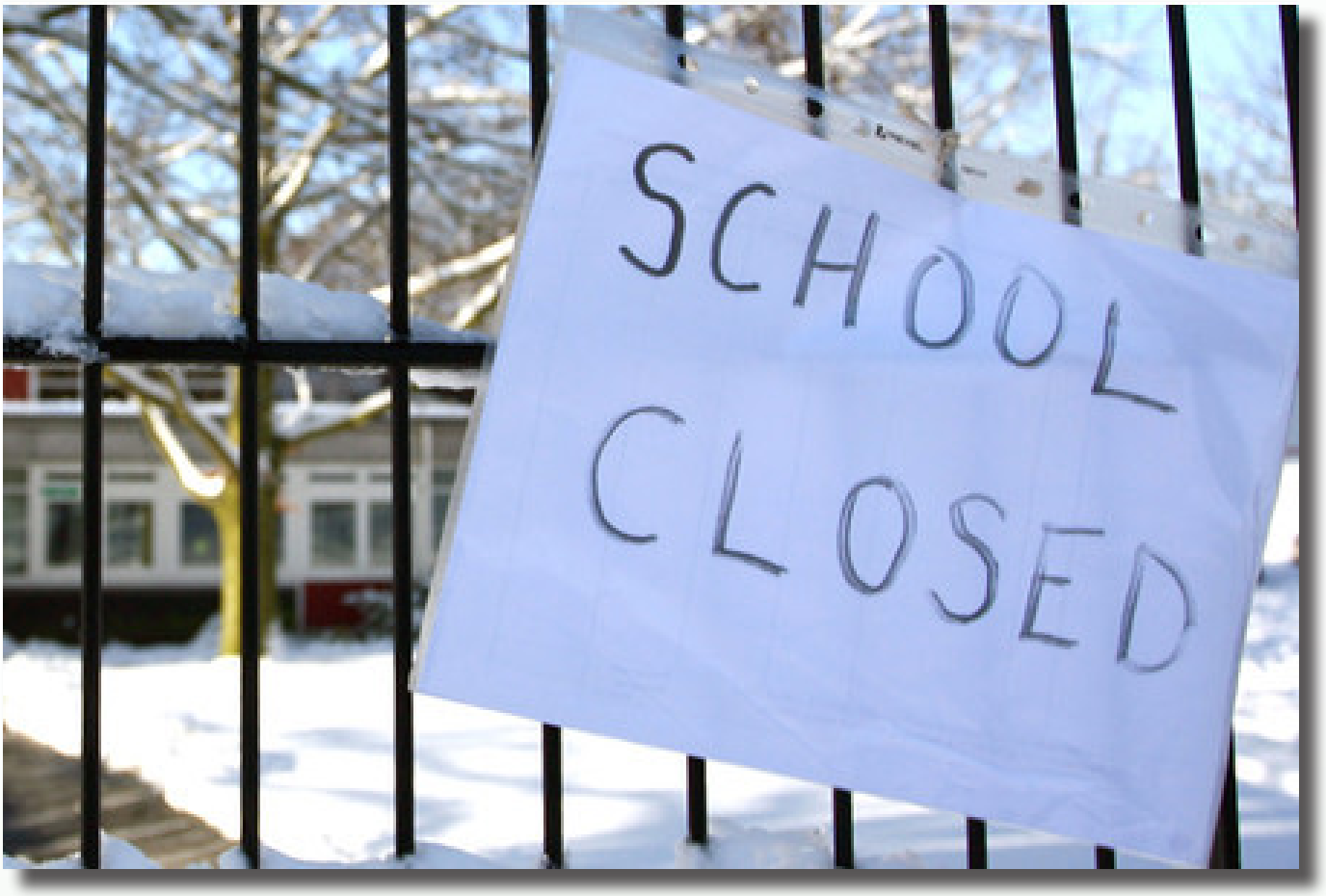
NEPC Review: Lights Off: Practice and Impact of Closing Low-Performing Schools (CREDO, August 2017)



Reviewers:

Matthew Gaertner, SRI International
Ben Kirshner, University of Colorado Boulder

October 2017

National Education Policy Center

School of Education, University of Colorado Boulder
Boulder, CO 80309-0249
(802) 383-0058
nepc.colorado.edu

Kevin Welner
Project Director

William Mathis
Managing Director

Alex Molnar
Publishing Director

Publication Type: NEPC Reviews are expert third-party reviews of selected non-peer-reviewed publications. Using academic peer review standards, reviewers consider the quality and defensibility of a report's assumptions, methods, findings, and recommendations. Written in non-academic language, reviews are intended to help policymakers, reporters, and others assess the social science merit of a reviewed report and to judge its value in guiding policy.

Funding: This review is one of a series made possible in part by funding from the Great Lakes Center for Education Research and Practice.

Suggested Citation: Gaertner, M. & Kirshner, B. (2017). *NEPC Review: "Lights Off: Practice and Impact of Closing Low-Performing Schools."* Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/thinktank/review-closures.

This material is provided free of cost to NEPC's readers, who may make non-commercial use of thematerial as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.

NEPC
NATIONAL EDUCATION POLICY CENTER

NEPC Review: Lights Off: Practice and Impact of Closing Low-Performing Schools (CREDO, August 2017)

Reviewers:

Matthew Gaertner, SRI International
Ben Kirshner, University of Colorado Boulder

October 2017

Summary of Review

This report provides an extensive analysis based on the most comprehensive dataset ever assembled for school closure research, including 1,522 low-performing schools that were closed across 26 states between 2006 and 2013. The report finds that even when holding constant academic performance, schools were more likely to be closed if they enrolled higher proportions of minority and low-income students. It also finds test score declines, relative to the comparison group, for two groups of students displaced by closures: those who transferred to schools with a prior record of relatively lower test-score performance and those who transferred to schools with equivalent past test-score performance. The slightly less than half of students who transferred to higher performing schools showed academic improvement relative to their matched peers. In general, although we found this to be a careful and rigorous study, we see a few missed opportunities. First, the report's focus on some tenuous analyses (involving pre-closure transfers) obscures its most important findings – disproportionality in school closures and inadequate numbers of higher quality receiving schools, leading to performance declines for most. Second, we are concerned about statistical modeling choices and matching challenges that may threaten the validity of subgroup analyses (charter school students). Finally, we would have liked to see the report acknowledge the inescapable moral dimensions of school closure: The communities most likely to be negatively affected are unlikely to have participated in closure decisions.

NEPC
NATIONAL EDUCATION POLICY CENTER

NEPC Review: Lights Off: Practice and Impact of Closing Low-Performing Schools (CREDO, August 2017)

Reviewers:

Matthew Gaertner, SRI International
Ben Kirshner, University of Colorado Boulder

October 2017

I. Introduction

School closures began to proliferate in the mid-2000s and, after about a decade, researchers began to respond to this controversial education policy issue with an emerging body of findings. The evidence base includes by our estimate 16 refereed journal articles and many more white papers, book chapters, and technical reports. This is our second review of a major study of the practice and impact of school closures; the first focused on a 2015 Fordham Institute evaluation of school closure outcomes in Ohio.[1] We are heartened that more scholars are taking up this question, given the intense debates about the rationales and consequences of school closure.

Two questions are at the forefront of debates about closure as a school reform strategy: do students from closed schools, on the whole, do better? And, are closures decided fairly, or is there a disproportionate burden on low-income communities of color?

Prior studies can be difficult to synthesize because they have vastly different datasets, but there appear to be some trends in the research: when displaced students attend higher performing schools, there appears to be evidence that their academic performance improves, more consistently in math than in reading.[2] However, displaced students often do not have the option of attending higher performing schools. And when they attend equivalent or lower performing schools, their scores either declined or do not improve.[3]

With regard to disproportionality in closures, a number of qualitative case studies have documented the politics of closure and described how powerful elites target schools in communities of color for closure.[4] To our knowledge there have not been larger-sample quantitative analyses testing the existence of racial bias in closure decisions, when controlling for income or prior academic performance.

We therefore welcome a new report: *Lights Off: Practice and Impact of Closing Low-Performing Schools*, authored by Chunping Han and colleagues at the Center for Research on Education Outcomes (CREDO).[5] The extensive report is divided into two volumes (Volume I and Volume II, with a total of 44 figures and 28 tables across both volumes). The report analyzes schools' academic performance prior to closure, the demographic differences between schools that are closed and schools that remain open, and student outcomes in the years prior to and after the closures. In this review we examine the data, the methods, and the assumptions that drive the report's key findings. We describe not only the extent to which the report has marshalled sufficient evidence to support its claims, but also the report's relevance for school closure policy and practice.

II. Findings and Conclusions of the Report

The report offers a comprehensive study of 1,522 low-performing schools across 26 states that were closed between 2006 and 2013. Of these, 1,204 were traditional public schools (TPSs) and 318 were charters. Volume I focuses on the landscape of school closures and examines how closures are distributed across sectors, years, states, grade spans, demographic groups, and locales. In the three years leading up to closure, schools that were ultimately closed showed steeper declines in academic performance than other low-performing schools. The report also found that, holding constant academic performance, schools with higher proportions of African American and Latino students, as well as schools with students from higher poverty families, were more likely to be closed. This second finding suggests racial bias in the application of school closure decisions.

Volume II focuses on the performance of students who attend closing schools, including their transfers before and after closure and their post-closure academic progress. More than 30 percent of students left closing schools in the year before the schools were closed, and students who left early ("early leavers") had worse pre-closure academic performance than students who remained until the official closure. Following a closure, slightly less than half of displaced students transferred to higher performing schools. The quality of the post-closure receiving school was a major predictor of student performance. Students who attended better schools showed improvement relative to the comparison group; those who attended equivalent or worse schools (the slight majority) showed declines relative to the comparison group.

III. The Report's Rationale for Its Findings and Conclusions

The report's rationale is based on empirical analyses of data collected from 26 states focused on two sets of schools: TPSs and charters. A number of empirical analyses examined features of closed and non-closed schools and the performance of students in both contexts.

IV. The Report's Use of Research Literature

The report provides a brief review of the research literature that identifies main trends in the field. The authors cite several relevant studies. They conclude that studies have come to different conclusions and that, broadly speaking, the literature base is still too fragmented and inconsistent to claim clear trends, particularly about effects on student performance.

V. Review of the Report's Methods

Strengths

This report benefitted from a dataset that was far more comprehensive than any that has been used for published school closure analyses in the past. The authors were able to draw on closures across 26 states and 7 academic years. In fact, many of the closure studies with which we are well acquainted focused on closures that were likely included in this report's dataset.[6] Data comprehensiveness is important because although some studies may suffer from generalizability limitations (deriving entirely from one school or one state), this study does not. It is reasonable to suspect this report therefore captured more variation – in state and district policies and in the quality of post-closure receiving schools – than any study that has come before. The report authors also made some judicious analytic choices that boost the credibility of their findings. First, the application of Virtual Control Records (VCRs) provided a way to ensure close matches between students who experienced a closure and students who did not. The procedures for creating VCRs are described clearly and appear superior in many respects to other matching methods, such as propensity scores. VCRs require close or exact matching on all pre-closure covariates, while propensity score matching requires close matching on the single propensity score, which may obscure large differences on the covariates that contributed to that propensity score. Based on our reading, the VCR procedure is nearly identical to coarsened exact matching – a method which has been applied sparingly in educational research[7] but is a widely cited tool in the statistics literature for minimizing bias in observational studies.[8] Finally, and perhaps less prominently given the expansive scope of the report, the authors wisely elected to exclude students in closing schools' top grade from their impact analysis (e.g., 8th graders moving on to high school). These students would be heading to a new school regardless, so a school closure would not impose forced mobility on them the way it would on other students.

Weaknesses

In this section we discuss methodological issues that may impact the validity of the report's conclusions. For the sake of clarity and ease of presentation, methodological issues are parsed according to the relevant report sections.

Are similarly low-performing schools treated equally? (Volume I, Section 4)

The report states that closure rates were highest for high-minority and high-poverty schools in the lowest ventile (each ventile contains five percent of the full distribution). That may signal inequitable treatment, but there are other plausible explanations. Specifically, were these high-minority and high-poverty schools still lower achieving than other schools within that ventile? It would seem as though restricting the analysis to a single ventile "controls for" prior school achievement, but there is substantial achievement variation within the ventiles furthest from the middle of the school achievement distribution. It would be useful to know whether – within the lowest achievement ventile – mean scale scores were still negatively correlated with school closures, and further, whether that association explains part of the relationship between minority rates, poverty rates, and closure.

Pre-closure transfer of students (Volume II, Section 5)

This report's "early leavers" analysis – focused on students who left a closing school in the year prior to the closure – relies on questionable assumptions. The report shows that early leavers generally had lower levels of achievement than those who stayed until the actual closure (Vol. II, p. 4), and that early leavers were more likely than other students to transfer to a superior school post-closure (Vol II, p. 10). The report then makes unwarranted logical leaps, suggesting some parents were "in the know" (p. 12) about the quality of their children's schools, and, because the percentage of early leavers was higher in charter schools, parents of charter school students are savvier about finding better educational options for their children. However, there may be many reasons students left a soon-to-be-closing school, perhaps unrelated to school quality or the closure itself. The authors' arguments would be more compelling if they had (1) examined students in schools that announced their closure a year early and then (2) tracked those students' progress, separately for stayers and early leavers. A closure announcement date would give readers some reassurance that the early leavers were leaving because they knew the school was closing. Moreover, closure announcement dates would support analyses comparing outcomes for students from schools that closed with advanced warning versus students from schools that closed suddenly. Unfortunately, closure announcement dates were not available, so the early leavers analyses ultimately provide counter-intuitive findings rather than policy-relevant insight.

Quality of post-closure schools (Volume II, Section 5)

In this section, the report summarizes the quality of post-closure schools by dividing the sample into achievement ventiles and reporting the proportions of students who transferred to an inferior, equivalent, or superior school (at least two ventiles above the closing school). However, the ventile approach to determining superior post-closure schools threw away some information. For the analysis of post-closure student flows, the classification scheme is understandable. The authors want to present proportions of students headed to a worse, equivalent, or better school, and presenting proportions requires categories. But is the loss of information worth it? Ventiles impose an arbitrary classification scheme on test scores,

and the bottom ventile will be much wider (in terms of mean scale score range) than the middle ventile, assuming mean scale scores are normally distributed. Moreover, a receiving school 12 ventiles above a closed school (e.g., at the 65th percentile relative to a closed school at the 5th percentile) was treated the same in terms of superiority as a receiving school two ventiles above the closed school (e.g., at the 15th percentile relative to a closed school at the 5th percentile). For the analysis of post-closure schools' quality and for the subsequent analysis of impact, why not measure the impact of receiving schools' distance (in terms of mean scale scores) from the closed school?

Post-closure student academic performance (Volume II, Section 6)

The report rightly highlights the post-closure impact analysis as one of the most important sets of research questions it addresses. The statistical analysis relies on two key steps – a matching procedure for creating VCRs and a statistical model, both detailed in the Technical Appendix of Volume II. Our review of these analyses covers both steps – matching and modeling.

Matching. The procedures for creating VCRs are described in detail, enabling a reader to essentially replicate the process. But the authors clearly encountered problems matching charter school students to virtual peers (Table 28, p. 48). Only 61% were matched. By contrast, 90% of TPS students were matched. Matched charter school students are slightly more than half as likely to be students with special needs, relative to the full (matched + unmatched) population. The authors also employ a minor sleight of hand by comparing matched students to the full sample; the reader should be able to compare matched students to *unmatched* students to understand the differences between who made it into the analysis and who did not. Furthermore, although the report shows demographic comparisons between matched students and the target sample, achievement comparisons (test scores) are conspicuously absent.[9] Given the differences between matched and unmatched charter school students, it would be helpful to know if the authors considered applying analysis weights so the matched sample would mirror the target population.

Modeling. The authors use a difference-in-differences approach (equation 1, p. 48), which models one-year test-score changes as a function of attending a closed school, the quality of the post-closure school, and other pre-closure covariates. The difference-in-differences method is appropriate for this analysis, but the authors did not appear to use multilevel modeling or any other correction for clustering (e.g., Taylor series linearization), nor did they include school fixed effects.[10] This analytic choice deserves an explanation, because standards for causal research generally require some statistical correction for clustering (that is, nesting of students within schools), unless between-school variance is negligible.[11] The consequences may be non-trivial. Without clustering corrections, standard errors and confidence intervals will be artificially small and the likelihood of a type I error will increase, which means researchers will find statistical significance where they should not.

VI. Review of the Validity of the Findings and Conclusions

For the most part, we find the authors have marshalled the necessary evidence and performed the appropriate analyses to back their claims, although for the impact analysis, some methodological questions remain.

Strengths

Throughout Volumes I and II, the report's conclusions are measured and focused; the authors tend not to reach beyond their data when making headline claims. For example, based on the authors' executive summary and our own understanding of the school closure literature, two of the most important claims in this report are (1) fewer than half of the students displaced by a closure enrolled in a superior school immediately after the closure, and (2) a closure's effects on student achievement depend heavily on the post-closure receiving school. We find no reason to suspect these claims, stated as such, are invalid.[12]

Weaknesses

In this subsection we address methodological issues and unwarranted claims that raise doubts about some of the report's conclusions.

Methodological issues

The authors did not report using or considering any corrections for clustering with a dataset that clearly contained students nested within schools. Therefore, the analyses that uncovered small effect sizes significant at the 0.05 level may in fact uncover null effects when appropriate clustering corrections are applied. This problem is compounded by the reporting of small effects for charter school students, because the VCR method failed to match nearly 4 in 10 charter school students. We are less concerned about the headline impact findings. Although clustering corrections will reduce statistical power, larger effect sizes require less power to detect, all else equal. So it is reasonable to assume the authors' claims about the importance of the quality of the receiving school (e.g., Vol. II, Figure 38, p. 23) are not sensitive to clustering corrections.

Unwarranted claims: Early leavers, post-closure flows, and charter school parents

The authors use their analyses of early leavers and students' post-closure schools to make unexpected and unwarranted claims about the dispositions and experience of charter school students' parents (for example "The proportion of superior placement was higher for charter closure students than for TPS closure students, an indication of the stronger experience

of parents of charter school students in navigating the school choice landscape" [Vol. II, p. 13]). There are four problems with claims like this. First, as discussed above, low match rates for charter students and the absence of clustering corrections raise doubts about the credibility of claims based on small effect sizes for charter students. Second, it is impossible to isolate the mechanism that caused early leaving, because the authors' data did not include the dates of closure announcements. Third, as the report acknowledges (p. 10), a relatively large proportion of charter school students left closing schools that were at the bottom of the achievement spectrum. In other words, in their post-closure destinations, these charter school students may have had few places to go but up. Fourth, and most importantly, the authors did not need to use their school closure analysis to make claims about the comparative knowledge and experience of charter school students' parents. Even if it is now received wisdom that charter school parents are comparatively savvy, this was not the report's focus, the authors did not have the data to sufficiently back up their claims, and by association these tenuous arguments undermine other (much more credible) findings.

VII. Usefulness of the Report for Guidance of Policy and Practice

This report's comprehensive analyses are quite useful for understanding three key facets of school closure as a policy lever: (1) which communities are typically targeted in closure decisions; (2) the set of post-closure options usually available to those affected; and (3) the typical student outcomes associated with each of those options. These insights alone are policy-relevant and novel, given the unprecedented size and scope of the report's dataset.

Based on this evidence we see two reasonable policy prescriptions, offered elsewhere but never before backed by such comprehensive analysis. First, be fair. If states and districts are going to close schools, they should make such decisions based on performance, not on schools' economic or demographic compositions, and ensure that certain communities are not disproportionately burdened by the disruption and displacement that can follow from closure. Second, do no harm. States and districts should ensure there is a surplus of higher performing nearby schools with adequate capacity to educate displaced students.

As we have noted in other writing, we also want to point out that debates about school closure should not be adjudicated solely on consequentialist arguments based on empirical findings. Decisions about school closure and broader questions of reform raise normative political questions about participation and rights – including questions about the voices of students and their parents. Although this study did not take up such normative questions, we believe they are an important part of policy discussions about closures.

Notes and Resources

1 Kirshner, B., & Gaertner, M. (2015). *Review of School Closures and Student Achievement: An Analysis of Ohio's Urban District and Charter Schools*. Boulder, CO: National Education Policy Center. Retrieved October 2, 2017, from http://nepc.colorado.edu/thinktank/review-school-closures

2 Carlson, D. & Lavertu, S. (2015). *School closures and student achievement: An analysis of Ohio's urban district and charter schools*. New York, NY: Thomas Fordham Institute;

de la Torre, M., & Gwynne, J. (2009). *When schools close: Effects on displaced students in Chicago Public Schools*. Chicago, IL: Consortium on Chicago school research.

3 Engberg, J., Gill, B., Zamarro, G., & Zimmer, R. (2012). Closing Schools in a Shrinking District: Do Student Outcomes Depend on Which Schools are Closed? *Journal of Urban Economics, 71*(2), 189-203.

4 See Finnigan, K., & Lavner, M. (2011). A political analysis of community influence over school closure. *The Urban Review, 44*(1), 133-151;

Johnson, A.W. (2013). "Turnaround" as shock therapy: Race, neoliberalism, and school reform. *Urban Education, 48*, 232-256.

5 Han, C., Raymond, M., Woodworth, J., Negassi, Y., Richardson, W.P., & Snow, W. (2017). *Lights Off: Practice and Impact of Closing Low-Performing Schools*. Stanford, CA: CREDO.

6 Kirshner, B., Gaertner, M., & Pozzoboni, K. (2010). Tracing transitions: Understanding the impact of a school closure on displaced students. *Educational Evaluation and Policy Analysis,* 32(30), 407-429;

Brummet, Q. (2014). The effect of school closings on student achievement. *Journal of Public Economics, 119*, 108–24;

Carlson, D. & Lavertu, S. (2015). *School closures and student achievement: An analysis of Ohio's urban district and charter schools*. New York, NY: Thomas Fordham Institute.

7 For an example of coarsened exact matching applied in an educational research context, see Gaertner, M., & Hart, M. (2015). From Access to Success: Affirmative Action Outcomes in a Class-Based System. *Colorado Law Review, 86*(2), 431-475.

8 On the validity of the VCR method for use in non-experimental observational studies like this one, our review differs somewhat from Maul, A. (2015). *Review of Urban Charter School Study*. Boulder, CO: National Education Policy Center. Specifically, we do not see the VCR method as inferior to more heavily-cited methods in the educational research literature, such as propensity score matching. We believe the VCR method is essentially identical to coarsened exact matching. In fact, it would be useful for this report to explicitly cite in its Technical Appendix the other matching methods to which the VCR method is closely related, because if the VCR method is quite similar to coarsened exact matching, then it is supported by empirical analyses in the statistics literature. For example, Iacus, S.M., King, G., & Porro, G. (2011), Multivariate Matching Methods that Are Monotonic Imbalance Bounding. *Journal of the American Statistical Association, 106*, 345-361 demonstrate that coarsened exact matching can yield better balance between treatment and comparison groups, and can maximize precision and minimize bias for causal estimates, relative to propensity score matching.

9 On the problem of low match rates for the VCR method, our views are more closely aligned with Maul, A. (2015). *Review of Urban Charter School Study*. Boulder, CO: National Education Policy Center. The dilemma this report faces is a classic problem in the matching literature: Should researchers ease matching requirements to allow more students in the analytic sample, or impose restrictions (e.g., exact matching)

and exclude some students in order to obtain better matches? More thorough reporting would help the reader understand how the authors addressed this dilemma. For example, readers could be presented with two candidate matching approaches in the Technical Appendix, one with priority placed on maximizing the quality of matches, and another with priority placed on maximizing the proportion of students included from the target sample. Next, the authors could compare the two approaches according to two criteria: (1) baseline equivalence, that is, similar means on pre-closure covariates for the matched students and their virtual twins, and (2) the percentage of the target sample included or excluded. Finally, the authors could make a case for choosing one approach over the other, for example, choosing the method that maximizes baseline equivalence to guard against imbalance and bias, or choosing the method that includes the largest proportion of the sample to maximize generalizability. Under either approach, to allow the reader and other researchers to assess generalizability and comparability between matched students and their virtual twins, the Technical Appendix should present two separate tables comparing means on pre-closure covariates: (1) matched versus unmatched students; and (2) baseline equivalence, that is, matched students versus virtual twins.

10 Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed.)*. Newbury Park, CA: Sage.

11 What Works Clearinghouse (2016). *What Works Clearinghouse Procedures and Standards Handbook (Version 3.0)*. Washington, DC: Institute of Education Sciences, US Department of Education.

12 This closure impact study did not involve the random assignment of students to treatment (closure) or control (non-closure) conditions. Therefore, closure impact estimates may be influenced by non-random, unmeasured factors (for example, self-selection into the "treatment" group). As Maul notes in his critique of a charter schools impact study, neither the VCR matching method nor a propensity score matching method can account for such imbalance in unmeasured covariates (Maul, A. [2015]. *Review of Urban Charter School Study*. Boulder, CO: National Education Policy Center.). In the charter schools-versus-TPS literature, self-selection bias is a real threat. However, we do not see the same self-selection threats in school closure analyses. To our knowledge, parents and students do not elect to be enrolled in closing or non-closing schools a priori, and if they did, such a selection mechanism would be addressed sufficiently by matching on student-level pre-closure academic, demographic, and socioeconomic covariates.