Finding the Best Fit: Building a Novel Student-School Compatibility Tool to Improve School Choice

Tyler Burleigh, PhD   *The HUMAN Project*
Paul Glimcher, PhD   *The HUMAN Project*

Abstract

There is a widespread conviction amongst educators, policy makers, and the public that increasing school choice will improve school outcomes. Offering public school students the opportunity to choose amongst a set of public and private/charter schools should result in improved rates of graduation and college completion. The existing data, however, suggests that this is not the case. This has led many commentators to conclude that we do not have, or do not provide students and families with, the critical information for effective school choice.

In this white paper we describe the development of a Next-Gen system for efficient school choice. By gathering extensive data about students and schools, and then by correlating these huge datasets with outcome measures, we should be able to develop algorithms that accurately predict graduation rates on a per-student basis. We show here that with the right data we can highly individualize graduation prediction; we can actually tell *an individual student her likelihood of graduation at each of the schools she is considering*. We can even search through a set of schools to rank order them for any given student in terms of any outcome measure. We could, for example, thus provide cost/performance trade-offs for individual school-student pairings.

If school choice is to improve outcomes we need to get the right students in the right schools. We show here how that can be done, empowering students and families to achieve higher graduation rates and better college preparedness.

Introduction

Over the course of the last decade there has been a growing interest in *school choice* as a tool for improving educational outcomes. This is reflected in the growing use of charter schools, in the development of school wide choice initiatives like the one in New York City, and in the policy of the U.S. Department of Education. This growing interest in school choice reflects the fact that in most parts of the U.S. today, children are assigned to schools based on the neighborhood in which their family lives. This results in unequal access to quality education and the social mobility that education provides: although children born into higher-income families might be able to pick up and move to a different neighborhood, or to enroll their children in a charter school, lower-income families are often stuck (and worse, stuck attending lower-quality schools; U.S. Department of Education, 2011).

The school choice movement has argued that there are many reasons why parents might want to choose a different school. Schools provide different learning environments through specializations, teaching methods, school culture, and other practices or resources. Likewise, children have different learning styles, needs, preferences, and capacities. These differences in students and schools give rise to instances where students and schools are fundamentally mismatched on key factors that would be necessary for a student's success. For example, a student who learns particularly well from hands-on learning experiences may perform poorly in a school that provides few opportunities for hands-on learning. In recognizing these differences, educators and policy-makers have set their sights on *school choice* programs that aim to empower parents and increase access to quality education, while creating an incentive for schools to improve (Pattillo,

2015).

Despite this attractive logic, however, school choice programs have for the most part failed to deliver on the promise of improving educational outcomes at scale (Teske and Schneider, 2001; Pattillo, 2015; Bell, 2009). Why? One of the major reasons is that, to put it simply: **choice is hard**. Consider an eighth grade student in New York City who must choose between 500 possible high schools. Although opening up school choice is intended to empower and increase individual agency, these themes do not describe experience of parents navigating the choice process. In interviews with parents in Chicago, Pattillo (2015) found that one of the key barriers that parents reported was the *amount of effort that making a choice required*. This is not to say that parents are lazy or apathetic—quite the contrary (Pattillo (2015) describes parents as having "quite heartily" engaged in search). But research has shown that the choice is cognitively demanding and resource-depleting (Schwartz, 2004)—the very same resources that individuals living in poverty are lacking due to life stressors (Shah et al., 2012). Similarly, Bell (2009) found that when parents used the resources offered to them by school districts to construct a "choice set" of schools, the average set size was only 3 to 4 schools, and the choice sets of lower-income parents included more schools classified as "failing" than those of middle-class parents.

So what kinds of information are being offered to students and parents as they choose? The answer is quite a bit. Details of school programs, locations and assets are all being passed to students. But are these the data people really need to choose a school. For a particular student at risk of failing to graduate high school what is the single most important fact she needs to know: The liklihood that she – personally – will graduate at each high school. It is not enough to know what the average graduation rate is at a school. **Parents and students simply cannot make informed choices if they do not know the likelihood that they, as individuals, will graduate.**

School choice fails to deliver because parents must decide on the basis of limited information. In an experimental field study, Hastings and Weinstein (2008) found that when parents were given the average test scores for schools within a five mile radius, more parents chose the better schools. These findings of sub-optimal school choice resonate with the larger literature on decision-making and bounded rationality (Gigerenzer and Selten, 2002). Imagine what would happen if parents and students had access to highly accurate and individualized probabilities of graduation for each school they were considering. *If we had that data and students believed the data were accurate because they came from an impartial source, we could transform school choice and unleash its potential.*

At the HUMAN Project we believe that parents should be given every possible chance to give their child the best education. In the context of school choice, we believe that the way to do this is to streamline the choice process and reduce the burden and uncertainty that parents face—to make it easier, especially for parents of underserved communities, to make the best choice for their children.

This endeavour requires deep data discovery: identifying the key properties of students *and* schools that combine generate positive and negative school outcomes. And it requires methods for reducing this complexity to simple predictive analytics, probabilities of graduation, probabilities of other outcomes, that anyone can use. These indicators would go well beyond the "ABCs" of Attendance, Behavior, and Course performance that are used in similar applications (Faria et al., 2017). We envision a tool that uses these key factors to project the likelihood of graduation (or other success metric), were a student to attend a given school. Using this tool, parents could readily identify quality schools based on a deep analysis of student-school compatibility—thereby overcoming the cognitive burden, uncertainty, and effort that currently undermines school choice. The HUMAN Project is positioned to build these tools using its database of hundreds of indicator variables on thousands of New York students.

The HUMAN Project

The HUMAN Project is a signature research initiative at New York University developed to solve our community's toughest challenges—from preventing diabetes and asthma to improving schools, ending poverty, and beyond. Started in 2016 with a $25MM investment, we have developed a quite a number of tools now deployed in a preliminary cohort which will be extended this winter to 10,000 additional New Yorker participants. These participant-partners share the tiny bits of information they create from moment to moment in their daily lives. This allows scientists discover hidden patterns and chart a course to making New York – and our greater society – a healthier, cleaner, happier place to live. As designed, the database will contain over 3000 children and will follow them for 20 years as they mature through school and secondary education. Our data is also representative, which allows us to draw robust and reliable conclusions. This provides an unparalleled dataset for relating a host of non-traditional measures to educational outcomes during data analysis.

A Case Study in Student-School Compatibility

Melissa has an extroverted personality: she's outgoing and socially confident, and she has a need for lots of social stimulation; she finds it easy and to speak in front of large groups of people and prefers to work in large groups. As a Black female, she has a unique perspective on the world, and she finds it difficult to relate to material when it doesn't speak to her own experiences. As someone who is impulsive and creative, she has a tendency to get off-task when she's not in an environment with clearly defined, and fairly enforced rules.

Sam has an introverted personality: he prefers not to make himself the center of attention and finds too much social stimulation to be tiring; although he can "put himself out there" when he needs to, he prefers a smaller, more tight-knit social environment, and he prefers to work alone or in small groups of people. As a White male, he has a perspective that is well-represented, but he also has a curious mind, and likes to read about the experiences of others. He is very self-disciplined, largely thanks to the strict, authoritarian style of his parents.

Because Melissa and Sam have different characters, needs, and capacities, they thrive in different kinds of learning environments. It is important to recognize that contemporary school choice systems (as opposed to parents and students) almost entirely ignore these differences between students. **Most school choice tools report a school's – not a student's – probability of graduation.** But it is the student and not the school that is graduating.

In our example, Melissa's extroverted personality allows her to thrive in schools with relatively large class sizes, while Sam's introverted personality means that he performs slightly better in schools with smaller class sizes. Being a minority female, Melissa does particularly well in schools that make an effort to represent diverse perspectives and create safe spaces for learning and exploring ideas. Although Sam likes learning about the experiences of people who are different from him, since he is a White male he has a fairly easy time with "traditional" course curriculum. Melissa's impulsive and creative nature means that she does really well in a structured environment, where the expectations are clear and discipline is reliable and fairly enforced. Sam, on the other hand, doesn't have a need for external structure or discipline. However, Sam's parents use a strict, authoritarian style of parenting and have unrealistic expectations. Because of this, Sam is at a disadvantage in schools where parents and administrators don't communicate well with parents. Sam and Melissa's parents may intuit some of these facts, but even for professionals it seems almost impossible to prioritize these school properties effectively.

With the right data, the kind of data gathered by the HUMAN Project, knowing how these factors quantitively impact outcomes is a simple matter of categorical regression. Figure 1 provides

a visual representation of the type of process involved in estimating student-school compatibility. At the bottom of the figure is a 5x5 grid that represents 5 different student and school parameters (though the number of parameters can be much higher than 5). Each cell of this grid represents the interaction between parameters in relation to a student's success in school. At the top of the figure is a magnification of where the Extraversion and Class Size cells intersect. In this magnified diagram there is a topographic map in which the height at a particular X,Y coordinate reflects the contribution that those factors would make to a student's likelihood of graduation. On this coordinate plane, we can see the class size Melissa and Sam do best in, as well as *the degree to which Melissa and Sam are expected to benefit from selecting a school on the basis of class size*. Melissa, being a highly extraverted person, would benefit greatly from picking a school with a large class size; whereas Sam, being an introverted person, would benefit from picking a school with a smaller class size, but would benefit MUCH LESS from picking on the basis of class size than Melissa.

Our Vision

We envision a future in which parents are not only empowered to make school choices, but where parents have the tools to make *informed* school choices on the basis of deep and personalized student-school compatibility analytics. These tools would not only be useful to parents, but they would also provide immeasurable value to school designers, school administrators, and policy-makers. Imagine a school district which benefits from existing data that has been collected from tens of thousands of children. When parents use the compatibility tool we have described here, it tells them the best schools for their children. And it gives them the tools to accurately trade-off commute distances or increased costs against completion probabilities.

One can also imagine the inverse. Looking at the data in aggregate, school designers who had a complete picture of the student population that they need to serve could optimize their schools. For each neighborhood a school designer would be able to identify the specific characteristics that are best for their student population. New schools could be designed to fill niches more efficiently, and existing schools could be modified to better serve the needs of their local population.

The HUMAN Project's Role

By the end of 2018, The HUMAN Project will already have the largest (0.5Pb) dataset about students in the world. In 10 years the New York Database alone will be 10Pb. We will have hundreds of biological, psychological, sociological, and environmental measures from 3,000 students in the NYC high school system, in addition to their siblings and parents. We will also have access to school-level data provided by the NYC Board of Education, which will include hundreds of additional data-points relating to schools themselves. Our dataset will not only be large, but it will also be unique in that it is representative sample of the population from which it derives. This will allow us to not only develop a deep understanding of student-school compatibility in NYC, but to develop an understanding that generalizes to other cities in the United States. Once we have this data, we will then develop robust models of student-school compatibility. By mid to late 2019 we will be in a position to deliver the foundational components that are required in these Novel Student-School Compatibility Tools.
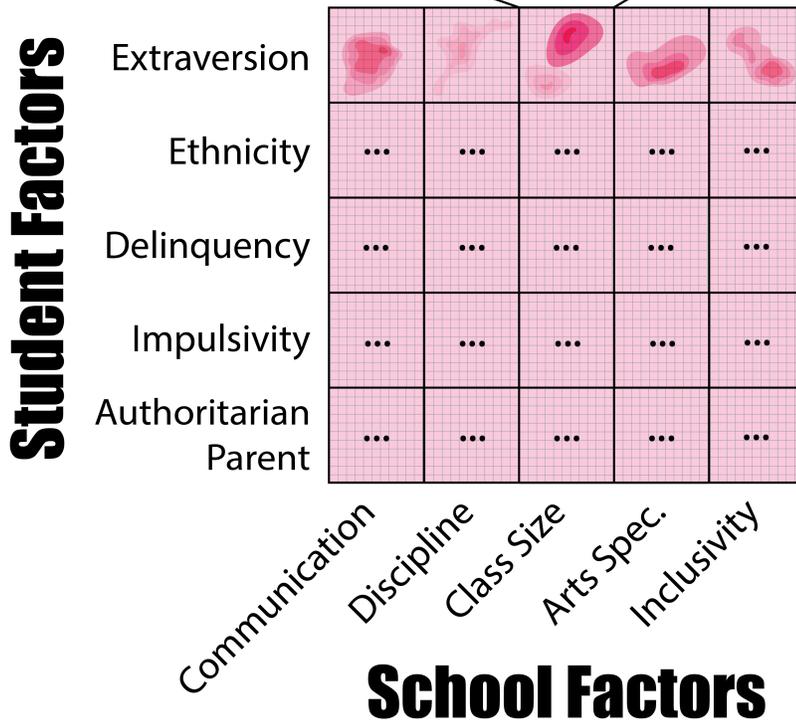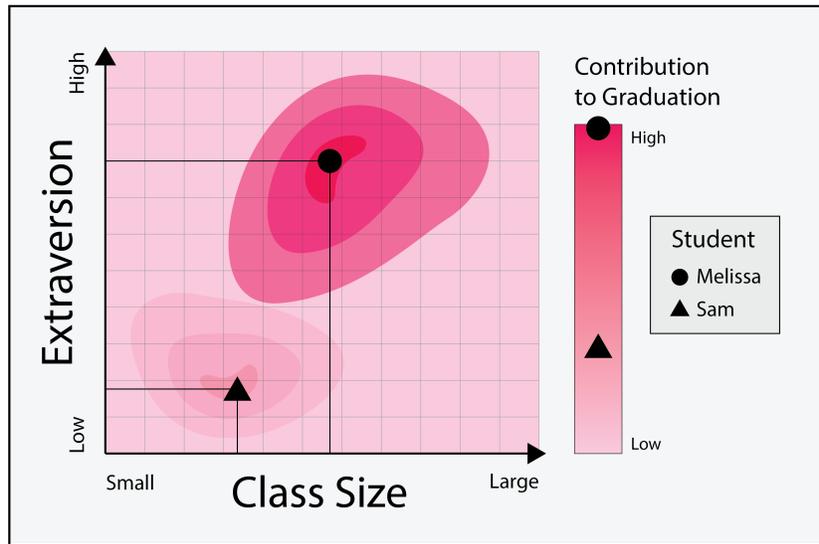
Figure 1: Student-School Compatibility Matrix.

References

Bell, C. A. (2009). All choices created equal? the role of choice sets in the selection of schools. *Peabody Journal of Education*, 84(2):191–208.

Faria, A.-M., Sorensen, N., Heppen, J., Bowdon, J., Taylor, S., Eisner, R., and Foster, S. (2017). Getting students on track for graduation: Impacts of the early warning intervention and monitoring system after one year.

Gigerenzer, G. and Selten, R. (2002). *Bounded rationality: The adaptive toolbox*. MIT press.

Hastings, J. S. and Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics*, 123(4):1373–1414.

Pattillo, M. (2015). Everyday politics of school choice in the black community. *Du Bois Review: Social Science Research on Race*, 12(1):41–71.

Schwartz, B. (2004). *The paradox of choice*. Harper Collins New York, NY.

Shah, A. K., Mullainathan, S., and Shafir, E. (2012). Some consequences of having too little. *Science*, 338(6107):682–685.

Teske, P. and Schneider, M. (2001). What research can tell policymakers about school choice. *Journal of Policy Analysis and Management*, 20(4):609–631.

U.S. Department of Education (2011). More than 40% of low-income schools don't get a fair share of state and local funds, department of education research finds. https://www.ed.gov/news/press-releases/more-40-low-income-schools-dont-get-fair-share-state-and-local-funds-department-education-research-finds.