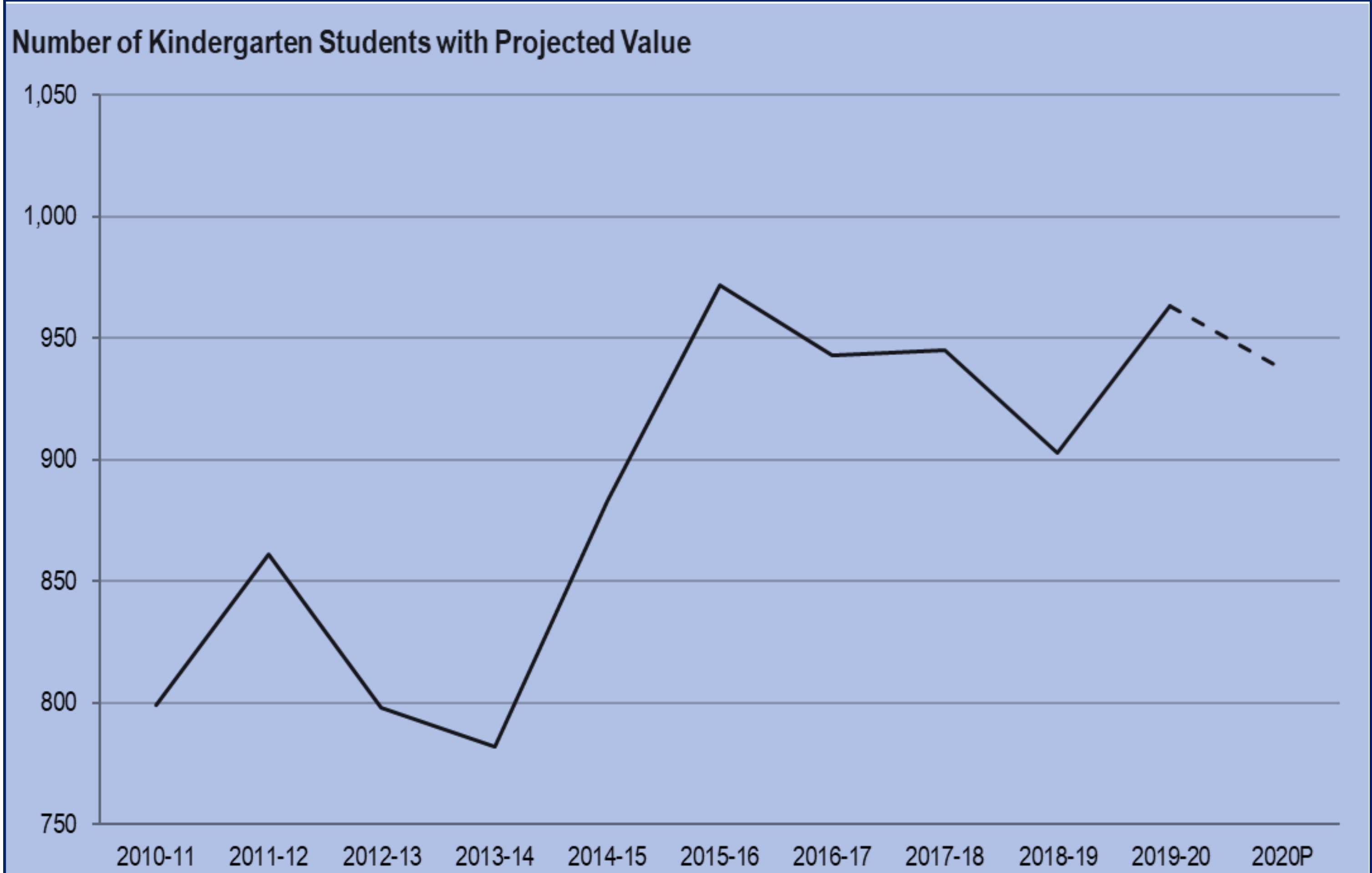
Enrollment Projection Methods

White Paper

June 23, 2020



PLANware

15850 Concord Circle, Suite B
Morgan Hill, CA 95037
408-776-7646 | www.eddata.com

PLANware

Creator of ONPASS® Pro, GIS software for educational planning.

Table of Contents

Tables

Figures

Introduction

Planware™, a division of Educational Data Systems, has prepared this white paper with three main goals related to enrollment projections:

1. Provide a guide to understanding the terminology used when generating enrollment projections.
2. Demonstrate how to do one-year projections using methods that do not require a statistical consultant and that require only historical enrollment data and common spreadsheet software.
3. Describe a method for determining the “best” projection method for your district.

Using a set of example kindergarten enrollment data, we evaluated six methods that require only historical enrollment data and commonly used spreadsheet software. This paper summarizes our findings and includes worked examples to demonstrate each method. Although each of the methods described is suitable for projecting enrollments for all grades and multiple years, we examined only kindergarten one-year projections.

It is, of course, not possible to predict the future. Trends in demographic variables may change abruptly, and the underlying relationships between and among demographic variables may shift through time. There is no guarantee that the most accurate methods using historical data will be the most accurate going forward. Planners in school districts must use their knowledge of local changes in trends and take any projection as a starting point for making decisions about the upcoming academic year.

This paper has three sections:

- Section I, **Terminology, Projection Methods and Data**, offers definitions of some key terms with descriptions of the six projection methods to clarify the language.

  Section II, **Worked Example Enrollment Projections**, contains “worked examples” of the six projection methods applied to example district-level historical enrollment data. Simple line charts show how the projections compare with actual enrollments over time. With some caveats, the same methods may be used to make individual school-level enrollment projections.
- Section III, **Determining the "Best" Projection Method**, explains how to find the projection method that is historically the most accurate and reliable for the example data set.

I. Terminology, Projection Methods and Data

Terminology

Projection Versus Forecast. Some analysts use the term “projection” and “forecast” interchangeably, and the terms are similar in that both refer to a prediction of something in the future. There is a subtle but important difference: a projection is a prediction made using only the past history of the variable of interest—in this case, kindergarten enrollment. A forecast includes other variables expected to be important for predicting the future—for example, live birth rates or student yield rates for housing developments.

This paper is about projections rather than forecasts since the methods described use only historical enrollment data to predict future enrollment. Graphs and tables show the actual and projected values, which may also be referred to in the text as “predicted” values.

Enrollment Data. “Enrollment” can mean that a family filled out a form or in some way took action to register, or enroll, their child(ren) to attend classes. In this paper, “enrolled” or “enrollment” refers only to the number of students who attend class(es) in a specified year.

Time Series Data. The enrollment numbers used for this paper are in the form of time series—enrollments collected for the district at a specific point in time, once each academic year, for multiple years in a row. The October average daily attendance from our example district is used for this report because it is collected in a consistent way and at the same time in each school year.

Time series enrollment data is influenced by many demographic trends (for example, birth rates, death rates, migration patterns, and inter- and intra-district transfers) and has the advantage of being readily available. But, it lacks information about nonenrollment variables, such as birth rates, and does not provide a way to answer questions about why trends emerge or change.

Standard Deviation. With the mean (or the average), the standard deviation is the workhorse of statistics. It measures the dispersion of values around the mean. It is calculated as the root mean squared difference between a set of values and their mean. A large standard deviation means that there is less certainty about the value of the true mean.

Error or Error Term. The error, or error term, is defined as the difference between the actual and projected value. The use of “error” does not imply that a mistake has been made; it only means that the projected value differs from the actual observed value. Nor is it quite the same as what is meant by "standard error" in statistics.

The error term is here defined simply as the (actual value – projected value). The error is positive if the projection is smaller than the actual value, and negative if the projection is larger.

Figure 1 shows example time series enrollment data with actual enrollments (solid blue line), projected enrollments (dashed red line), and the error terms (vertical yellow lines). For the 2017–18 school year, the actual and projected values were almost the same so the error term is not visible.

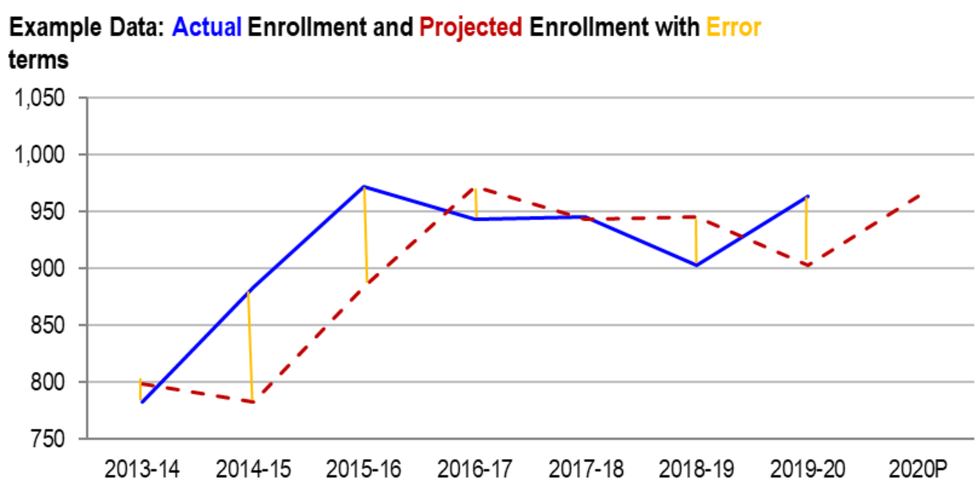


Figure 1: Actual and Projected Enrollment with Errors

The formula bars at the top of **Figure 2** show the Excel formulas used in a two-step process to calculate the error (the left panel) and absolute error (the right panel) for 2012–13. The formula is copied down the relevant column (D or E) to fill in all the error and absolute error values.

Computing the error.

=$B7-C7

	A	B	C	D
3	Year	Enroll	Projection	Error
4	2009-10	1491		
5	2010-11	1578		
6	2011-12	1664		
7	2012-13	1708	1672	36
8	2013-14	1732	1742	-10
9	2014-15	1782	1749	33
10	2015-16	1726	1818	-92
11	2016-17	1691	1724	-33
12	2017-18	1746	1670	76
13	2018-19	1709	1757	-48
14	2019-20	1759	1719	40
15	2020P		1766	

Computing the absolute error.

=ABS($B7-C7)

	A	B	C	D	E
3	Year	Enroll	Projection	Error	Abs Error
4	2009-10	1491			
5	2010-11	1578			
6	2011-12	1664			
7	2012-13	1708	1672	36	36
8	2013-14	1732	1742	-10	10
9	2014-15	1782	1749	33	33
10	2015-16	1726	1818	-92	92
11	2016-17	1691	1724	-33	33
12	2017-18	1746	1670	76	76
13	2018-19	1709	1757	-48	48
14	2019-20	1759	1719	40	40
15	2020P		1766		

Figure 2: Computing Errors and Absolute Errors

Summary error statistics are used to estimate the historical accuracy of enrollment projections. In this paper we focus on the root mean square error. For completeness other common summary methods—mean absolute error and mean absolute percentage error—are described in the appendix.

Root Mean Square Error (RMSE). The most commonly used statistic (for obscure mathematical reasons) is the root mean square error, which is particularly sensitive to the presence of large discrepancies between the actual and predicted enrollment. To compute the RMSE, square each error, compute the mean of the squared errors (the sum of squares, divided by the number of error terms), and take the square root of the mean.[1]

The formula bar in **Figure 3** shows the RMSE formula in cell D18.

[1] http://canworksmart.com/using-mean-absolute-error-forecast-accuracy/

=SQRT(SUMSQ(D9:D14)/COUNT(D9:D14))

	A	B	C	D	E
3	Year	Enroll	Projection	Error	
4	2009-10	1491			
5	2010-11	1578			
6	2011-12	1664			
7	2012-13	1708	1672	36	
8	2013-14	1732	1742	-10	
9	2014-15	1782	1749	33	
10	2015-16	1726	1818	-92	
11	2016-17	1691	1724	-33	
12	2017-18	1746	1670	76	
13	2018-19	1709	1757	-48	
14	2019-20	1759	1719	40	
15	2020P		1766		
16					
18	**RMSE:**			58	

Figure 3: Computing the Root Mean Square Error (RMSE)

The RMSE is in units of the variable being projected, in this case, the number of students. Because they are squared, a few relatively large errors will increase the RMSE relative to the mean absolute error (see appendix).

This white paper relies on the RMSE as it is the most commonly used and gives greater weight to large misses. The most accurate projection method is defined to be the one with the lowest RMSE based on historical data.

Accuracy. Accuracy is the inverse of the summary error (defined to equal 1/RMSE in this white paper); accuracy is greatest when error is smallest. Because there is no way to know the difference between the predicted and actual value in a future year, historical error values are used to estimate the accuracy of each projection method. The assumption is that errors for future-year projections will be similar to historical errors, but that cannot be known in advance.

Reliability. Reliability in this context refers to the expected consistency of projections over time. This depends heavily on the average enrollment size of the school or district under consideration. It is difficult to find a reliable method for projecting enrollments when the numbers are less than 1,000 students,[2] which is why it is easier to make reliable projections at the district level than the school level.

In large districts, random factors cause enrollments to fluctuate in unpredictable ways, but the fluctuations are more likely to cancel out. In small districts, even a small change in the number

[2] The discussion of reliability depends upon *K-12 School Enrollment Projections Study Final Report*, p. 19, published online at http://www.k12.wa.us/SchFacilities/Publications/pubdocs/EnrollProjectionMethodologiesFinalReport2008.pdf.

of students causes a relatively large percent change in enrollment levels, and there is less cancelation of effects, causing the projections to be less reliable.

It bears repeating that even with many years' worth of historical data, it is not possible to predict the future. All enrollment projection and forecasting methods depend on assumptions about the future that may or may not be correct. It must be assumed that future error rates will be similar to historical rates.

Projection Methods

This white paper presents five basic methods for projecting enrollments and a sixth "ensemble" method. The text here serves as an introduction to each method. Worked examples are provided in Section II, Worked Example Enrollment Projections.

Previous Year. The simplest method for projecting enrollment is to use the previous year's enrollment as the projected value. This method requires only one year of historical data to make the projection and has the advantage of "recovering" quickly from large errors in projection. It incorporates current demographic information but fails to take into account longer-term trends.

Growth Rate. One method for extrapolating longer-term trends is to apply the average year-to-year growth rate to the current year's enrollment to project the next year's enrollment. The growth rate is found by calculating the (current_year / previous_year) ratio for each of the previous three years and averaging them[3]. Then multiply the current year's enrollment by this ratio to get the projection for the next year. When the average growth rate is exactly 1, the projection will match the Previous Year method.

3-Year Average. Another way to take into account longer trends and smooth the projections is to use the mathematical average of enrollment in recent years (this report settled on three years) as the prediction for the next year. This is called a 3-year unweighted moving average. Moving averages smooth the curve but have an important disadvantage. When large errors occur, they distort projections not just for the next year but for the next several years.

Moving averages can be extended to longer time periods, say five years, or shortened to two, causing more or less smoothing. However, like the Previous Year method, they are not able to extrapolate a trend.

3-Year Weighted Average. Weighted moving averages are the same as moving averages except that they give more weight to recent years. This makes it easier for projections to "catch up" when there has been a big change in enrollments. Like unweighted moving averages, they do not extrapolate a trend.

The literature on projections and forecasts describes many methods for determining what weights to use in weighted moving averages. The projections in this report use a very simple scheme in which the weights are based on the number of years to be included in the weighted average. In a 3-year weighted average, the weights [3, 2, 1] are applied to recent enrollments. So if the three most recent enrollments are (in chronological order) 500, 600, 700, the weighted sum is 3*700 + 2*600 + 1*500 = 3,800; then divide by the sum of weights: 3,800 / 6 = 633.3.

Ordinary Least Squares (OLS) Regression. OLS regression is a technique that fits a line or "trendline" to data points in a way that minimizes the sum of squared differences between the observed data values and the line. Because OLS regression identifies a trend and fits a line to

---

[3] The time period of three years may be varied depending on local circumstances. We find that it is a good starting point, but your district enrollment trends may make a shorter or longer time period more reasonable.

it, it yields a simple straight-line formula for making predictions into the future: just plug a future date (as an integer or "time period") into the formula and get a prediction. Of the projection methods detailed in this report, OLS regression requires the most historical data.

There are limitations. Reliable OLS requires more data points than are usually available for enrollment projections. And because it is based on squared error terms, a single large outlier can have a big effect on the trendline. Most important, OLS regression assumes a straight-line linear trend, and many demographic trends underpinning school enrollments are not linear. Nonetheless, OLS and Growth Rate are the only two of the basic methods used here that can be said to actually extrapolate a trend into the future.

Ensemble. Individual forecasting methods tend to perform quite differently with different datasets. Sometimes the previous year is the best basis for a prediction, sometimes the OLS trendline is most accurate, and sometimes another method. The Ensemble method is a way to *combine* the various projection methods, putting extra weight on those methods that happen to be most successful with the current dataset.

It is assumed that each projection method has been applied to a sequence of years for which there exist actual enrollment figures, making it possible to calculate a summary RMSE error term for each projection method. The projections for the different methods are then combined using a weighted average where the weight is the accuracy of the method, given by the inverse of its RMSE (1/RMSE). We multiply each projection by 1/RMSE and divide the sum by the sum of the 1/RMSE terms.

This ensemble method takes advantage of trend-based methods like OLS and Growth Rate when they work and shifting to history-based methods like the 3-Year Average when they don't.

Table 1 summarizes the projection methodologies included in this paper. A numerical worked example for each method is provided in the Worked Example Enrollment Projections section. To facilitate explanation, a Planware analyst used an Excel spreadsheet to evaluate the six projection methods for their ability to predict enrollments with district-level example data.

Table 1: Projection Methods Summary

Projection Method	Abbreviation	Description
Previous year enrollment becomes the projected value	Previous Year	The enrollment from the previous year becomes the projection for the current year.
Year-to-year kindergarten enrollment growth rate, 3-year average	Growth Rate	Calculate the (current_year / previous_year) ratio for each of the previous three years and average them. Then multiply the current year's enrollment by this ratio to get the projection for next year. For school-level projections, use the district average growth rate to prevent unusually high/low results.
Moving average enrollment, 3-year	3-Yr Avg	Average the current enrollment with the two previous. That value is the projection for the next year.
Weighted moving average enrollment, 3-year	3-Yr Weight Avg	The current enrollment gets a weight of 3, next most recent gets a weight of 2, and next most recent gets a weight of 1. Multiply each enrollment by its weight, add the 3 products, and divide by the sum of the weights (that is 6 for 3 years). The result is the projection for the next year.
Ordinary Least Squares (OLS) regression	OLS	Use all previous kindergarten enrollment data and the “forecast” function in Excel (or “forecast.linear” function, depending on the Excel version) to produce the OLS projection. The first projected value reported here uses five years worth of data.
Weighted average of the projected values from all five methods	Ensemble	Weighted average of the projections from each of the other five methods. The weight is the accuracy of the method (1/RMSE).

Data File Description

The enrollment data files (2009–10 through 2019–2020) used in this paper are examples of data files that a typical school district would have as a result of state reporting requirements. In this paper we refer to our example district as the “District”.

- The historical data includes District enrollments by academic year and grade.
- The enrollment data is not disaggregated by special education demographic categories.
- Enrollment data includes students who attend public school in the District irrespective of where they live.

II. Worked Example Enrollment Projections

This section provides worked examples of each projection method. It displays line graphs containing year-by-year District-level kindergarten enrollments with their corresponding projections for each of the six projection methods. Screenshots of the relevant Excel tables, with cell formulas, show exactly how each method's projections are calculated. Projections for 2020–21 are labeled as "2020P."

The graphs in this paper show the number of students enrolled on the y-axis, and time (academic year) on the x-axis. In each graph, the y-axis is truncated; it does not show the entire 0 to 1,050 range for number of enrolled students. The truncated scale makes differences between lines more clear, but also makes year-to-year fluctuations in enrollment seem larger than they would if the entire scale were presented.

Enrollment information from previous time periods (not always shown) is used to calculate projected values and we show acutal and projected time series in the same graph. This shows the historical accuracy (the difference between acutal and projected values) of each method. In each line graph the actual enrollment is represented by a solid blue line and the projected enrollment is represented by a dashed line of another color.

Figure 4 includes only the actual number of kindergarten students enrolled for each academic year, 2009–10 through 2019–20. We want to use the information summarized in this trendline to project enrollment for the 2020-21 academic year.

The line shows abrupt increases in the number of kindergarten students starting in 2014–15, the kind of change that is most challenging for statistical forecasting as it comes out of nowhere. More recently, a downward trend was reversed in 2019–20.

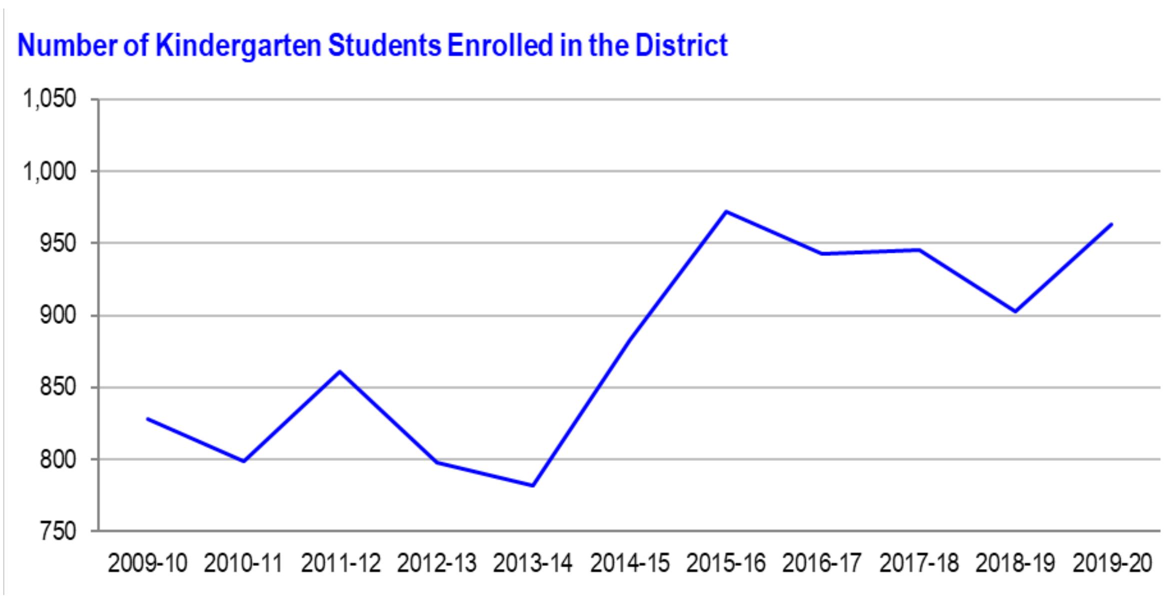


Figure 4: District Kindergarten Enrollment, 2009–10 to 2019–20

Previous Year Method

Figure 5 compares the actual enrollment in a given year (solid blue line) with a projection based on the previous year's enrollment (dashed red line). The chart starts at 2013–14 to enable comparisons between methods; this is the first year for which projections are available for most methods.

The projected enrollment line looks almost like a shadow of the actual enrollment line because it is just shifted by one time period.

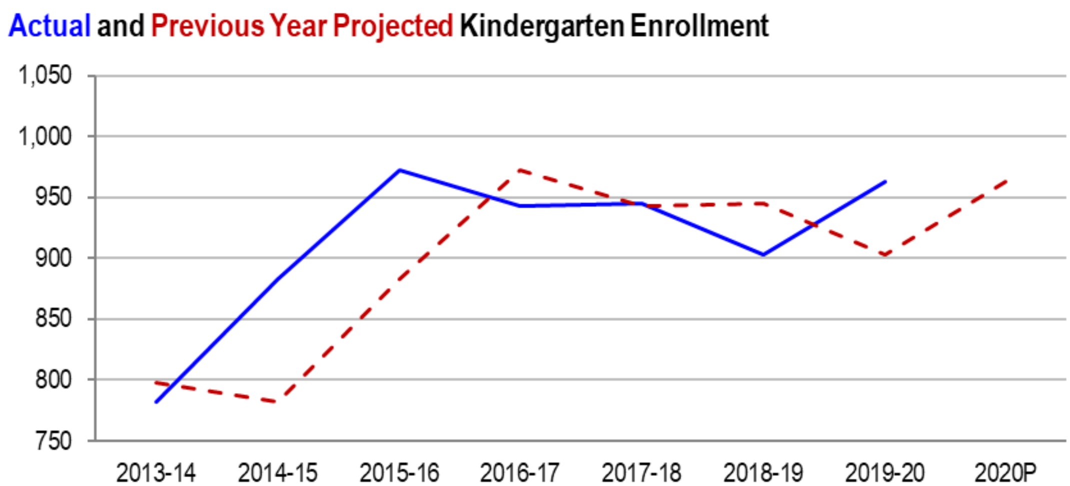


Figure 5: Previous Year Chart

Figure 6 shows the corresponding Excel data and formula. The formula bar at the top shows the projection formula typed into cell E7 (with the black border), which in this case is a simple reference to the actual enrollment from the previous year (cell B6 with the green border). The Excel formula is copied down the cells in column E and leads to a projection for 2020 in cell E17.

fx =B6

	A	B	D	E
5	Year	Enroll		Project
6	2009-10	828		
7	2010-11	799		828
8	2011-12	861		799
9	2012-13	798		861
10	2013-14	782		798
11	2014-15	883		782
12	2015-16	972		883
13	2016-17	943		972
14	2017-18	945		943
15	2018-19	903		945
16	2019-20	963		903
17	2020P			963

Figure 6: Previous Year Calculation

Growth Rate Method

Figure 7 compares the actual enrollment (solid blue line) with the growth rate (dashed orange line) projections. The growth rate reflects recent demographic trends and, when averaged, smooths out effects of departures from those trends. The projection line starts with 2013–14 as the formula requires four previous years of enrollment data.

The high growth rates between 2013–14 and 2015–16 lead to large "misses" in projections for the following years when those growth rates are not sustained.

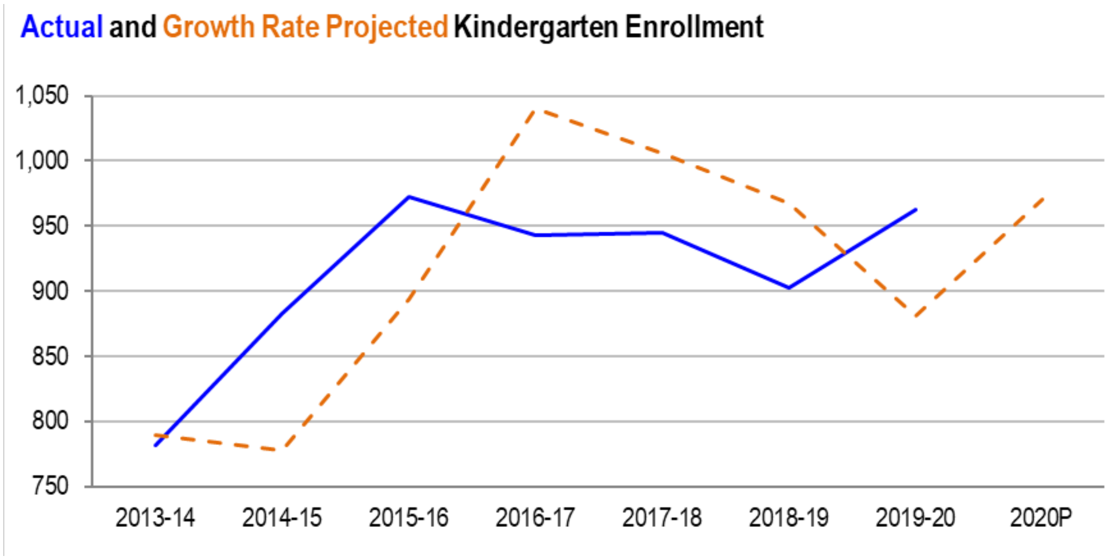


Figure 7: Growth Rate Chart

Growth rates are calculated in three steps, each step corresponding to an extra column in the following spreadsheet tables[4]. Step 1, shown in **Figure 8**, is to calculate year-to-year growth rates, where the growth rate is the ratio of the enrollment in the current year to that of the previous year (cell J7). The formula is repeated for each year down column J. The "$" in front of the "B" tells Excel to "lock" the column as you are copying the formula.

[4] Excel experts can combine these steps, but the process is laid out here for clarity.

=$B7/$B6

	A	B	D	J
5	Year	Enroll		Growth Rate
6	2009-10	828		
7	2010-11	799		0.96
8	2011-12	861		1.08
9	2012-13	798		0.93
10	2013-14	782		0.98
11	2014-15	883		1.13
12	2015-16	972		1.10
13	2016-17	943		0.97
14	2017-18	945		1.00
15	2018-19	903		0.96
16	2019-20	963		1.07

Figure 8: Growth Rate Calculation, Step 1

Step 2 in **Figure 9** calculates the 3-year average growth rate (cell K9), averaging the cells from K7 to K9. The formula is copied down column K.

=AVERAGE(J7:J9)

	A	B	D	J	K
5	Year	Enroll		Growth Rate	3-year Avg
6	2009-10	828			
7	2010-11	799		0.96	
8	2011-12	861		1.08	
9	2012-13	798		0.93	0.99
10	2013-14	782		0.98	0.99
11	2014-15	883		1.13	1.01
12	2015-16	972		1.10	1.07
13	2016-17	943		0.97	1.07
14	2017-18	945		1.00	1.02
15	2018-19	903		0.96	0.98
16	2019-20	963		1.07	1.01

Figure 9: Growth Rate Calculation, Step 2

Step 3 in **Figure 10** multiplies the most recent enrollment figure (cell B9) by the average growth rate up to that year (cell K9) to yield the next year's (rounded) projection (cell L10). Because the method requires four years of historical data, projections only start in 2013–14; the previous years in column L are empty. The “$” in front of the “B” tells Excel to “lock” the column as you are copying the formula.

=$B9*K9

	A	B	D	J	K	L
5	Year	Enroll		Growth Rate	3-year Avg	Project
6	2009-10	828				
7	2010-11	799		0.96		
8	2011-12	861		1.08		
9	2012-13	798		0.93	0.99	
10	2013-14	782		0.98	0.99	790
11	2014-15	883		1.13	1.01	778
12	2015-16	972		1.10	1.07	894
13	2016-17	943		0.97	1.07	1040
14	2017-18	945		1.00	1.02	1006
15	2018-19	903		0.96	0.98	968
16	2019-20	963		1.07	1.01	881
17	2020P					971

Figure 10: Growth Rate Calculation, Step 3

3-Year Average Method

Figure 11 compares the actual enrollment (solid blue line) with the 3-year moving average (dashed purple line), the average enrollment of the three previous years. This method shows a substantial lag in projected enrollments starting in 2014–15 because the low enrollment in 2013–14 pulls down the average for each of the following three years.

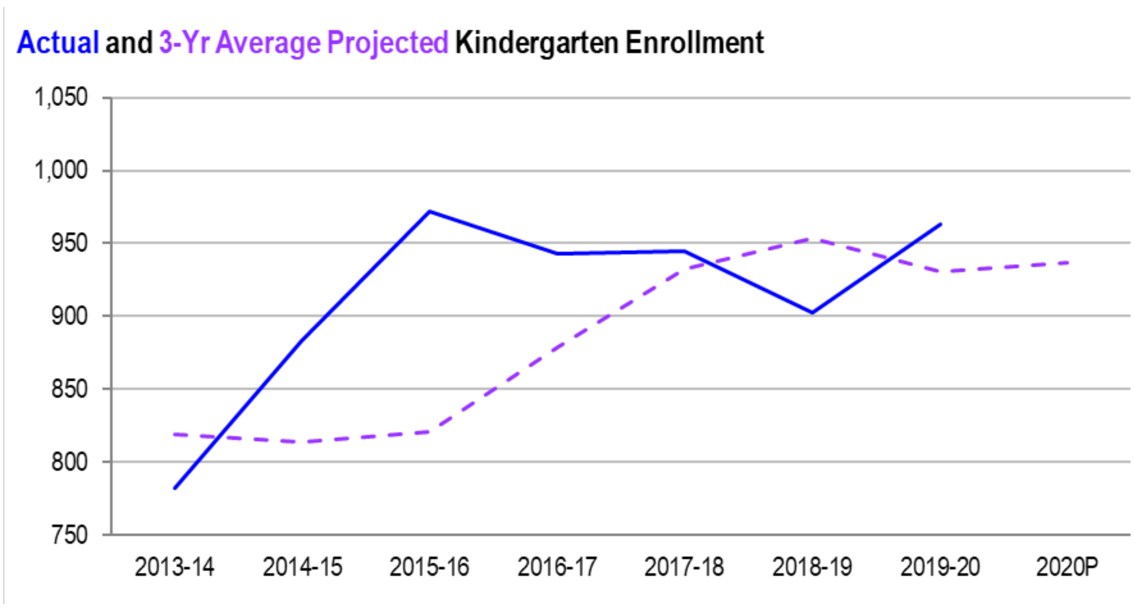


Figure 11: 3-Year Average Chart

The 3-Year Average calculation is shown in two steps. The first panel on the left (step 1) of **Figure 12** shows that cell Q8 is the average of the cells from B6 to B8. The second panel on the right (step 2) simply carries this (rounded) average forward as the projection for the next year, 2012–13 (cell R9). The "$" in front of the "B" tells Excel to "lock" the column as you are copying the formula.

=AVERAGE($B6:$B8)

	A	B	D	Q
5	Year	Enroll		3-Yr Avg
6	2009-10	828		
7	2010-11	799		
8	2011-12	861		829.3
9	2012-13	798		819.3
10	2013-14	782		813.7
11	2014-15	883		821.0
12	2015-16	972		879.0
13	2016-17	943		932.7
14	2017-18	945		953.3
15	2018-19	903		930.3
16	2019-20	963		937.0
17	2020P			

=Q8

	A	B	D	Q	R
5	Year	Enroll		3-Yr Avg	Project
6	2009-10	828			
7	2010-11	799			
8	2011-12	861		829.3	
9	2012-13	798		819.3	829
10	2013-14	782		813.7	819
11	2014-15	883		821.0	814
12	2015-16	972		879.0	821
13	2016-17	943		932.7	879
14	2017-18	945		953.3	933
15	2018-19	903		930.3	953
16	2019-20	963		937.0	930
17	2020P				937

Figure 12: 3-Year Average Calculation, Steps 1–2

3-Year Weighted Average Method

Figure 13 compares the actual enrollment (solid blue line) with the 3-Year Weighted Average (dashed brown line), in which the projected value gives more weight to the most recent enrollments. This method responds more quickly to the multi-year 2014–15 through 2015–16 surge in enrollment because the low 2013–14 enrollment is successively less influential in the average.

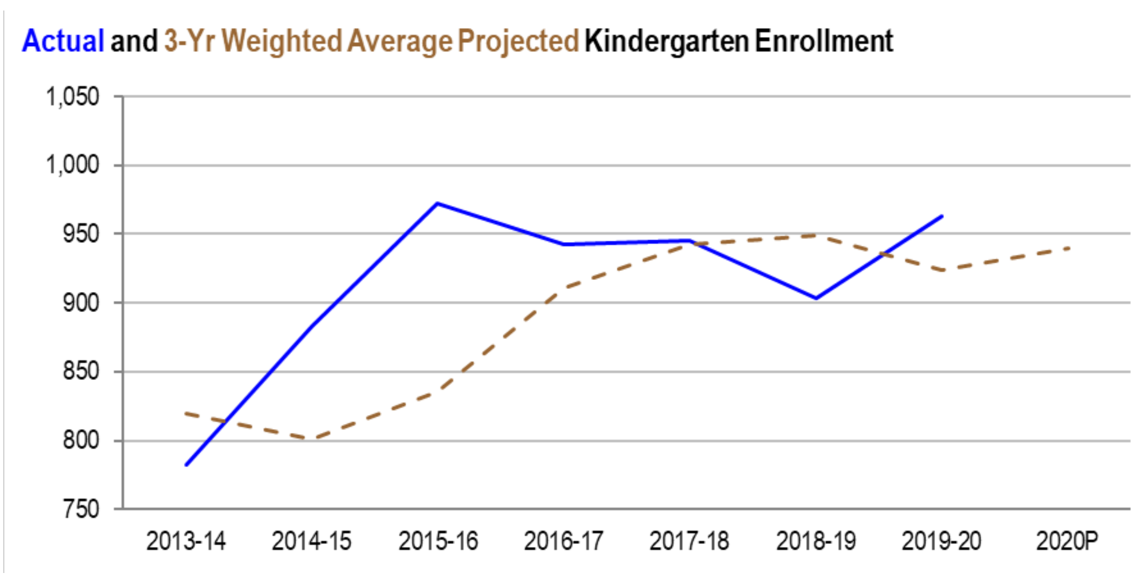


Figure 13: 3-Year Weighted Average Chart

The first panel (step 1) of **Figure 14** (cell W8) shows how the weighted average is calculated, with more recent years getting more weight (weights = [1, 2, 3]), remembering to divide by the sum of the weights. The second panel (step 2) carries the (rounded) weighted average to the next year as the projection (cell X9).

fx =((B6*1)+(B7*2)+(B8*3))/6

	A	B	D	W	X
5	Year	Enroll		3-Yr Weight Avg	
6	2009-10	828			
7	2010-11	799			
8	2011-12	861		834.8	
9	2012-13	798		819.2	
10	2013-14	782		800.5	
11	2014-15	883		835.2	
12	2015-16	972		910.7	
13	2016-17	943		942.7	
14	2017-18	945		948.8	
15	2018-19	903		923.7	
16	2019-20	963		940.0	
17	2020P				

fx =W8

	A	B	D	W	X
5	Year	Enroll		3-Yr Weight Avg	Project
6	2009-10	828			
7	2010-11	799			
8	2011-12	861		834.8	
9	2012-13	798		819.2	835
10	2013-14	782		800.5	819
11	2014-15	883		835.2	801
12	2015-16	972		910.7	835
13	2016-17	943		942.7	911
14	2017-18	945		948.8	943
15	2018-19	903		923.7	949
16	2019-20	963		940.0	924
17	2020P				940

Figure 14: 3-Year Weighted Average Calculation, Steps 1–2

Ordinary Least Squares (OLS) Regression Method

Computing OLS regression coefficients involves minimizing the sum of squared errors when fitting a straight line to a set of data points. The OLS regression coefficients describe the slope and intercept of the line. The projection is made by plugging the known information (the year as an integer "time period") into the formula of the line to get the next expected value.

Excel includes a function called "FORECAST" (in newer versions of Excel, "FORECAST.LINEAR") that uses OLS regression to extrapolate the next value in a series.

Figure 15 compares the actual enrollments (solid blue line) with OLS regression projections (dashed light blue line). The projections start in 2014–15 as we allocate five years of historical data to calculate the regression line.

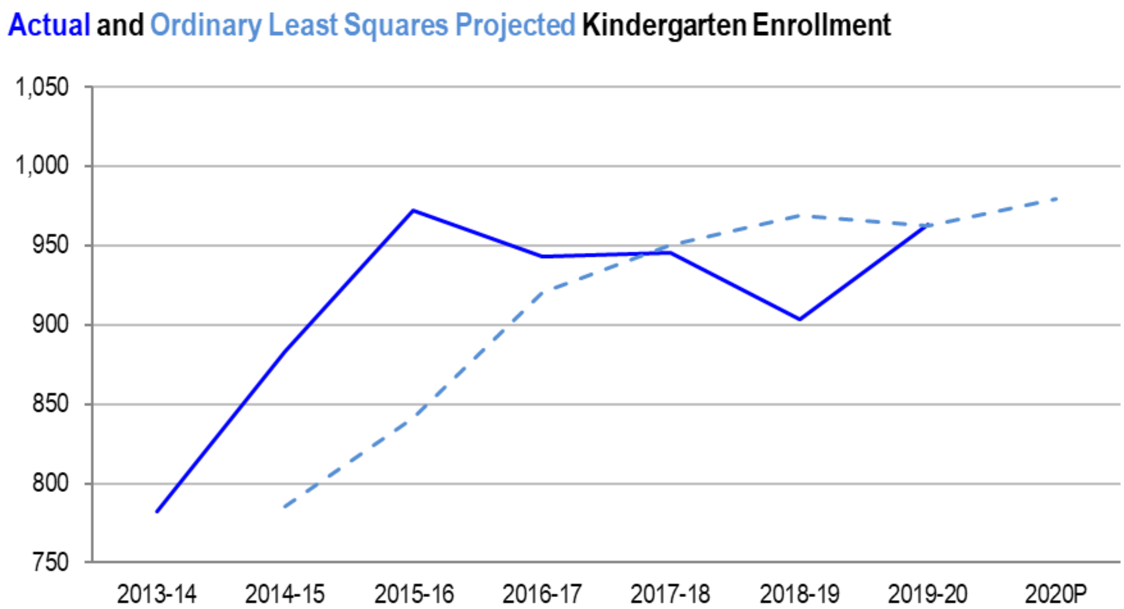


Figure 15: OLS Regression Method Chart

The first panel of **Figure 16** (cell AC11) shows how the FORECAST function in Excel calculates the regression line by relating a time period integer (in column C) to enrollment data from 2009–10 through 2013–14 (in column B).

The formula then applies it to the desired time period integer, in this case "6," to get a 2014–15 projection. The second panel (cell AC17) shows how this process is carried forward to get a 2020 projection. You want to be sure the entire range is included in your formula. The “$”s in front of the “B”, “C” and the “6” tell Excel to “lock” the column and the starting cell as you are copying the formula.

=FORECAST(C11,$B$6:B10,$C$6:C10)

	A	B	C	D	AC	AD
5	Year	Enroll	Time period		Project	
6	2009-10	828	1			
7	2010-11	799	2			
8	2011-12	861	3			
9	2012-13	798	4			
10	2013-14	782	5			
11	2014-15	883	6		786	
12	2015-16	972	7		841	
13	2016-17	943	8		921	
14	2017-18	945	9		950	
15	2018-19	903	10		968	
16	2019-20	963	11		962	
17	2020P		12		979	

=FORECAST(C17,$B$6:B16,$C$6:C16)

	A	B	C	D	AC	AD
5	Year	Enroll	Time period		Project	
6	2009-10	828	1			
7	2010-11	799	2			
8	2011-12	861	3			
9	2012-13	798	4			
10	2013-14	782	5			
11	2014-15	883	6		786	
12	2015-16	972	7		841	
13	2016-17	943	8		921	
14	2017-18	945	9		950	
15	2018-19	903	10		968	
16	2019-20	963	11		962	
17	2020P		12		979	

Figure 16: OLS Regression Method Calculation

Ensemble Method

One way to use information from all of the projection methods explained here is to take an average of the projected values and use it as the next projected value. A plain average would give equal weight to each projection method.

For our Ensemble method, we weight the average of the other five methods.The weight for each projection method in the average is the historical accuracy (1/RMSE) for that method.

Figure 17 compares the actual enrollments (solid blue line) with Ensemble projections (dashed brown line). The projection line starts at 2014–15 as this is the most recent year for which projections exist for all five other methods.

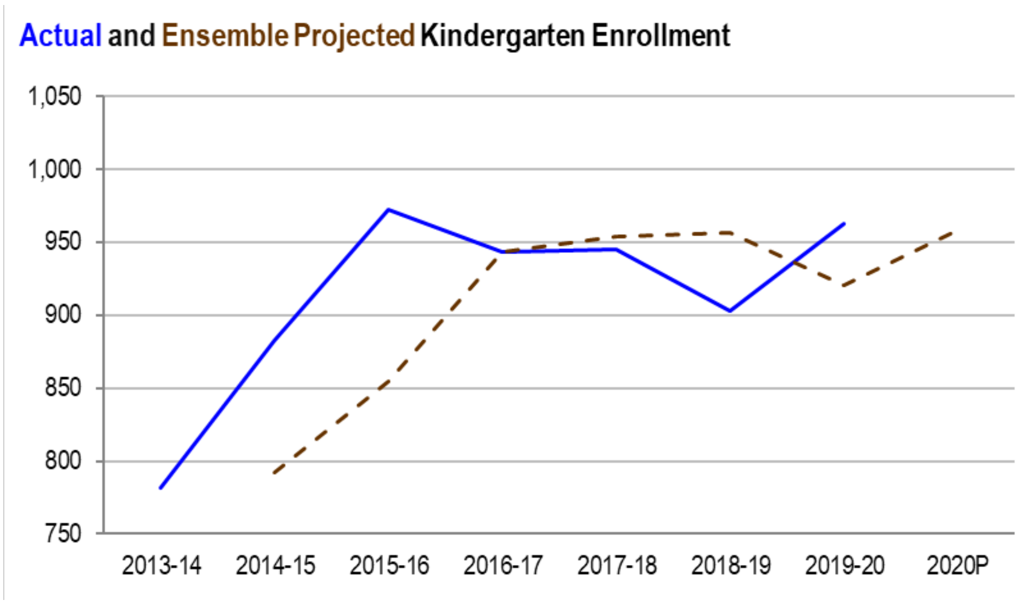
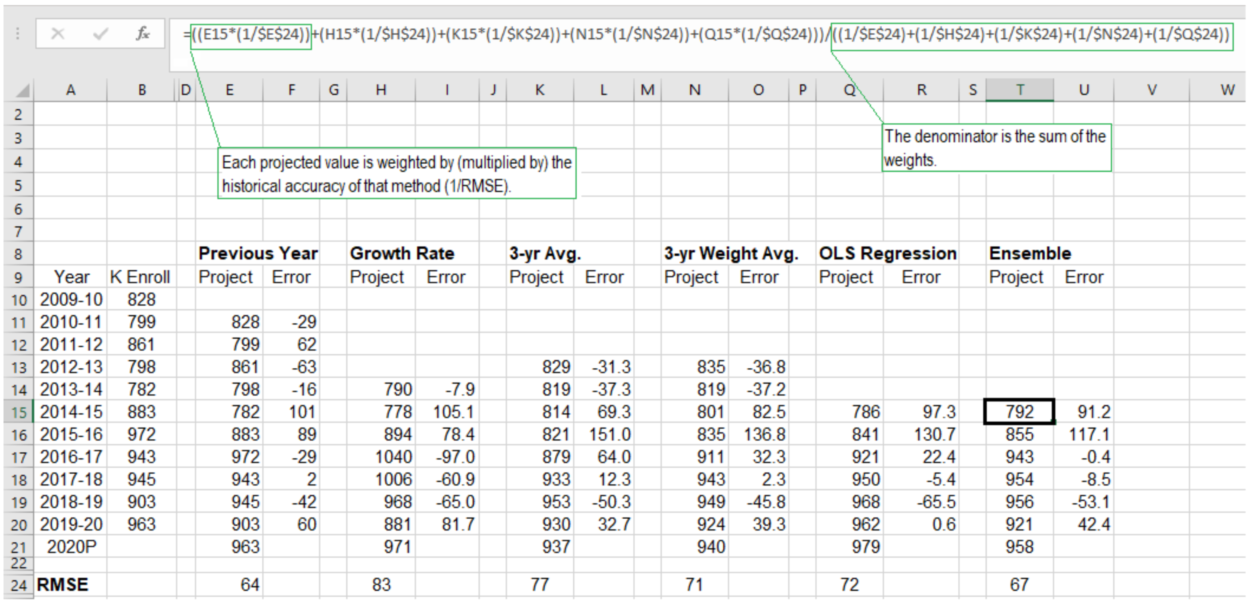


Figure 17: Ensemble Method Chart

Figure 18 shows the formula in cell T15 for the Ensemble projection, which is the weighted average of all five projections (E15, H15, K15, N15 and Q15). The weight for each projection is the historical accuracy of that projection method, defined as the inverse of the RMSE (which is in row 24).



		Previous Year		Growth Rate		3-yr Avg.		3-yr Weight Avg.		OLS Regression		Ensemble	
Year	K Enroll	Project	Error	Project	Error	Project	Error	Project	Error	Project	Error	Project	Error
2009-10	828												
2010-11	799	828	-29										
2011-12	861	799	62										
2012-13	798	861	-63			829	-31.3	835	-36.8				
2013-14	782	798	-16	790	-7.9	819	-37.3	819	-37.2				
2014-15	883	782	101	778	105.1	814	69.3	801	82.5	786	97.3	792	91.2
2015-16	972	883	89	894	78.4	821	151.0	835	136.8	841	130.7	855	117.1
2016-17	943	972	-29	1040	-97.0	879	64.0	911	32.3	921	22.4	943	-0.4
2017-18	945	943	2	1006	-60.9	933	12.3	943	2.3	950	-5.4	954	-8.5
2018-19	903	945	-42	968	-65.0	953	-50.3	949	-45.8	968	-65.5	956	-53.1
2019-20	963	903	60	881	81.7	930	32.7	924	39.3	962	0.6	921	42.4
2020P		963		971		937		940		979		958	
RMSE		64		83		77		71		72		67	

Figure 18: Ensemble Method Calculation

All Methods Shown Together

A graph with all the projections (dashed color lines) and the actual enrollment (solid blue line) shows the relative differences in projection results. Because the projection methods require more or less historical data, some of the lines shown start in later years than the others. For example, the OLS line (dashed light blue) does not begin until 2014–15, even though it uses all available previous data to generate that projected enrollment.

Figure 19 shows how the projections that depend upon average values take longer to catch up to the enrollment increases of 2014–15 through 2015–16 and to the lower enrollments that followed.

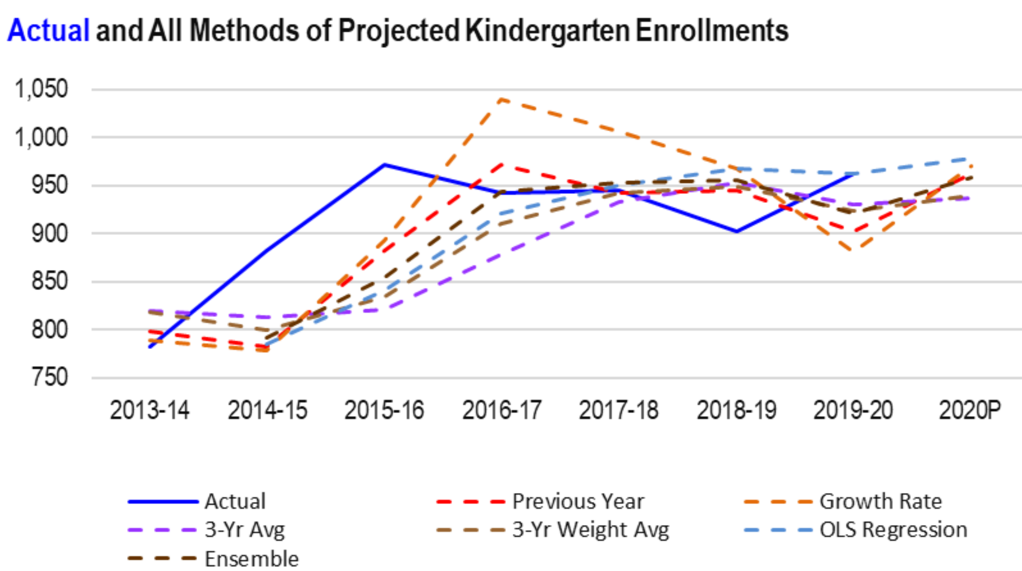


Figure 19: District Projections, All Methods

Table 2 shows the actual enrollment and (rounded) projections of all six methods, along with their RMSEs. Since some methods require more historical data to make a projection, they have more empty cells leading up to the first projection.

To ensure that the RMSE statistic is comparable across methods, it is calculated from the same range of years for each method, 2014–15 through 2019–2020.

Table 2: Historical Enrollment, Projections, and RMSE for Each Method

Year	Kindergarten Enrollment	Projections Previous Year	Growth Rate	3-Yr Average	3-Yr Weighted Average	OLS	Ensemble
2009–10	828						
2010–11	799	828					
2011–12	861	799					
2012–13	798	861		829	835		
2013–14	782	798	790	819	819		
2014–15	883	782	778	814	801	786	792
2015–16	972	883	894	821	835	841	855
2016–17	943	972	1040	879	911	921	943
2017–18	945	943	1006	933	943	950	954
2018–19	903	945	968	953	949	968	956
2019–20	963	903	881	930	924	962	921
2020P		963	971	937	940	979	958
RMSE		64	83	77	71	72	67

The worked examples include only district-level data for one grade. The projection methods can be used for school and/or grade-level data, however some important cautions must be noted.

First, the reliability is likely to be highest at the district level for the reasons explained in the Terminology section. School-level projections are thus typically less accurate than district-level projections.

Second, if you add up school-level projections they often yield different (and sometimes obviously inaccurate) district totals than if you make projections using district-level data[5]. It is important to be sure the school-level projections are constrained by reasonable district projections.

For example, one method to constrain school-level projections is to calculate the growth rate at the district level and apply it to all schools within the district. The problem is that sometimes schools really are growing at different rates. Determining the best projection method across schools requires a nuanced approach.

[5] For a detailed discussion of this topic as it relates to cohort-survival models, see Kelley D. Cary, *School District Master Planning: A Practical Guide to Demographics and Facilities Planning*, pp. 94-100.

III. Determining the "Best" Projection Method

The charts and tables in this report display actual and projected district-level enrollments for six methods across 10 years, with an RMSE statistic to indicate the success of each method. To ensure that the RMSEs are comparable across methods, they were calculated using only the years for which all methods had valid projections (no missing cells).

In principle, one could do school and district-level projections and create a table such as Table 2 for each set of projections, scan the charts, look for the method with the lowest RMSE, and call that the "best" method. But, it soon becomes clear that this is not always practical, especially for projecting enrollment for schools within a district. Different methods may pop up as "best" for each school, and it is hard to compare the RMSEs from schools that have different enrollment sizes. We need a statistical procedure.

First, we need to be clear what "**best**" means. We want just one method that can be applied with confidence to the district and all schools, over time. It should not only be the most *accurate* overall but the most *consistently* accurate (reliable, robust). To get the most accurate method, one might think to simply average the RMSE statistic for each method across schools (treating the district as simply a large school). But this doesn't quite work because the size of the school in part drives the size of the RMSE statistic; the RMSEs are not comparable.

To get around this problem, we *standardize* each RMSE statistic with a formula commonly used in statistics:

$$\mathrm{std}[x] = (x - \mathrm{mean}(x)) / \mathrm{stdev}(x)$$

where x in this case is the RMSE. So for each school (or district), we first calculate the average RMSE across all methods as well as its standard deviation (see Terminology). Then for each method, we subtract the mean from the RMSE of the method and divide by the standard deviation. The result is a number that tends to fall somewhere between -2 and 2 (where zero is the mean). This is the standardized RMSE. The smaller and more negative the standardized RMSE for a given method, the more accurate it is.

Figure 20 shows how this is done in Excel. The top panel formula bar shows how to calculate the standard deviation using the STDEV.P formula[6] in cell J16 (the AVERAGE formula in cell I16 is not shown). The second panel shows how the standardized RMSE is calculated for the Ensemble method with the formula bar showing the contents of cell H17. The formula is copied across cells C17 to H17.

[6] Older versions of Excel use the formula “STDEVP”.

J16 =STDEV.P(C16:H16)

	A	B	C	D	E	F	G	H	I	J
1										
2	**District**									
3	**Year**	**Kindergarten Enrollment**	**Previous Year**	**Growth Rate**	**3-yr Average**	**3-yr Weighted Average**	**OLS Regression**	**Ensemble**		
4	2009-10	828								
5	2010-11	799	828							
6	2011-12	861	799							
7	2012-13	798	861		829	835				
8	2013-14	782	798	790	819	819				
9	2014-15	883	782	778	814	801	786	792		
10	2015-16	972	883	894	821	835	841	855		
11	2016-17	943	972	1040	879	911	921	943		
12	2017-18	945	943	1006	933	943	950	954		
13	2018-19	903	945	968	953	949	968	956		
14	2019- 20	963	903	881	930	924	962	921		
15	2020P		963	971	937	940	979	958	**Mean**	**Std. Dev.**
16	**RMSE**		64	83	77	71	72	67	72.24	6.32
17	**Std_RMSE**		-1.35	1.68	0.73	-0.20	0.01	-0.87		
18										

H17 =(H16-$I$16)/$J$16

	A	B	C	D	E	F	G	H	I	J
1										
2	**District**									
3	**Year**	**Kindergarten Enrollment**	**Previous Year**	**Growth Rate**	**3-yr Average**	**3-yr Weighted Average**	**OLS Regression**	**Ensemble**		
4	2009-10	828								
5	2010-11	799	828							
6	2011-12	861	799							
7	2012-13	798	861		829	835				
8	2013-14	782	798	790	819	819				
9	2014-15	883	782	778	814	801	786	792		
10	2015-16	972	883	894	821	835	841	855		
11	2016-17	943	972	1040	879	911	921	943		
12	2017-18	945	943	1006	933	943	950	954		
13	2018-19	903	945	968	953	949	968	956		
14	2019- 20	963	903	881	930	924	962	921		
15	2020P		963	971	937	940	979	958	**Mean**	**Std. Dev.**
16	**RMSE**		64	83	77	71	72	67	72.24	6.32
17	**Std_RMSE**		-1.35	1.68	0.73	-0.20	0.01	-0.87		
18										

Figure 20: Standardized RMSE

Once the RMSEs have been standardized, the process for finding the "best" projection method is straightforward. Calculate the mean and standard deviation of the standardized RMSEs for each method. The method with the smallest (most negative) standardized RMSE is the most accurate overall. The method with the smallest standard deviation is the most consistent (reliable, robust).

The example in Figure 20 is for a district-level projection, but the same process would be carried out to produce school level projection tables (one table for each school) with RMSE and standardized RMSE.

Table 3 shows an example of a district with 11 elementary schools. The table is based on enrollment projections (not shown for brevity) that have been made using each method for each school. Each cell contains the standardized RMSE for the given school and enrollment projection method.

In this example, the most accurate projection method is the Ensemble method, which has a mean standardized RMSE of -0.57. As it happens, the Ensemble method also is the most consistent with a standard deviation of 0.18. The runner-up is the 3-Year Weighted Average method with a mean of -0.36 and standard deviation of 0.55.

Table 3: Example District Kindergarten Standardized RMSE

District/School	Previous Year	Growth Rate	3-Yr Average	3-Yr Weight Average	OLS	Ensemble
Elementary One	-1.35	1.68	0.73	-0.20	0.01	-0.87
Elementary Two	0.21	2.11	-0.78	-0.70	-0.24	-0.61
Elementary Three	0.01	-0.34	-0.33	-0.67	2.16	-0.83
Elementary Four	-1.48	-0.62	1.47	0.36	0.88	-0.60
Elementary Five	0.71	1.71	0.21	-0.85	-1.23	-0.55
Elementary Six	-0.56	0.48	0.04	-1.11	1.91	-0.76
Elelmentary Seven	0.67	1.93	-0.76	-0.42	-0.89	-0.52
Elementary Eight	-0.72	-1.39	1.72	0.57	0.25	-0.43
Elementary Nine	1.01	1.48	-1.41	-0.82	0.04	-0.30
Elementary Ten	-1.02	-1.23	1.62	0.42	0.67	-0.43
Elementary Eleven	-1.33	0.25	0.01	-0.54	1.94	-0.33
Mean	-0.35	0.55	0.23	-0.36	0.50	-0.57
SD	0.87	1.24	1.00	0.55	1.09	0.18

How do we interpret this finding? First, it is important to note that the simplest projection method, Previous Year, works quite well. This is due in large part to the fact that Previous Year responds most quickly to unpredictable jumps and drops in enrollment as occurred with the district in 2014–15 and 2015–16, and enrollments *always* contain unpredictable jumps and drops over time. However, there is often a genuine long-term trend component in enrollment that Previous Year necessarily lags.

Thus, we see two types of methods: historical and trend-based. Historical methods (Previous Year, 3-Year Average, and 3-Year Weighted Average) imitate what occurred recently. Trend-based methods (Growth Rate and OLS) try to extrapolate a trend. The Ensemble method is clever enough to know when to emphasize the historical methods, when to emphasize the trend-based methods, and when to shift its weighting as necessary. That is why it performs **best** overall in this example district.

Appendix: Alternative Ways to Summarize Projection Error

Enrollment projections may report indicators of accuracy based on other (than RMSE) ways to summarize the historical errors. This appendix provides definitions and examples for how to compute the mean absolute error and mean absolute percentage error using example kindergarten enrollment data.

Mean Absolute Error (MAE). The MAE is the mathematical average of the absolute values of the error terms. The unit of the MAE is the same as the variable being projected; for enrollment projections, it is number of students. The MAE may underestimate the effect of large but infrequent errors, sometimes called “outliers.”

Figure A1 shows how to calculate the mean absolute error over time. The formula is shown in cell E17.

f_x =AVERAGE(E9:E14)

	A	B	C	D	E
3	Year	Enroll	Projection	Error	Abs Error
4	2009-10	1491			
5	2010-11	1578			
6	2011-12	1664			
7	2012-13	1708	1672	36	36
8	2013-14	1732	1742	-10	10
9	2014-15	1782	1749	33	33
10	2015-16	1726	1818	-92	92
11	2016-17	1691	1724	-33	33
12	2017-18	1746	1670	76	76
13	2018-19	1709	1757	-48	48
14	2019-20	1759	1719	40	40
15	2020P		1766		
16					
17	**MAE:**				53

Figure A1: Computing Mean Absolute Error (MAE)

Mean Absolute Percentage Error (MAPE). This is used to get a unit-free estimate of historical accuracy. Calculate individual percentage error terms (((actual – projected) / actual) * 100) and average their absolute values. Because the MAPE is unit-free, it is often used to compare the projection performance between data sets.

Figure A2 demonstrates how to first calculate the absolute percentage error for each year (cell F7 on the left). Average those values across all years (cell F19 on the right) to get the MAPE.

Computing the absolute percentage error.

=((ABS($B7-C7))/$B7)*100

	A	B	C	D	E	F
3	Year	Enroll	Projection	Error	Abs Error	Abs % Error
4	2009-10	1491				
5	2010-11	1578				
6	2011-12	1664				
7	2012-13	1708	1672	36	36	2.09
8	2013-14	1732	1742	-10	10	0.59
9	2014-15	1782	1749	33	33	1.83
10	2015-16	1726	1818	-92	92	5.31
11	2016-17	1691	1724	-33	33	1.94
12	2017-18	1746	1670	76	76	4.34
13	2018-19	1709	1757	-48	48	2.79
14	2019-20	1759	1719	40	40	2.29
15	2020P		1766			
16						
19						

Computing the mean absolute percentage error (MAPE).

=AVERAGE(F9:F14)

	A	B	C	D	E	F
3	Year	Enroll	Projection	Error	Abs Error	Abs % Error
4	2009-10	1491				
5	2010-11	1578				
6	2011-12	1664				
7	2012-13	1708	1672	36	36	2.09
8	2013-14	1732	1742	-10	10	0.59
9	2014-15	1782	1749	33	33	1.83
10	2015-16	1726	1818	-92	92	5.31
11	2016-17	1691	1724	-33	33	1.94
12	2017-18	1746	1670	76	76	4.34
13	2018-19	1709	1757	-48	48	2.79
14	2019-20	1759	1719	40	40	2.29
15	2020P		1766			
16						
19	**MAPE:**					3.08%

Figure A2: Computing Mean Absolute Percentage Error (MAPE)

One important difficulty with MAPE is that, due to how it defines the denominator, it overemphasizes negative errors relative to positive errors so that even when two error terms have the same absolute value, their corresponding percentage errors will be unequal.[7] Therefore, choosing a projection method that relies only on minimizing the MAPE may yield a method that systematically underpredicts actual enrollment values.

Figure A3 demonstrates how a negative error (an overprediction) can carry more weight in the MAPE. In this school-level kindergarten enrollment data the absolute error was six in both 2011 (cell H8) and again in 2018 (cell H15), but in 2018 the value was negative (cell G15). As shown in cell I15, the negative percentage error has a larger absolute percentage error than the positive error shown in cell I8.

=((ABS($D15-F15))/$D15)*100

	C	D	F	G	H	I
5	Year	Enroll	Project	Error	Abs Error	Abs % error
6	2009	147				
7	2010	145	147	-2	2.0	1.38
8	2011	**151**	**145**	**6**	**6.0**	**3.97**
9	2012	154	151	3	3.0	1.95
10	2013	132	154	-22	22.0	16.67
11	2014	143	132	11	11.0	7.69
12	2015	182	143	39	39.0	21.43
13	2016	157	182	-25	25.0	15.92
14	2017	153	157	-4	4.0	2.61
15	2018	**147**	**153**	**-6**	**6.0**	**4.08**
16	2019	152	147	5	5.0	3.29
17	2020P		152			

Figure A3: MAPE and Negative Errors

[7] An explanation for this result is given at https://robjhyndman.com/hyndsight/smape/.

About Planware

For over 40 years, Educational Data Systems' focus has been on the collection, processing, and reporting of data for education. The assessment division prepares, manages, analyzes, and reports on data for small to very large survey projects, program evaluations, local benchmark testing programs, and statewide large-scale assessments. The planning division, Planware, provides geographical information system (GIS) services and software to school districts.