# A Review

of the

# Housing Unit and Enrollment Projection Methodology in Prince William County

September 2019

# Executive Summary

Statistical Forecasting LLC and RLS Demographics Inc. ("Consultants") were hired by the Prince William County Public Schools ("PWCS") to review and evaluate the current methodology used to project enrollments in the school district, as well as the process undertaken to compute future housing units by the Prince William County Government ("PWCG"). In addition to a thorough methodological review, the Consultants provided recommendations for improvements to the forecasting process and an assessment of best practices in housing unit and enrollment projection methodology was completed through a literature review.

To conduct the investigation, a multitude of documents were reviewed, including: enrollment history, enrollment projections, Metropolitan Washington Council of Governments ("MWCOG") forecasting methods and results, PWCG demographic estimates and projections, student generation factors, comparisons of projected to actual enrollments, and Prince William County birth and migration data. Excel files and databases were analyzed as well as output generated from the statistical software package SPSS. Follow-up conversations occurred with Division of Planning Services staff to clarify elements of the enrollment projection methodology. In the following sections, recommendations are made based on the telephone interviews conducted and after examining the aforementioned documents.

This evaluation makes 19 recommendations, which are addressed more thoroughly in the report. The recommendations are as follows:

**Recommendation #1: PWCG should consider the development of comprehensive methodology documentation. It is difficult to get a single, detailed description of the methods employed.**

**Recommendation #2: Use of current American Community Survey (ACS) microdata for the county's three PUMA's to update is important and necessary. However, margins of error in the ACS estimates can be large and should be evaluated.**

**Recommendation #3: The age composition of the population is an important driver of housing needs, development and school enrollment and should be incorporated in the estimates and projections process.**

**Recommendation #4: Analysis of the ACS migration can be used to better understand migration flows and characteristics. This, along with on-going analysis of age-specific migration patterns will update inputs to the demographic projections**

**Recommendation #5: Age-specific fertility patterns by age of mother should be analyzed on an annual basis to understand changes in fertility levels and patterns.**

**Recommendation #6: A Cohort-Component Demographic forecast model should be developed to complement the housing based forecasts and add age distribution characteristics detail to the projections.**

**Recommendation #7: Use the Cohort-Survival Ratio method, which is the standard in the field of school demography, to project enrollments at the school level rather than the Housing Unit method.**

**Recommendation #8: Eliminate the Top-Down approach in producing projections and solely utilize the Bottom-Up approach when projecting enrollments. This would avoid the arbitrary reconciliation of the two projections.**

**Recommendation #9: Hold Student Generation Factors constant throughout the projection period and consider them to be a "snapshot" in time.**

**Recommendation #10: Compute Student Generation Factors at the elementary, middle, and high school attendance area levels rather than the planning zone, which may be too small in size to provide a reasonable estimate.**

**Recommendation #11 Continue using the Multi-family Student Generation Factor for townhouse-style condos and 2-Over-2s.**

**Recommendation #12: Update the Students Generation Factors by attendance area on an annual basis.**

**Recommendation #13: Project kindergarten students using birth-to-kindergarten survival ratios instead of using the Housing Unit Method.**

**Recommendation #14: Project future birth counts in Prince William County using the Cohort- Component Method, which will aid in projecting kindergarten students in years 5-10 of the projection period.**

**Recommendation #15: Perform an in-depth analysis of the possible reasons for the differences in enrollment for schools with an error rate of greater than 5%.**

**Recommendation #16: Conduct a longitudinal analysis of error rates over time at the school level and by grade configuration (K-5, 6-8, and 9-12)**.

**Recommendation #17: Consider taking into account length of ownership when computing Student Generation Factors as current values may be underestimated.**

**Recommendation #18: Create two sets of enrollments projections, Baseline Projections and Adjusted Projections, which take into account new students from housing developments.**

**Recommendation #19: As a long-term goal, use confidence intervals to provide an enrollment projection range of the K-5, 6-8, 9-12, and K-12 totals for a one-year projection.**

# Introduction

Statistical Forecasting LLC and RLS Demographics Inc. ("Consultants") were hired by the Prince William County Public Schools ("PWCS") to review and evaluate the current methodology used to project enrollments in the school district, as well as the process undertaken to compute future housing units by the Prince William County Government ("PWCG"). In addition to a thorough methodological review, the Consultants provided recommendations for improvements to the forecasting process and an assessment of best practices in housing unit and enrollment projection methodology was completed through a literature review.

To conduct the investigation, a multitude of documents were reviewed, including: enrollment history, enrollment projections, Metropolitan Washington Council of Governments ("MWCOG") forecasting methods and results, PWCG demographic estimates and projections, student generation factors, comparisons of projected to actual enrollments, and Prince William County birth and migration data. Excel files and databases were analyzed as well as output generated from the statistical software package SPSS. Follow-up conversations occurred with Division of Planning Services staff to clarify elements of the enrollment projection methodology. In the following sections, recommendations are made based on the telephone interviews conducted and after examining the aforementioned documents.

*a) Overview*

Prince William County is located in northern Virginia just outside of Washington D.C. The Census Bureau estimates show that the population of Prince William County has grown from 402,000 in the 2010 Census to just over 468,000 as of July 1, 2018. The total growth of more than 66,000 persons is made up of both natural increase (births minus deaths) and net migration, both international and domestic. Natural increase accounts for nearly two-thirds of the growth while the remaining growth is migration. International migration makes up fully 80 percent of the net in-migration total. Population change due to migration and fertility are two important components driving school enrollment forecasts.

Prince William's most rapid period of growth was during the 1960's and 70's when the population more than doubled in each decade. That growth has continued at rates between 30 and 50 percent in each subsequent decade. Growth has been strong for the last two decades as metropolitan Washington has expanded. Prince William is a suburban area that is highly attractive to a young and mobile population and this is reflected in the fact that fully 30 percent of the population is between the ages of 25 and 44.

Following is a review of the methods used by 1) the PWCG in the development of population, housing unit and household estimates and MWCOG forecasts to 2045, and 2) the PWCS in development of enrollment projections.

*b) Housing Unit Projection Methodology*

Data from the Decennial Census is used as the benchmark for housing units and households. This is important for the evaluation of methods because Prince William County and various third party sources of housing data also rely on Census data in compiling their estimates and forecasts.

The number of households is equal to the count of occupied housing units, so estimates of occupancy rates can have a significant impact on estimates of households. A household is simply defined as the people who occupy a housing unit. Each household includes the householder and related and unrelated individuals living in the household. If the household contains members related to the householder, the household is defined as a family household. If there are no individuals related to the householder, unrelated students living together in off-campus apartments for example, the household is a non-family household comprised of unrelated individuals.

In addition to the Decennial Census, there are a number Census Bureau sources that provide data on housing units, households and occupancy including:

- American Community Survey (ACS)
- Current Population Survey (CPS-ASEC)
- American Housing Survey (AHS)
- Housing Vacancy Survey (HVS)
- Census State and County Total Housing Unit Estimates[1]

These benchmark counts and survey estimates of housing units and households are supplemented by a variety of other non-census sources, especially at the local level. This includes local building permits, estimates of non-permitted construction, mobile home shipments and estimates of changes in the housing stock due to demolition and other losses. Each of the Census Bureau survey sources noted above are based on samples of all housing units and require controls to represent the total universe. The control counts for the ACS, AHS and HVS are all based on the current housing unit estimates.

Conceptually, the estimating process is relatively simple. Beginning with the count of housing units from the latest Decennial Census, newly constructed units are added to the base housing stock, estimates of new mobile home placements are added, and estimates of units lost are subtracted.

· New residential construction is estimated by taking the number of building permits issued and applying a completion rate. At the local level it is desirable to use certificates of occupancy, if available, as the completion rate is simply a national estimate of permits that are not completed derived from the Survey of Construction ("SOC").

Estimates of non-permitted construction are developed for areas that do not issue building permits. These estimates are derived from the SOC at the regional level and are applied to the area based on the area share of the region's total housing units.

[1] U.S. Census Bureau, "Methodology for State and County Total Housing Unit Estimates (Vintage 2018):April 1, 2010 to July 1, 2018", https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2010-2018/2018-hu-meth.pdf

New mobile home placements come from the Manufactured Homes Survey ("MHS"). State level shipments of mobile homes are distributed to each place based on the place's share of the state's total mobile homes.

Housing loss is based on regional level data from the American Housing Survey. A matrix of loss rates by region, age, and type of unit is used to estimate local housing losses.

This standard methodology used by the Census Bureau has a number of limitations when used to estimate local housing unit and household change. There are a number of differences between each of the Census surveys including: sampling methods, definitions of occupancy and vacancy status, and time periods. When being used to develop local estimates, the use of national and regional survey estimates requires a significant assumption that each local area is a reflection of the national or regional patterns.[2]

A high-level summary of the forecasting process:

The MWCOG regional and county forecasting program utilizes forecasts of population, housing, and employment developed by IHS Global Insights. These are used as control totals for county forecasts within the region.

Prince William County forecasts of population and housing are reconciled with the MWCOG forecasts (as are other counties within the region) to be within a three percent range of each other. The adopted forecasts at the county level become the inputs to the PWCS enrollment forecasting model.

This is a common process and methodology for the development of county and regional forecasts and is the general process used by large metropolitan regions where regional forecasts are distributed to counties, sub-county areas, census tracts, and traffic analysis zones. The following are just a few examples:

- The Twin Cities Metropolitan Council
- The New York Metropolitan Transportation Planning Council
- The San Diego Association of Governments
- The Miami-Dade Transportation Planning Organization

[2] McCue D., Masnick G., Herbert C., "Assessing Households and Household Growth Estimates with Census Bureau Surveys", Joint Center for Housing Studies, Harvard University, July 2015.

**PWCG Housing Unit Methodology**

As noted above, the housing unit method begins with a baseline count of housing units. Housing units are adjusted for occupancy resulting in the number of households. The method is more detailed as households must be characterized by type for use in the school enrollment model and the application of student generation factors. Updating the baseline count of housing units is where the general methods become more specific to Prince William County.

The Geographic Information Systems Division within the Department of Information Technology (DoIT) is responsible for development of population and housing estimates. These updated estimates are produced quarterly and also forecast to the year 2040. This methodology follows the general, and accepted, process of estimating housing units by type. This process adjusts for occupancy to generate households and applies an estimate of the average household size to generate the household population. The total group quarters population is added to the household population resulting in the total resident population.

The Department maintains the Premise Address Database, which includes all residential addresses in the county by type. This database is updated on a quarterly basis with new housing occupancy permits entered into the Development Management System ("DMS"). The DMS is an application that supports the County's development application and monitoring process. Demolitions are not included in the method as the county estimates that there are few demolitions each year and have a negligible impact on the estimates.

**Current Estimates**

Housing occupancy rates are derived from the Census Bureau's American Community Survey at the county and PUMA level and the Decennial Census at the census block level. The ACS allows for annual updating and is specific to unit type. The data used comes from the Public Use Microdata Sample 5-year data. This is an appropriate source and the 5-year data is preferred over the 1-year data due to the increased sample size, smaller margin of error and greater stability. However, the sample variability on these estimates can be quite high and there is no provision of measuring the effect of wide confidence intervals on the estimates.

Average household size is derived from the same source and carries the same benefits of characteristic detail versus the effect of sampling error measurements. The margins of error on the ACS data is estimated at the 90 percent confidence level.

Population is derived by applying the average household size to the estimate of households and including the current estimate/count of group quarters persons. Data on group quarters population is obtained by the Planning Department through annual and biannual reviews of existing facilities and tracking special use permits for new facilities.

PWCG should consider the development of comprehensive methodology documentation. There are general methods statements and presentations that have been provided by PWCG staff and on the website, but it is difficult to get a single, detailed description of the methods employed. The county's population estimates and application of occupancy rates and average household estimates is quite detailed and staff has in-depth knowledge of the data sources and interaction. Transparency of their data and methods would be improved by system documentation that identifies the process, data sources, and assumptions.

It is understood that the handling of current and potential (pipeline) development is rarely a certainty. Housing needs and development opportunities and resources can change, which affect the future projections of residential housing. However, this seems to be an undocumented and ad-hoc process of discussion between county planners and the development community and is driven by expectations rather than hard data. While outside of the County's scope and this project, this is also true for the IHS Global Insight forecasting process.

Housing occupancy rates and average household size for the County are based on the 2010 Census at the census block level and current ACS microdata for the county's three PUMA's. This continual update process is important and necessary. However, margins of error in the ACS data can be large for small areas and should be evaluated to determine the confidence interval surrounding occupancy rates and the potential impact on the estimates.

**Recommendation #1: PWCG should consider the development of comprehensive methodology documentation. It is difficult to get a single, detailed description of the methods employed.**

**Recommendation #2: Use of current American Community Survey (ACS) microdata for the county's three PUMA's to update is important and necessary. However, margins of error in the ACS estimates can be large and should be evaluated.**

**Demographic Analysis and Forecasts**

Forecasts of future household change are obviously dependent upon assumptions about the course of future events. While these assumptions are guided by analysis of historical trends, housing is dependent on a variety of factors including economic trends, migration, land use, zoning, and others. The County's Comprehensive Plan is an important guide for future development and addresses meeting the housing needs of future residents.

An annual "Build-Out" analysis includes assessment of current zoning, long-range land use of undeveloped territory and potential redevelopment. Analysis of permitting activity provides a short-term view of development but future development requires assessment of current and proposed zoning changes, special use permits and Comprehensive Plan amendments. These help guide development consistent with the community's desired growth patterns.

These planning concepts come together with the county and regional forecasts through the fitting of a logistic curve equation to the existing historical trend of household growth. The regional forecasts provide an adopted limit to growth that then controls while maintaining consistency with the historical data.

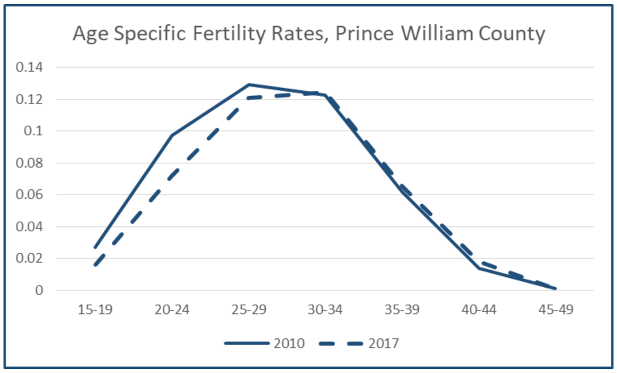
As noted earlier, the housing-based forecasting method used by PWCG is quite appropriate and is carried out with high quality housing and development tracking systems. It is ultimately guided by the County's Comprehensive Plan, which addresses development potential and limits to growth. However, this method cannot generate population characteristics such as age and sex which are important components of population analysis and have direct bearing on school enrollment forecasts.

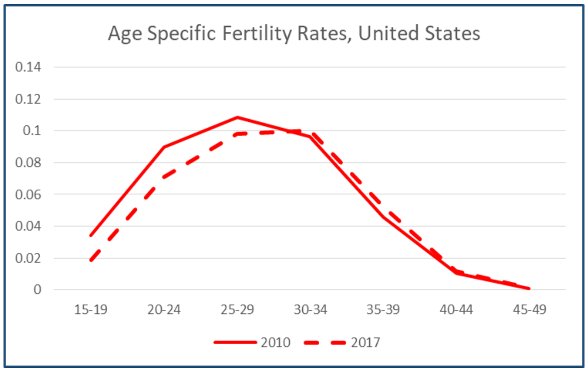
Of particular concern here is the impact of migration on characteristics of the population and changing levels of fertility. The housing development data provides an important guide related to residential and commercial capacity but it does not address the changing composition of the population – particularly age characteristics. Population change in any area is a function of fertility, mortality, and migration.

The strength of the demographic cohort-component method is that it captures each of these processes and allows for the "aging" and migration of the population. Mortality is generally the least important process as survival rates at each age are relatively stable. Longevity continues to improve but it tends to not have immediate impacts.

Fertility

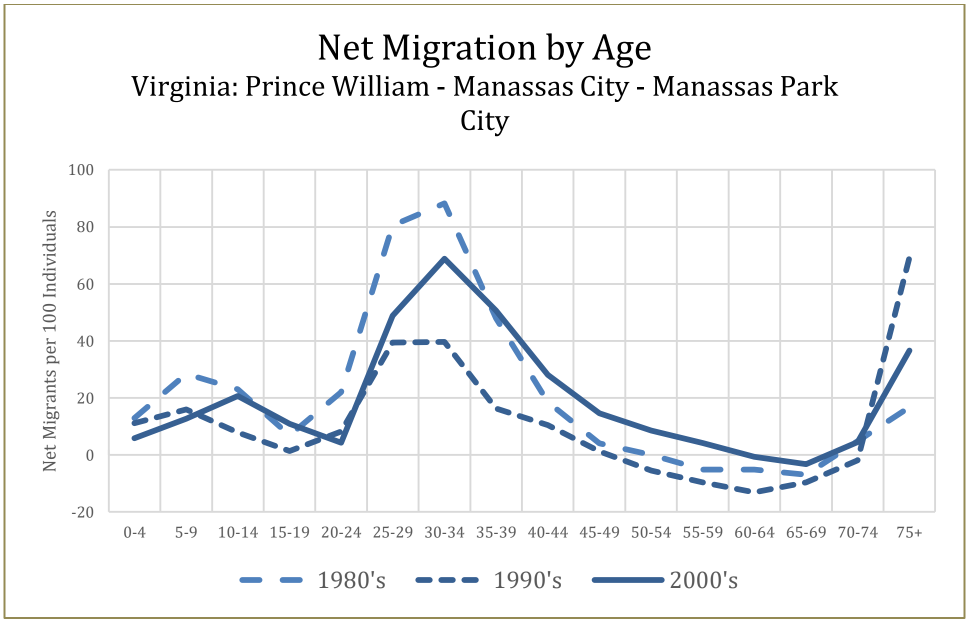
Fertility is more important, especially considering its link to school enrollment. Fertility rates in the U.S. have been at very low levels for decades and continue to fall. The timing of fertility among women of childbearing age has been delayed such that the peak years are in their late 20s and early 30s. Prince William County has a fertility rate higher than the nation at about 2.1 children per woman – right at the replacement level of 2.1, but this rate has also declined in the last decade. Fertility in Prince William County is higher in the younger ages with peak fertility occurring earlier than in the nation and can have a greater impact on future births and enrollment.




Migration

Migration on the other hand is the most volatile component of population change. One only needs to look at the recent recessionary period to understand the impact of economic crises on migration. Migration rates plummeted in states of high in-migration and moderated in states of high out-migration. The figure below illustrates the importance of demographic methods and the impact of migration in Prince William County.




*Winkler, Richelle, Kenneth M. Johnson, Cheng Cheng, Jim Beaudoin, Paul R. Voss, and Katherine J. Curtis. Age-Specific Net Migration Estimates for US Counties, 1950-2010. Applied Population Laboratory, University of Wisconsin – Madison, 2013. Web.*

This migration pattern by age shows that Prince William County clearly attracts population through their 20s and 30s with peak attraction occurring in the late 20s and early 30s. It also shows that in-migration drops sharply through their 40s and continues at a slight negative level throughout the older ages. A similar pattern is seen in the 1990 to 2000 period. High in-migration of the young population is characteristic of many urban areas attracting young workers. However, in many areas this population then moves out when they start to form families and look for alternative housing arrangements.

The age composition of the population is an important driver of housing needs and development and school enrollment, yet this is an area that receives limited attention in the estimates methodology and is not incorporated in the forecast process at all. This initial review of migration by age indicates that more detailed analysis is necessary to understand both the migration and aging effects in Prince William County. Prince William's high proportion of population in the 25 to 44 year age categories warrants detailed analysis of aging in the county – particularly as it affects women of childbearing age.

Analysis of the American Community Survey, Public Use Microdata Samples can identify migration flows (both in- and out-migrants) and the demographic characteristics of migrants. Prince William County is defined by three Public Use Microdata Areas and this would allow for some additional geographic detail within the county. The PWCG uses these data for monitoring occupancy and household size and should expand their analysis to better understand migration streams and characteristics.

Migration analysis should also include development of decade-to-decade age-specific migration patterns as illustrated above and incorporating group quarters adjustments to account for special populations. This analysis is a basic component of the demographic cohort-component projection method discussed below.

The fertility rate data presented above is a snapshot of fertility experience in 2010 and 2017 based on birth data from the National Center for Health Statistics. More detailed historical data is likely available from Prince William County vital records office and/or the Virginia Department of Health and should be analyzed by single year of age of mother and for a longer historical period. Third party access to these more detailed data is often restricted by state and county health departments. It is also desirable to base fertility rate calculations on a 3-year average of births by age centered on the Decennial Census dates to reduce the effect of year-to-year variation.

**Recommendation #3: The age composition of the population is an important driver of housing needs, development and school enrollment and should be incorporated in the estimates and projections process.**

**Recommendation #4: Analysis of the ACS migration can be used to better understand migration flows and characteristics. This, along with on-going analysis of age-specific migration patterns will update inputs to the demographic projections**

**Recommendation #5: Age-specific fertility patterns by age of mother should be analyzed on an annual basis to understand changes in fertility levels and patterns.**

**Development of Cohort-Component Demographic Forecasts**

As noted below in the discussion of the use of birth-to-kindergarten ratios, the PWCG and PWCPS do not attempt to explicitly forecast future births based on the demographic composition of the population. A demographic model accounts for the interaction of migration and fertility in the projection of women of childbearing age. Adopting such a method will improve the projection of future births and student enrollment.

Analysis of aging, migration, and fertility points to the need for adoption of a demographic cohort-component method to be integrated with the existing housing based model. As noted earlier, strict demographic estimating methods have no explicit controls for housing capacity, which the county's housing method incorporates through data on land use and development plans. These are not accounted for in a demographic method. However, the housing method cannot be used to generate age distributions and it can be seen from the migration patterns that Prince William County's population, and age composition, is greatly influenced by migration. Integrating the fertility component is critical to forecasting births and future school enrollment.

While not a requirement of the MWCOG Regional Cooperative Forecast process, the integration of an economic demographic model with the current housing/commercial space forecast should be considered. The development of a demographic forecast would be valuable as a supplement and check on the housing unit methodology. The County's housing data is a valuable resource and the methods used in calculation are appropriate to the task. Integration of a demographic method provides another measure that captures demographic change, particularly related to fertility and migration. Such integration can provide a "control" process that accounts for housing and land development and constraints in the Comprehensive Plan, employment growth by economic sector, and implied population change. Employment, labor force, and population change all impact commuting patterns, which are also driven by workplace employment. Currently, capacities in housing and commercial space are the controls but there is no resulting analysis of those impacts on population by age and sex.

**Recommendation #6: A Cohort-Component Demographic forecast model should be developed to complement the housing based forecasts and add age distribution characteristics detail to the projections.**

*c) PWCS Enrollment Projection Methodology*

**Introduction**

The PWCS Planning and Financial Services Administration ("Planning") is responsible for producing the enrollment projections for the school district. In the last 20 years, PWCS has experienced significant growth. In 2018-19, there were 90,203 students, which is a gain of more than 38,000 students from the 1998-99 enrollment. As a result, 37 new schools were opened from 2000-2018. To account for the rapid growth in its student population, Planning has modified its enrollment projection methods over the last two decades. In 1997, Planning used the Cohort-Survival Ratio method, which progresses students from one grade to the next based on historical patterns (this will be discussed later in the report in greater detail). Over time, due to significant housing growth in the County, there was a methodological shift to include new housing starts in the projections. By 2003, Planning created planning zones, which are small areas that are coterminous with all elementary, middle, and high school boundaries with a target size of 100-150 students in each zone. From 2003-2008, Planning's methodology became more of a hybrid in nature, as it used elements of the Cohort-Survival Ratio Method as well as taking into account new housing units being constructed in the County. Refinement of this process continues to occur in the present time.

**Overview**

To project enrollments, Planning employs the Housing Unit method. In this method, the total number of housing units in the County is multiplied by the number of children per housing unit (student generation factor) to determine future enrollments. Planning produces two sets of projections using both a "Top-Down" approach and "Bottom-Up" approach. In the Top-Down projections, the Housing Unit method is employed on the division-wide grade-level counts and does not consider lower levels of geography, such as enrollments at the school building level.

After the Top-Down projections are completed, Planning uses the Housing Unit method to project enrollments at the planning zone level, whereby enrollments are aggregated to determining the division-wide enrollments. When the Bottom-Up projections are aggregated to the division-wide total, they will not match the total from the Top-Down approach since the geographical unit of analysis is different.

Planning uses the Top-Down approach to initially forecast the division-wide ten-year grade-level projections, rather than the more labor-intensive Bottom-Up approach, which can take up to three weeks to complete. The Bottom-Up projections are completed shortly after the Top-Down projections to produce student counts at the school level by aggregating the projections from the planning zones. Since the aggregate of the projections from the Bottom-Up approach does not match the Top-Down division-wide projections, Planning reconciles the differences between the two projections so that the enrollments of both approaches match identically.

To complete the projections, Planning uses Geographic Information System ("GIS") software, ArcGIS, and the statistical software package SPSS. Beginning in 2001, Planning wrote syntax for SPSS so that the process could be replicated year after year, as well as writing a detailed overview of the formulas used in SPSS and the outputs they produce. In addition, Planning has documented, for quality assurance purposes, the process of geocoding students, as well as bringing in new housing unit information from PWCG. PWCS is to be commended for the steps it takes to ensure replication of the process it has created for future enrollment projections, as staff may change over time.

**Top-Down Approach**

To compute enrollments using the Top-Down approach, PWCS maintains a database with the actual number of housing units in the County since 1990. An estimate of the number of housing units is computed for the next ten years based on estimates obtained from the PWC Finance Department's Revenue Committee Reports. In conjunction with the housing unit data, PWCS maintains a database of division-wide historical enrollment data by grade level so that student generation factors ("SGF") can be computed for each grade (K-12, including Special Education), as well as for elementary (K-5), middle (6-8), high (9-12), and division-wide (K-12). In its simplest form, SGFs are computed by dividing the number of children by the number of housing units at the unit of analysis desired (grade, configuration level, etc.). SGFs have been computed by PWCS from 1990-2018 to show the changes over time. To estimate the number of future students, the projected number of housing units is multiplied by the SGF in each individual grade. The SGFs used in the projection years are computed by multiplying the SGF in the previous year and previous grade by a weighted three-year average survival ratio (based on the SGFs historical progression patterns) and an Adjustment Factor Matrix ("Matrix")**.** The values in the matrix, which are by grade and projection year, are not the product of a formula and are either slightly greater or less than 1. The values are subjectively used to increase or decrease the projected number of students based on previous patterns in the school district. As an example, to compute the SGF for grade 5 in 2019-20, the SGF for grade 4 in 2018-19 is utilized, with the assumption that the SGF advances with the cohort from $4^{th}$ to $5^{th}$ grade, and is multiplied by a weighted three-year average survival ratio and then by an adjustment factor in the Matrix. In short, the values of the SGFs are changing throughout the ten-year projection period. The number of students computed at each grade is then aggregated to derive the division-wide enrollment for each projection year.

**Bottom-Up Approach**

In the "Bottom-Up" approach, a similar process is undertaken by PWCS to project enrollments. Using GIS software, residential properties at the parcel level in Prince William County are spatially joined to the geocoded student point file containing PWCS student data. Student information such as grade level, school attending, and residing attendance area are included in the database, as well as student demographic characteristics (race, free or reduced lunch status, Limited English Proficiency, etc.). By joining the PWCS student database to the County property database, SGFs can be computed by housing type at the planning zone level. In 2018-19, there were 988 planning zones in the school district. Planning zones can then be aggregated to comprise a school attendance area.

PWCS computes its SGFs at the planning zone level by the three types of housing classifications provided by the PWCG: single-family, townhouse, and multi-family. In a separate research study, PWCS analyzed whether townhouse-style condos and 2-Over-2s, which were historically categorized for enrollment projection purposes as townhouses, had student yields similar to those of townhouses or those of multi-family units. The results showed that the student yields from the townhouse-style condos and 2-Over-2s were similar to those of multi-family units but not of townhouses. While these types of units are classified as townhouse units, PWCS uses the multi-family classification for townhouse-style condos and 2-Over-2s. In 2015-16, the difference between the townhouse SGF (0.572) and multi-family SGF (0.334) was fairly significant, which could lead to overestimation of future enrollments if the townhouse SGF were to be used for townhouse-style condos and 2-Over-2s.

In preparation of projecting enrollments at the planning zone level, PWCS builds future housing into its model. PWCS uses a housing pipeline database of approved developments and collaborates with the PWCG in August or September to ensure accuracy. Information in the database includes the name of the development, the number of total approved units by housing unit type (e.g., single-family detached, townhouse, and multi-family), number of built units and the number remaining to be built, the elementary, middle, and high school attendance areas, and projected number of students upon build-out for each grade configuration. Developments are also categorized by their status: active, planned, or rezoning submitted[3]. "Active" is where rezoning has been approved by the Board of County Supervisors ("County Supervisors") and a site development permit has been issued. "Planned" is where rezoning is approved by the County Supervisors and proposed plans are submitted. "Rezoning submitted" is when a rezoning application has been submitted or no recent progress has occurred. Using the build-out rate (the number of units constructed over a period of time) by building type from the past year, the build-out of each projected development is extrapolated forward by the ten geographic areas that comprise the school district (e.g., Cardinal, Cherry Hill, Ferlazzo, etc.). The sum of the total number of housing units projected in each year must agree with the projected housing units per year used in the Top-Down approach. Housing units are then aggregated at the planning zone level by type.

To project enrollments at the planning zone level, the following steps are taken:

1. Students in each grade are multiplied by a transfer rate, which is based on historical transfer rates in the school district. In most instances, the transfer rate follows the grade cohort as it progresses through the district. However, the transfer rates are changed slightly during the high school years.

2. The number of students from each new housing development is computed by multiplying the SGF by housing type and grade level by the number of new

[3] In the Bottom-Up approach, these classifications are further disaggregated into Stages, which are not discussed here.

housing units by type. Transfer rates by grade are also applied to these cohorts as described above.

3. Students are aggregated by grade to determine the number of students at the planning zone level.

Students are then aggregated to determine the projected counts in each school's attendance area, which are defined by the planning zones that are located in the attendance area. Enrollment can be aggregated further to determine the division-wide projections.

It should be noted that PWCS does not use the Cohort-Survival Ratio ("CSR") method to project enrollments, which is also known as the Grade Progression Ratio method. CSR is the preferred method to project enrollments by school demographers across the country. This is discussed in further detailed in the Appendix, as an assessment of best practices in housing unit and enrollment projection methodology was completed through a literature review.

In essence, CSR looks at historical grade-to-grade progression ratios of students and assumes the historical progression ratios will continue into the future, which essentially provides a linear projection of the population. In this method, a survival ratio is computed for each grade, which essentially compares the number of students in a particular grade to the number of students in the previous grade during the previous year. The survival ratio indicates whether the enrollment is stable, increasing, or decreasing. A survival ratio of 1.00 indicates stable enrollment, less than 1.00 indicates declining enrollment, and greater than 1.00 indicates increasing enrollment. If, for example, a school district had 100 fourth graders and the next year only had 95 fifth graders, the survival ratio would be 0.95.

The CSR method assumes that what happened in the recent past will also happen in the future. The CSR method is most appropriate for districts that have relatively stable trends without any major unpredictable fluctuations from year to year. In school districts encountering rapid growth not experienced historically (i.e., a change in the historical trend), the CSR method must be modified and supplemented with additional information (e.g., additional children from new housing starts). Due to the fluctuation in survival ratios from year to year, it is appropriate to calculate an average survival ratio, which is then used to calculate future grade enrollments.

When projecting enrollments for small schools (200 or fewer students and grade sizes less than 30-35 students), districts utilize the Grade Progression Differences ("GPD") method. In the GPD method, the change in the number of students, as opposed to the ratio, is computed for each grade progression from one year to the next. A positive value indicates an inward migration of students while a negative value indicates an outward migration of students. Differences are computed over a number of historical years (up to five years like CSR) and averaged to project grade-level enrollments.

As compared to a ratio, a numerical change is less sensitive to the movement inward or outward of a few students and is preferred when grade sizes are small. In the CSR method, small grade cohorts can lead to greater fluctuation of the survival ratios with the entering or exiting of just a few students. To prevent this, GPD should be used when cohort sizes are less than 30 students.

PWCS should consider using CSR to project enrollments, instead of the Housing Unit method, as CSR is the standard in the field of school demography. CSR would be used for projecting enrollments at the school attendance area level rather than the planning zone, which is too small a unit of analysis to produce reliable results. In addition, it is difficult for PWCS to base its enrollment projections on the accuracy of projected housing units from PWCG, which itself is a projection, and housing pipeline data that or may or may not come to fruition.

Loosely speaking, PWCS uses the terminology of grade progression ratios for computing SGFs for each grade, as each grade level has a unique SGF. PWCS advances SGFs, not students, from one grade to the next and applies a weighted three-year grade progression ratio. However, in the strict definition of the term, grade progression ratios relate to the advancement of students from one grade to the next, not SGFs, which is a computed ratio (students/housing unit). PWCS should be commended in its efforts to create a model that incorporates the significant number of new houses and children resulting from the growth that the County has experienced in the last 20 years. However, instead of advancing SGFs through the grade levels, PWCS should use the most current SGFs in each grade throughout the enrollment projection period and consider the values a “snapshot” in time, which is the common practice in the field. The SGFs can then be updated annually during the enrollment projection process to capture the most recent values for each housing type.

In addition, rather than computing SGFs at the planning zone level, which are small and may not have a significant number of housing units of each type, PWCS should consider computing SGFs at a larger level of geography such as by school attendance area (elementary, middle, and high school zones) and use these values to estimate future children from new housing units. PWCS currently uses the countywide SGF in situations where there are not enough students or housing units of a particular type in a planning zone. This situation can be avoided if PWCS expands its SGF calculation to the school attendance area. Therefore, while PWCS uses the terminology of grade progression ratios, it does not employ the calculation and procedures as employed in the field of school demography in projecting students.

While the Top-Down projections have some utility for internal use, the Bottom-Up projections are the values that should be utilized to project the district and school-level enrollments. From a methodological standpoint, both approaches have their merits. The Top-Down approach utilizes larger student counts, which produces less variable SGFs due to the larger sample size. On the other hand, the Bottom-Up projections are the most viable, since they provide a much richer analysis of information at the school level. Bottom-Up projections consider patterns that may be unique to specific areas of the County, which are masked when using the countywide Top-Down approach. In addition, reconciling the two projections is arbitrary and lowers the integrity of the enrollment projection process.

**Recommendation #7: Use the Cohort-Survival Ratio method, which is the standard in the field of school demography, to project enrollments at the school level rather than the Housing Unit method.**

**Recommendation #8: Eliminate the Top-Down approach in producing projections and solely utilize the Bottom-Up approach when projecting enrollments. This would avoid the arbitrary reconciliation of the two projections.**

**Recommendation #9: Hold Student Generation Factors constant throughout the projection period and consider them to be a "snapshot" in time.**

**Recommendation #10: Compute Student Generation Factors at the elementary, middle, and high school attendance area levels rather than the planning zone, which may be too small in size to provide a reasonable estimate.**

**Recommendation #11 Continue using the Multi-family Student Generation Factor for townhouse-style condos and 2-Over-2s.**

**Recommendation #12: Update the Students Generation Factors by attendance area on an annual basis.**

**Live Births and Kindergarten Projections**

PWCS obtains Prince William County live birth data from the Virginia Department of Health, which tabulates the number of births to women who reside in Prince William County. Surrounding counties are observed by Planning as well (Loudoun, Arlington, Stafford, Fairfax, and Fauquier). Using countywide birth data lagged five years behind its kindergarten enrollment, birth-to-kindergarten survival ratios are computed. For instance, in the 2012-13 school year there were 6,873 births, which is tabulated from October 1 to September 30, as children may attend kindergarten in PWCS if they are five by September 30$^{th}$. Five years later in the 2017-18 year, there were 6,435 kindergarten students, which is a birth-to-kindergarten survival ratio of 0.936. Unlike most school districts, PWCS does not use an average birth-to-kindergarten ratio to project future kindergarten students. In many larger districts with pre-school programs, it is also used to project the number of pre-kindergarten students either three or four years later. Typically, school districts use an average birth-to-kindergarten ratio throughout the entire projection period to project future kindergarten students. Instead, Planning uses the Housing Unit method, as described previously, to project future kindergarten students. PWCS does compute the future birth-to-kindergarten ratios based on its projected kindergarten enrollment and compares the ratios to historical values. Based on my observations, the future birth-to-kindergarten ratios computed using the Housing Unit method are higher than historical birth-to-kindergarten ratios. While births are not being explicitly used in the PWCS enrollment projections, Planning does track the data so birth-rate patterns can be qualitatively considered.

PWCS has performed a research study internally indicating that births are not a great predictor of kindergarten students five years later at the planning zone level. In that study, the dependent variable, kindergarten (at time t), was regressed on the independent variables, births five years prior (t-5) and the previous year's kindergarten cohort (t-1). There was considerably more variance in the kindergarten cohort (t) explained by the variance in the previous year's kindergarten cohort (t-1) than by the variance in births five years prior (t-5) at the planning zone level. As this is a very small level of geography and may not generalize to larger levels, future research at larger scales (e.g., attendance area, district-wide, etc.) is necessary before any significant conclusions can be drawn.

In the consultant's opinion, there is some merit in finding an alternative method to project kindergarten students. Of all of the survival ratios computed (birth-to-kindergarten, K to 1, 1 to

2, etc.) in the CSR method, the birth-to-kindergarten ratio has the greatest variability since there is a five-year time lag between a child being born and when the child enters kindergarten. In that five-year period, a number of things may happen. A child born in Prince William County may attend PWCS five years later, or may attend a private or parochial school, or may move out of the school district's attendance area. On the flip side, children born in other locations other than Prince William County may move into the area and attend PWCS. For these reasons, projecting future kindergarten counts is often difficult due to the mobility of the population under the age of 5. By using the housing unit method, PWCS is attempting to control for these factors as the number of future houses in the County (and planning zone in the Bottom-Up approach) defines the projected number of kindergarten students.

Despite not using births to project kindergarten students, PWCS should consider doing so to be in line with the industry's standard, as PWCS' research on its effectiveness was limited in scope (planning zones) and may not generalize to larger geographical areas. To project future births, which would assist in projecting kindergarten students in years 5-10 of the projection period, PWCG can assist PWCS by projecting the number of women of childbearing ages (15-49) for five-year intervals (15-19, 20-24, etc.) for 2020 and 2025 using the Cohort-Component method that was described previously. Births can then be projected by multiplying the age-specific fertility rates (based on historical birth data) by the number of women in each age class. Kindergarten counts in years 5-10 can then be computed by using average birth-to-kindergarten ratios that were used to project kindergarten students in years 1-4.

In addition, PWCS could geocode historical births in Prince William County so that birth counts can be aggregated by each elementary attendance area. Historical birth-to-kindergarten survival ratios can then be computed using birth counts and kindergarten students unique to each elementary attendance area. An average or weighted-average birth-to-kindergarten survival ratio would be used to project future kindergarten enrollments.

**Recommendation #13: Project kindergarten students using birth-to-kindergarten survival ratios instead of using the Housing Unit Method.**

**Recommendation #14: Project future birth counts in Prince William County using the Cohort- Component Method, which will aid in projecting kindergarten students in years 5-10 of the projection period.**

**Enrollment Projection Accuracy**

Planning currently assesses the accuracy of its projections by computing the numerical and percent differences both division-wide and by school building for a one-year period. Division-wide error rates are also computed for two, four, and five years prior to the year being analyzed. For instance, if the projections were computed in 2014-15 for the 2018-19 school year, the error rates computed would be for the projections completed four years prior to the year being analyzed. Mean Percent Errors ("MPE") and Mean Absolute Percent Errors ("MAPE") were also computed for the one, two, four and five years before the projected year to determine the effectiveness of the projections over time. The MPEs compute the average of the percent errors, whereby positive and negative error rates (under-projecting and over-projecting respectively) can cancel each other out, in effect masking the error rate. To avoid that, absolute values are taking of the error rates and then averaged to compute a MAPE.

In a survey of educational planners who complete enrollment projections that was conducted by Schellenberg and Stephens, two-thirds believe that an error rate of 1% per year for the total enrollment is acceptable.[4] For a ten-year projection, this would mean that a 10% error rate in the tenth year would be acceptable. At the school level, half of the respondents thought an error rate of 3-5% was acceptable in a one-year projection while approximately two-thirds of the respondents aimed for grade-level error rates of less than 4%.

In relation to these industry accuracy expectations, Planning aims for a 99% accuracy rate, i.e., a 1% error rate for its one-year projections, which is comparable to the expectation from the Schellenberg and Stephens survey. In the last nine years, PWCS has hit its target accuracy rate of 99%. At the school level, Planning computes the percentage of schools with a 5% error rate, which is at the upper bound of industry expectations for school-level enrollment projection accuracy. Planning aims for a target rate of 80% of its schools having less than a 5% error rate. In the last five years, the school-level error rates have been below the target rate of 80%, as they have ranged from 67.0%-79.8%. While PWCS does perform a numerical analysis of percent errors, it does not perform an in-depth review of the reasons why the error rates exceeded 5% in those respective schools. It is recommended that PWCS perform an in-depth analysis of the possible reasons for the differences in enrollment for schools with an error rate of greater than 5%.

In addition, Planning performed an in-depth longitudinal error-rate analysis in 2015. In this study, the five-year projections that were computed in the 2009-10 school year for 2010-11 through 2014-15 were compared to actual enrollments. Numerical errors were computed (observed-projected) at the grade level and for the K-5, 6-8, and 9-12 totals for each year. To avoid canceling the positive and negative differences in the errors when the error values are aggregated, the absolute values of all the values were computed. The absolute mean error (a numerical value, not a percentage) was then computed for each grade and plotted to demonstrate the grade levels that had the greatest error rates over time. Further analysis was also conducted at the school level by plotting error rates as well as noting the schools that had undergone boundary changes in the five-year projection period that limited the effectiveness of the projection. Planning is to be commended for the detailed analysis that was performed in 2015 as it can be used to help shape decision-making in its future enrollment projections. In the future, Planning should re-institute its in-depth longitudinal error analysis as part of the process of refining its enrollment projection methodology.

**Recommendation #15: Perform an in-depth analysis of the possible reasons for the differences in enrollment for schools with an error rate of greater than 5%.**

**Recommendation #16: Conduct a longitudinal analysis of error rates over time at the school level and by grade configuration (K-5, 6-8, and 9-12)**.

---

[4] Schellenberg, S. J., & Stephens, C. E. (1987). Enrollment projection: variations on a theme. Paper presented at the Annual Meeting of the American Educational Research Association, Washington D.C., (ERIC Document Reproduction Service No. ED 283 879)

**Use of Projections**

PWCS compares the enrollment projections by school to a building's program capacity, where the difference is computed. Positive values indicate a surplus in seating and negative values indicate inadequate seating. Finally, the percent utilization is computed, which is the ratio of the projected number of students to building capacity. Values exceeding 100.0% reflect buildings that are over-capacity and help drive decisions on future building additions, new schools, or using portable classrooms.

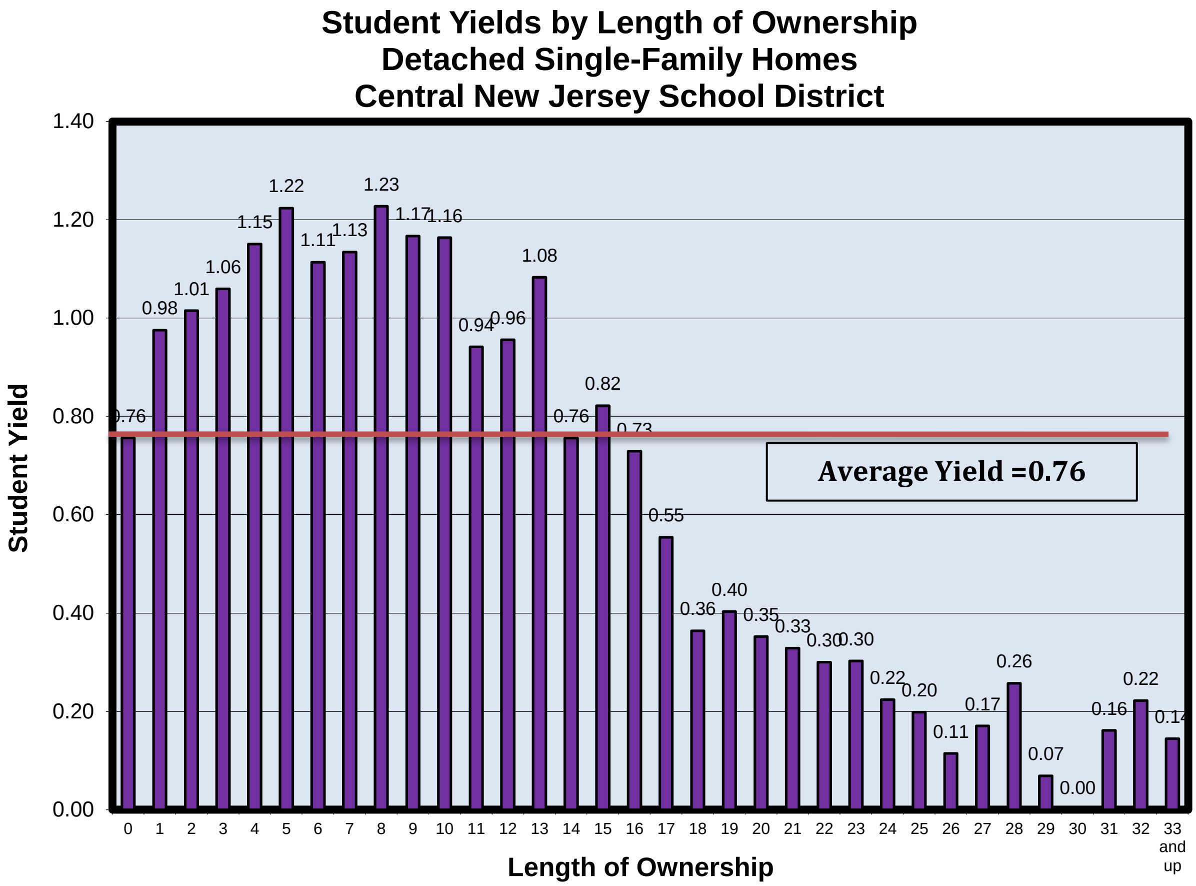
**Length of Ownership**

PWCS may also want to consider length of ownership in computing SGFs, only for housing units that are owned, as research has shown that student yields vary with length of ownership. Apartments would be excluded since it is difficult to obtain length of residency data for tenants. Currently, PWCS computes the SGF by dividing the number of students in a particular housing unit type (single-family, townhouse, and multi-family) by the number of housing units of that type. The SGFs therefore include homes owned by all age segments of the population, which lowers the overall student yield.

To compute student yields by length of ownership, it is necessary to know the year of the most recent sale, which may be available from the County Assessor's office. "Paper sales," which are sales between members of the immediate family for a low price (e.g., $1 or $100) and result in a change in title but often not a change of the occupant, need to be excluded and the next most recent sale date used instead. If a home had never been sold, the length of ownership could be computed by simply subtracting the year that the home was built from the current year.

An example of student yields by length of ownership for detached single-family homes is shown on the following page for a community in New Jersey. Long-held homes (20 or more years) will have fewer children, as they would have graduated from the district. Typically, yields gradually increase with length of ownership, peaking at around 8-10 years of ownership. Student yields at these lengths of ownership are higher than when considering average yields for the entire population. Communities with excellent school districts often have new homes that are purchased with greater frequency by families with children. Computing the average over the entire length of ownership period underestimates the number of children, since there are so few children at longer lengths of ownership as children graduate from the school district. Unfortunately, there is no research-based metric to determine what part of the distribution should be used to estimate future schoolchildren. One possible estimate is to compute an average using all of the years up to the <u>peak</u> student yield (1.23 students at eight years of ownership in the following example), which estimates the maximum impact before student yields begin to decline. In the following example, if an average student yield is computed for the first eight years of ownership when the peak student yield occurs, the yield increases to 1.07. This is likely a better estimate of the student yield of detached single-family homes than the average (0.76) using all years of ownership.

PWCS has performed exploratory research on computing SGFs by length of ownership. In its study, SGFs were computed by length of ownership for single-family homes and townhouses by occupancy type (owner-occupied vs. renter-occupied). With the exception of renter-occupied townhouses, the shapes of the distribution in the PWCS study were fairly similar to the New Jersey distribution. The findings suggest that there is utility in controlling for length of ownership in PWCS enrollment forecasting. PWCS should continue to explore using SGFs by length of ownership, provided that parcel-level sales data is available from the County, and eventually implement these values into its enrollment projection model. To be consistent with the recommendations provided above, SGFs by length of ownership should be computed at the attendance area level (elementary, middle, and high school).




**Recommendation #17: Consider taking into account length of ownership when computing Student Generation Factors as current values may be underestimated.**

**Implementing New Housing with CSR**

If PWCS chooses to employ a Bottom-Up approach using CSR at the school-level to project enrollments going forward, it will still need to account for new housing that has been approved in the County.

When determining the impact of future new housing on enrollment, some school planners assume that the number of future students should be directly added into the enrollment projections. However, the enrollment projections utilize survival ratios that do take into account prior new home construction growth. Children who move into new homes during the historical period are captured by the survival ratios. Therefore, it is not appropriate to add all of the new children generated from new housing units without considering the historical period, as double counting would occur, since the survival ratios have already increased, due to the new children. For instance, if PWCS computes survival ratios based on the last five years (2014-2018), the number of children from newly constructed homes needs to be identified, which can easily be done through its enrollment management system. It is these new units that have affected the survival ratios when additional children enroll in the district. If PWCS uses survival ratios based on only the last three years, it should only consider new housing in that historical period. Next, using the SGFs computed by PWCS, the number of students estimated in the next ten years (the enrollment projection period) can be computed. The *difference* in the number of students from the projected and historical period should be added into the projections, where the number of students is then spread out evenly across grade levels. To provide a simple example, if 500 children entered a school in the last five years from new housing units, the survival ratios would have captured these students entering the school and would have increased accordingly. If 700 detached single-family homes are planned in the next ten years and are estimated to produce 700 new children using the SGFs, only the difference, 200 children, should be added to the projections since the ratios have already increased due to the children from new housing.

PWCS should continue to use its build-out rate of proposed new housing developments to estimate timeline of occupation. In general, estimating the timeline of occupancy is difficult and somewhat arbitrary. Developments may take several years to build, where the number of new students would need to be spread out in the projections for perhaps a two- or three-year period. In addition, some developments that have received approval may never be built due to changing market conditions or lack of financial resources. Due to the uncertainty of the timeline of occupation and whether some residential developments will ever be constructed, PWCS may want to consider performing two sets of projections: a baseline (“Baseline”) set of projections and an adjusted (“Adjusted”) set of projections due to housing growth, which may be considered a worst case scenario if all the proposed housing comes to fruition. The Baseline projection would essentially assume that future housing growth in a school attendance area would be similar to that which occurred in the historical period used to compute the survival ratios. It does not necessarily mean that there is zero growth. The Adjusted projections allow for additional modifications of enrollments in light of housing development that is greater than recently experienced.

**Recommendation #18: Create two sets of enrollments projections, Baseline Projections and Adjusted Projections, which take into account new students from housing developments.**

**Enrollment Projection Range**

Rather than giving a single number in an enrollment projection, some districts/stakeholders wish to be given a range, with the understanding that projections are not an exact science. There is limited research available on this topic. In 2019, I published a paper[5] that used two separate methods (confidence intervals and stochastic forecasting) to provide lower and upper bounds of enrollment projections for three school districts of varying size (small, medium, and large). PWCS would be considered a large school district in the context of the study. Confidence intervals were computed districtwide (K-12) and by grade configuration (K-5, 6-8, 9-12). While there was limited utility in using stochastic forecasting to provide enrollment ranges, confidence intervals were fairly effective in providing a range for larger districts, particularly in the short-term (1-2 years).

As a long-term goal, PWCS could construct confidence intervals for its one-year or two-year projections, which would provide an enrollment range for stakeholders. The confidence intervals would be performed for the division-wide projections, as well as by grade configuration (K-5, 6-8, 9-12). Confidence intervals should not performed at the individual school level as the upper and lower bounds will be too wide to provide a meaningful range. At the division-wide level, it is not expected that the confidence intervals will be narrow enough to be used for staffing. However, it may be helpful in making decisions for future school additions and new school buildings.

**Recommendation #19: As a long-term goal, use confidence intervals to provide an enrollment projection range of the K-5, 6-8, 9-12, and K-12 totals for a one-year projection.**

[5] Grip, R.S. & Grip M.L. (2019). Using Multiple Methods to Provide Prediction Bands of K-12 Enrollment Projections. Population Research and Policy Review, DOI 10.1007/s11113-019-09533-2,1-22.

# Appendix

## Assessment of Current Best Practices and Academic Research Regarding Housing Unit and Student Enrollment Projection Methods

*Overview*

Despite the existence of a long history of academic literature concerning school planning, both the theory and the practice of enrollment projection have remained essentially constant since at least the '70s (Sweeney & Middleton, 2005, p. 1). The cohort survival method has been the chief tool of administrators, school planning boards, and consultants tasked with predicting how many students would attend public school in the upcoming years. Also known as grade retention, this method uses historical data to calculate the proportion of students advancing to the next grade and is used nearly universally for enrollment projections (Schellenberg & Stephens, 1987, p. 1). Other techniques fall into two broad subcategories: multiple regression and numerical simulation. These methods are generally employed in conjunction with cohort survival calculations in order to refine projections or to provide more specific predictions concerning individual school attendance rather than district-wide figures (Bernhardt, 1980, p. 5). All projection techniques in general rely on discovering patterns in historical data and extrapolating trends into the future, and this is especially evident in the context of enrollment forecasting.

Since cohort survival methodology works under the assumption that trends in enrollment today will continue tomorrow, proper identification of events that may vastly alter school enrollment patterns is of particular importance. One major example of such an event is the introduction of new housing units into the community during projection years, since these often bring school-age children (Holley, 1966, p. 2). In this case, multiple regression is extremely useful. Instead of merely using historical survival ratios to predict enrollment, variables such as number of new housing units are inserted into the equation. These additional variables are vital for analysis of districts that are experiencing rapid growth and often make or break an enrollment projection process (Grip, 1998, p. 35).

*Cohort Survival Methods*

Perhaps the most intuitive technique for district enrollment projections is the student flow model. The flow model for projecting one year is exhaustive in that every single student entering and exiting district school systems is accounted for in the variables of the flow equation:

$$S(I, t) = S(I-1, t-1) - D(I-1, t-1) + A(I-1, t-1) - W(I-1, t-1)$$

where student enrollment for grade $I$ in year $t$ is given by $S(I, t)$, deaths of a member of grade $I$ during year $t$ is given by $D(I, t)$, entrances into grade $I$ (ascensions) and withdrawals from grade $I$ after school year $t$ starts given by $A(I, t)$ and $W(I, t)$, respectively (Sweeney & Middleton, 2005, p. 2). The model attempts to aggregate individual student behavior to come up with a total enrollment figure for each grade. In other words, the number of students in grade $i$ this year should be equal to the number of students in grade $i - 1$ last year, accounting for deaths, drop-outs, withdrawals, and entrances during the year. While this model is often not of practical significance due to the difficulty of accurately estimating the various flow terms, it is the perfect introduction to understanding the cohort survival method.

One may think of the cohort survival method as a simplification of the flow model. Instead of trying to explicitly find all the probabilities associated with students entering and exiting the school district, the cohort model simply takes historical data as a proxy for terms *D*, *A*, and *W*, thus simplifying the equation to become

$$S(i, t) = \alpha * S(i-1, t-1).$$

$\alpha$ represents the grade retention ratio, or survival ratio. The straightforward nature of the equation is clear, and the only data requirement to calculate the survival ratio is historical enrollment figures. Once the ratio has been calculated, it is then multiplied by current enrollment figures in the grade below to produce that cohort's enrollment projections for the next school year. Kindergarten enrollment projections rely on a slight variation of this method. Since there is no grade level before kindergarten, one must use birth data to estimate enrollment. Comparing kindergarten enrollment to district births from five years prior produces a similar ratio as above.

With this technique, predicting more than one year of data consists of multiplying the previous year's prediction by the grade retention ratio again until the final time period has been predicted. As long as the data on previous years' enrollment is available, this method is very reliable; survey data indicates projectors aim for one percent error or less per year when predicting total district enrollment (Schellenberg & Stephens, 1987, p. 7). To increase accuracy, forecasters generally consider multiple years of historical data, calculate the cohort survival ratio for each year, and then use the average ratio for projection calculations.

The mean survival ratio calculation is shown below for *n* years of historical data, with $\alpha(t-1, t)$ representing the survival ratio from year $t-1$ to year $t$:

$$\alpha = \alpha(t-1, t) / n \;+\; \alpha(t-2, t-1) / n \;+\; \ldots \;+\; \alpha(t-n-1, t-n) / n.$$

A forecaster may instead feel that it is more appropriate to give more weight to recent years, compared to the equal weights of the arithmetic mean. This slight modification is known as exponential smoothing and is endorsed in the *Projections of Education Statistics to 2027*, the most recent annual report by the Department of Education. Instead of giving equal weight to all ratios from the historical data, "projection is a weighted average based on exponentially decreasing weights":

$$\alpha = c[\,\alpha(t-1, t) * (1-c) \;+\; \alpha(t-2, t-1) * (1-c)^2 \;+\; \ldots \;+\; \alpha(t-n-1, t-n) * (1-c)^{n-1}],$$

where $0 < c < 1$ (Hussar & Bailey, 2019, p. 72). In this variation, most recent survival ratios are thought to have the most predictive power. Comparing the two formulas, the $1 / n$ weight in the former equation is replaced by $c*(1-c)^t$ in the latter equation for data that occurred in year $T-t$, which is the same as saying data from $t$ years ago. Since this weight is fractional, it is clear that the quantity decreases as data from less recent time periods are used. This is one of the many basic modifications that school forecasters might use to personalize the method to the district in question. It is at this point that many local officials use common-sense estimations to "fine-tune" the numbers in the right direction according to their intuition or inside knowledge of the school district (Schellenberg & Stephens, 1987, p. 10).

*Multiple Regression*

The cohort survival method and its variations are best suited for districts experiencing relatively stable enrollment trends. Using these tools, districts that violate the assumption that enrollment trends in the coming years are a continuation of trends in the recent past will not have accurate forecasts. What is to be done for areas of rapidly increasing growth or significant demographic changes that affect school enrollment? Additional variables must be modelled using regression techniques or accurate forecasting is simply impossible.

Multiple regression comes in endlessly many flavors, since different equation forms and additional variables can always be utilized. Ultimately, in a linear regression one wants to predict some quantity $y$ based on observations of $p$ variables $x_1$ through $x_p$ by finding optimal coefficients $\beta_1$ through $\beta_p$ and y-intercept $\beta_0$ that minimize prediction error $\varepsilon$:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon.$$

This is easily adapted to the flow model equation introduced earlier. The authors of this report have previously used essentially the same technique in research on school districts in central New Jersey, constructing the model:

$$S_{i+1} = \beta_{i0} + \beta_{i1} S_i + \beta_{i2} M_i + \beta_{i3} H + \varepsilon_t,$$

where $S_{i+1}$ represents the predicted number of students in grade $i + 1$, $S_i$ represents the number of students in grade $i$, $M_i$ represents the net migration of students from grade $i$ to grade $i + 1$ who did not start the year in that class, and $H$ represents the number of new local housing units (Grip, 1998, p. 29). Note that this equation must be constructed for each grade level. Then, each coefficient $\beta_{ij}$ is the model's best estimate of the respective impacts from each independent variable $S_i$, $M_i$, and $H$ on the following year's enrollment for each grade $i + 1$. It is important to stress that the values of $M_i$ and $H$ should be inputted to the model after they have been separately predicted. This is where the strength of the model lies, because it is a way to formalize forecasters' intuition about exogenous variables. Pertinent data may be collected

> by those persons most familiar with each of the separate variables or by one person who gathers information from….U.S. Census publications, local school censuses, birth records, marriage licenses, building permits, utility company records and projections, mail route changes, U. S. Office of Education, National Education Association, Chambers of Commerce, planning boards, zoning commissions, school enrollment reports, building and occupancy reports, tax records, real estate developers, and government and news publications which discuss general business conditions and the activity of home and industrial construction industries (Denham, 1971, p. 29).

More detail on the housing unit projection process and methodology will be presented in the following section.

Alternative formulations to the basic regression model are used. Consider again the cohort survival equation with simplified notation, $S_{i+1} = \alpha * S_i$. One might postulate that new housing developments will also have an impact on the grade retention ratio $\alpha$, so the equation reads $S_{i+1} = \alpha * S_i * H$. The equation is now a multiplicative model, and the "coefficients" are now exponents:

$$S_{i+1} = \beta_{i0} * S_i \wedge (\beta_{i1}) * H \wedge (\beta_{i2}).$$

Notice that after taking the natural logarithm of both sides and simplifying using log rules, the equation reverts back to a basic multiple linear regression:

$$\ln S_{i+1} = \ln \beta_{i0} + \beta_{i1} * \ln S_i + \beta_{i2} * \ln H.$$

Now the same closed-form solution can be solved for the multiplicative model the same way it is for the linear model. In a linear model, the coefficients indicate how much the dependent variable changes in response to a unit change in the independent variable, while in a multiplicative model, the coefficients indicate how much the *percentage* of the dependent variable changes in response to a unit *percent* change in the independent variable, and this measure is known as elasticity. There may be no *a priori* reason to favor a multiplicative or linear model over the other, but a forecaster may observe better or more intuitive results with one equation and just use that model for the district (Hussar & Bailey, 2019, p. 73). There is usually no right answer for the best model to use, so it is up to the discretion of the forecaster to use the tools that best apply to the situation.

*Data Collection*

The academic world has little to offer in terms of actually computing the effect of additional housing units on school enrollment, as the literature tends to speak of this process very abstractly. Instead, information is sourced from reports created by different districts publishing their implementation of factoring new housing into enrollment projections. For simplicity, information from three case studies from the Mid-Atlantic are referenced in this section: Alexandria City Public Schools, Arlington Public Schools, and Montgomery County Public Schools.

The main question to be answered is, how many new students will attend district schools if new housing units are constructed? This may be answered simply by taking the number of students in the district and dividing by the number of dwelling units in the district; this quantity is known as student yield or generation rate (Alexandria City Public Schools, 2013, p. 1). Student addresses should be accessible through a school database or other electronic storage, but generally schools do not have access to housing data. This requires districts usually partner up with county or municipal organizations to get access to data if not readily available. Once this data is secured, each student address should be mapped to a housing type (White, 2019, p. 11).

This mapping may be done using a programming language like R or Python, or special geospatial software like ArcGIS. Then it is a simple matter to sum up the number of students living in each type of dwelling to find appropriate student yields, using Excel or continuing to use more custom programming techniques. This analysis may be more detailed, if one wants to break down, for example, the distribution of student grade levels per unit type. Indeed, this level of depth would be necessary to plug back into the *H* term of the multiple regression model presented in which each grade level has its own equation and unique coefficients. In Arlington, student migration is not taken into account, but students are differentiated by living in garden apartments, high-rises, single-family detached dwellings, and townhouses (White, 2019, p. 10). Montgomery, on the other hand, does take student migration into account, looking at transfer student history, and similarly to Arlington the county divides housing into high-rise, low- to mid-rise, single-family attached, and single-family detached (Lazdins & Ruiz, 2014, p. 3). Ultimately, the appropriate student generation for new housing units is calculated for various housing types and grade levels depending on the desired granularity for the county and this information is incorporated into the models introduced above.

*Monte Carlo Methods*

In the context of school enrollment, Monte Carlo methods are usually applied on top of a cohort survival model, although these methods can also be used to estimate more complex relationships such as those given in the multiple regression section. In general, a Monte Carlo simulation involves approximating the results of either a probabilistic process (i.e., forecasting school enrollment) or deterministic process (i.e., calculating the area under a curve) by using random sampling of input variables. In other words, the creator of the simulation specifies the distribution that the independent variables take on and the relationships between the independent variables and the dependent variable. The sample distributions of input variables should be either simple distributions from built-in computer functions (uniform, normal, etc.) or based on historical data. After samples are drawn from each independent variable, the value of the dependent variable is calculated, and this process is repeated until enough simulated data points have been created to approximate a distribution for the dependent variable.

Although methods in the previous sections are often enough to forecast an average enrollment figure, it is wise to have lower and upper bounds for the forecast. Even more useful is a complete probability distribution of enrollment figures, so administrators and other officials may glean a more holistic understanding of the different enrollment scenarios that could play out. Indeed, Denham recommends Monte Carlo simulations for this exact reason (1971, p. 6). Simulations may also be used to approximate more complex relationships between input variables. Take for example the problem of estimating housing units and student migration in the multiple regression model. It seems likely these variables impact each other. However, using a Monte Carlo simulation, one must only sample from the years of historical data to approximate enrollment using the relationship specified in the original equation after estimating coefficients.

A more concrete example may be found in a recent paper by the writers of this proposal. Twenty years of cohort survival ratios from three school districts were used, and in fact the *changes* between the cohort survival ratios were sampled and then multiplied by the most recent cohort survival ratio and the number of students enrolled for each grade:

$$S_{i+1} = \alpha * S_i * \Delta\alpha.$$

The $\Delta\alpha$ term is one of the 19 historical cohort survival changes from which one may randomly sample, and this process is repeated five times for each five years into the future and done for each grade. When that is done, the process restarts until there are significant simulations to draw summary statistics from and construct a distribution (Grip & Grip, 2019). This takes thousands upon thousands of iterations and results in comparable estimations to the basic cohort survival and multiple regression methods. While Monte Carlo methods need a simpler model on which to draw from, they can do better than those models in terms of constructing different enrollment scenarios and providing an easy way to simulate complex relationships between variables.

*Conclusion*

Recently, it seems that public and private entities alike have benefited from contemporary advances in raw computing power, statistical and machine learning techniques, and data availability, yet this trend has not revealed itself in enrollment projection methodologies. This is not due to ineptitude of local government or an aversion to major change from the education industry but rather fundamental constraints on the problem at hand. Every school district is different, and local officials, administrators, and teachers are most knowledgeable about trends in variables affecting enrollment that may not be easily quantifiable by a computer program or outside consultant.

Extremely localized patterns and complex variable interactions means “the factor of human judgment probably cannot be replaced by any computational means” (Schellenberg & Stephens, 1987, p. 13). For example, how would one predict family decisions to enroll children in private school? This is just one of many factors contributing to final enrollment numbers. The authors do not wish to say that this is impossible for a specific district but instead wish to express the lack of a general methodology that can accurately trace out the complex relationships between community variables and enrollment in public school across heterogeneous school districts. Furthermore, economic data that might provide additional information may not be available or may be too costly in terms of time, resources, or monetary value than is worth it to include in the project analysis. This is why the dominant projection technique is to extend trends in historical enrollment data, occasionally using regression techniques to incorporate additional information such as anticipated housing unit expansion during periods of instability or significant district feature changes. Furthermore, computational methods may be utilized in the creation of probability distributions to assist school boards, administrators, and municipal planners in making contingency plans for different enrollment scenarios. Ultimately, forecasting school enrollment is unique to the district in question and a highly differentiated plan should be constructed to achieve optimal results.

**References**

Alexandria City Public Schools (2013). *Enrollment Forecasts/Demographics Subcommittee of the Long-Range Educational Facilities Plan Work Group.* Retrieved from https://www.alexandriava.gov/uploadedFiles/planning/info/LREFP/Student%20Generation%20Rates_Summary_10_25.pdf

Bernhardt, Victoria L. (1980). *Projecting Student Enrollments: A Basic Step in Comprehensive School District Planning for Declining Enrollment.* Retrieved from https://files.eric.ed.gov/fulltext/ED187020.pdf.

Denham, Carolyn H. (1971). *Probabilistic School Enrollment Predictions Using Monte Carlo Computer Simulation. Final Report.* Retrieved from https://files.eric.ed.gov/fulltext/ED062729.pdf

Grip, R. S. (1998) Prediction of public school enrollments using the Modified Regression Technique (Doctoral Dissertation, Rutgers University, 1998).

Grip, R. S. & Grip M. L. (2019). Using Multiple Methods to Provide Prediction Bands of K-12 Enrollment Projections. *Population Research and Policy Review*, DOI 10.1007/s11113-019-09533-2,1-22.

Holley, Paul (1966). *School Enrollment by Housing Type* (Information Report No. 210). Retrieved from American Society of Planning Officials: https://www.planning.org/pas/reports/report210.htm

Hussar, W. J., & Bailey, T. M. (2019). *Projections of Education Statistics to 2027* (NCES 2019-001). U.S. Department of Education, Washington, DC: National Center for Education Statistics.

Schellenberg, Stephen J. & Stephens, Charles E. (1987). *Enrollment Projection: Variations on a Theme.* Retrieved from https://files.eric.ed.gov/fulltext/ED283879.pdf

Sweeney, Stuart H. & Middleton, Erin J. (2005). *Multiregional cohort enrollment projections: Matching methods to enrollment policies.* Retrieved from http://www.geog.ucsb.edu/~sweeney/research/Sweeney_Middleton_PSP.pdf

Lazdins, Valdis & Ruiz, Roberto (2014). *Subdivision Staging Policy: Revised Student Generation Rates*. Montgomery County Planning Department, Silver Spring, MD. Retrieved from https://www.montgomeryplanningboard.org/agenda/2014/documents/item8bRecalculationofStudentGenerationRatesv7.pdf

White, Lionel, Ruiz, Robert & Mimberg, Kathy (2019). *Annual APS Enrollment Projections Report.* Arlington, VA. Retrieved from https://www.apsva.us/wp-content/uploads/2019/02/Fall-2019-28-APS-Enrollment-Report_v12_.pdf