Thinking about Useful School Metrics: Comparing 'Like' Schools

Christopher Wohn, M.D. and Liepa Boberiene, Ph.D.

December 2018

Large urban school districts, such as Baltimore City Public Schools, are often faced with many data points and contexts in which to evaluate their students, schools, and programs. Schools can have widely varying proportions of students of different races, socioeconomic statuses, and disability levels. There are also many risk factors that can be associated with outcomes such as assessment performance, graduation, or college readiness. Given the large number of characteristics that can be measured about schools and students, it can quickly become difficult to recognize patterns or relationships within educational data sets.

A major issue in ranking schools based on student performance is that it fails to take into account underlying demographic or economic issues. Schools with high proportions of low-income students will tend to perform poorly in comparison to more economically advantaged schools. Given this, how can districts evaluate school performance while considering demographic and economic conditions that may explain differences in student performance?

The goal of this paper is to outline available and viable techniques for school evaluation given unique demographic conditions. While no individual technique is perfect, each one can help adjust for underlying factors that schools may not have direct control over. We will examine how Baltimore City Public Schools took student demographic characteristics into consideration when implementing its School Performance Measure (SPM).

Baltimore City Schools' School Performance Measure (SPM)

With Race to the Top, Maryland redesigned teacher and principal evaluations with multiple performance measures that combine into a total effectiveness rating. The evaluation design was intended to capture the multifaceted work of educators in schools with measures that allow for performance to be differentiated and to provide insight to improve staff and school performance which would have a positive impact on student outcomes.

Baltimore City's School Performance Measure (SPM) is a snapshot of a whole school's year-long performance that serves as one component in teacher and principal evaluations. SPM includes learning environment (school surveys from parents and students, attendance and chronic absence rates), achievement and student growth on assessments. Additionally, for high schools, college and career readiness (SAT, ACT, AP and IB test taking rates and performance, Career and Technology Education course taking and success, and dual enrollment in college coursework) and graduation rates are used as student growth components. SPM scores range from 0 to 100 on each of the components, and an average of all component scores is computed as the overall SPM score used in teacher and principal evaluations.

As a part of teacher and principal evaluations, SPM was designed to norm scores within our district rather than compare schools to state or national standards. We looked at each school's absolute performance and annual growth and only used the higher ranked of these two scores for each school. This meant that even if performance was still low, schools could achieve high scores by showing annual progress in the right direction.

Getting Stakeholder Feedback

In the past few years, stakeholders, teachers, principals, and district administrators, have shared feedback that has been important in improving both the indicators and the calculation approach of SPM. In past years, we prioritized the most meaningful indicators, omitting cohort retention and staff survey responses about their school. We also improved indicators; for example, teachers and principals helped us select 14 specific items from the school survey that they felt responsible for.

More recently, stakeholders have discussed the ways that school-specific context did not feel represented in SPM. There was concern that even when our scoring included the higher score of absolute performance or growth, school circumstances could make it difficult for hard work and improvements to be captured in the score. Like other school reporting systems based on ranking performance, SPM did not control for student background characteristics.

To respond to these concerns, we looked to other states, particularly California and Oregon. California created a "Schools Characteristics Index" that used eight variables to calculate expected outcomes with linear regression and score schools against those expectations. In Oregon, the state report card used a grouping technique based on demographic similarities to identify comparison schools for ranks. Using the feedback provided by our own teachers and principals and state models led us to examine our options.

Comparing Similar Schools

Our goal was to capture the impact of educators on students, while recognizing that student characteristics play a role in absolute performance. Therefore, we set out to determine how schools could be evaluated fairly while capturing their actual performance in light of the unique circumstances of their diverse student populations. This would help us understand relative performance of demographically similar schools and ensure that our accountability measures did not reinforce the influence of poverty or other demographic factors. Exemplar schools could then be examined or used as models to identify strategies and practices that led to their higher performance serving similar students.

We examined three options to define 'like schools': linear regression, K-means clustering, and Nearest Neighbors.

- **Linear regression** - Uses school variables to make predictions on outcomes. *Can be used to adjust school ranks for the influence of demographics.*
- **K-means Clustering** - Finds patterns in a dataset based on school characteristics. *Can be used to create defined groups based on demographics for ranks.*

- **"Nearest Neighbors"** - Creates unique virtual groups for each school. *Can be used to rank schools only against those that are most similar.*

As seen in Table 1, each of these methods has advantages and disadvantages.

Table 1. Pros and Cons of Different Similar School Comparison Techniques

	Linear Regression	K-Means Cluster Analysis	Nearest Neighbors
Pros	More accurately accounts for the relationship between demographic factors and performance outcomes	Easier to conceptualize Does not adjust performance standard; rather, ranks schools against similar peers	Does not adjust performance standard; rather, ranks schools against similar peers Not a forced distribution of scores in distinct groups
Cons	Sets different performance standard based on population composition	Close neighbors may not be available for all types of schools Performance goal is not clear Distinct groups can create large differences for schools on the cusp	Close neighbors may not be available for all types of schools Performance goal is not clear Difficult to visualize

Our Nearest Neighbor Model

In the end, we chose to use a Nearest Neighbor model to rank similar schools within grade band. This method compares schools with similar proportions of students in economic disadvantage (ED), special education (SWD), and English learners (EL). In Baltimore City and nationally, these characteristics impact instruction and school resources, and have a strong relationship with school performance. The Nearest Neighbor approach provides every school its own virtual comparison group of similar schools for comparisons. Groups are unique for each school— Nearest Neighbor schools are those that are the closest distance, as seen in Figure 1 below.

Figure 1. Nearest Neighbors based on indicators of Economic Disadvantage, Students with Disabilities, and English Learners.
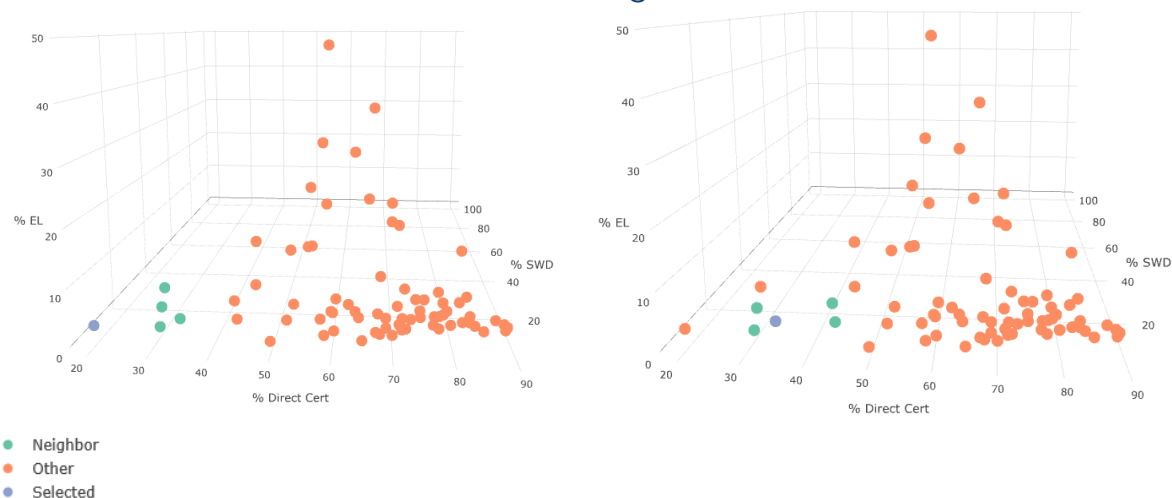
This approach is easier to understand than linear regression— because it is based on actual schools in Baltimore City versus relying on average trends, it is much easier to interpret. Similarly, it is more transparent, and educators can plainly see where their performance falls in relation to others in our district. Standards are determined by peer schools rather than set as a function of demographics. Finally, our approach creates groups of peer schools that school leaders can leverage for improvement efforts. They can learn which schools are most like theirs in terms of student population and learn from those schools about interventions and professional development.

Feedback from educators encouraged us that we were providing a metric that:
1. *Felt fairer*. Considering school differences makes the comparisons more meaningful, especially when schools with nearly 100% of students in poverty are compared to schools that serve students from higher social economic backgrounds who arrive at school with different attendance and performance behaviors.
2. *Increased transparency*. It becomes very easy for a school to replicate their score by checking their raw values and those of their nearest neighbors.
3. *Improved data literacy*. School leaders thought that ranks among peer schools helped them better understand the underlying data for their school. When compared to similar schools, where does their performance fall?
4. *Created peer groups*. Finally, principals were excited to know which schools were similar to theirs, so they could connect with other school leaders and see what interventions they implement and what professional development their offer.

We felt that the most important difference was identifying schools performing well given their populations, so best practices could be identified and shared.

Practically, the shift to Nearest Neighbor scoring weakened the relationship between SPM and economic disadvantage. Looking at Baltimore City schools in four quartiles of poverty, we see that average school scores increased across economic disadvantage groups, and they increased much more in the highest poverty quartile, so that schools with the most concentrated poverty increased by 15 points, as shown in Figure 2. The trend lines show the negative relationship between SPM score and poverty rate flattened, so that poverty is less correlated with SPM score when Nearest Neighbors scoring is used.

Figure 2. Changes seen in average school SPM score by quartiles of economic disadvantage.
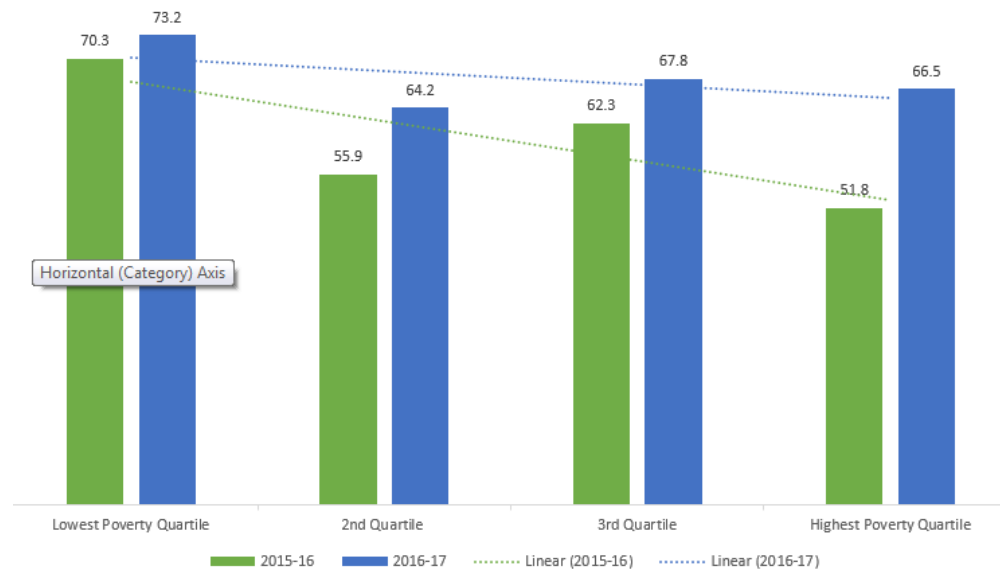
Limitations and Next Steps

Moving from accountability to support and improvement led to the redesign of a tool that provides more nuanced data to district leaders and to staff. Although many educators are happy with the results, there are still several limitations to consider:

1. For some schools, nearest neighbors are further away than for others. Including all Maryland schools could increase the size of groups and provide more similar comparisons for our more unique school populations.
2. Because we only compare schools on economic disadvantage, students with disabilities, and English learners, schools may not be comparable in other factors such as management type or school size. This is part of the trade-off between using a Nearest Neighbors and regression approach, which could accommodate many more variables.
3. Another concern is that small and large differences in an indicator's raw value lead to the same change in rank. For example, in groups of five with rank scores of 100, 80, 60, 40, and 20, a school may receive 20 fewer points for having a raw value difference of as little as .1 (e.g., 90.2 and 90.3 attendance rates). Again, this is the trade-off between simplicity/transparency of comparison groups and accuracy/precision of a regression.
4. Finally, some stakeholders felt the results could encourage competition rather than collaboration between schools.

Our next steps involve raising awareness of the benefits of like school comparisons. As we help educators understand school scores in terms of their impact versus the absolute performance of students, we can facilitate support efforts. We acknowledge that our schools have room to grow, and Nearest Neighbors can be a useful tool in developing intervention strategies and professional development plans that have worked for similar populations of students.

The authors are employees of Baltimore City Schools.