	0.018 (0.004)***	0.025 (0.005)***
LEP x Grades 9-12	-0.278 (0.005)***	-0.284 (0.005)***	0.040 (0.005)***	0.063 (0.005)***
Prior year math x prior year LEP	-0.033 (0.007)***	-0.033 (0.007)***	0.022 (0.009)**	0.022 (0.009)**
prior year reading x prior year LEP	0.038 (0.007)***	0.037 (0.007)	0.041 (0.009)***	0.037 (0.009)***
Spanish Language current year test	0.634 (0.005)***	0.644 (0.005)***	0.555 (0.007)***	0.566 (0.007)***
Spanish language prior year reading	-0.427 (0.004)***	-0.436 (0.004)***	-0.078 (0.003)***	-0.083 (0.003)***
Spanish language prior year math	-0.170 (0.005)***	-0.160 (0.005)***	-0.291 (0.005)***	-0.290 (0.005)***
Spanish language prior year reading x reading prior year test score	0.036 (0.003)***	0.033 (0.003)***	-0.006 (0.002)**	-0.010 (0.002)***
Spanish language prior year math x Math prior-year test score	-0.031 (0.004)***	-0.028 (0.004)***	-0.028 (0.004)***	-0.028 (0.004)***
Gifted class	0.189 (0.001)***	0.189 (0.001)***	0.181 (0.002)***	0.177 (0.002)***
Special Education class	-0.352 (0.001)***	-0.349 (0.001)***	-0.214 (0.001)***	-0.210 (0.002)***
Model includes current grade x current test interactions?	yes	yes	yes	yes
Model includes prior-year grade x prior-year test interactions, by subject?	yes	yes	yes	yes
Random effects for districts?	yes	yes	yes	yes
Random effects for campuses?	yes	no	yes	no
Number of Observations	2,465,864	2,465,864	1,972,791	1,972,791

Standard errors in parentheses.

The asterisks indicate that a coefficient is significant at the *** 1%, **5% and * 10% levels.

ESTIMATION AND RANDOM EFFECTS

The model was estimated using maximum likelihood. Unadjusted campus effects, r_{0jk}, and district effects, μ_{0k}, were predicted based on estimated variance components. These campus and district effects were constructed to minimize the expected mean-squared error and were reliability-weighted composites of, essentially, the ordinary least squares estimate for the relevant group (campus or district) and an estimate for the overall model.[5] For convergence reasons, variables with very small interaction cell sizes - those less than 200 - were omitted. However all remaining observations were included in the estimation and then, prior to release, any campuses or districts with less than ten observations were deleted for FERPA compliance.

These calculated effects were best linear unbiased predictions, often called empirical Bayes residuals, and formed the basis for estimating campus (or district) effects in most of the models previously cited. The unadjusted campus effect is relative to its district. The campus effect was summed with the district effect to compare across all campuses. Standard errors were also calculated for both the (adjusted) campus and district predictions.

CONSTRUCTING THE ACADEMIC PROGRESS MEASURES

The Academic Progress Score is based on a combination of value added in reading and math. Each year, we add the math value added measure to the reading value added measure and divide by two to get the combined value added measure for a campus or district. If the value-added measure is missing for one of the two subjects (as would occur if there are fewer than 10 useable student records for that campus or district) then the composite is also set to missing. The Academic Progress Score is a three-year moving average of the annual values for the composite value added measure. If there are fewer than three years of composite value added measures (as would be the case for a new campus), the academic progress score is set to missing.

To construct the Composite Academic Progress Percentile, we convert the academic progress scores into percentiles. The academic progress percentiles range from 0 to 99. Schools in the 99th percentile had academic progress scores that were better than 99 percent of Texas schools.

Spending Measures

Schools that operate in high cost-of-living communities must spend more dollars to provide the same level of real resources as other schools. Similarly, schools that serve more challenging student bodies must deploy more real resources to accomplish the same results as other schools. Economies of scale make the per-pupil cost of education lower in large school districts than in small ones. All of these factors—labor cost, student need, and size—combine to form an educational environment that shapes the decisions school districts make.

Any evaluation of school district efficiency must take differences in this educational environment into account. TXSmartSchools accounts for the educational cost environment by evaluating the fiscal performance of each school or district in comparison to that of its fiscal peers. Fiscal peers are schools or districts that operate in a similar labor market, are of similar size, and serve similar students.

THE DATA

The key to identifying fiscal peers is developing reliable data on the fiscal environment in which each school district operates. Guided by conversations with Texas stakeholders and the scholarly literature on educational productivity, the TSS research team matched school and districts on the basis of two labor cost indicators, two size measures, and five measures of student needs.

Labor Costs. The education sector is labor-intensive, requiring professional staff such as teachers and administrators as well as nonprofessional staff such as clerks, educational aides, and maintenance workers. To capture regional differences in the prices paid for professional staff, the TSS research team used a Comparable Wage Index (CWI) that measures regional variations in the prevailing wage for college graduates. The basic premise of a CWI is that all types of workers demand higher wages in areas with a higher cost of living or a lack of amenities. Thus, if Dallas accountants are paid 15 percent more than the state average accounting wage, Dallas engineers are paid 15 percent more than the state average engineering wage, Dallas nurses are paid 15 percent more than the state average nursing wage, and so on, then a CWI predicts that Dallas teachers would need to be paid 15 percent more than the state average teacher salary, and that Dallas principals would need to be paid 15 percent more than the state average principal salary.

Figure 1 illustrates the ACS-CWI used in the 2015-2016 analysis. See the appendix for details on the estimation of the 2015 ACS-CWI.

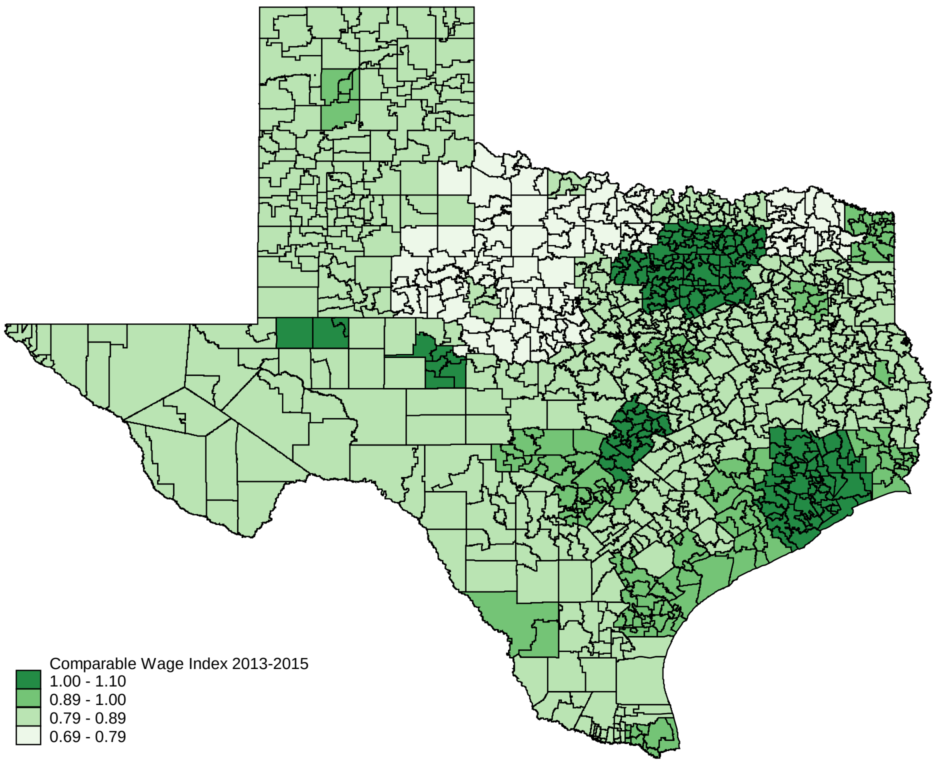

FIGURE 1: AMERICAN COMMUNITY SURVEY COMPARABLE WAGE INDEX 2013-2015

Because the wages of workers without a college degree may have a different geographic pattern than do the wages of college graduates, the TSS research team used a CWI that measures regional variations in the prevailing wage for high school graduates who do not have a bachelor's degree as the indicator for regional differences in the prices paid for non-professional staff. The appendix also presents estimation details for the construction of the High School CWI.

Size Measures. Differences in school district size are a primary determinant of variations in the cost of education. Districts with small enrollments are much more expensive to operate than are larger school districts, for a host of reasons. Small enrollment districts have higher administrative costs per pupil and may have classrooms that are too small to be cost effective, simply because there aren't enough students in a grade level to fill all the seats. On the other hand, districts with large geographic areas may be more expensive to administer because the students, teachers, and schools are highly dispersed. The school finance formula of the state of Texas recognizes the inherent cost differences of small enrollment districts by providing additional revenue to small and midsized school districts. Additional funding adjustments are also provided to small districts that serve a geographic area of more than 300 square miles. To reflect these potential cost drivers, our analysis includes two measures of school district size—

the number of students in fall enrollment, and the number of square miles in the district. TEA provided the data for these indicators.

Student Need. To capture variations in costs that derive from variations in student needs, districts were matched based on five measures of student demographics, the percentages of students in each district who were:

- High needs special education students (available only at the district level)
- Other special education students
- Limited English proficient (LEP) students
- Economically disadvantaged students
- High mobility students (those who missed six or more weeks at a particular school)

Schools are expected to need more resources (for example, specialized teachers and supplies, or smaller required class sizes) as the share of students in each category increases. Data on these school and district characteristics come from TEA's AEIS and TAPR reports and the individual student records housed in the Education Data Center at UTD.

Schools and districts are matched to their fiscal peers based on a three-year average of school and district characteristics. Using a three-year average reduces the influence of one-time events on the selection of fiscal peers.

In addition to the matching variables, the TSS spending index also requires data on actual expenditures by schools and districts, which come from TEA's Public Education Information Management System (PEIMS) and were reported to TEA by the districts themselves.

The key financial indicator for the TSS methodology is core operating expenditures. Core operating expenditures are current operating expenditures (as defined by TEA) on

- Instruction (function 11)
- Instructional resources (function 12)
- Curriculum and staff development (function 13)
- Instructional leadership (function 21)
- School leadership (function 23)
- Guidance counseling (function 31)
- Social work (function 32)
- Health services (function 33)
- Extracurricular activities (function 36)
- General administration (function 41)
- Facilities maintenance and operations (function 51)
- Security and monitoring (function 52)
- Data processing services (function 53) and

- Fund raising (function 81—charter schools only).

Unlike TEA's definition of current operating expenditures, our definition of core operating expenditures excludes student transportation (function 34), food service (function 35), the incremental costs associated with the chapter 41 purchase or sale of Weighted Average Daily Attendance (WADA) related to school district wealth sharing (function 92), and payments to juvenile justice alternative education programs (function 95). These categories of spending are not considered core operating expenditures because they represent additional functions of local school districts not directly related to student achievement. Notably, neither definition of operating expenditures includes spending on construction or debt service.

To reflect differences in school district purchasing power, the payroll component of core operating expenditures has been adjusted for regional differences in labor cost using the ACS-CWI. Adjusting payroll expenditures for differences in the ACS-CWI ensures that the Fiscal Index reflects the real resources each district is using to produce academic progress.

The core operating expenditures used to construct the TSS Fiscal Index are also adjusted for the fact that some school districts act as a fiscal agent for another district or group of districts. Fiscal agents collect funds from the member districts in a shared service agreement, and make purchases or pay salaries with those shared funds on behalf of the other member districts. As a result, the spending of fiscal agents is artificially inflated while the spending by member districts is artificially suppressed.

To correct for this pattern, we rely on TEA data about shared service arrangements (SSAs). School districts that serve as fiscal agents are required to indicate the amounts they spent on behalf of the member districts each year. We use this information to allocate the spending by fiscal agents to the member districts on a proportional basis. For example, in 2015-16, Hudson ISD spent $421,891 from shared service funds on instruction, $162,739 on school leadership, $92,939 on facilities maintenance and operations, and $82,692 on miscellaneous other functions. Hudson's SSA report indicates that it spent 19.4% of those funds ($147,795) on its own behalf, 57.1% ($434,185) on behalf of Lufkin ISD, 13.1% ($99,290) on behalf of Diboll ISD, and 10.4% ($78,991) on behalf of Central ISD. Therefore, we allocate 19.4% of Hudson ISD's shared service spending for instruction, 19.4% of its shared service spending for school leadership, 19.4% of its shared service spending on maintenance, and 19.4% of its shared service spending for other functions to Hudson ISD. We similarly allocate 57.1% of Hudson ISD's shared service spending in each category to Lufkin ISD, 13.1% to Diboll ISD, and 10.4% to Central ISD.

Unfortunately, the SSA distribution reports from roughly two-thirds of the fiscal agents are either missing or do not balance with their actual financial reports (Table 2). For example, South San Antonio ISD reported on the Public Education Information Management System (PEIMS) actual financial report for 2015-16 that it spent a total of $225,813 from shared service fund

435 on behalf of its member districts. However, South San Antonio ISD's SSA distribution report for the same year indicates that it spent a total of $610,162 from shared service fund 435 on behalf of 8 member districts. Either the actual financial report or the SSA distribution report must be wrong. Because the actual financial report is audited and the SSA report is generally not, we treat the actual financial report as the more reliable source of information. Whenever the SSA data are off by more than 2% and by more than $2,000, we conclude that it was not possible to reliably determine how those funds should be distributed and do not allocate the shared service spending. This means that despite our best efforts, total spending will be overstated for fiscal agents that file inconsistent SSA reports (or fail to file any SSA report at all), and will be somewhat understated for their corresponding member districts.

TABLE 2: DISTRICTS WITH INCONSISTENT SSA FINANCIAL DATA

	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16
NUMBER OF DISTRICTS OR CHARTER SCHOOLS SERVING AS FISCAL AGENTS	308	298	275	262	253	237	230	228
NUMBER OF FISCAL AGENTS FAILING TO FILE SSA DISTRIBUTION REPORTS	93	66	28	27	30	26	23	22
NUMBER OF DISTRICTS FILING AN INCONSISTENT SSA DISTRIBUTION REPORT	121	155	150	150	135	132	120	115
NUMBER OF DISTRICTS FILING A CONSISTENT SSA DISTRIBUTION REPORT	94	77	97	85	88	79	87	91

Note: An inconsistent SSA distribution report diverges from the PEIMS actual financial report by more than 2% and by more than $2,000.
Source: Texas Education Agency and TXSmartSchools.

For the campus-level measures, we rely on a narrower definition of core operating expenditures—campus-related core operating expenditures—which is defined as operating expenditures for instruction, instructional resources, instructional leadership, school leadership, and student support services (the total of all operating expenditures in functions 11-33). Unlike district core operating expenditures, campus-related core operating expenditures exclude extracurricular activities, general administration, facility maintenance and operations, security and monitoring services, and data processing services.

IDENTIFYING FISCAL PEERS

TSS uses a well-regarded research strategy to identify the fiscal peers for each school district—propensity score matching. Propensity score matching is a statistical strategy used to construct a control group (comparison group) for experiments that do not use random assignment.[6] For example, if you want to know the effect of a jobs training program, you need to compare the program participants to a group of nonparticipants who are as similar as possible to the participant group, so that you can be reasonably confident that differences in employment outcomes are the result of the training, and not a result of some other difference between the two groups. Propensity score matching identifies the best available potential controls for any given member of the treatment group. The TSS research team used propensity-score matching to identify the up to 40 school districts that are most similar to each Texas school district with respect to the common determinants of school district cost—labor costs, school district size, and student demographics. The team used a similar methodology and campus-level data to identify the fiscal peers for individual campuses.

District-Level Matches

Some Texas school districts are unusual enough in at least one cost dimension to limit their number of potential peers. For example, seven Texas districts are designated by the TEA as residential treatment facilities. Arguably, these seven districts should be matched only with one another. Similarly, while most school districts serve a full range of grade levels, some have no high school and others have no elementary schools. It seems most appropriate to match these restricted grade-level districts only to districts offering similar grade ranges.

Still another group, districts in the Alternative Education Accountability (AEA) system serving at-risk youth, seems to match poorly with other K-12 districts. Finally, a small number of districts in Texas are very large — more than 1,000 times larger than some other districts. It seems inappropriate to match a very large district with a very small one, no matter how similar they are in other respects.

To accommodate these unusual cases, the districts were stratified before applying the propensity score matching technique (Table 3). Each district was assigned to one of seven strata based on various student population characteristics, and propensity score matching was used as needed to identify fiscal peers within each stratum. If the stratum contained no more than 40 districts, then all districts in the stratum were designated as potential fiscal peers, and propensity score matching was not used.

Although the propensity score matching technique identifies up to 40 school districts that are a statistically valid comparison group for each district or campus, some of the matches are obviously better than others. And some of the districts within strata that were not propensity score matched are highly dissimilar. To improve the internal consistency of the comparison groups, the TSS team trimmed out potential fiscal peers that differed from the focus district by more than six standard deviations with respect to any of the cost factors.

The smallest stratum contained the seven districts designated by TEA as residential treatment facilities.[7] These seven districts represent an independent stratum in which each school has the same six potential fiscal peers. However, one member of the stratum is not like the others in multiple dimensions. Boys Ranch ISD (a special-purpose, traditional public school district that serves a residential facility for at-risk youth) is the only district in the category with a below average share of special education students. As a result, it has no fiscal peers. The other districts in the stratum have between two and five fiscal peers.

The 24 smallest K-12 districts — those with no more than 125 students on average over the last three years — comprised their own stratum and were matched accordingly. It seems unreasonable, however, to exclude possible matches with slightly more than 125 students; after all, the best possible match for a district with 124 students could be a district with 126 students. Therefore, districts with 125 or fewer students were matched with any K-12 district having no more than 140 students. Thirty-four K-12 districts had a three-year average of no more than 140 students in fall enrollment, so each of the smallest K-12 districts had 33 potential fiscal peers. After trimming, 20 of the 24 districts in this stratum has 33 fiscal peers and the remaining four districts had 31 fiscal peers.

The 18 largest Texas school districts — those with an average of more than 50,000 students over the last three years — also comprised their own stratum. These districts also were matched with any district having at least 40,000 students. Therefore, each of the largest districts had 28 fiscal peers. Trimming was not necessary for any of the districts in this stratum.

TABLE 3: TEXAS SCHOOL DISTRICTS BY STRATUM, 2015-16

	TOTAL NUMBER OF DISTRICTS	PROPENSITY SCORE MATCHED?
RESIDENTIAL TREATMENT FACILITIES (AEA)	7	no
VERY SMALL K-12	24	no
VERY LARGE K-12	18	no
AEA DISTRICTS	12	no
ALL OTHER K-12 DISTRICTS	983	yes
NO ELEMENTARY GRADES	22	no
NO HIGH SCHOOL GRADES	120	yes
TOTAL	1,186	

Note: Districts that opened after the 2012-13 school year and districts without PEIMS financial data are not included. “Very small” K-12 school districts have no more than 125 students. “Very large” K-12 districts have more than 50,000 students. Alternative Education Accountability (AEA) school districts are not residential treatment facilities and serve both elementary and secondary grade levels.
Source: TXSmartSchools.

AEA districts serve students at high risk of dropping out and are subject to different accountability standards. Twelve districts that were not residential treatment facilities served both elementary and secondary grades and were classified as AEA districts by TEA. These twelve districts represent an independent stratum in which each school has the same 11 potential fiscal peers. After trimming, nine districts had 10 potential fiscal peers and two districts had nine potential fiscal peers. The twelfth district—the University of Texas University Charter—was highly dissimilar from the other members of the stratum, and like Boys Ranch ISD has no fiscal peers.

Twenty-two school districts have no elementary grade levels. All of them are charter school districts except for South Texas ISD, the state's only all-magnet school district. Most of them are AEA districts. Each of the districts in this stratum had exactly 21 potential fiscal peers. After trimming, the districts in this stratum had between 14 and 21 fiscal peers.

The largest stratum consists of districts serving both elementary and secondary school children. Propensity score matching was used to identify fiscal peers for each of the districts in this stratum, "All Other K-12." To estimate the propensity scores, districts were divided into two groups based on size (those with fewer than 1,600 students and those with at least 1,600 students).[8] The two groups were then subdivided into a total of 20 similarly sized subgroups (twelve small and eight non-small) based on core operating expenditures per pupil, whether or not the district was above average with respect to the percentage of economically disadvantaged students, and whether or not the district was located in a metropolitan area.[9] By grouping campuses and districts by size, student poverty and metropolitan status, and then by core operating expenditures per pupil, the designated fiscal peers are ensured to be similar to one another with respect to the three primary determinants of educational cost—economies of scale, student poverty and geographic variations in labor costs.

Each of the 20 subgroups then was assigned to a treatment group. The research team estimated the corresponding probability model using the nine cost factors, their squares and the cube of log enrollment as control variables. Regardless of size, all non-AEA K-12 school districts are eligible matches and included in the set of possible control schools for each of the 20 subgroup analyses. Therefore, while there were 985 possible treatment districts in the stratum, there were 1,025 possible matches for each district.

For each model, a corresponding distribution of propensity scores was calculated. These 20 sets of propensity scores were used to identify fiscal peers for all but the smallest and largest of the state's K-12 school districts. The research team identified the 40 school districts with the nearest propensity scores to that of each treatment district. Thus, propensity scores from model 1 were used to find the nearest neighbors for districts in the first subgroup, while the propensity score from model 20 identified the nearest neighbors for the districts in the last subgroup.

It is important to note that each district's peers were drawn from the other 1,025 districts. Each district can have a unique peer group, so that the peer groups of a particular district's peers will not necessarily be the same.

Potential matches with propensity scores more than two standard deviations away from the district's own score were discarded, as were potential matches that were more than six standard deviations away from the district itself in any cost dimension. If 40 neighbors were not within a two-standard- deviation radius of the propensity score and six standard deviations of each variable, then the district has fewer than 40 fiscal peers. Ninety percent of the 983 districts have 40 fiscal peers, and 97 percent have more than 20 fiscal peers, but there are a few districts that are sufficiently unique in one or more dimensions to have only a handful of fiscal peers, and one charter school—John H. Wood Public Charter District—with no viable fiscal peers.

The final remaining stratum contains the 120 school districts with no high school. None of these districts are AEA districts. Because the stratum is not small, we used propensity score matching to find fiscal peers for each of these districts. The stratum is not large enough, however, to be divided into 20 subgroups, as was done with the All Other K-12 Districts stratum. Therefore, the districts were divided into five groups based on their metropolitan status and core operating expenditures per pupil.

As with the stratum of 983 K-12 districts, each of the five subgroups was assigned as a treatment group, and the corresponding probability model was estimated using the nine cost factors and their squares as control variables.

Again, the 40 school districts with the nearest propensity scores to those of each designated treatment district were identified, and potential matches outside of a two-standard-deviation band were discarded. Potential matches that were more than six standard deviations different from the district were also discarded. With the exception of Doss Consolidated ISD (which had only 12 students in 2016), all of the districts in this stratum had at least 13 viable propensity score matches.

Assessing Improvements in Match Quality

In the end, the goal is to identify up to 40 peer districts that are highly similar to each individual district. Match quality is evaluated based on the extent to which the designated peers differ from the district itself with respect to each of the nine cost factors. The mean squared error (MSE) for each cost factor measures the sum of squared differences between the district value for a cost factor and the peer values for that cost factor.[10] It represents the average deviation from baseline for the districts in the peer group. Lower MSEs indicate better matches; higher MSEs indicate poorer matches.

Table 4 illustrates the distribution of mean squared errors for each of the nine cost factors for 2015 and 2016. As the table makes clear, refinements in the matching methodology for 2016 have systematically improved the internal similarity of the fiscal peer groups with respect to student need, while having an only negligible impact on internal similarity with respect to the other determinants of district cost.

TABLE 4: COMPARING MATCH QUALITY, 2015-16 AND 2016-17

COST FACTOR	MEAN SQUARED ERROR 2014--15	MEAN SQUARED ERROR 2015-16
ENROLLMENT	13.08	13.69
SQUARE MILES	45.71	47.37
ACS-CWI	1.24	1.39***
HS-CWI	0.84	0.76***
ECONOMICALLY DISADVANTAGED STUDENTS	11.58	6.58***
HIGH NEEDS SPECIAL EDUCATION STUDENTS	1.31	0.99**
OTHER SPECIAL EDUCATION STUDENTS	1.28	1.18
LIMITED ENGLISH PROFICIENT (LEP) STUDENTS	22.29	20.26
HIGH MOBILITY STUDENTS	5.78	4.48**
NUMBER OF DISTRICTS	1,188	1,182

Note: The asterisks indicate that the difference between 2014-15 and 2016 is statistically significant at the 1% (***), 5%(**) or 10% (*) levels.
Source: TXSmartSchools.

Campus-Level matches

The Texas public school system includes more than 8,000 campuses that differ widely with respect to size and student demographics. We focus on those with at least 25 students in fall enrollment (on average over the three year period from 2013-14 through 2015-16).

It seemed most appropriate to match schools that serve similar grade levels. Therefore, the campuses were stratified according to the grade levels served in 2016 (early elementary, elementary, middle, secondary, and multi-level).[11] The secondary campuses also were divided into very large high schools and other high schools. (The very large high schools have at least 2,000 students, and are roughly analogous to the division 5A high school classification used for interscholastic athletics. No other type of campus is this large.) Finally, the model separated out AEA residential campuses, AEA nonresidential campuses, juvenile justice campuses, and special education campuses (those serving more than 75 percent special education students). Table 5 describes the number of campuses in each stratum.

TABLE 5: TEXAS PUBLIC SCHOOL CAMPUSES BY STRATUM, 2015-16

TYPE OF CAMPUS	NUMBER OF CAMPUSES	PROPENSITY SCORE MATCHED?
EARLY ELEMENTARY SCHOOLS*	330	Yes
ELEMENTARY SCHOOLS	4,073	Yes
INTERMEDIATE SCHOOLS	94	Yes
MIDDLE SCHOOLS	1,581	Yes
VERY LARGE SECONDARY SCHOOLS*	262	Yes
OTHER SECONDARY SCHOOLS	958	Yes
MULTI-LEVEL SCHOOLS	314	Yes
AEA RESIDENTIAL SCHOOLS		
SECONDARY SCHOOLS	22	No
OTHER SCHOOLS	33	No
AEA NON-RESIDENTIAL SCHOOLS		
ELEMENTARY AND EARLY ELEMENTARY SCHOOLS	0	No
MIDDLE SCHOOLS	12	No
SECONDARY SCHOOLS	194	Yes
MULTI-LEVEL SCHOOLS	15	No
JUVENILE JUSTICE SCHOOLS	71	Yes
SPECIAL EDUCATION SCHOOLS	19	No
TOTAL	7,978	

Note: Early elementary schools serve students up through the second grade. Intermediate schools are elementary schools that only serve grades 5 and up. Very large secondary schools have more than 2,000 students. Juvenile Justice schools are either Juvenile Justice Alternative Education Program (JJAEP) or Disciplinary Alternative Education Program (DAEP) schools. Special education schools serve at least 75 percent special education students.
Source: TXSmartSchools.

Propensity score matching then was applied within each stratum containing more than 40 members. As with the district-level analysis, campuses were sorted into expenditure subgroups within each stratum. In this case, however, the sorting was based on operating expenditures per pupil for campus-related activities instead of the broader definition employed in the district-level analysis.[12] Operating expenditures for campus-related activities (instruction, instructional services, instructional leadership, school leadership, and student support services) are more consistently defined across campuses due to the way districts allocate administrative costs. Some districts allocate most of their central administration activities to specific campuses, while others do not. Virtually all districts allocate their campus-related operating expenditures.

The elementary, middle, and secondary campuses then were divided into two groups — metropolitan and nonmetropolitan schools — and then subdivided into subgroups based on their instructional operating expenditures per pupil and whether or not they were above average with respect to the percentage of economically disadvantaged students. There were too few nonmetropolitan schools in the early elementary schools, large secondary schools, and AEA strata, so these strata are not divided into regional groups before subdividing by instructional expenditures per pupil.

Once divided into strata and subgroups, propensity score matching was used to identify the fiscal peers for each stratum with more than 40 campuses. The matching analysis used campus-level versions of most of the cost factors included in the district-level analysis. Geographic size is not relevant at the school level and was not included. High-needs special education students and other special education students cannot be differentiated at the campus level, and so those two groups were combined. The other six cost factors from the district-level model, as well as their squares, remained. In addition, we included an indicator for the highest grade level served in the matching model for multi-level schools.

To increase the quality of the potential matches for schools near the dividing line between very large and other secondary schools, we allowed very large secondary schools (those with at least 2,000 students) to match with any secondary school with at least 1,000 students, and other secondary schools (those with fewer than 2,000 student) to match with any secondary school with fewer than 3,000 students.

Again, the 40 campuses with the closest propensity scores (i.e. the 40 nearest-neighbor matches) within two standard deviations of the campus's own propensity score were designated as its potential fiscal peers. If 40 neighbors were not within a two-standard-deviation radius, the campus has fewer than 40 fiscal peers. As with the district-level analysis, we also trimmed out potential fiscal peers that were more than six standard deviations removed from the district itself in any dimension. The vast majority of campuses, however, have 40 viable, nearest-neighbor matches.

Assessing Improvements in Match Quality

As with the district-level analysis, the goal is to identify up to 40 peer campuses that are highly similar to each individual campus. Match quality is evaluated based on the extent to which the designated peers differ from the campus itself with respect to each of the campus-level cost factors. The mean squared error (MSE) for each cost factor measures the sum of squared differences between the campus value for a cost factor and the peer values for that cost factor. It represents the average deviation from baseline for the campuses in the peer group. Table 6 illustrates the distribution of mean squared errors for each of the seven cost factors for 2015 and 2016. Lower MSEs indicate better matches; higher MSEs indicate poorer matches. As the table makes clear, refinements in the matching methodology for 2016 have systematically improved the internal similarity of the fiscal peer groups with respect to student need and campus size, while having an only negligible impact on internal similarity with respect to labor cost.

TABLE 6: COMPARING MATCH QUALITY, 2015-16 AND 2016-17

CAMPUS COST FACTOR	MEAN SQUARED ERROR 2014--15	MEAN SQUARED ERROR 2015-16
CAMPUS ENROLLMENT	6.94	6.57**
ACS-CWI	1.22	1.25**
HS-CWI	0.90	0.87***
ECONOMICALLY DISADVANTAGED STUDENTS	16.73	6.26***
SPECIAL EDUCATION STUDENTS	2.57	1.44***
LIMITED ENGLISH PROFICIENT (LEP) STUDENTS	31.78	20.57***
HIGH MOBILITY STUDENTS	5.69	3.94***
NUMBER OF CAMPUSES	7,955	7,966

Note: The asterisks indicate that the difference between 2014-15 and 2016 is statistically significant at the 1% (***), 5%(**) or 10% (*) levels.
Source: TXSmartSchools.

CONSTRUCTING THE SPENDING MEASURES

A district's Fiscal index is based on its core operating expenditures, adjusted for labor cost differences and shared service agreements. A three-year average of the adjusted core spending of a school district is compared with a three-year average of the adjusted core spending of its fiscal peer group. Districts that spend more than 80% of the districts in their peer group are identified as very high spending districts. Districts that spend more than 60% of the districts in their peer group are identified as high spending districts, and so on. Districts in the lowest spending quintile are identified as very low spending districts. Districts with fewer than four fiscal peers have no Fiscal Index.

The Fiscal Index for a campus is constructed the same way as the Fiscal Index for a district, except that the campus-level index is based on a narrower definition of core operating expenditures—campus-related core operating expenditures. Again, campuses with fewer than four fiscal peers have no Fiscal Index.

ENDNOTES

[1] W.L. Sanders, A. Saxton, and S.P. Horn, "The Tennessee Value- Added Accountability System: A Quantitative, Outcomes-Based Approach to Educational Assessment," in *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* by Jason Millman, ed., (Thousand Oaks, California: Corwin Press, 1997); Dale Ballou, William Sanders and Paul Wright, "Controlling for Student Background in Value-Added Assessment of Teachers," *Journal of Educational and Behavioral Statistics* (Spring 2004), pp. 37-65, http://web.missouri.edu/~podgurskym/Econ_4345/syl_articles/ ballou_sanders_value_added_ JEBS.pdf; Carnegie Corporation of New York, *Evaluating Value-Added Models for Teacher Accountability*, by Daniel F. McCaffrey, J.R. Lockwood, Daniel M. Koretz and Laura S. Hamilton, Rand Corporation (New York, New York, 2003), http://www.rand.org/pubs/monographs/2004/R AND_MG158.pdf; Steven Ponisciak and Anthony Bryk, "Value-Added Analysis of the Chicago Public Schools: An Application of Hierarchical Models," in *Value Added Models in Education: Theory and Applications*, by Robert Lissitz, ed., (Maple Grove, Minnesota: JAM Press, 2005); Robert Lissitz, Harold Doran, William Schafer and Joseph Willhoft, "Growth Modeling, Value Added Modeling, and Linking: An Introduction," in *Longitudinal and Value-Added Models of Student Performance*, by Robert Lissitz, ed., (Maple Grove, Minnesota: JAM Press, 2005) ; and Stephen W. Raudenbush, "Adaptive Centering with Random Effects: An Alternative to the Fixed Effects Models for Studying Time-Varying Treatments in School Settings," *Education Finance and Policy* (Fall 2009), pp. 468-491.

[2] Students who took TAKS/STAAR-Alternative tests were not included, due to significant differences in these versions of the test. Students who took algebra 2 or English 3 were not included as these tests were optional and so would not allow a fair comparison with campuses not requiring these tests.

[3] William J. Webster and George T. Olson, *An Empirical Approach to Identifying Effective Schools*, presented at the Annual Meeting of the American Educational Research Association, (New Orleans, Louisiana, April 23-27, 1984), pp. 1-35, http://www.dallasisd. org/eval/research/articles/Webster-An-Empirical-Approach-to- Identifying-Effective-Schools-1984.pdf ; William J. Webster, Robert L. Mendro, and Ted O. Almaguer, *Effectiveness Indices: The Major Component of an Equitable Accountability System*, presented at the Annual Meeting of the American Educational Research Association, (Atlanta, Georgia, April 12-16, 1993), pp. 1-40, http://www.eric.ed.gov/PDFS/ED358130.pdf ; William J. Webster, Robert L. Mendro, Karen L. Bembry and Timothy H. Orsak, *Alternative Methodologies for Identifying Effective Schools*, presented at the Annual Meeting of the American Educational Research Association, (San Francisco, California, April 17-21, 1995), pp. 1-78, http://www.dallasisd.org/eval/research/articles/Webster-Alternative-Methodologies-For-Identifying-Effective Schools-95. pdf ; Robert L. Mendro, William J. Webster, Karen L. Bembry and Timothy H. Orsak, *An Application of Hierarchical Linear Modeling in Determining School Effectiveness*, presented at the Annual Meeting of the American Educational Research Association, (San Francisco, California, April 17-21, 1995), pp. 1-44, http://www.dallasisd.org/ eval/research/articles/Mendro-Application-of-HLM-in-Determining-School-Effectiveness-1995.pdf; William J. Webster, Robert L. Mendro, Timothy H. Orsak and Dash Weerasinghe, *An Application of Hierarchical Linear Modeling to the Estimation of School and Teacher Effect*, presented at the Annual Meeting of the American Educational Research Association, (San Diego, California, April 13-17, 1998), pp.1-27, http://www.dallasisd.org/eval/research/articles/Webster- An-Application-of-Hierarchical-Linear-Modeling-1998.pdf; Dash Weerasinghe and Timothy Orsak, *Can Hierarchical Linear Modeling Be Used to Rank Schools: A Simulation Study with Conditions under which Hierarchical Linear Modeling is Applicable*, presented at the Annual Meeting of the American Educational Research Association, (San Diego, California, April 13-17, 1998), pp. 1-12, http://www.

dallasisd.org/eval/research/articles/Weerasinghe-Can-Hierarchical- Linear-Modeling-Be-Used-to-Rank-Schools.pdf; Dash Weerasinghe, Mark Anderson, and Karen Bembry, *Precision of Measures of Central Tendency: Computing an Effectiveness Index for Teachers*, presented at the Annual Meeting of the Southwestern Educational Research Association, (New Orleans, Louisiana, February 2001), pp. 1-12, http://www.dallasisd.org/eval/research/articles/Weerasinghe- Precision-of-Measures-Computing-an-Effectiveness-Index-for- Teachers-2001.pdf (last visited November 28, 2010); and William J. Webster and Robert L. Mendro, "The Dallas Value-Added Accountability System," in *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* by Jason Millman, ed., (Thousand Oaks, California: Corwin Press, 1997).

[4] Tom A.B. Snijders and Roel J. Bosker, *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Thousand Oaks, California: Sage Publications, 1999); Stephen W. Raudenbush and Anthony S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed., (Thousand Oaks, California: Sage Publications, 2002); Anders Skrondal and Sophia Rabe-Hesketh, *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models* (Boca Raton, Florida: Chapman & Hall/ CRC, 2004); and William H. Greene, *Econometric Analysis*, 6th ed. (Upper Saddle River, New Jersey: Prentice Hall, 2007.)

[5] Tom. A.B. Snijders and Roel J. Bosker, *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*; Stephen W. Raudenbush and Anthony S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd ed*.; and Anders Skrondal and Sophia Rabe-Hesketh, *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*.

[6] Rajeev H. Dehejia and Sadek Wahba, "Propensity Score Matching Methods for Nonexperimental Causal Studies," The Review of Economics and Statistics (February 2002), pp. 151-161, http://www.personal.ceu.hu/staff/Gabor_Kezdi/Program-Evaluation/ Dehejia-Wahba-2002-matching.pdf; Rajeev H. Dehejia, "Practical Propensity Score Matching: A Reply to Smith and Todd," Journal of Econometrics (No. 125, 2005), pp. 355-364, http://www- personal.umich.edu/~econjeff/Papers/dehejia_practical_pscore. pdf; and Marco Caliendo and Sabine Kopeinig, Some Practical Guidance for the Implementation of Propensity Score Matching (Bonn, Germany: IZA, May 2005), pp. 1-29, http://www.acoes.org.co/pdf/ Documentos%20HFTF/30.pdf. (Last visited November 29, 2010.)

[7] All of the districts with more than 35 percent special education students (which would have been designated as "special education districts" in the 2016 analysis) are included in this stratum.

[8] Traditional school districts with fewer than 1,600 students in average daily attendance are eligible for the small district adjustment under Texas' school finance formula.

[9] Metropolitan school districts are those located in a county that is part of a metropolitan statistical area as defined by the U.S. Office of Management and Budget. For a list of metropolitan counties, visit http://www.census.gov/population/www/metroareas/metroarea.html.

[10] We calculate the mean squared error for school district j as

$$MSE_j = 100 * \frac{\sum(x_i - x_j)^2}{n\bar{x}}$$

where x_j is the value of the cost factor for school district j, x_i is the value of the cost factor for peer district i, $\bar{x}$ is the statewide mean value of the cost factor and n is the number of school districts in the peer group. Dividing the squared errors by the statewide mean makes the scaling consistent across the nine cost factors, allowing for comparisons among them.

[11] Early elementary campuses serve students up through the second grade.

[12] Campus-related activities are all operating expenditures in functions 11-33, and objects 6100-6499

Appendix A: Estimating the Comparable Wage Indices

The ACS-CWI and HS-CWI for this analysis are based on analyses of public use micro-data from the 2013, 2014, and 2015 American Community Surveys (ACS).[1] The ACS, which is conducted annually by the U.S. Census Bureau, has replaced the decennial census as the primary source of demographic information about the U.S. population. It provides information about the earnings, age, occupation, industry, and other demographic characteristics for millions of U.S. workers. The ACS-CWI measures earnings differences for college graduates; the HS-CWI measures earnings differences for high school graduates who do not have a bachelor's degree. In both cases, the analysis is modeled after the baseline analysis used to construct the National Center for Education Statistics' (NCES) CWI.[2]

Like the NCES CWI, the ACS-CWI and HS-CWI are derived from regression analyses of individual earnings data. Workers with incomplete data and workers without a high school diploma were excluded from the estimation sample, as was anyone who had a teaching or educational administration occupation or who was employed in the elementary and secondary education industry. Self-employed workers were excluded because their reported earnings may not represent the market value of their time. Individuals who reported working less than half time or for more than 90 hours a week were also excluded, as were workers under the age of 18 and over the age of 80. Finally, individuals employed outside the United States were excluded because their earnings may represent compensation for foreign travel or other working conditions not faced by domestic workers.

The ACS-CWI and HS-CWI are estimated from nationwide data because the national sample is much larger and yields much more precise estimates of wages by industry and occupation than could be generated using only the ACS data for the state of Texas. For similar reasons, the analyses combines data from the three most recent administrations of the ACS.

Table A-1 presents the results from the two regression analyses (one for the ACS-CWI, one for the HS-CWI). The dependent variable in each case is the log of annual wage and salary earnings. Key independent variables include the age, sex, race, educational attainment, language ability, and amount of time worked for each individual in the national sample. The model includes the interaction between sex and age, to allow for the possibility that men and women have different career paths, and therefore different age-earnings profiles. In addition, the estimation includes indicator variables for occupation and industry for each year.[3] This specification allows wages to rise (or fall) more slowly in some occupations or industries than it does in others. Such flexibility is particularly important because the analysis period includes the period immediately after the "Great Recession" and some industries and occupations are recovering more slowly than others. Finally, each regression includes indicator variables for each labor market area.

The labor markets are based on "place-of-work areas" as defined by the Census Bureau. Census place-of-work areas are geographic regions designed to contain at least 100,000 persons. The place-of-work areas do not cross state boundaries and generally follow the boundaries of county groups, single counties, or census-defined places (Ruggles et al. 2012). Counties in sparsely-populated parts of a state are clustered together into a single Census place-of-work area. All local communities in the United States are part of a place-of-work area. Individuals can live in one labor market, and work in another. Their wage and salary earnings are attributed to their place of work, not their place of residence. The labor markets used in these analyses are either single places of work, or a cluster of the places-of-work that comprise a metropolitan area.[4]

As Table A-1 illustrates, the estimated model is consistent with reasonable expectations about labor markets. Wage and salary earnings increase with the amount of time worked per week and the number of weeks worked per year. Earnings also rise as workers get older, but the increase is more rapid for men than for women (perhaps because age is not as good an indicator of experience for women as it is for men). Workers with advanced degrees earn systematically more than workers with a bachelor's degree (in the ACS-CWI model) while workers with an associate's degree earn significantly more than workers with a GED (in the HS-CWI model). Whites earn systematically more than apparently comparable individuals from other racial groups. Workers who do not speak English well earn substantially less than other workers, all other things being equal.

The predicted wage level in each labor market area captures systematic variations in labor earnings while controlling for demographics, industrial and occupational mix, and amount of time worked.[5] Dividing each local wage prediction by the corresponding national average yields the ACS-CWI, and the HS-CWI, respectively.

TABLE A-1: ESTIMATING THE ACS-CWI AND HS-CWI

EXPLANATORY VARIABLES	HS-CWI MODEL		ACS-CWI MODEL	
	ESTIMATE	STD. ERROR	ESTIMATE	STD. ERROR
USUAL HRS. WORKED PER WEEK	0.977	0.002	0.942	0.003
WORKED 27-39 WEEKS	-0.455	0.002	-0.558	0.004
WORKED 40-47 WEEKS	-0.222	0.002	-0.246	0.003
WORKED 48-49 WEEKS	-0.096	0.003	-0.105	0.004
FEMALE	0.294	0.007	0.284	0.014
AGE	0.060	0.000	0.086	0.000
AGE, SQUARED	-0.001	0.000	-0.001	0.000
FEMALE*AGE	-0.020	0.000	-0.015	0.001
FEMALE*AGE, SQUARED	0.000	0.000	0.000	0.000
NOT AN ENGLISH SPEAKER	-0.289	0.006	-0.509	0.021
REGULAR HIGH SCHOOL DIPLOMA	-0.039	0.001		
GED	-0.094	0.002		
LESS THAN 1 YEAR OF COLLEGE	0.000			
SOME COLLEGE, NO DEGREE	0.016	0.001		
ASSOCIATE'S DEGREE	0.046	0.001		
BACHELOR'S DEGREE			-0.219	0.003
MASTER'S DEGREE			-0.103	0.003
PROFESSIONAL DEGREE			0.000	
DOCTORAL DEGREE			0.059	0.004
HISPANIC	-0.080	0.001	-0.098	0.003
AMERICAN INDIAN	-0.066	0.005	-0.128	0.002
BLACK	-0.095	0.001	-0.076	0.011
CHINESE	-0.161	0.005	-0.087	0.004
JAPANESE	-0.002	0.009	-0.085	0.008
OTHER ASIAN/PACIFIC ISLANDER	-0.113	0.003	-0.078	0.002
OTHER RACE, N.E.C.	-0.049	0.002	-0.067	0.005
MIXED RACE	-0.042	0.003	-0.064	0.004
WHITE	0.000		0.000	
INDUSTRY*YEAR INDICATORS?	Yes		Yes	
OCCUPATION * YEAR INDICATORS?	Yes		Yes	
LABOR MARKET INDICATORS?	Yes		Yes	
NUMBER OF OBSERVATIONS	1,312,686		831,457	

Source: Ruggles et al. (2015) and author's calculations.

ENDNOTES

[1] The analysis is based on annual files for each survey administration, and not on the combined three-year file. The ACS for 2012, 2013 and 2014 were used to construct the ACS-CWI and HS-CWI used in the TSS analyses for 2013-14 and 2014-15. See the TxSmartSChools Methodology for 2016.
[2] Taylor and Fowler (2006).
[3] The model also includes random effects for states. Treating state effects as random rather than fixed ensures that the predicted wage is the same in Kansas City, Kansas as it is in Kansas City, Missouri, while allowing for a correlation in the errors among labor markets within any given state.
[4] Place of work areas were matched to counties and aggregated into core based statistical areas using data from the Missouri Census Data Center's MABLE/Geocorr12: Geographic Correspondence Engine.
[5] Formally, the predicted wage level in each market is the least-squares mean for the market fixed effect. The least-squares mean (or population marginal mean) is defined as the expected value of the mean for each effect (in this context, each market) that you would expect from a balanced design holding all covariates at their mean values and all classification variables (such as occupation or sex) at their population frequencies