Pupil Transportation Allocation Study:

How the state funds K-12 student transportation

Report to the Governor and Legislature
Section 129, Chapter 357, Laws of 2020

Forecasting and Research
Office of Financial Management
December 2020

Table of Contents

Executive summary

The 2020 supplemental operating budget (Section 129, Chapter 357, Laws of 2020) directed the Office of Financial Management to study the state's pupil transportation funding model. OFM collaborated with the Office of Superintendent of Public Instruction, and those findings are presented in this report. We would like to thank the staff at OSPI for their cooperation and expertise. They shared data used in this analysis and patiently answered our questions.

This study was requested before the coronavirus pandemic forced school closures, so this analysis is only relevant in a pre- and post-pandemic world. Funding issues related to the pandemic aren't discussed, but it is important to note that the pupil transportation funding model is severely impacted by the pandemic. The model will not work until the 2022-23 school year, assuming in-person learning begins by the 2021-22 school year. Since the model relies on prior year trends, it will not work until the school year following a typical school year. Alternate funding strategies are needed until then.

What we found

1. We found that some districts experience significant funding gaps. This means some districts do not equally benefit from the state's transportation funds.
2. While we did not find a pattern to why or where this happens, we found that the gaps negatively impact these districts.
3. We found that if we made additional steps to the allocation model, we could increase funding.
4. We could not determine if OSPI is aligned with RCW 28A.160.192(1)(b).
5. The total amount of money that we put into pupil transportation is less than what school districts spend. (Page 20).

What we recommend

1. We need to close the transportation funding gaps by increasing the overall amount that school districts receive (Pages 20-25).
2. While we did not find a pattern to why these funding gaps happen, we recommend fixing this issue.
3. We recommend that OSPI add steps to the allocation calculation that increase the overall funding amount.
4. We need more data to find out if OSPI is in line with state law. To do that, we recommend that OSPI leads a study that samples school districts to get more data. From there, we can determine if this issue impacts transportation funding (Pages 9-10).

How the model is working, where the dollars are going

We need more data to know if the model works for all school districts

Some school districts have consistently fared worse than others. Our analysis shows there are some school districts whose allocation was at 90% or less of their costs. This happened for at least three out of the past four years. **It might be worth tracking this statistic in the future to make sure the same school districts aren't affected year after year. If they are, then the allocation model may not be working as well for this subset of school districts.**

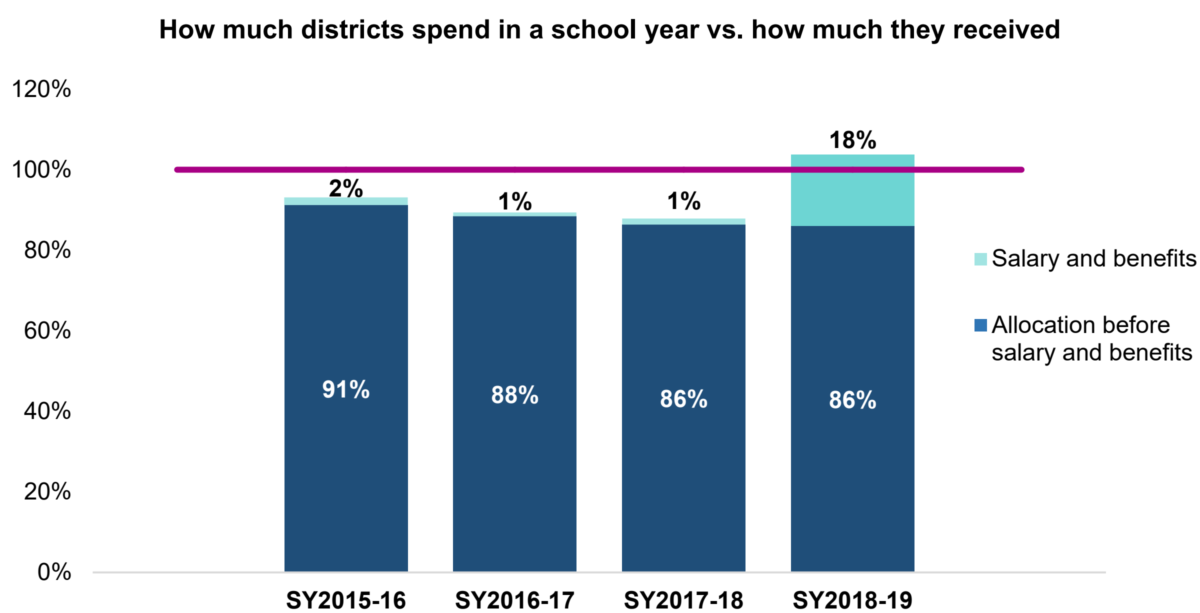
On a positive note, the overall allocation system improves the distribution of pupil transportation funds. It does this by reducing many of the high allocation numbers that STARS calculates. (You can find information on STARS on Page 26). On the flip side, certain school adjustments can lift up some of the low amounts that STARS allocates for certain districts. But not all districts qualify for adjustments in a given year. So, many districts still get less than 90% of their costs.

School districts typically spend more than they are given

During the last three of four years, districts got less money to cover their expenditures (costs) than they needed. The $100 million-plus legislative appropriation during the 2018-19 school year means that most got just a bit more than they needed (see table below).

Here's the most important thing to take away from this graph:

- *Without* the huge influx resulting from the McCleary decision, school districts – for the fourth year – *would not* have enough allocated money to cover their costs.



How the system calculates how much to give a school district

The state displays two allocation amounts for each school district.

The first amount is what STARS calculates, plus any adjustment that a district already qualified for (i.e., if a district qualifies for "low ridership" or is part of a transportation co-op).

The second amount is last year's costs plus additional state resources that OSPI calculates using a federal rate (known as the "federal restricted indirect rate"). It based this amount on district level costs that benefit the transportation program. These district level activities include payroll services, insurance, utility fees, etc. **The state then gives the school district the smaller amount,** and then we add the salary and benefit appropriated by the Legislature. This means that many school districts get an allocation less than what STARS calculated.

Here are the most important takeaways from the table below:

- For the first three years, how much districts got was *less than* the STARS model calculated. However, higher (and unprecedented) compensation that came from legislation means the past two years show that districts got *more than* what the STARS model calculated.
- Even after we get the STARS amount, it can change because of two factors. The first factor comes about when we add any qualifying district adjustments. The second factor comes about when we add how much a district spent last year on pupil transportation (which we then add to a small, federal funding amount). Finally, we look at the two numbers and use the lesser (smaller) number as the final allocation number (we call this "lesser than provision").

How the state calculates district's final amount					
	2015-2016	**2016-2017**	**2017-2018**	**2018-2019**	**2019-2020**
STARS (expected allocation)	430,608,406	440,280,916	477,241,262	521,477,289	575,466,780
Adjustments	6,461,362	7,080,836	6,443,332	7,829,226	8,202,939
Lesser than provision	(28,059,673)	(23,449,218)	(25,589,521)	(28,597,182)	(33,286,551)
Legislative compensation	8,361,267	4,751,986	7,792,629	103,255,459	25,489,698
Final Allocation	417,371,362	428,664,521	465,887,701	603,964,792	575,872,866

The process that OSPI uses after it gets the STARS calculation helps the final numbers land closer to actual district costs (see the table below). This helps school districts with a very low STARS calculation (that would've gotten less money), get more money.

How the STARS calculation compares to how much the school districts actually get	2015-16	2016-17	2017-18	2018-19
Within +/- 10% of costs				
STARS	34%	38%	39%	34%
Actual Allocation	63%	60%	61%	54%
More than +/- 25% of costs				
STARS	28%	25%	26%	31%
Actual Allocation	7%	10%	9%	10%

When this happens, the system works to "trim" many of the allocations that are 25% above or below a district's costs. This is because STARS gives more than a quarter of school districts an allocation that is 25% higher or lower than their actual costs. The "trimming" also increased (about 20 percentage points) the percent of school districts that got funds within 10% of their actual costs.

Overall, the process of calculating allocation works for most districts. Considering that STARS is part of a whole allocation system, **we can argue that STARS works well for most districts** because STARS isn't designed to work in a vacuum.

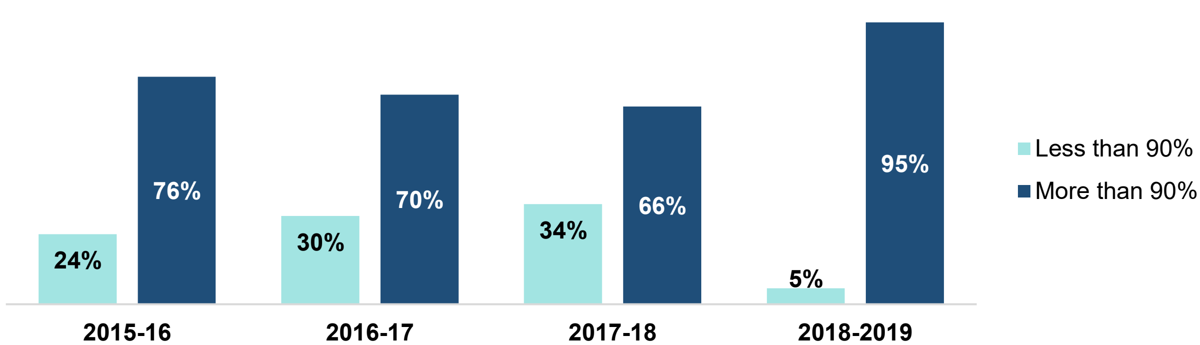
Allocation by school district

During the last four years, many school districts received an allocation below their costs. This is a minor concern when low allocations are just under expenditures (costs). But many school districts get an allocation less than 90% of their costs. The 2018-19 school year was the only year with a low number of school districts who got an allocation less than 90% of expenditure. During the three previous school years, school districts with an allocation of less than 90% of costs ranged from 23% to 33%.

Here is the most important thing to take away from this graph:

- Many school districts get less than 90% of their costs.

The percentage of districts that got allocation above and below 90% of their costs

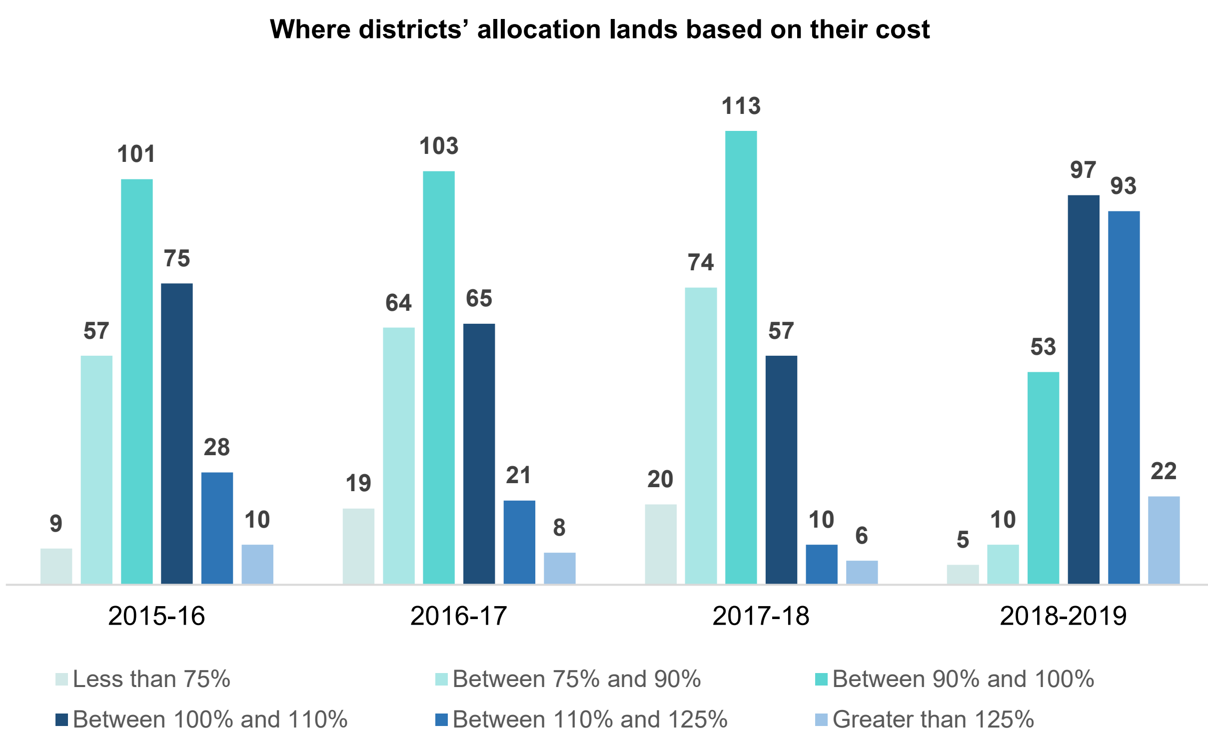


The next chart shows school districts that fall into different categories of allocation. We divided school districts into six categories. Then, we organized these categories by where a district landed with their share of the costs and their allocation:

- Less than 75% of expenditure
- Between 75% and 90% of expenditure
- Between 90% and 100% of expenditure
- Between 100% and 110% of expenditure
- Between 110% and 125% of expenditure
- Greater than 125% of expenditure

Here's the most important thing to take away from this table:

- Usually, school districts got less than they spent. **But they were rarely given more or less than 25% of costs.**



The distribution during the first three years is similar. Few school districts have very large or very small allocations as a percent of expenditure. But it was more common for school districts to have an allocation that was below expenditure. For the 2018-19 school year, many more school districts got an allocation that was more than the expenditure amount and very few that had an allocation of less than 90%.

How we determine allocation consistency

Historically, a significant share of school districts got an allocation below 90% of their costs. But how often does that happen to the same school district? Too often. The allocation model is equitable if different school districts cycle in and out of that group, but there may be a bias if that doesn't occur. A way to determine this is to count the number of years each school districts falls into an allocation category.

There were 217 school districts that got an allocation of 90% or more of costs for at least three out of the last four years. And, there were 36 school districts that had an allocation of less than 90% of costs during three out of the last four years. Even though so many school districts fared well, a high share (13%) were given an allocation below 90% of costs.

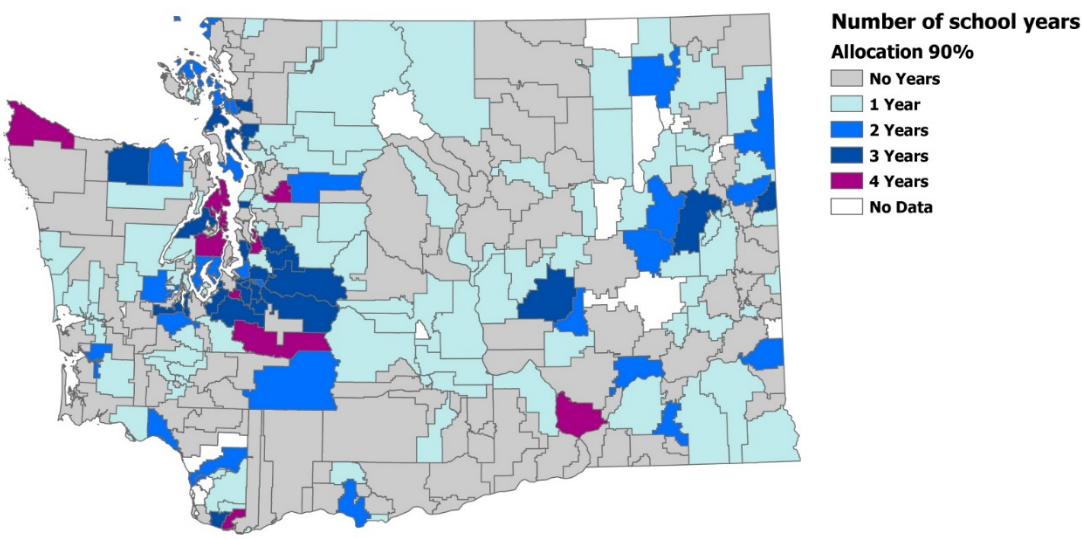
The maps on the next two pages show school districts and the number of years they got an actual allocation in the lower two categories (less than 75%; between 75% and 90%). This is when school districts are most adversely affected because they spend more than they get. It's worthwhile to investigate what school districts they are and if there are commonalities.

Here are the most important takeaways from the first map below:

- Thirty-six school districts were funded at less than 90% of costs during three out of the last four years. These large and small school districts are scattered throughout the state.

In this map, the maroon areas show districts that got 90% of costs each year during the last four years. Dark blue areas show those that got 90% of costs during the same timeline.

School districts with allocation less than 90% of expenditure



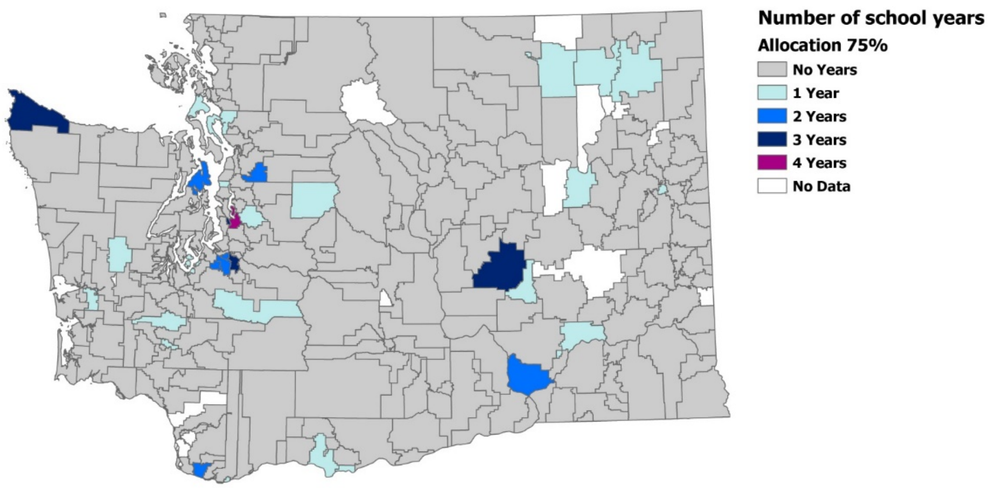
School districts with three years: Auburn, Bethel, Central Kitsap, Clover Park, East Valley, Enumclaw, Evergreen, Fife, Highline, Issaquah, Moses Lake, Mount Vernon, Oak Harbor, Olympia, Orting, Port Angeles, Puyallup, Reardan-Edwall, Shoreline, Stanwood-Camano, Sumner, Tahoma, Tukwila, University Place, White River

School districts with four years: Bainbridge, Camas, Cape Flattery, Eatonville, Franklin Pierce, Mercer Island, Monroe, North Kitsap, Pasco, Renton, South Kitsap.

Here is the most important takeaway from the second map below:

- Six school districts were funded less than 75% of costs during three out of the last four years. This represents a small number of districts, and the vast majority were never funded at that level.

School districts with allocation less than 75% of expenditure



School districts with three years: Cape Flattery, Moses Lake, Sumner, Tukwila

School districts with four years: Mercer Island, Renton

The regression model

Using the regression model in the right way

Here's the problem with using a regression model: It comes up with one number for each coefficient for the whole state. So, OSPI uses that number to determine how much money a district should get. But using this model (that spits out an average) means that some districts will be really close to the average and some will land too far above or below the average – **sometimes coming up with an incomplete picture of the true need.**

The problem is that some districts have a lower than average relationship with these variables than the statewide average.

So, since we can't find a pattern to why this happens, it means we can't yet fix the problem. While it could be that the districts are just on the losing end of this regression model, we can't yet rule out if the model is throwing off what OSPI gives to those districts. A regression model can produce spurious results in the wrong hands, so it's a good idea to "check under the hood."

The regression model and how well it follows the law

The STARS model must follow [RCW 28A.160.192](RCW 28A.160.192).

The STARS model **aligns with** RCW 28A.160.192(1)(a). All of the cost factors are used as independent variables in the first run of the regression model. Only those that are statistically significant at the 95% confidence interval are used in the final run.

The STARS model **aligns with the *first part* of** RCW 28A.160.192(1)(b) but **not the second part.** We base estimated expenditures on average relationships defined by coefficients because we use them to calculate average costs. The model does not align to the part that says we need to limit the previous year's cost variable to certain compensation. The previous year's cost variable would have to be reduced to remove certain compensation amounts. We don't have expenditure detail that explicitly breaks down the cost components, which includes the separation of base salary and compensation. The model is not aligned with statute in this regard, but we may not be able to mandate this because we don't have the full data.

We encourage OSPI to lead a study to address 28A.160.192(1)(b)

We don't know a lot about how much the previous year's expenditure variable would be reduced if we limited it to "the base salary or hourly wage rate, fringe benefit rates, and applicable health care rates provided in the omnibus appropriations act" (RCW 28A.160.192(1)(b)). However, a study involving a sample of school districts could determine this. Some school districts could collect salary information throughout a school year and report the information to OSPI so that we better know the statute's magnitude.

The formula must allocate funds to school districts based on the average predicted costs of transporting students to and from school, using a regression analysis. That means we should only use statistically significant factors in the regression analysis.

It's a bit outside the study's scope to recommend statutory changes, but we want to point out an error in this statute:

> "Employee compensation costs included in the allowable transportation expenditures used for the purpose of establishing each school district's independent variable in the regression analysis shall be…"

The term *independent variable* should read *dependent variable*. The independent variables are the explanatory variables, such as basic ridership, land area, etc. The limitation is meant to be placed on the previous year's expenditures variable, which is the dependent variable in the regression model.

Top factors to think about when you run a regression model

This section explores if the regression is mathematically sound (which it is).

The regression model should show strong "goodness of fit" qualities, be theoretically sound, and meet the seven classical assumptions of ordinary least squares regression. Doing so will produce the best data. The regression model we are using for this section is based on the 2019-20 school year.

The model fits 'goodness of fit' qualities

One of the most important numbers we care about in this regression model is called the R^2. This number came in very high at 0.9679, which is good news. That indicates a good fit.

The variables we use in the model are theoretically sound

The independent variables used in the regression model should intuitively make sense and fit with theory. In this case, they do. School districts with higher basic and special ridership and more destinations *should* have higher costs. Also, school districts with more land area and greater average distances would have longer bus routes and cost more. The dummy variable make sense, too. School districts that don't have a high school and don't transport high school students should have lower costs. The variables in the model cover much of what should drive costs for the average district.

The model meets the seven assumptions

Ordinary least squares regression is the best type of linear estimation, but only under certain conditions. We call these conditions the classical assumptions of OLS. The pupil transportation regression is an OLS regression, so it must meet these assumptions. **This matters because an OLS regression that does not follow these rules might produce bad results.**

The regression meets the conditions of the classical assumptions with just one issue worth noting. There is a significant linear relationship between basic ridership and special ridership, which could cause problems. But there is enough data variance and the sample is large enough that the regression performs adequately. We present an in-depth description of the OLS assumptions in Appendix A.

How appropriate is the model?

The model falls short if the goal is to get an allocation close to actual expenditure (cost), **but it performs adequately if** we want to provide an allocation based on statewide averages.

However, it's a hard sell to say the model is appropriate when it consistently provides expenditure estimates well below actual costs (not counting the McCleary influx year).

In most situations, the STARS model is not going to provide adequate resources to school districts for pupil transportation funding. No amount re-specifying the regression model is going to change that.

We found two reasons for that:

- First, we are using coefficients from one year to predict the next year. This would work if coefficients remain stable across years, but they do not. While they do not move much, small movements can have large impacts when adding up all of the school districts. They change because the relationship between the independent variables and expenditure is a little different every year, as they should be. Some costs may incrementally increase each year because of factors such as inflation or higher fuel prices. And, increasing costs lead to higher expenditures, which, in turn, leads to different coefficients.
- Second, the "lesser than provision" concept (that we mention on Page 4) reduces allocations more than added adjustments increase it. This reduced the allocation in each of the last five years.

Historically, both of these factors contributed to a gap between allocation and expenditure, where allocation landed significantly lower. It's likely that this gap will persist unless we make changes to the allocation model and provide more dollars to districts (see our recommendations at the start of this report).

The model's data analysis is appropriate

OSPI did a good job selecting variables for the model. Plus, the natural log transformation (a common regression method) makes sense and is empirically supported.

A basic principle of building a statistical regression model is that the dependent variable (cost) is statistically dependent on the explanatory variables (such as basic ridership, special ridership and land area). Basically, we need to make sure the explanatory variables are linearly independent of each other. You can see the relationship between variables (using correlation statistics and correlation plots) in Appendix B.

We use the natural log of some variables in the regression model. So, this analysis compares the relationship of nontransformed and natural log transformed.

The main finding of correlation statistics is the land area variable is insignificant in both the transformed and nontransformed versions. This means **we cannot say** with confidence that this variable is linearly related to the dependent variable.

Why we may want to change the timeline

OPSI runs the allocation model *after* school districts submit their winter ridership and destination counts to OSPI. The legislative session is well underway by this point and we know that budget writers could benefit from knowing the pupil transportation allocation amount sooner.

If we want to change this timeline to work better with session dates, then we would need to remove or substitute the current year winter count. That's because we currently collect ridership and destination counts from the previous year spring count *and* the current year fall and winter counts (and then combine the data to determine a weighted average number).

There are two new ways OSPI could count calculations going forward. One, they could drop the winter survey. Or two, they could estimate the winter survey. We describe these options in the next two sections.

While these are just a few solutions, we wanted to offer some proposals and analyze how they would have changed allocations in the past. OFM ran these two new ideas using actual data from school years 2015-2016 to 2019-2020 and compared against actual cost.

OSPI could run this model twice. First, to calculate a total allocation amount to give budget writers before the legislative session starts. Then run again using the current method when we finally have the actual winter count. If history is any guide, the difference between running this twice will be less than 1% and this small change should not surprise the budget writers.

What happens if we drop the winter survey?

Under this scenario, OFM calculated counts as the weighted average of the previous year's spring survey and the current year's fall survey. The spring is weighted at 3/8 and the fall is weighted as 5/8. This means that current year counts matter more (see table below).

Here's the most important thing to take away from this table:

- The change to total allocation would never have been more than 0.3%.

Drop winter count					
	SY2015-16	**SY2016-17**	**SY2017-18**	**SY2018-19**	**SY2019-20**
Allocation percent change	-0.06%	-0.06%	0.30%	-0.11%	0.09%
Impact to individual school districts					
Percent no change	42.86%	42.03%	44.13%	42.75%	41.30%
> 1% change	33.93%	28.99%	29.54%	34.78%	30.07%
1–3% change	18.57%	26.45%	18.86%	19.20%	23.55%
3–5% change	2.50%	2.54%	5.34%	1.81%	2.90%
> 5% change	2.14%	0.00%	2.14%	1.45%	2.17%

What happens if we estimate the winter survey?

In this scenario, OSPI would substitute an estimated current winter value for the actual winter count. **That means we can still use the previous year's spring and current year fall counts, and keep the same weighted average formula.** We based the estimate of the current year count on a ratio of the previous year's winter enrollment count. We applied this ratio against the current year enrollment to get the final winter count (see table below).

Here's the most important thing to take away from this table:

- The change to total allocation would never have been more than 0.51%.

Estimate winter count					
	SY2015-16	**SY2016-17**	**SY2017-18**	**SY2018-19**	**SY2019-20**
Allocation percent change	-0.30%	0.51%	-0.43%	-0.25%	0.00%
Impact to individual school districts					
Percent no change	41.07%	41.30%	42.35%	40.58%	39.86%
> 1% change	22.50%	25.00%	22.06%	26.09%	24.64%
1–3% change	23.57%	21.01%	22.78%	22.10%	24.28%
3–5% change	7.50%	9.78%	5.34%	6.88%	7.97%
> 5% change	5.36%	2.90%	7.47%	4.35%	3.26%

Our conclusion on changing the timeline

The final allocation will change little if we use either of these alternate methods. This shows that we can run the model before the winter count is available. This has extremely low impact at the school district level, too. Most school districts won't be affected or will experience a less-than 1% allocation change. Only a handful would experience a 5% or more allocation change.

Extra data about ridership and expenditure trends

This analysis highlights the extremely wide range of school district size and ridership experience over the study period. It is likely that this range of disparate characteristics made the cost of student transportation funding at the district level a difficult exercise.

It covers the 283 school districts that we included in the last five STARS allocation efforts. It also combines the Basic Program and Special Program rider counts.

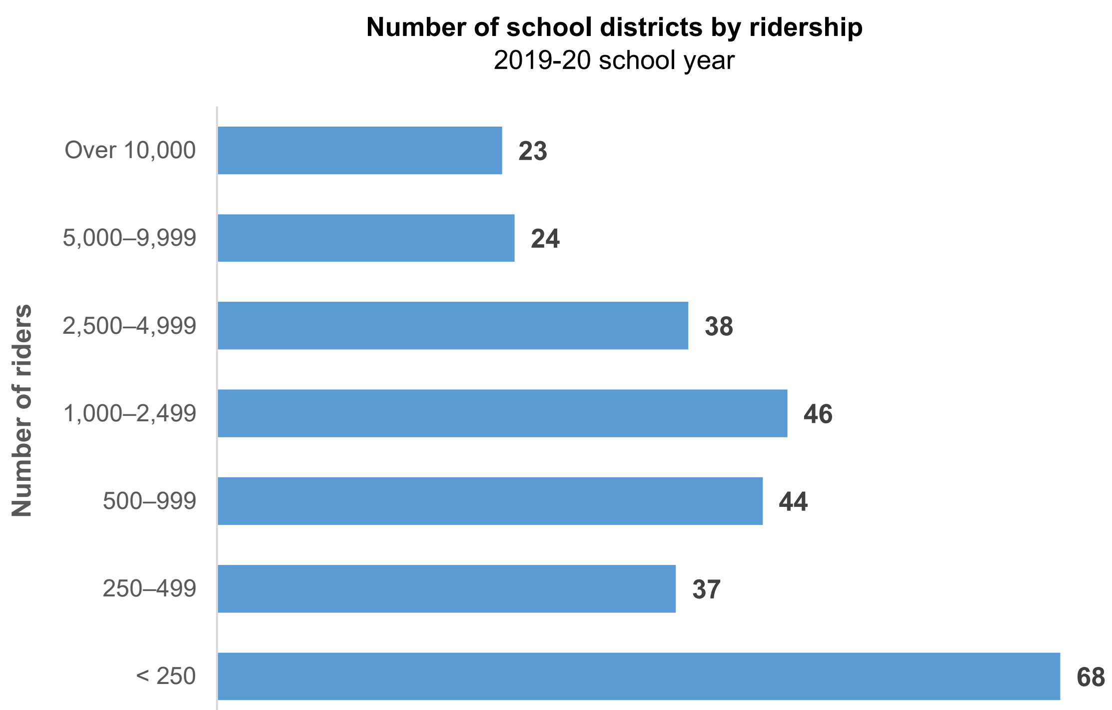
Ridership trends

Ridership counts

As of the 2019-20 school year, Washington school districts served an estimated 794,012 students with transportation services. School districts vary dramatically in size, as does their ridership. For instance, the Seattle School District provides transportation services to over 23,000 students, while the Star School District serves just 17. We illustrated the number of school districts by ridership counts in the graph below.

Here's the most important thing to take away from this graph:

- Districts with fewer than 250 riders are the biggest category.



These ridership numbers, compared to the total K-12 fall headcount enrollment (1,115,946 for the 2019-20 school year), results in a very consistent ratio of students that had transportation services. Just over 71% of all enrolled students across all five school years in this examination got transportation services. No doubt this ratio varies across school districts based on school proximity to student residences, with the likelihood that rural districts have higher ratios of riders within their total enrollment.

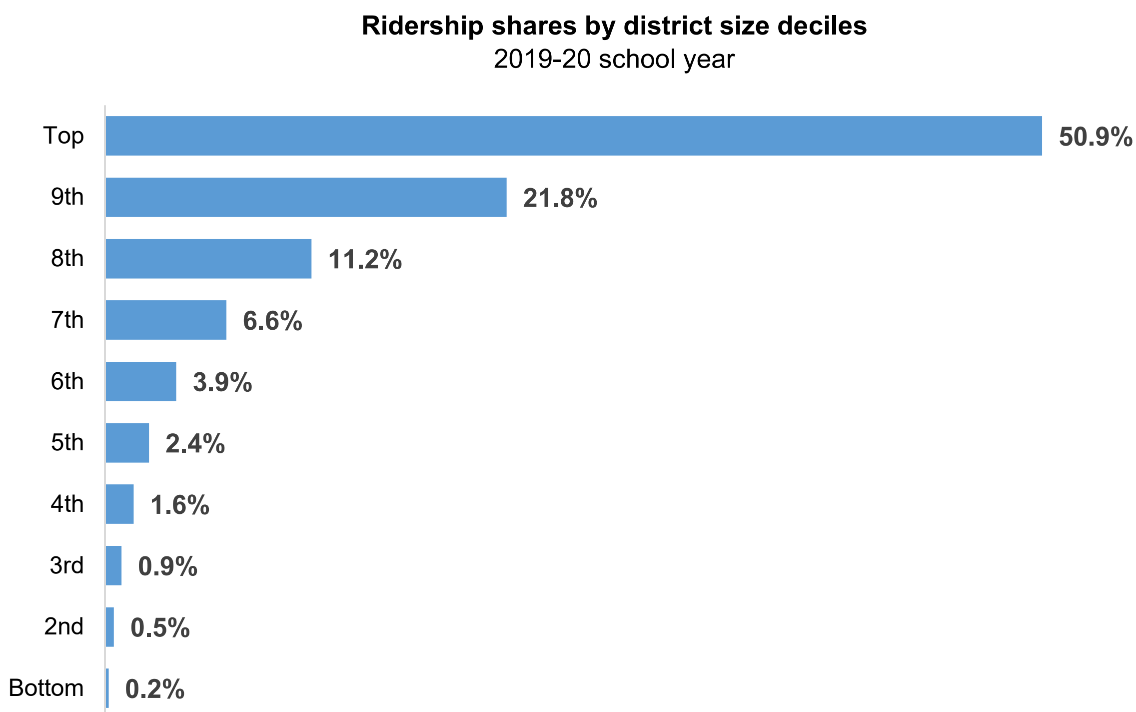
Ridership shares

When we examine the above data, it is apparent there are a relative handful of very large districts, and many smaller ones. Because of the disparate size differences, a small share of school districts represent a significant share of enrollment and student ridership.

As we illustrate in the figure below, during the 2019-20 school year, the top 10% of the total number of districts provided transportation services for over half the total ridership statewide. At the other end of the spectrum, the bottom 10% of the total number of districts provided transportation services to 0.2% of the total ridership statewide.

Here's the most important thing to take away from this graph:

- Most riders are concentrated in just 10% of districts.

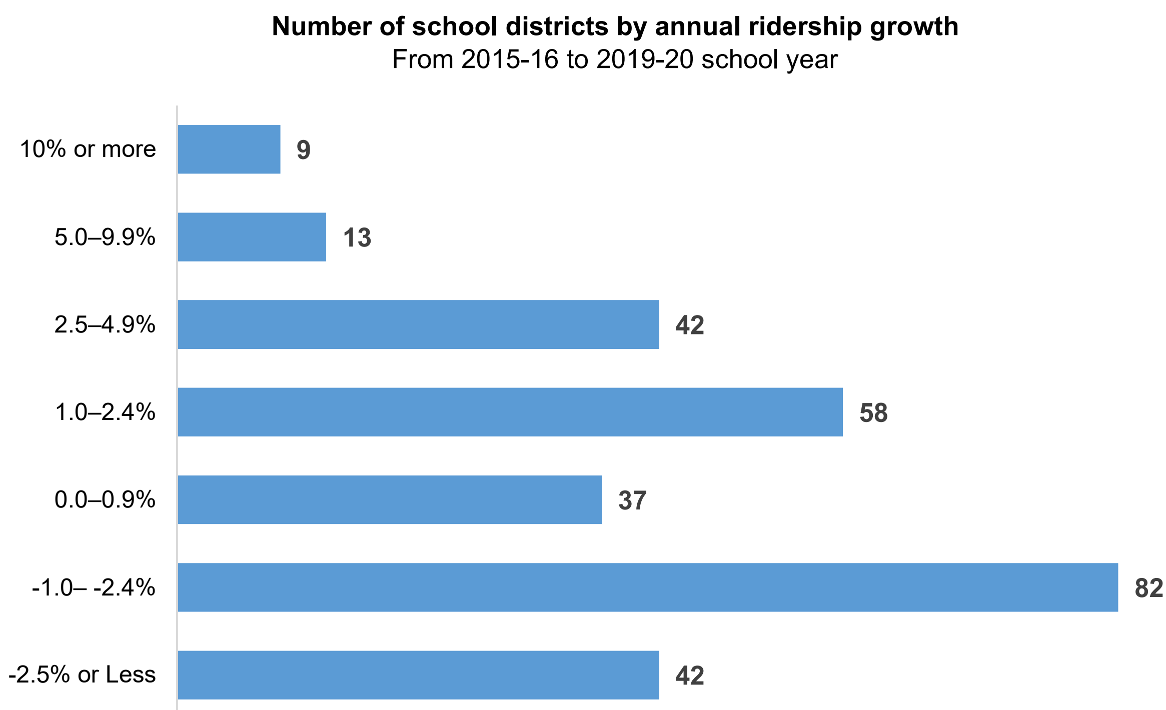


Ridership growth

Over the past four school years (ending with the 2019-20 school year), the number of K-12 students who need transportation services has grown 0.6% each year to reach 794,012. Across the 283 school districts in this analysis, some districts' ridership growth was much quicker, while others experienced negligible growth and some even declined. We portray ridership growth rates by the number of school districts in the following graph.

Here's the most important thing to take away from this table:

- Many districts had a decline in ridership.



About 124 school districts experienced declining ridership over the examination period. Seattle Public Schools, the state's largest school district, experienced a ridership decline of 2.8% per year over the past four years. The burgeoning cost of housing in Seattle was one of several factors that contributed to this decline. This likely incentivized many households with children to relocate to lower cost areas.

Expenditure trends

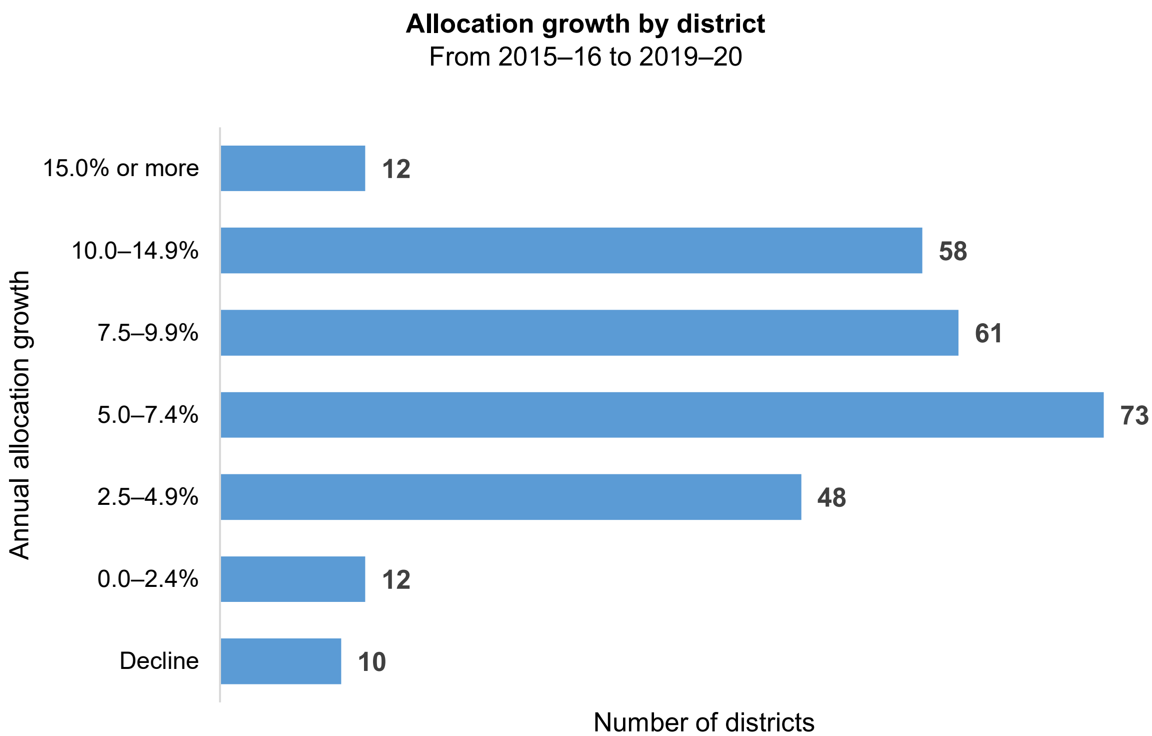
Allocation growth by district

The total actual allocation amounts (calculated by the STARS method) grew from $416.7 million in the 2015-16 school year to $575.9 million in the 2019-20 school year. This is an annual growth rate of 8.4% per year. While the school districts vary widely in ridership growth, they also vary widely in cost growth.

Examples include a high of 38.6% per year in the Wishram School District to a low -9.7% per year in the Keller School District. However, these are very small districts, so their unusual rates are more illustrative than representative. Below is a graph that shows annual cost growth.

Here's the most important thing to take away from this graph:

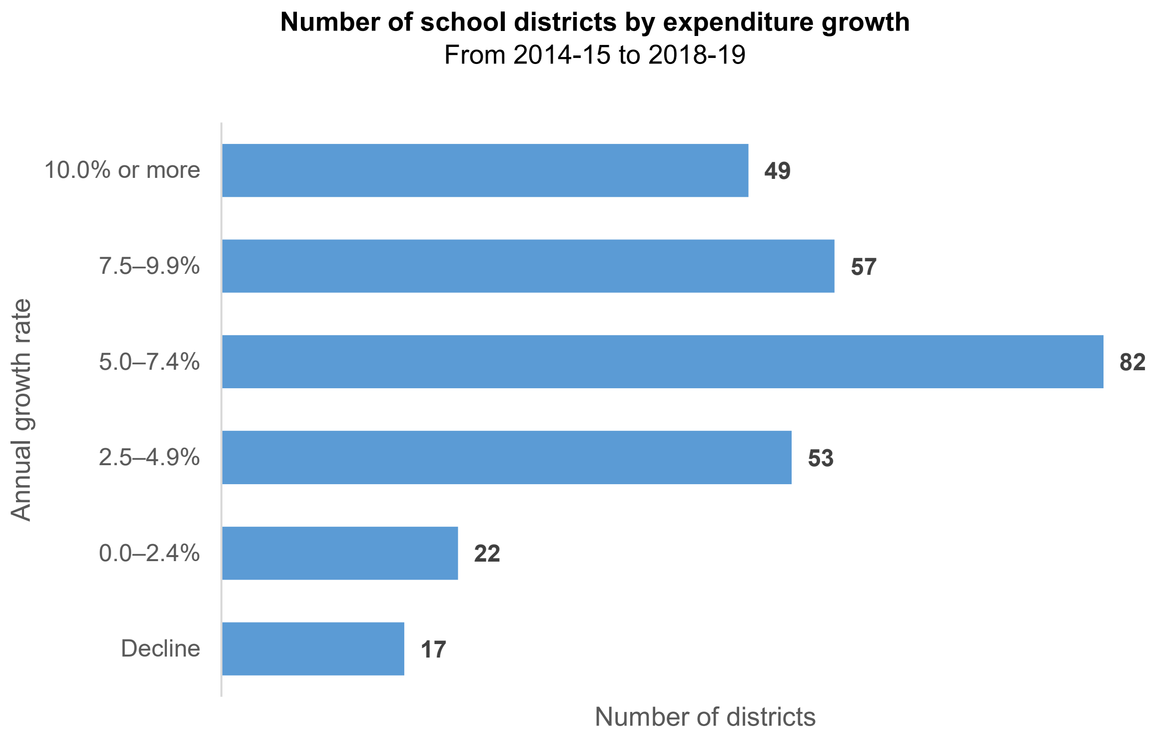
- Annual allocation grows most of the time.



The median growth rate, where half the districts had higher growth rates and half had lower growth rates, was 7.0% per year. The *interquartile* range of growth, (which means you remove the top 25% and lower 25%, and then measure the remaining 50%) ranged from 4.6% per year to 9.9% per year.

Growth in transportation expenditures

Over the examination period, total transportation expenditures grew from $428.3 million in the 2014-15 school year to $582.0 million in the 2018-19 school year. This is a compound growth rate of 8% per year. Like our ridership analysis, a school district's expenditure growth by varies widely, from an increase of 35.6% per year in the Wishram School District, to a decline of 12.3% per year in the Keller School District. The rule of small numbers tells us that these outsized gains and losses – while they help to illustrate the broad range of experience – carry little weight in the overall trends. We show those trends in the following graph.

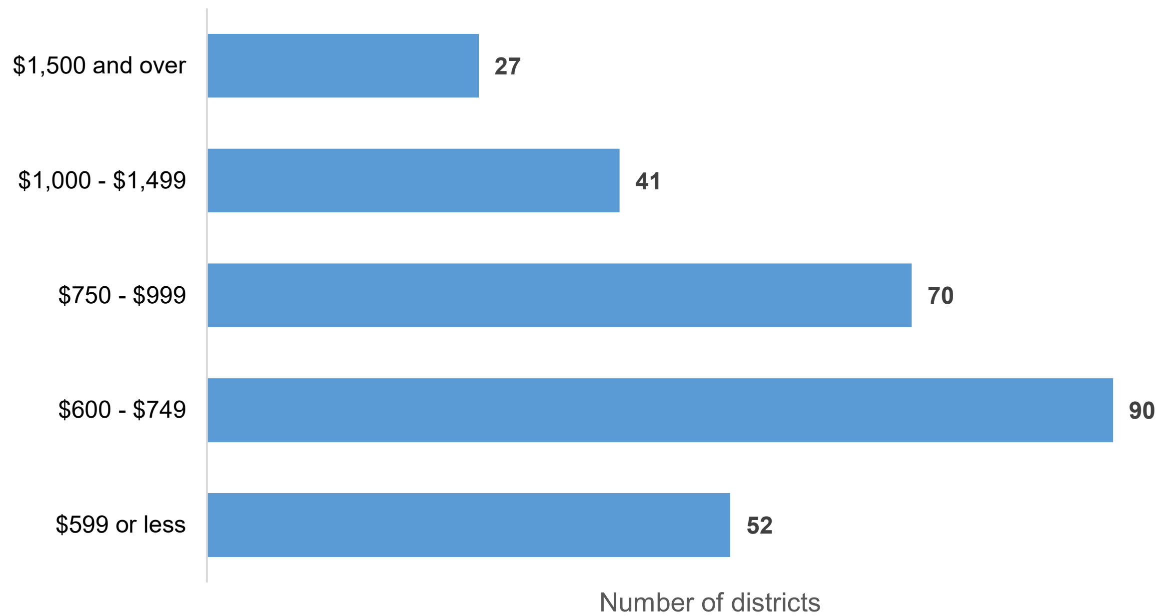


The median expenditure growth rate, where half the districts had higher growth rates and half had lower growth rates, was 6.3% per year. The interquartile range of expenditure growth rates, which captures the middle 50% of districts by excluding the highest and lowest quarters, ranged from 4.1% per year to 9.1% per year. Again, the growth rate at the upper bounds of the interquartile range is more than double the lower bounds.

Expenditures per rider

Though we don't use expenditures per rider as part of the allocation method, the cost per rider is yet another measure to contrast the experiences among the school districts. Comparing the 2019-20 corrected prior year expenditures with the 2018-19 ridership, the average cost per student across all districts was $735.04 per school year, or $4.08 per school day (180 days per school year). Expenditures per rider ranged from a low of $306.62 in the Evaline School District, to $7,582.54 in the Star School District. The following graph shows the number of school districts by expenditures per rider.

Number of districts by expenditures per rider
2018-19 school year



The median cost per rider in the 2018-19 school year was $746.72. This number is not significantly different from the mean average we referenced above. The interquartile range of costs per rider extended from $631.78 to $967.73. Though still wide, this was a more compressed range than in earlier sections.

It's reasonable for us to assume that smaller, rural districts would not have access to the economies of scale (this means that bigger organizations can do things more cheaply than smaller ones) of larger districts, and thus trend toward a higher cost per student. However, we see mixed evidence to support that assumption. While Star and Lacrosse School Districts are small and rural, and have the highest per rider expenditures in the state, Evaline and Crescent School Districts are also small and rural but have some of the lowest per rider expenditures. On the other end of the size spectrum is Seattle Public Schools with per rider expenditures in the top decile, while Evergreen, the second largest school district, has expenditures per rider less than half that of Seattle. In addition, North Thurston Public Schools, in the top decile of ridership, had rider expenditures in the lowest decile.

Why pupil transportation costs fall short

The total amount of money that we put into pupil transportation **is less than what we need**. The top reasons we fall short are:

- Timing of cost factors
- Regression skew
- The final numbers from the STARS model (this is when we use the "lesser than provision" term we mentioned on Page 4)

Timing of cost factors

We updated three of the STARS factors that calculate the expected allocation each year. They are basic ridership, special ridership and number of destinations. We don't fully know the growth or decline in ridership or number of destinations from one year to the next because, in part, we have to base them on the previous year's figures. We also don't use the complete annual count because of timing constraints. Instead, we use an average count of the current fall, current winter and previous spring (as shown below). Changing this timeline could help us get more accurate numbers.

School year 2018-2019 count					
2017-18			2018-19		
Fall	Winter	Spring	Fall	Winter	Spring

Fiscal impact

If we calculated the ridership counts and destination counts using the *complete* year, there would have been an allocation increase during the last four years. The table below shows the counts that would make up the weighted average.

School year 2018-2019 count					
2017-18			2018-19		
Fall	Winter	Spring	Fall	Winter	Spring

The dollar impact is relatively low though, not peaking at more than $1.8 million and as low as about $400,000 for any given year. This means that this recommendation has a small effect.

Additional allocation Use complete annual count			
2015-2016	**2016-2017**	**2017-2018**	**2018-2019**
$587,917	$1,494,896	$1,767,450	$383,175

What we recommend

If we change this timeline to follow the *complete* school year, then the fiscal impact is quite small. But the priority to change to the allocation model is low, and budget writers need a pupil transportation allocation well before the same year spring count is ready, so including it is impossible. However, we could calculate counts on other ways that may come closer than the current method. Ideas include:

- Estimate the same year spring count.
- Use the fall and winter count without the spring count.

Regression skew

There is a consistent skew in the regression result. While this skew is a natural occurrence, it gives us a number that is *below* actual expenditure. It tells us that this model is going to come up with a lower number than we need.

This matters because the Legislature wants to make sure it's allocating money fairly and effectively but it will not be able to do that if the number we use is too low to begin with.

The reader needs background knowledge of OLS regression to completely understand this issue. The error (residual) of the regression is the difference of predicted value and observed value, and the total of residuals equals zero. In our case, this is the difference of the predicted natural log of expenditure and the natural log of actual expenditure. But we use the STARS model to calculate allocation, *not the natural log* of allocation, so we care about the difference of unlogged transformed variables.

This is where things get interesting. The sum of unlogged predicted expenditure compared to the sum of unlogged observed expenditure can show us whether the model is likely to over or underestimate expenditure. That's because the STARS allocation amount is the unlogged predicted expenditure. In our case, the unlogged predicted amount is much lower than the unlogged observed amount. **This suggests that the STARS amount model underestimated during those years.** This is not an indictment of the regression model, but is just a natural consequence of using a natural log transformed model to predict values.

Fiscal impact

For the 2019-20 school year, the predicted expenditure was about $20 million less than actual expenditures. This is from the regression, not the STARS model. The difference is quite large during the year, with the exception of the 2015-16 school year. There is no reason to believe this should be negative every year. It could very well skew in the other direction.

Why we recommend adding more money and how we can quantify this amount

This phenomenon will carry through to the STARS model and under allocate to school districts. Using the 2017-18 school year as an example, the STARS amount for that year could be about $12.4 million below costs because of this issue. A solution is to add $12.4 million to the (expected) STARS allocation and equitably distribute it among school districts.

The amount in the table below shows the difference between the actual expenditure and the predicted amount from the regression model. The regression predicted amount has been much less than the actual expenditure, so the STARS amount for those years will be off by a similar amount.

Here's the most important takeaway from this table:

- The regression predicted an expenditure amount that was less than actual expenditure for each of the last five years.

Regression predicted expenditure compared with actual expenditure					
	2015-16	**2016-17**	**2017-18**	**2018-19**	**2019-20**
Prior Year Expenditure	428,640,787	447,305,356	479,665,865	530,299,889	581,951,884
Predicted Prior Year Expenditure	428,636,744	438,585,246	467,242,254	516,431,637	561,973,948
Difference	(4,043)	(8,720,110)	(12,423,611)	(13,868,252)	(19,977,935)

How the "lesser than provision" outweighs the positives of the STARS adjustments

One of the components of the allocation model is that no school district can receive an amount above their corrected prior year expenditures (this is last year's amount plus the federal rate indirect). This acts as a cap on how much a school district may receive - there is no offsetting floor. Some districts qualify for more money in the form of adjustments, but the influence of these adjustments is smaller than the cap.

If we agree that the STARS model does a good job of calculating the average impact, then the total STARS allocation should be a reasonable guess of costs. But we allocate a lesser amount because the allocation cap effect is larger than the floor.

In a way, when we artificially limit the upper bounds of the allocation, we end up with a kind of surplus, or residual amount. We could use that surplus/residual amount to address school districts that receive allocations well below their previous year's expenditures.

Fiscal impact

For the 2019-20 school year, adjustments added $8.2 million to the STARS allocation. But the "lesser than provision" reduced it by $33.3 million. That means the STARS allocation was cut by $25.1 million because of the additions and subtractions. For the same school year, the STARS estimate was $575 million, but was reduced to $550 million. This effect played out the same during the last five years.

Here's the most important takeaway from this table:

- The net effect of:
  - school districts getting an adjustment *and*
  - districts being capped at last year's cost, plus federal rate indirect,
    was to lower the STARS model amount by $16.4 million to $25.1 million.

Impact of additions and subtractions to STARS (expected) allocation					
	2015-16	**2016-17**	**2017-18**	**2018-19**	**2019-20**
Adjustments Added	6,461,362	7,080,836	6,443,332	7,829,226	8,202,939
Lesser Than Provision Subtracted	(28,059,673)	(23,449,218)	(25,589,521)	(28,597,182)	(33,286,551)
Net of Additions and Subtractions	(21,598,311)	(16,368,382)	(19,146,190)	(20,767,956)	(25,083,612)

What we recommend

OSPI needs to increase the total allocation by the difference so that it equals the STARS amount. Some districts could benefit if OSPI has a way to direct these funds to districts with an allocation below cost.

Fiscal impact of all recommendations

We illustrate the dollar impact of these three recommendations in the next table and compare them against the actual allocation.

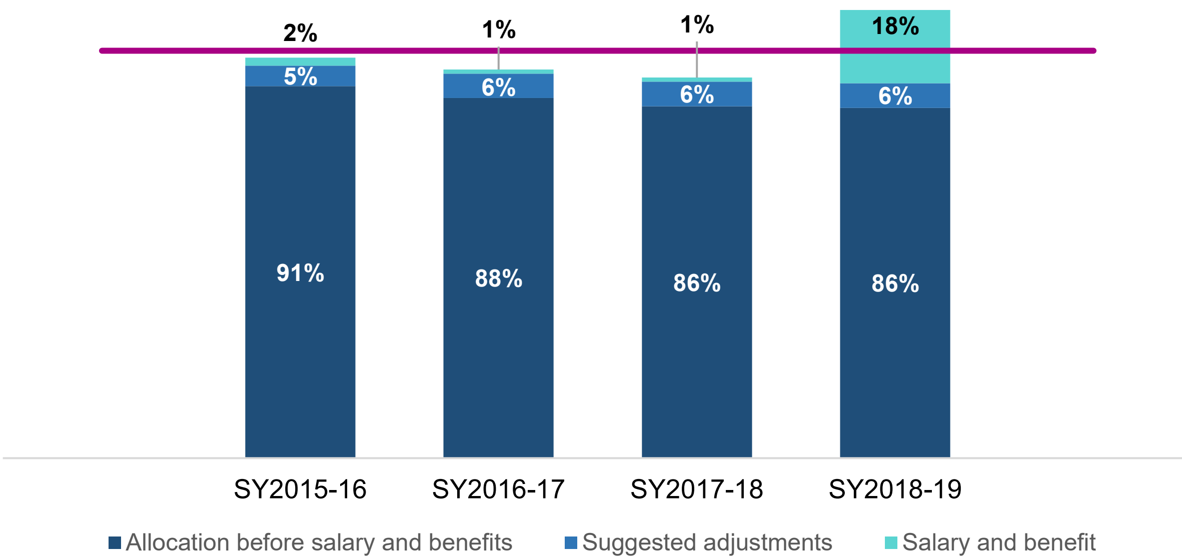
Here's the most important takeaway from this table:

- This hypothetical table illustrates what the final allocation would have been with extra steps. The table shows that these additions pull the allocation closer in line with costs.

Allocation with recommendations				
	2015-16	**2016-17**	**2017-18**	**2018-19**
Use full STARS allocation	21,598,311	16,368,382	19,146,190	20,767,956
Complete year count	587,917	1,497,309	1,767,450	383,185
Regression skew fix	4,061	8,753,824	12,689,477	14,003,748
Recommended adjustments	22,190,289	26,619,515	33,603,116	35,154,889
Lesser of adjusted allocation or adjusted prior year expenditures	409,010,095	423,912,534	458,095,072	500,709,333
Recommended adjustments	22,190,289	26,619,515	33,603,116	35,154,889
Allocation with recommended adjustments	431,200,385	450,532,049	491,698,188	535,864,222
Legislative salary and benefit	8,361,267	4,751,986	7,792,629	103,255,459
Final allocation with recommended adjustments	439,561,651	455,284,035	499,490,817	639,119,681

Compare				
	2015-16	**2016-17**	**2017-18**	**2018-19**
Actual expenditure	448,213,083	479,396,313	530,299,889	581,951,884
Actual allocation	417,371,362	428,664,521	465,887,701	603,964,792
Proposed allocation	439,561,651	455,284,035	499,490,817	639,119,681

How suggested recommendations would impact the allocations



The background of why we use STARS

How our old system collected transportation data

Before we used STARS, we used the Legacy Funding System. On average, the system funded about 62% of a school district's cost.

Here are some issues and concerns with the old system that OSPI used:

- The system counted students only once per year student count. This means we didn't capture ridership trends throughout the school year.
- The system only counted students who rode the bus in the morning before school and didn't compare this number to afternoon ride trends. Tracking both sets of data could have changed how much money each district received.
- The system distributed more funds to districts if the districts had low average busloads. While OSPI hoped this would help small rural districts, larger districts learned it was a way to increase their funding by using more buses to do the work.
- It was extremely time consuming and cumbersome for district staff to collect and process this data.

How the current system collects transportation data

The state started using the new funding model, STARS, in 2010. OSPI fully implemented it during the 2015-16 school year.

We built STARS to better address how we collected and processed data. The state hired a contractor to develop the Students Transportation Allocation Reporting System after the Legislature passed a 2007 law to create a new system. The Legislature wanted the state to center the new funding on how much it actually costs to provide transportation to-and-from school. There was a specific focus on resource efficiency and funding predictability.

Here are some benefits from the new system:

- Simplified, web-based, reporting process.
- More consistent student counts at three times per year. This includes morning and afternoon.
- Includes last year's expenditures so we can:
    - Tie allocation to actual operation costs.
    - Prevent districts from being funded more than their costs from the prior year.
- Includes alternate funding methodology for unique operations.

On average, today's system funds 94% of a district's transportation needs, which is a great improvement from the 61% in the old model.

While STARS increased how much districts received when compared to their costs, we know it has concerns or weaknesses. OSPI and school districts have expressed the following concerns:

- The system can't forecast allocation with credible accuracy.
- The system calculates the final allocation in February. This timeline can create budgeting issues for school districts and is almost too short of a timeframe to provide information to the Legislature.
- Districts typically seem to be one year behind in recouping transportation costs because of how the system designates funds. This means, a district has to use local money if it needs more transportation funds.
- Student transportation (mandated by the McKinney Vento Act) does not receive the full benefit of the STARS formula. That's because much of the transportation comes from nondistrict vehicles. This transportation can cause a spike in current year expenditures and result in spending local money.
- The efficiency rating in the STARS model isn't that valuable because it provides a false sense of understanding. Districts that operate more efficiently actually *risk getting less funding* in the STARS model. In fact, when districts try to improve their efficiency rating by reducing their costs, it could potentially decrease how much district and districts get statewide. That's why we don't recommend using this efficiency rating.

Appendix A

Classical Assumptions of Ordinary Least Squares Regression

Assumption 1: The dependent variable is linearly related to the coefficients.

The regression equation is a linear regression, even though some variables are log transformed, the relationship between the dependent variable and the coefficents is linear. The regression equation is given below.

$$\ln \boldsymbol{PriorYearExpenditure} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} \ln \boldsymbol{BasicRider} + \boldsymbol{\beta_2 lnSpecialRider} + \boldsymbol{\beta_3} \ln \boldsymbol{Land} + \boldsymbol{\beta_4 NumberOfDestinations} + \boldsymbol{\beta_5 AverageDistance} + \boldsymbol{\beta_6 NonHighNo} + \boldsymbol{\varepsilon}$$

Assumption 2: None of the independent variables have a perfect linear relationship with any of the other independent variables.

A correlation matrix displays the correlation coefficients between variables. We use it to test for collinearity among independent variables. A correlation coefficient of one or negative one indicates perfect positive or negative correlation, but a figure close to one or negative can indicate problems, too. The next table shows the person correlation coefficients of the independent variables used in the regression model. As expected, there isn't perfect linear collinearity between any independent variables. But at 0.92, there is very strong collinearity between the natural log of regular ridership and the natural log of special ridership.

Pearson Correlation Coefficients, N = 280 Prob > |r| under H0: Rho=0					
	ReLn	SeLn	LandLn	AverageDistance	NumberOfDestinations
ReLn	1.00000	0.92480 <.0001	-0.06449 0.2822	-0.46586 <.0001	0.68669 <.0001
SeLn	0.92480 <.0001	1.00000	-0.08409 0.1605	-0.47072 <.0001	0.67802 <.0001
LandLn	-0.06449 0.2822	-0.08409 0.1605	1.00000	0.36553 <.0001	-0.18706 0.0017
AverageDistance AverageDistance	-0.46586 <.0001	-0.47072 <.0001	0.36553 <.0001	1.00000	-0.29768 <.0001
NumberOfDestinations NumberOfDestinations	0.68669 <.0001	0.67802 <.0001	-0.18706 0.0017	-0.29768 <.0001	1.00000

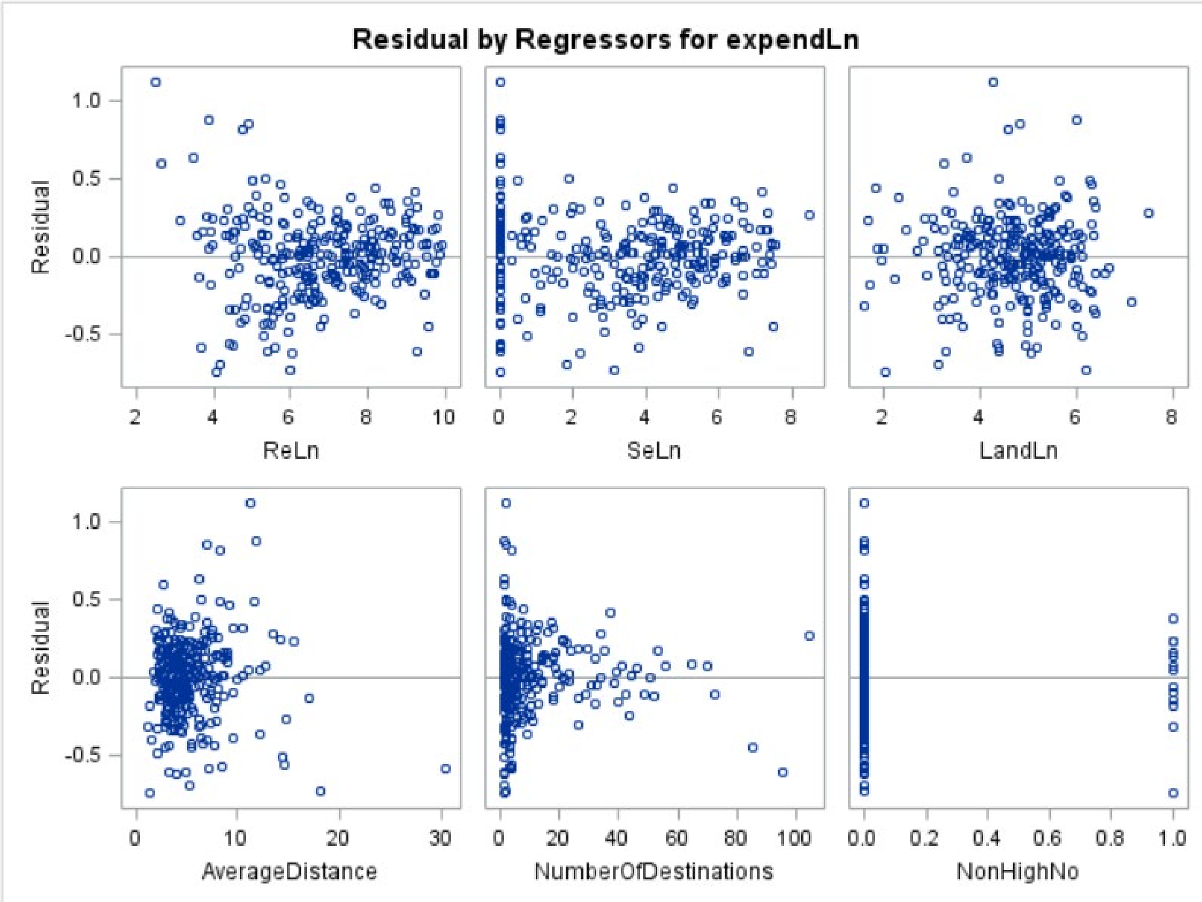
We use a test statistic (called the variance inflation factor) to measure multicollinearity. Multcollinearity happens when independent variables are highly correlated in a linear way. None of the variance inflation factors are above 10, which we consider a threshold for high multicollinearity. So, this test supports evidence that we can leave basic and special ridership in the model.

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	Intercept	1	7.92704	0.15521	51.07	<.0001	0
ReLn		1	0.68497	0.02692	25.44	<.0001	7.57610
SeLn		1	0.10280	0.01820	5.65	<.0001	7.19718
LandLn		1	0.06025	0.01786	3.37	0.0009	1.34843
AverageDistance	AverageDistance	1	0.04130	0.00624	6.62	<.0001	1.54764
NumberOfDestinations	NumberOfDestinations	1	0.01251	0.00148	8.47	<.0001	2.05015
NonHighNo	NonHighNo	1	-0.33045	0.08100	-4.08	<.0001	1.21889

Sometimes high correlation among two independent variables causes one of the variables to lack statistical significance or the coefficient sign (positive or negative) to swap. In this case, both variables in question (basic and special ridership) are statistically significant and their coefficients have a positive sign, as expected. So it doesn't appear the high correlation is having a negative effect.

Assumption 3: No independent variables are correlated with the error term.

The next six scatter plots illustrate the relationship between the error term (residual) and the independent variables. Correlation between the two would show up in these plots as a pattern, such as linear, log or squared. There are no obvious patterns in any of these graphs, so this condition is met.



Assumption 4: The error term observations are not correlated with each other.

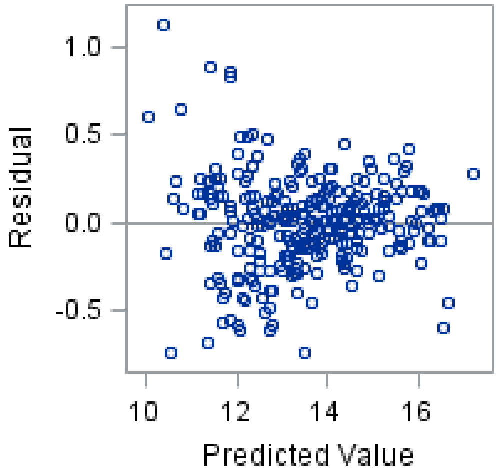
Nonrandom error terms cause problem with OLS, but is usually an issue with time series regression. Since the OLS regression used is a cross section regression model, this assumption is satisfied.

Assumption 5: The mean of the error term is zero.

The error term is the value not explained by the model, in other words, the difference between the actual value of the dependent variable and the predicted value. The presence of a constant variable (the y-intercept) forces the mean value to zero. The regression equation has a constant (β_0), so this condition is met.

Assumption 6: Error term has a constant variance.

Error terms should not fit a pattern, such as having wider variance for large school districts. The scatter plot of the residuals and predicted value doesn't show a pattern, so the variance of the error term looks constant.

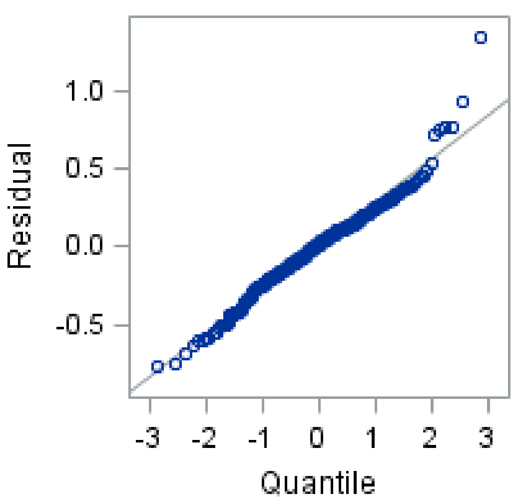


White's test is a statistical test used to detect errors that have constant variance. The results of White's test is the null hypothesis - that the variance of the errors is equal - is rejected, providing more evidence this assumption is met.

Heteroscedasticity Test					
Equation	**Test**	**Statistic**	**DF**	**Pr > ChiSq**	**Variables**
expendLn	White's Test	108.0	26	<.0001	Cross of all vars

Assumption 7: Error term is normally distributed.

For the most part, the residual plots follow the reference line, with the exception of some outliers in the upper quartile. If the distribution is normal, plots should closely follow this line. They do with just a few exceptions on the upper end, so this assumption is deemed met.



Appendix B

Correlation Statistics and Correlation Plots

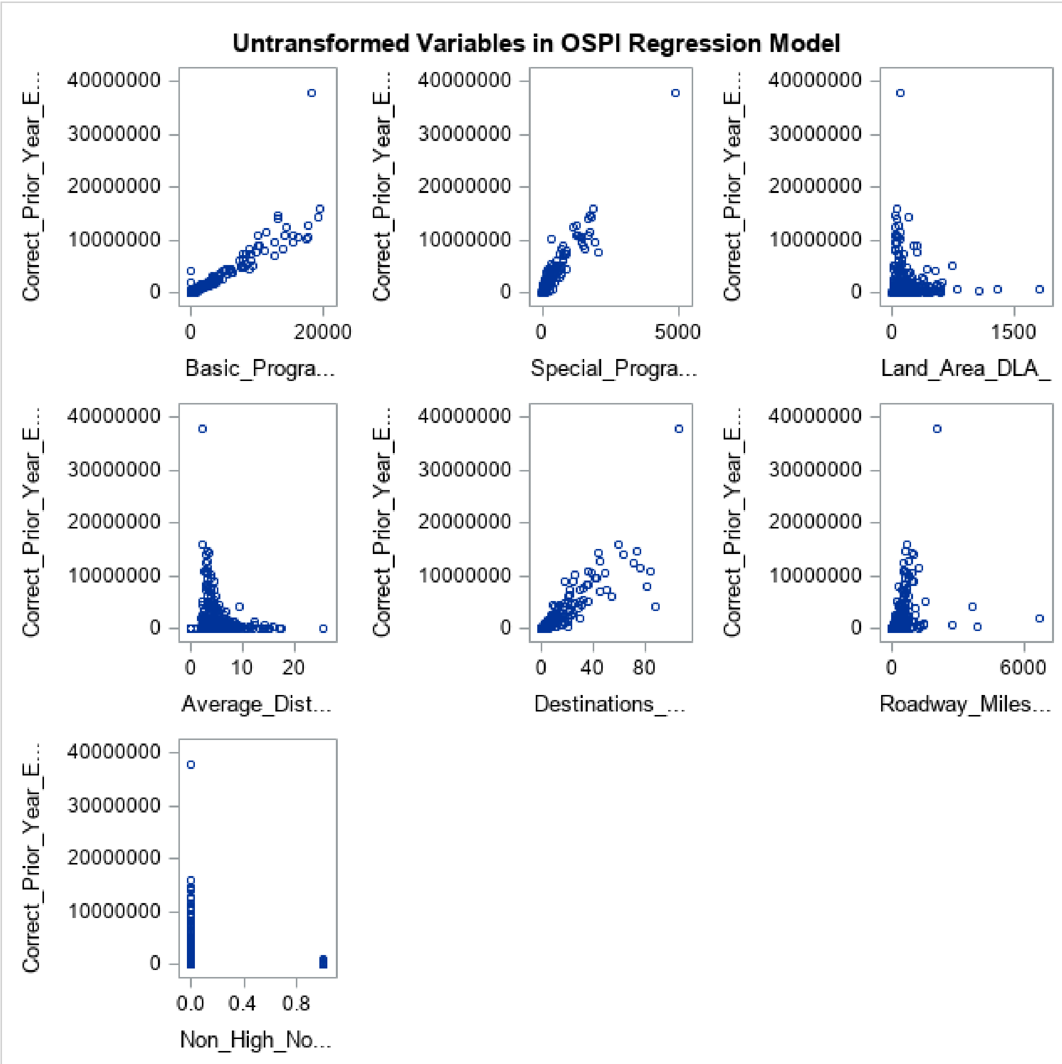
Table 1: Correlation Statistics of Untransformed Regression Variables

Correlation Statistics for Prior Year Expenditures by 7 OSPI Regression (Predictor) Variables		
****These Variables are in levels form and untransformed****		
	r	**p-value**
Basic_Program_CBPC_ (basic riders)	0.91	<.0001
Special_Program_CSPC_ (special riders)	0.96	<.0001
Land_Area_DLA_	-0.07	<.2008
Average_Distance_AAD_	-0.15	<.0009
Destinations_AND_	0.87	<.0001
Roadway_Miles_TRM_ (Roadway miles)	0.28	<.0001
Non_High_No_NHN_ (No high school)	-0.12	<.0343
Note:		
A p-value of ≤ 0.05 is statistically significant.		
A p-value of > 0.05 is not statistically significant and indicates strong evidence of no statistical correlation between two variables		
The final OSPI regression model contains 6 variables. The Roadway Miles is dropped from the regression		

Table 1 displays correlation coefficients and p-values associated with prior year's expenditures relative to the initial seven explanatory variables in the OSPI regression model. The correlation statistics indicate that of the seven independent variables, four are strongly significant at <.0001 percent and the fifth and sixth variables (average distance, and high school indicator variable) are significant at <.0009 and <.0343 respectively. The final regression model contains six variables after dropping the roadway miles variable. It should be noted that these variables are in their levels form and untransformed to their log counterparts.

Correlation plots of the untransformed dependent variable against the initial seven untransformed explanatory variables in the OSPI regression model are displayed in Figure 1. Scatter plots are the first step to discerning relationships; if there is a relationship it will be evident in the picture graph.

Figure 1: Correlation Plots of Untransformed Variables in the OSPI Regression Model (note: roadway miles were dropped out of the final regression model)



The scatter plots above indicate strong linear statistical relationships between prior year's expenditures for basic riders, special riders, average distance, number of destinations, roadway miles, and the high school indicator variable. The land area variable is insignificant with a p-value of <.2008. Further examination of the scatter plots all indicate existence of outlier observations which most likely will exert influence on parameters of the model that represent allocation coefficients or weights.

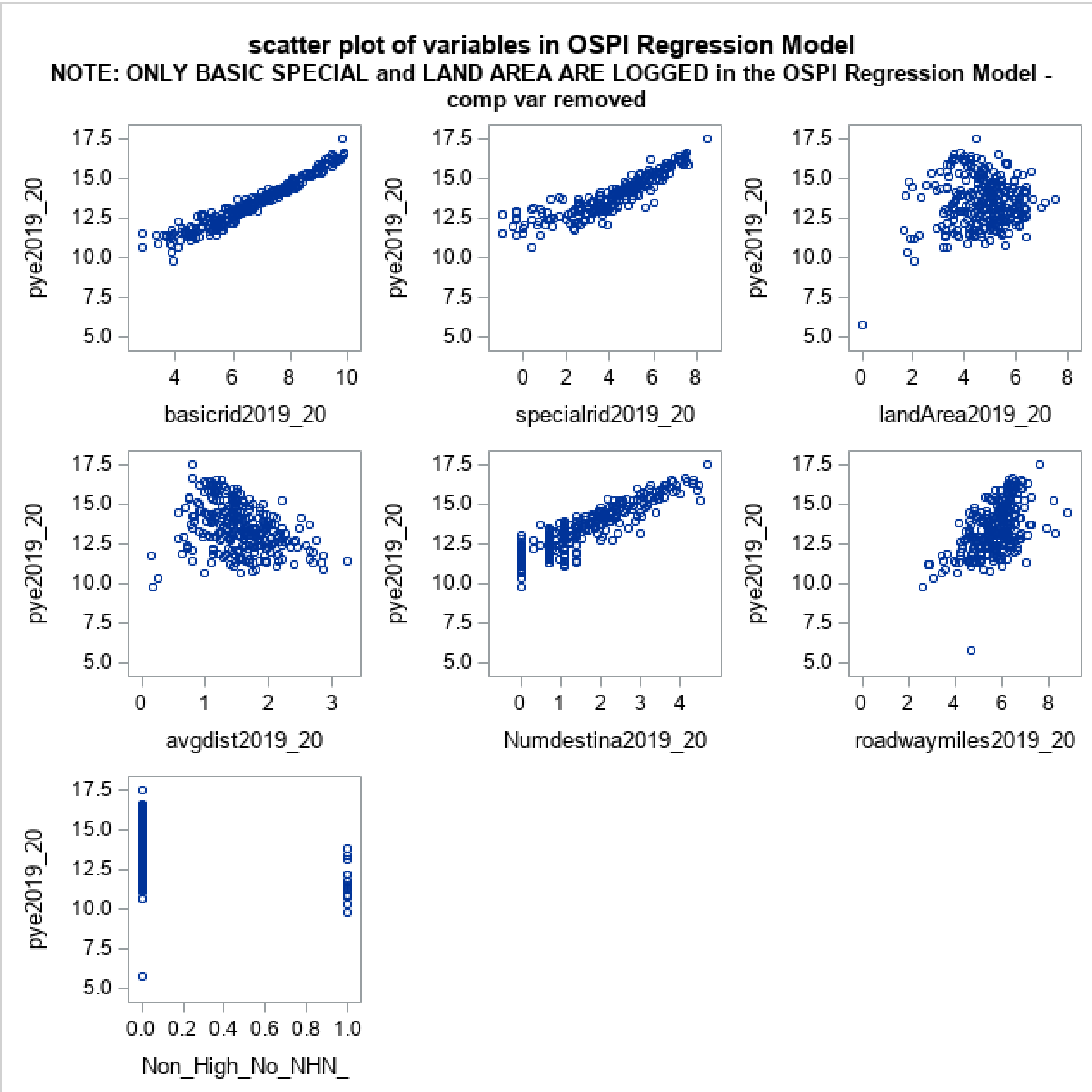
Table 2: Correlation Statistics of Log-transformed Regression Variables

Correlation Statistics of Prior Year Expenditures by 7 OSPI Predictor Variables		
*****Log Transformed Variables*****		
	r	**p-value**
Basic_Program_CBPC_ (basic riders)	0.97	<.0001
Special_Program_CSPC_ (special riders)	0.89	<.0001
Land_Area_DLA_	0.09	<.1438
Average_Distance_AAD_	-0.33	<.0001
Destinations_AND_	0.90	<.0001
Roadway_Miles_TRM_ (Roadway miles)	0.54	<.0001
Non_High_No_NHN_ (No high school)	-0.29	<.0001
Note:		
A p-value of ≤ 0.05 is statistically significant.		
A p-value of > 0.05 is not statistically significant and indicates		
The final OSPI regression model contains 6 variables. The Roadway Miles is dropped from the regression		

Table 2 above displays correlation coefficients and p-values associated with prior year's expenditures relative to the initial seven explanatory variables in the OSPI regression model. The main difference between the correlation values in Table 1 and Table 2 is that in Table 2 we are now looking at log transformed variables which is the form that the OSPI regression equation is formulated. The land area variable continues to be insignificant in both transformed (Table 2) and nontransformed (Table 1) iterations. The correlation statistics indicate that all but one (land area) of the seven independent variables are strongly significant at <.0001 percent. The land area variable is insignificant at <.1438.

Figure 2: Correlation Coefficients and Scatterplots for Log-transformed OSPI Regression Model Data

Note: All variables in the scatter plots are log transformed



Office of Financial Management
P.O. Box 43124
Olympia, WA 98504-3124
360-902-0599