Forecasting Student Populations

Arlington Analytics uses its Arlington Model to forecast student populations. In this paper, we:

1. Present an overview of the Arlington Model
2. Explain how we use it to estimate current student generation factors (SGFs)
3. Explain how Arlington Analytics forecasts future SGFs
4. Explain Arlington Analytics forecasts student enrollment
5. Cover how to use the website to understand our forecasts and make your own forecasts

The Arlington Model

The Arlington Model is a property-level model of Arlington. Arlington Analytics keeps track of all of the properties available on the Arlington Open Data website. We combine this data on properties with information on:

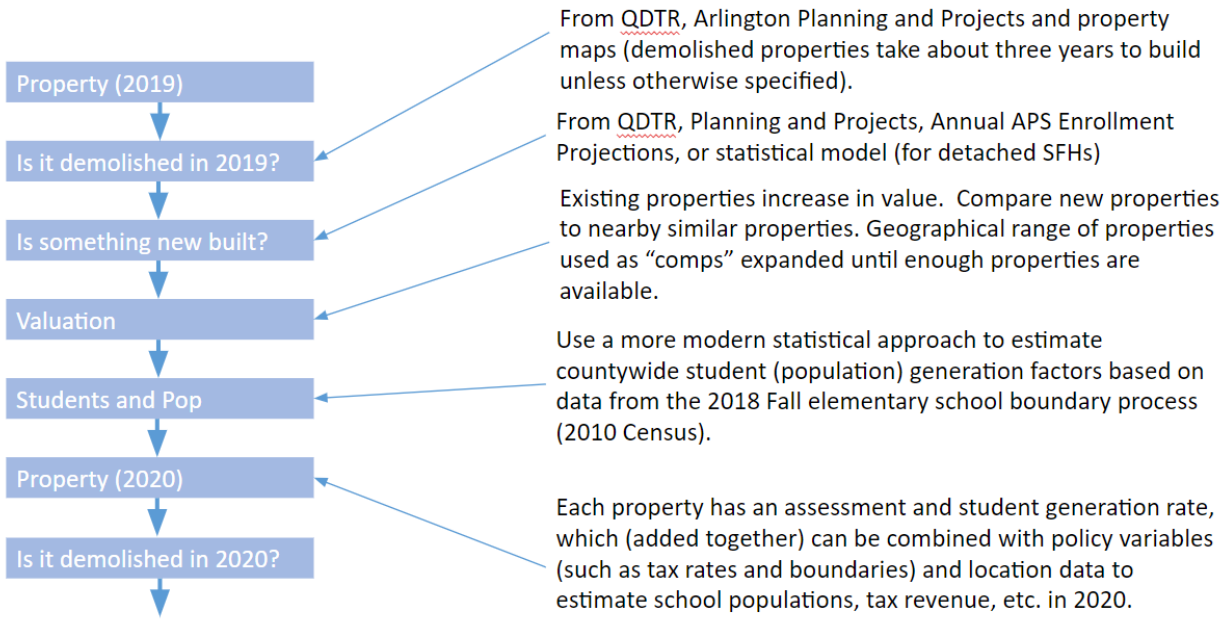
- Assessments
- School data such as planning unit student populations, walk zones,
- Maps of planning units, census blocks, and planning units
- Census population data
- Real estate data and property information (particularly from realtor.com, apartments.com, property websites, county websites on affordable housing, the white pages, Google Maps satellite pictures, and more).
- Crime and service requests
- and other data sources

We use this information to build a database of all current properties in Arlington. This allows us to understand how the county looks today and how it is going to grow in the future. In order to forecast what the county is going to look like in the future, we combine this database with data from:

- Quarterly Development Tracking Reports (QDTR)
- APS Housing Unit Forecasts (see Attachment B)
- Development Project Pages (example)
- Our in-house statistical model of detached single-family home (SFH) redevelopment
- Our own rough estimate of additional building starting in six or more years likely being developed, but not approved or listed in the reports above

Then we run property-by-property simulations with all of this data to develop an extremely granular picture of the county, typically going about 10 years into the future. Each property is simulated in each year for about 10 years using the procedure shown in Figure 1.

Figure 1: Forecasting the Effects of Properties



This procedure provides a full, year-by-year census of properties in Arlington. Each residence, on average, generates a certain number of students that are enrolled in Arlington Public Schools (APS). In order to figure out how many APS enrollments, on average, a type of residence produces, we need to estimate SGFs.

Estimating Student Generation Factors

We estimate student generation factors using a statistical model that takes, as inputs, the number of elementary students from each planning unit and the census of residences per planning unit[1] generated by the Arlington Model. We estimate how each type of residence contributes to APS enrollment.[2] Reflecting findings in a 2016 consultant's study, we estimate the effect of newly-renovated / newly-redeveloped detached single family homes separately from the existing stock. Furthermore, reflecting trends in published student generation reports, we separate out the enrollment contributions from apartments near Columbia Pike, which generate enrollment at a rate that far exceeds the enrollment rates of apartments elsewhere in the county.

Unlike APS, we do not estimate our SGFs by school. Although looking at SGFs by school can and does highlight some geographical trends in enrollment, we are concerned that estimating SGFs by school and using those in our model would overfit the data leading to spurious relationships. Spurious relationships limit the SGFs' usefulness when changing school boundaries, and limit the predictive power of the SGFs when types of housing are not available in a school's boundary.

[1] A planning unit is the base geographical unit used by APS during its planning processes. Each planning unit is assigned to an elementary school, a middle school, and a high school. During boundary planning periods, whole planning units are attached to schools. For example (with one exception in which a large planning unit was split into two), during the Fall 2018 boundary planning process, each planning unit was reassigned in its entirety to an elementary school to rebalance current and future school enrollment.

[2] We use a non-linear, least-squares model with a few restrictions. We restrict houses, condos, duplexes, and townhomes to generate at least as many students on average as the same properties with fewer bedrooms. For example, a 3-bedroom detached single-family home generates at least as many kids as a 2-bedroom detached single-family home on average.

Although they are constantly being improved and revised, our current student generation factors are presented in Table 1.

Table 1: Elementary School Student Generation Factors

Type		Garden	Mid	Elevator
Apt (Market)		0.066	0.014	0.016
Apt (CAF)		0.251	0.247	0.230
Apt (Market, Pike)		0.200	0.086	0.137
Condo (1 bed)		0.017	0.000	0.000
Condo (2 bed)		0.094	0.062	0.066
Condo (3 bed)		0.183	0.500***	0.066

Type	2- bed	3 bed	4 bed	5+ bed
SFH	0.073	0.250	0.250	0.250
Town	0.130	0.130	0.130	0.130
Duplex	0.213	0. 213	0. 213	0. 213
SFH (less than 15 y.o.)	n/a	n/a	n/a	0.528

* Middle school factors, divide by 2.17.

** High school factors, divide by 1.78.

*** There are only 135 of these units, and the statistical variance is very high in this estimate. In future versions we will combine these with the 2-bedroom mid-rise condos.

We do not have adequate planning unit enrollment data for middle schools and high schools in order to perform similar estimates. Nonetheless, SGFs in existing county documents are fairly proportional across different housing types, so we simply divide each number by 2.17 to get middle school SGFs and 1.78 to get high school SGFs.

We make several observations. First, our estimates are broadly in line with the average estimates from APS. Second, we find that newly-constructed 5+ bedroom detached single-family homes generate nearly three times the number of students as any other house. Third, we find that market rate apartments on the pike generate much higher enrollments: Garden apartments generate twice as many, mid-sized apartments generate nearly four times as many, and elevator apartments generate about eight times as many enrollments as their market equivalents elsewhere in the county.

Forecasting Future SGFs

SGFs change over time. In Table 2 below, we present SGFs for all 12 grades calculated in six different reports: 2019, 2017, 2015, 2013, 2010, and 2005.

Table 2: Historical SGFs

Year	SFH	Town-home	Dup-lex	Elev. Mkt	Elev. CAF	Elev. Condo	Garden Mkt	Garden CAF	Garden Condo
2019	0.497	0.194	0.358	0.061	0.617	0.054	0.258	0.5784	0.11
2017	0.489	0.182	0.359	0.078	0.633	0.054	0.295	0.646	0.105
2015	0.469	0.267	0.267	0.06	0.58	0.049	0.268	0.614	0.086
2013	0.42	0.13	0.38	0.08	n/a	0.03	0.29	n/a	0.09
2010	0.413	0.246	0.246	0.051	0.447	0.035	0.236	0.501	0.07
2005	0.36	0.1	0.45	0.06	n/a	0.04	0.25	n/a	0.07

Notes: Townhomes and duplexes were estimated together in 2015 and 2010.
Sources: 2005, 2013 (2015 Community Facilities Study), 2010 (Methodology and Process Improvements: Phase 2), 2015 (Annual APS Enrollment Projections: 2016), 2017 (School Year 2017-2018 School Generation Factors), 2019 (Annual APS Enrollment Projection: 2019)

We identify several trends, which we summarize in Table 3:

- Across 2005 to 2019, market rate apartments (both elevator and garden) have generated roughly the same number of enrollments per unit. There is significant variation, however, reflected in a significant increase in 2013 and 2017 and a significant decline in 2010.
- SGFs for detached SFHs have increased in each of the reports, from 0.36 in 2005 to almost 0.50 in 2019. Townhomes SGFs have nearly doubled, increasing at a yearly rate of nearly 4.8 percent per year.
- Duplexes have decreased slightly over time.
- Both types of condominiums have increased student enrollment.
- Affordable housing factors have increased across each type of housing from 2010 to 2019 (no data was available in the 2005 report).

Table 3: Historical and Forecast Student Generation Factor Growth

Residence Type	Avg. Historical Growth Rate (2005-2019)	Baseline Forecast
Apartment (Market)	0.1% (elevator) to 0.2% (garden)	0.2%
Apartment (CAF)	1.6% (garden) to 3.6% (elevator)**	2.0% tapering to 0.5%
Condominium	2.2% (garden) to 3.3% (elevator)	2.0% tapering to 0.5%
Detached SFH	2.3%*	1.7% tapering to 0.0%
Townhome	4.8%	1.8% tapering to 0.0%
Duplex	-1.6%	0.0%

* The 2.3 percent increase in the student growth generation factor from detached SFH is 2.3 percent, which includes house redevelopment. The forecast SGF excludes redevelopment, which occurs at a rate of about 200-250 houses per year.
** CAF historical growth rates calculated from 2010 to 2019.

We increase our SGFs as shown in Table 3. Over time trends change and student growth per residence is likely to respond to changing demographics. The growth rates in these factors is determined judgmentally, looking at factors such as recent and long-term history, birth rates, and other factors.

SGFs are a product of two factors: kids born in Arlington who enter the school system and net migration of kids into the school system. Over the next five years, we project based on historical student enrollment that SGFs will increase in line with history as the large 2014 through 2017 cohorts of prospective students enter kindergarten and become elementary students. Large classes of elementary school kids will continue on to be large middle school classes and large high school classes. Moreover, kids will continue to migrate into the system at the high school level at a relatively as they have since 2010. The number of kids graduating over the next five years is relatively small, but grows quickly around 2024.

APS currently predicts that student enrollment will peak in 2024 as a result of two factors: much larger graduating high school classes leaving the system and much smaller kindergarten classes coming from a lower birth rate. APS predicts that the birth rate will drop from 12.3 births in 2018 to about 11.0 births in 2024 while the share of births reaching kindergarten will remain the same. Both of these cause a slight decline in the SGFs.[3]

We predict that the SGFs will level off, which we attribute to additional migration into the system. A renewed demand for office- and commercial-space across the county, reflecting Amazon's build-out as well as a broad decrease in the office vacancy rate, will increase demand for housing in Arlington, particularly for families.

Computing Student Enrollment

To compute student enrollment in each planning unit, we need the projected number of residences described in the section "The Arlington Model" and the projected SGFs described in the section "Forecasting Future SGFs". We multiply the number of each type of residence by that residence's SGF in the year we're forecasting and add the statistical residual from last year observed years.[4]

[3] Projected 2024 birth rate calculated from number of births provided by APS and a projected population from Arlington Analytics.
[4] The residual is the difference between the number of observed enrolled children and the number of predicted children, scaled by a factor that makes the number of predicted students add up to the number of observed students in the most recent year for which data is available. The residual does not change from year to year: if a planning unit has a residual of 10 kids---we observe 10 more children than the statistical model predicts--then it will have 10 kids through the entire forecast. As we improve the statistical model and get new data over time, we look forward to more accurately modeling the residual over time, which should improve forecasts at a planning unit level.

In addition to these values, we consider the possibility that some units will be constructed in the out years that are not currently listed in the county documents. These are a number of proposed projects—condos and apartments—that have not yet received final approval. Time from approval to construction can vary wildly, but we assume that these are all projects that fall beyond a five-year time horizon. Therefore, we assume a number of apartments units and condos will be built starting in year six. As we cannot anticipate where most of these projects will be undertaken, we do not tie the new residences to specific planning units. The students in these unanticipated projects only affect the total enrollment statistics.

The Results

We publish our forecasts for student enrollment based on the methodology above through 2028 on Arlington Analytics. For comparison, we also publish the APS forecasts—the low, medium, and high—as points of comparison. Drag the cursor over the point on the graph to get the exact value of the forecast.

The default options at the bottom of the webpage are set so that they approximately mimic the Arlington Analytics forecast.[5] There are three types of options that can be set in the model:

- The growth rate of the SGFs
- Turning off development of certain types of residences
- Adding additional apartments and condos starting in year six

One exercise to determine the relative importance of various factors is to set all of the SGF growth rates to zero, turn off all development, and set all additional unapproved development to zero. Turning a single option on at a time highlights the relative importance of each option toward projected growth in student enrollment.

Based on prior student forecasts, Arlington Analytics believes in far greater uncertainty in the enrollment projections than is captured by the APS paths presented in the 2019 10-Year Projections Report (see Attachment A). Forecasting enrollment rates is challenging, especially with rapid changes in population migration across and beyond the region. As seen through the app on the website, even modest and believable changes in certain options can lead to some substantial swings in student enrollment. Using the Arlington Analytics tool can give a sense as to the types of risks—for over- and under-predicting enrollment—embedded in the projections.

Notes and Acknowledgments

Authored by Jon Huntley. Last updated, February 2020.

For additional tools, data, and analysis, please visit our site at https://www.arlington-analytics.com/.

Special thanks to Kody Carmody, who coded parts of the Arlington Model. Special thanks to all of the people who have provided advice, particularly members of the APS Advisory Council on School Facilities and Capital Programs Steven Leutner and Lois Koontz. Thanks to Duke Banks, Lars Florio, Peter Rousselot, and John Vihstadt for providing feedback, advice, and recommendations. Thanks to Arlington Public Schools for providing a very rich and helpful dataset on student enrollment from the 2018 Boundary Process. Thanks to Jamie Lees, Arlington County's Chief Data Officer, and other Arlington county staff for filling various data requests, particularly for mapping information. All errors are the author's own.

[5] The full Arlington Model has many more options---including time-varying growth rates for SGFs---than are available on the webpage. The default options are selected to be similar to, but not identical to the default Arlington Analytics enrollment forecast.

Feedback is always welcome, and can be sent to analytics.arlington@gmail.com or jon.huntley@gmail.com.