The Impact of a District Turnaround Model: Applying Multilevel Matching to a Clustered Observational Study

October 1, 2019

1 Introduction

Initiatives designed to improve chronically underperforming schools have a mixed record of effectiveness. A wide range of federally-funded school turnaround efforts going back decades have spawned book-length chronicles and scores of recommendations (Bryk et al., 2015; Mehta et al., 2012; McGuinn, 2006). In the aftermath of the Great Recession, the federal government allocated $3.5 billion for states and districts to implement turnaround strategies. Whether funded through these School Improvement Grants (SIGs) or financed by local dollars, such efforts are as varied in their strategies as they are in their effects (Dee, 2012; Dickey-Griffith, 2013; Dragoset et al., 2017; Fryer, 2014; Sun et al., 2017).

In an effort to improve a group of chronically underperforming schools, the Wake County Public School System implemented its own turnaround strategy, dubbed the Elementary Support Model (ESM). In 2015-16, the district formed ESM as a new "area" that consisted of 12 schools ranked lowest on a district-developed index comprised of academic, human capital, behavioral, and socioeconomic indicators. ESM's three-year implementation period was funded primarily through federal Title I allocations, which supported governance reforms, additional staff, professional development, resources, leadership and instructional coaching, and support for calendar modifications.

Since ESM was non-randomly assigned at the school level and all students enrolled in ESM schools benefit from its supports, we classify it as a clustered observational study (COS) emulating a clustered randomized trial (CRT). Using a multilevel matching strategy, we leverage Wake County's full set of school- and student-level data in order to construct a control group that is balanced on observables. We report on a similar set of outcomes used to construct the ranking index, including student achievement in math and reading, attendance, teacher quality, and measures of teacher working conditions. This study complements prior work that describes multilevel matching for COSs (Page et al., 2019) and compares estimates to empirical benchmarks drawn from existing CRTs (Keele et al., 2019).

2 Elementary Support Model

In 2015-16, Wake County, the largest school district in North Carolina and the 14th largest in the nation, followed many other large districts by implementing a school turnaround strategy. The Elementary Support Model (ESM) provided supports across a wide range of dimensions to the district's 12 chronically lowest performing elementary schools. To identify these schools, district staff developed an index that accounted for academic, human capital, behavioral, and socioeconomic indicators and was averaged over the previous three years. The 12 elementary schools ranking at the bottom of this index were assigned to the ESM treatment condition and received a range of supports over the next three years, including governance reform, additional staffing, and instructional coaching. The three-year model cost $10.5 million, an average of nearly $300K annually per school.

Our analytic sample consists of a three-year panel, 2015-16 to 2017-18, of all schools and students in the district. In the fall prior to ESM's launch, the district had 106 elementary schools—12 ESM and 94 non-ESM. Table 1 shows that ESM schools had significantly larger populations of Black and Hispanic students, as well as more students receiving free or reduced price lunch, classified as Limited English Proficient, and receiving special education services. Given the considerable imbalance between ESM schools and non-ESM schools, we require a robust multilevel matching approach in order to derive control groups that are balanced on observables.

3 Method

To estimate the causal effect of ESM, we follow the framework outlined in Keele et al. (2019). To identify a control group that is balanced on observables, we first conduct student matches using every possible pairing of an ESM school and a non-ESM school. We then assess the quality of these matches by examining (1) covariate balance and (2) sample size. After matching at the student-level, we match at the school level using a distance matrix that was stored during student matching. Finally, we use refined covariate balance (Pimentel et al., 2015) in order to optimize school-level matches. The optimal match is drawn from, at most, the 94 non-ESM schools identified in Table 1.

At the analysis stage, we use a number of different specifications to estimate the impact of ESM. Our preferred specification follows Keele et al. (2019), who demonstrate in their validation of multilevel matching using experimental benchmarks that matching plus regression reduces bias more so than either method in isolation.

4 Conclusion

Despite the fact that cluster randomized trials (CRTs) constitute a gold-standard approach to estimating causal effects, such approaches are not always feasible. School districts generally assign treatment at the group level using non-random assignment. This may be due to a number of

Table 1: Pre-treatment balance between non-ESM and ESM schools, 2015-16

	(1) ESM Schools		(2) Non-ESM Schools		T-test Difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)
Male %	12	0.514	94	0.513	0.000
		(0.008)		(0.002)	
Asian %	12	0.026	94	0.078	-0.051
		(0.006)		(0.012)	
Black/African American %	12	0.467	94	0.213	0.255***
		(0.043)		(0.013)	
Hispanic/Latino %	12	0.351	94	0.177	0.175***
		(0.036)		(0.009)	
White %	12	0.117	94	0.495	-0.377***
		(0.023)		(0.019)	
FRPL %	12	0.760	94	0.359	0.401***
		(0.020)		(0.020)	
LEP %	12	0.222	94	0.110	0.112***
		(0.030)		(0.006)	
Special Education %	12	0.123	94	0.107	0.016**
		(0.006)		(0.003)	

Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

factors, including an insufficient number of schools within which to randomize or a predetermined assignment process that targets a select group of schools. As a large district, Wake County did have a sufficient number of schools. However, district leadership intentionally designed ESM for the dozen chronically low-performing schools that were identified on the basis of an internally-developed ranking index. As such, no suitable control group existed on the basis of the ESM index itself. By applying a multilevel matching strategy to the ESM model in a clustered observational study setup, we uncover arguably causal estimates for a comprehensive school turnaround effort in one of the nation's largest school districts.

References

Bryk, A. S., Gomez, L. M., Grunow, A., and LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.

Dee, T. (2012). School turnarounds: Evidence from the 2009 stimulus. Technical report, National Bureau of Economic Research.

Dickey-Griffith, D. (2013). Preliminary effects of the school improvement grant program on student achievement in texas. *The Georgetown Public Policy Review*, 2:21–39.

Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., Boyle, A., Upton, R., Tanenbaum, C., and Giffin, J. (2017). School improvement grants: Implementation and effectiveness. ncee 2017-4013. *National Center for Education Evaluation and Regional Assistance*.

Fryer, R. G. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *The Quarterly Journal of Economics*, 129(3):1355–1407.

Keele, L., Lenard, M., Miratrix, L., and Page, L. (2019). Matching methods for clustered observational studies in education.

McGuinn, P. J. (2006). *No Child Left Behind and the transformation of federal education policy, 1965-2005*. Univ Pr of Kansas.

Mehta, J., Schwartz, R. B., and Hess, F. M. (2012). *The futures of school reform*. Harvard Education Press.

Page, L. C., Lenard, M., and Keele, L. (2019). The design of clustered observational studies in education.

Pimentel, S. D., Kelz, R. R., Silber, J. H., and Rosenbaum, P. R. (2015). Large, sparse optimal matching with refined covariate balance in an observational study of the health outcomes produced by new surgeons. *Journal of the American Statistical Association*, 110(510):515–527.

Sun, M., Penner, E. K., and Loeb, S. (2017). Resource-and approach-driven multidimensional change: Three-year effects of school improvement grants. *American Educational Research Journal*, 54(4):607–643.