American Institutes for Research Working Paper

Networked Improvement Community for Improving Access to Effective Teachers: Initial Evidence of Effectiveness

Natalya Gnedko-Berry, Trisha Borman, Candace Hester, and David Manzeske

Work on this paper was supported by the National Board for Professional Teaching Standards with funding from the Supporting Effective Educator Development grant by the U.S. Department of Education in 2013 (U367D130002) and 2015 (U367D150019). Send all inquiries about this study to Natalya Gnedko-Berry at ngnedko-berry@air.org. All opinions and conclusions in this paper are those of the authors.

Introduction

Decades of reform and innovation aimed at persistent challenges in prekindergarten through Grade 12 (PK–12) classrooms, such as inequitable access to effective teachers or achievement gaps among student subgroups, have not produced a lasting change. Even when successful approaches are identified, scaling these approaches to different contexts remains problematic. A group of scholars argues that innovation alone is not enough to transform education; instead, innovation must be combined with careful attention to the conditions that shape implementation in particular contexts (e.g., Bryk, Gomez, & Grunow, 2011; Bryk, Gomez, Grunow, & LeMahieu, 2015). Networked Improvement Communities (NICs) have emerged to that end. NICs—a type of a research–practice partnership—bring together stakeholders from different settings in a structured collaboration to address common problems of practice, allowing for learning from the implementation in different contexts, potentially at an accelerated pace. NICs are grounded in the principles of improvement science: They implement small tests of change, analyze results, and repeat as necessary before an intervention is defined and a larger scale implementation takes place (Bryk et al., 2011; Bryk et al., 2015; Coburn, Penuel, & Geil, 2013; Proger, Bhatt, Cirks, & Gurke, 2017; Russell et al., 2017).

Championed by the Carnegie Foundation for the Advancement of Teaching, NICs have grown in popularity, as have the private and federal resources to support them, including the U.S. Department of Education and large foundations, such as the Spencer Foundation, the William T. Grant Foundation, and the Bill & Melinda Gates Foundation. With this increase in investment, the research on NICs has begun to accumulate. For example, the theoretical work by Bryk and colleagues is largely credited with the introduction of the NIC concept to the field of education (Bryk & Gomez, 2008; Bryk et al., 2011; Bryk et al., 2015). Additional early papers describe NICs and offer guidance on ways to launch a NIC successfully. For instance, using the case study approach, Hannah, Russell, Takahashi, and Park (2015) examined the Building a Teaching Effectiveness Network—a NIC comprising three school districts in partnership with the Carnegie Foundation, researchers, and a number of health and education agencies. This NIC was established to address the high rates of novice teacher turnover. The study of this NIC led to the authors' conclusion that leadership, relationships, structured measurement, flexibility in implementation, and continuous engagement of members were some of the essential components for a NIC to address a problem of practice in concrete and measurable ways, as other authors have also pointed out (e.g., Baron, 2017; Martin & Gobstein, 2016). Drawing on four network case studies, Russell and colleagues (2017) have synthesized guidance for launching a NIC by proposing an initiation framework that identifies five essential components: a theory of improvement, the use of improvement research methods (i.e., improvement science), a measurement infrastructure, strong leadership, and shared culture and norms. The authors argue that these components allow a fledgling NIC, which is likely to face early challenges of uniting individuals from different organizations and with different priorities, to develop into a mature and sustainable NIC.

Research on the outcomes of NICs is also beginning to emerge. For example, Hoang Huang, Sulcer, and Yesilyurt (2017) descriptively examined 5-year results of the Carnegie Math Pathways, structured as a NIC of college faculty and administrators, researchers, and curriculum developers representing up to 30 colleges and organized around the problem of low rates of completion of remedial math courses in community colleges. This NIC facilitated the implementation of alternative remedial math courses, Statway and Quantway. By Year 5 of implementation, the proportion of students completing remedial course requirements in 1 year was as much as 44% greater for students who took the alternative courses, compared with students who took the traditional math courses (the comparison group consisted of students from the same colleges prior to the implementation of alternative courses). The Centre for the Use of Research and Evidence in Education conducted a review of 19 evaluations of the Networked Learning Communities, the definition of which in many respects resembles that of NICs. That is, similar to NICs, the Networked Learning Communities are groups of people and organizations, typically schools, that come together to address shared problems of practice by learning together. The review found strong to medium evidence (researchers determined the weight of evidence on the basis of the type and strength of data) of a positive impact of the Networked Learning Communities on student outcomes in nine studies, including academic attainment and engagement. The review also found strong to medium evidence of a positive impact of Networked Learning Communities on teacher outcomes in 11 studies, including gains in teacher knowledge and skills, such as improvements in classroom instruction and use of research. Of the 19 studies included in the review, seven had a comparison group, although it is not clear how the comparison groups were determined. All networks included in the review consisted of at least three schools with additional external collaborators (Bell et al., 2006).

Although the body of research on NICs is growing, the prevalence of theoretical and descriptive papers leaves many unanswered questions about the effectiveness of this approach—a gap in knowledge that the current study begins to address by examining the effectiveness of a NIC implemented on a large multistate and multidistrict scale.

Background and Study Setting

The current study is grounded in the work conducted by the National Board for Professional Teaching Standards (National Board) to increase the proportion of U.S. teachers who are National Board certified. Since 1987 the National Board has been providing a voluntary certification for teachers who can demonstrate that they meet rigorous standards of accomplished teaching. These standards are based on the five core propositions that underscore the teachers' commitment to and responsibility for student learning (National Board, 2016). Research shows that National Board–certified teachers (NBCTs) have a positive impact on student learning (Cantrell et al., 2008; Clotfelter, Ladd, & Vigdor, 2007; Cowan & Goldhaber, 2015; Manzeske et al., 2017) and that the positive impact can be greater for minority and low-income students compared with their mainstream peers (Cavalluzzo, 2004; Goldhaber & Anthony, 2007). However, only about 3% of the nation's teaching workforce is currently Board certified.

With funding from the Supporting Effective Educator Development grant by the U.S. Department of Education for the period from 2013 to 2018, the National Board launched the Network to Improve Teaching (Network) with an overarching goal to improve student access to effective teachers, defined by the Network as teachers who are National Board certified. To achieve this goal, the Network is guided by three aims: (1) strengthen the support of new teachers to enable their development into accomplished teachers who are ready to pursue National Board certification, (2) increase the number of teachers who are National Board certified, and (3) leverage the instructional expertise of existing NBCTs by increasing their representation in instructional leadership roles.

Structured as a NIC, the Network advanced in two phases. In Phase 1, which launched in 2013–14, the Network included six sites: the states of Arizona, Kentucky, New York, and Washington; and public school districts of San Francisco, California, and Albuquerque, New Mexico. In phase two, which launched in 2015–16, the states of Alabama, North Carolina, and Illinois, and the public school district of Clark County, Nevada, joined the Network, and the public school district of Albuquerque, New Mexico, transitioned to a statewide implementation. Currently, the Network comprises 10 sites: eight states and two public school districts. A "home team" leads the Network at each site. A home team includes state representatives, who guide the work (district representatives at the two district-based sites). These representatives are senior leaders from each participating organization, usually a state education agency, a state union, and an education nonprofit. A home team also includes an NBCT, who brings classroom expertise, and district representatives from a select group of districts that pilot the Network activities. A site director at each site oversees the day-to-day work of the Network, working with and across the agencies to ensure that planning, testing, data collection, reporting, and operations are implemented as intended. The site director oversees the work on a full-time basis, with the remaining home team members contributing to various degrees, often combining their Network responsibilities with employment in a school, district, or another education agency. In its current state, the extended Network, including "home teams" and additional members, includes approximately 385 individuals and spreads across 120 schools, 50 districts, and 10 states.

The work of the National Board's Network is grounded in semiannual 2.5-day in-person learning sessions attended by a "travel team" from each site that includes seven or eight core team members from the state agencies, districts, and schools, as well as the site director. The learning sessions serve as a foundation for cross-site collaboration and learning, as well as opportunities for the relationship and identity building. Action periods between the learning sessions are the time for participating sites to engage in testing, learning, and improvement cycles. The work at each site during these periods is specific to the local context, taking place at the school, district, and state levels, although the emphasis depends on the direction that each site has chosen. In the action periods, the sites report progress monthly to the leadership team, attend virtual network meetings, and engage in unstructured communication with one another. Twice a year, the site directors attend in-person meetings organized and facilitated by the National Board project team, to engage in focused discussion about the individual and shared needs of their sites, Network operations, and as another opportunity to strengthen relationships.

The sequencing of the Network's main activities has not changed since it launched: 2.5- day learning sessions, separated by the 6-month action periods. However, as the Network advanced from Phase 1 to Phase 2, some of its activities and structures were altered or refined. For example, while some Network activities had taken place in schools since its inception, the focus on school-level implementation became more intentional in Phase 2, when pilot districts—and specific pilot schools within the districts—were identified. Using the National Board's professional development tools and frameworks, such as the content-specific National Board standards, had always been part of the Network's guidance for candidate learning. However, in Phase 2, the work of pilot schools became more directly focused on embedding the National Board Accomplished Teaching Body of Knowledge, of which the standards are a part, as the set of professional language and expectations to which the school adhered throughout its professional learning structure. The travel teams were also established in Phase 2. Previously, attendance at the learning sessions allowed some flexibility, although a core team of members from each site, including the site directors, attended all sessions. Supports provided by the Network have also matured. For example, beginning in Phase 2, the sites began receiving support for using the National Board certification data in the form of formal meetings with a data expert.

The National Board project team, in collaboration with content experts and researchers (other than AIR researchers who serve as external evaluators), and stewardship from the

Carnegie Foundation, remained the leader of the Network in both phases. The leadership team creates and refines the theory of action that guides the Network. This theory is grounded in a driver diagram that defines change ideas and measurement parameters. The leadership team provides ongoing support to the Network sites for identifying local goals and Network membership, sourcing information on relevant subject matters, use of improvement methods, communication, operations, and any additional needs that emerge. The leadership team also organizes and leads the learning sessions. Some of the leadership team functions have been refined over time. For example, the role of improvement advisor (specialist in improvement science methodology) was added to the team during the Network's second year.

The current study examines the relationship between the sites' participation in the National Board's Network and the number of teachers pursuing National Board certification at these sites, an increase in which is one of the Network's primary aims. First, we describe the study's scope and methods, including participants, data, our approach to creating a comparison group, and our approach to analysis. Next we present the study's results, followed by conclusions and the discussion of limitations.

Methods

The study examined the effectiveness of the National Board's Network during Phase 1 for 2 academic years, 2014–15 and 2015–16. We did not study effects for 2013–14, although the Network launched then, because the sites were engaged in planning and not the recruitment activities that would have been able to produce a change in the study's outcome. Below, we describe the study's participants and outcome in more detail, in addition to data, our approach to constructing a comparison group, and analysis.

Participants and Outcome

Participants in the study were the six sites that were part of the National Board's Network in Phase 1: the states of Arizona, Kentucky, New York, and Washington; and public school districts of San Francisco, California, and Albuquerque, New Mexico. The outcome in the study was persistence in earning the National Board certification. We operationalized persistence as the number of educators who submitted at least one certification component by the end of an academic year, which was the minimum achievement within that time frame required to advance to the next year. We examined persistence for educators who initiated the certification process in 2014–15 or 2015–16. We examined persistence for these educators' first year of pursuit only (that is, we did not follow these educators over time). Our outcome was persistence instead of full certification because the full Board certification requires a submission of four certification components and takes up to 4 years—a time frame beyond the scope of the current study. The National Board implemented the 4-year certification process in 2014–15. Before that, educators could submit all certification components and become certified within 1 year.

Data

We used extant data collected by the National Board from educators who had initiated the certification process. These data were self-reported, and included information such as age, gender, race, and school assignment. On the basis of the school information provided by candidates, we merged additional school characteristics into the data files from the Common Core of Data, such as the proportion of students per school receiving free or reduced-price lunch, the proportion of non-White students, and school enrollment. The years examined spanned 2000–01 to 2015–16. To examine the effectiveness of the Network, we studied the years 2014–

15 and 2015–16. The number of raw individual records was 113,177. We removed 33,009 records for one or a combination of three reasons: duplicate records, records that were missing information necessary to determine whether an educator was from a Network or non-Network site (i.e., state or city), and records from the five sites that joined the National Board's Network in 2015–16. For this last condition, we reasoned that, although the additional sites were not yet implementing candidate recruitment activities in 2015–16, planning activities could have prompted some emphasis on recruitment, which would have made these sites inappropriate candidates for inclusion in a possible comparison group pool. Of the 80,168 remaining records, 18,880 were from educators from the six Network sites, and 61,288 were from educators from the non-Network sites.

Comparison Group

To select a comparison group of non-Network sites, we used propensity score matching. This technique, which is commonly used in studies that cannot accommodate random assignment, allows for identifying nonparticipating units that, on the basis of their observable characteristics, are most similar to participating units. Before matching, we needed to aggregate the individual records to the level of analysis, which we intended to be the site level. However, the Network sites consisted of two different geographic units: a state and a public school district. The public school districts included in the Network—San Francisco, California, and Albuquerque, New Mexico—encompass an entire city. Therefore, we decided to reconfigure states into cities to have a uniform geographic unit for all sites. With this approach, we treated state sites as a collective of cities. Our city-level data file included all cities that appeared in the National Board data at least once. For a city to "appear," it had to have had at least one educator captured in the National Board data in any year available to us. The number of Network and non-

Network cities in the city-level data, by year, is shown in Table 1. To ensure that the data were balanced across time, any city that had appeared in any year was preserved in the data in all years, even if that city had zero educators pursuing or achieving certification for one or more years. Maintaining observations for cities in years in which there were no educators necessarily introduced missing values in our data. To incorporate data for the cities' characteristics in the years when they had no educators pursuing or achieving certification (and thus no corresponding school or district information), we generated values for these missing data using an average of any values for that city within a 5-year window (i.e., a 5-year moving average imputation).

Table 1: Number of Network and Non-Network Cities, by Year

Year	Network cities	Non-network cities
2000–01	144	1,635
2001–02	226	1,773
2002–03	239	1,775
2003–04	262	1,679
2004–05	252	1,470
2005–06	291	1,463
2006–07	305	1,497
2007–08	323	1,493
2008–09	364	1,400
2009–10	381	1,215
2010–11	364	987
2011–12	296	914
2012–13	286	794
2013–14	297	684
2014–15	403	819
2015–16	393	845

We used propensity score matching to identify a comparison group of non-Network cities that, on the basis of their observable characteristics for 5 years, from 2009–10 to 2013–14, were most similar to the Network cities. We were unable to include data from 2000–01 to 2008–09 in the matching component of the analysis because of the high degree of multicollinearity for each characteristic across time. For example, the number of students served by NBCTs varied little across years, and thus, including all years of data for each of these variables prevented the estimator from converging. Variables that we used in matching included educator, school, and student characteristics, such as the educator's age and highest degree, school enrollment, and the proportion of students receiving free or reduced-price lunch. We also used patterns of National Board certification, such the number of certified teachers at the city and school level. Table A1 in the appendix describes all data used in the propensity score matching.

As the first stage in our analysis, we used a logistic regression model to predict participation in the Network (Equation 1):

$$Network_c = \sum_{t=2009}^{T} \left(\sum_{n=1}^{N} \alpha_{nt} x_c^n\right)^t + \sum_{t=2009}^{T} \left(\sum_{m=1}^{M} \beta_{mt} z_c^m\right)^t + \vartheta_c \quad (1)$$

In this model, *Network* takes on a value of 1 if a city is part of a Network site and a value of 0 otherwise. Each variable is indexed according to city, c. The models' parameter $\sum_{n=1}^{N} x_c^n$ and $\sum_{m=1}^{M} z_c^m$ represent vectors of the educator and school/student characteristics of each city's candidates. Each parameter is presented separately for the years 2009–10 through 2013–14, the final year prior to when the effects could be anticipated. The error term, ϑ_{ct}, captures residual variation that remains unexplained by the parameters in the model.

Results of the logistic regression estimated the "propensity" for each city to have been a Network city after 2013–14, on the basis of the available educator and school/student characteristics. We used these estimates to identify one "nearest neighbor," non-Network comparison city for each Network city if the non-Network city demonstrated the closest propensity to that of the Network city. These cities were selected without replacement, using the Stata psmatch2 nearest neighbor algorithm.

We investigated a number of variations in the matching model—for example, matching each Network city to nine non-Network cities—and evaluated the relative benefits of each of these models, using a measure of the bias reduction associated with each model and as compared with the full sample of comparison cities. The 1:1 nearest neighbor match, without replacement, produced the best overall reduction in the standardized bias estimate of 10 percentage points between the Network cities and matched non-Network cities, as compared with the Network cities and the full sample of non-Network cities. Table A2 and Table A3 in the appendix provide complete results for evaluating the final matching model, all of which are within the recommended boundaries (Rosenbaum & Rubin, 1983). Because the final matching model was a 1:1 match, the number of matched non-Network cities in our comparison group was the same as the number of Network cities shown in Table 1.

Analysis

Our approach to the analysis was a comparative interrupted time series (CITS; Bloom, 2003; Shadish, Cook, & Campbell, 2002). With this approach we examined the change in the number of educators persisting in earning the National Board certification in the Network cities, compared with those in a set of matched non-Network cities, in 2014–15 and 2015–16, while accounting for preexisting differences between the cities' educator and school characteristics. The model we used to estimate the effect of the Network participation is shown below:

$$Candidates_{ct} = \beta_1 Network_c * 2014_t + \beta_2 Network_c * 2015_t + \beta_3\, EducatorCharacteristics_{ct} + \beta_4\, SchoolCharacteristics_{ct} + Year_t + Year_t * Post_t + \delta_c + \varepsilon_{ct} \quad (2)$$

In this model, $Candidates_{ct}$ represents the number of educators pursuing National Board certification in a particular city, *c*, and year, *t*. *Network* is an indicator that is set to 1 if the city was part of the six original Network sites; it is set to 0 otherwise; therefore, $Network_c * 2014_t$ is 1 for the Network cities in 2014 and $Network_c * 2015_t$ is 1 for the Network cities in 2015. $EducatorCharacteristics$ represents the educators' age, gender, and race, whether the educator received his or her certification in a STEM field, and a dummy variable for whether the educator has a masters' degree or higher.[1] $SchoolCharacteristics$ represents student enrollment, the number of educators pursuing National Board certification at the same school, and the percentage of students who are eligible for free or reduced-price lunch. $Year_t$ is a year trend, used to account for any linear variation in the number of educators pursuing or achieving board certification over time. Our model also allows this trend to vary after the Network implementation, with the $Year_t * Post_t$ parameter. City fixed effects, δ_c, control for persistent differences between cities, and the error term, ε_{ct}, captures residual variation that remains unexplained by the parameters in the model. We cluster the standard errors at the state level to account for systematic variation in outcomes by state.

We estimated results for a full-time series based on all data available to us (2000–2016), and a limited-time series (2009–2016). We estimate results for a limited-time series to isolate the postrecession component of analysis and also to check robustness of our estimates.

Results

Our results suggest a positive effect of the Network on the number of teachers persisting toward the National Board certification. We estimated a full-time series model (2000–2016) and a limited-time series (2009–2016). Findings are consistent between these models, but the estimates of significance vary (perhaps because of the variance-bias trade-off associated with

---

[1] The one covariate difference between these estimates and the ones used with propensity score matching is with teacher experience, which we did not receive for 2014–15 and 2015–16.

including fewer data in the limited-time series model, which should both reduce our power and increase our bias). In specific terms, the full-time series results suggest that, on average, an increase in the number of educators persisting toward National Board certification in their first year of certification pursuit was significantly greater in the Network cities than in the matched comparison, non-Network cities for the second year of outcomes, 2015–16 (β = .607, $p < .01$). The limited-time series results for 2015–16 suggest this same finding but with a lower degree of confidence, associated with the marginally significant (p < .10) threshold (β = 1.102). The results for the first year of outcomes, 2014–15, were positive, but did not reach the level of statistical significance across both models. These results are summarized in Table 2. Table A4 in the appendix presents a complete summary of results, including the covariates.

Table 2: Results of the Comparative Interrupted-Time Series (CITS) Analysis, Abbreviated

	Full-time series: since 2000	Limited-time series: since 2009
Treated	2.955*** (1.048)	0.898 (1.718)
Treated*YR1Post (2014–2015)	**0.587 (0.778)**	**0.896 (1.200)**
Treated*YR2Post (2015–2016)	**0.607*** (0.182)**	**1.102* (0.556)**
Linear year trend	-0.137* (0.0755)	-0.183 (0.122)
Linear year trend*Post	0.234** (0.0940)	0.281 (0.205)
Controls for educator and school characteristics	Y	Y
City fixed effects	Y	Y
Observations	7,588	3,119
R-squared	0.844	0.921

Note. Robust standard errors, clustered at the state level, in parentheses.
*$p < .10$; **$p < .05$; ***$p < .01$.

To understand the magnitude of findings, it is useful to aggregate the results of our analysis to the site level from the city level. The total number of cities that comprised the

Network sites in 2015–16 was 393 (see Table 1). The limited-time series model, which provides a better fit with data (R-squared of .92, as opposed to R-squared for the full-time series model of .84), estimates an increase by 1.102 educators per Network city persisting toward National Board certification, which translates into the total estimated increase of 433 educators in the Network sites above and beyond the increase that we would have expected in the absence of the Network. This increase attributable to participation in the Network, 433, is 20% of the total 2,112 educators at the Network sites in 2015–16 who persisted toward the National Board certification that year (Table A5 in the appendix shows the numbers of persisters and fully certified teachers in the Network and non-Network sites). The same calculation for the full-time series model yields a 234, or 11% increase in new candidates at the Network sites attributable to their participation in the Network.

Conclusions and Limitations

This working paper describes results of the study that examined the relationship between participation in the National Board's Network to Transform Teaching, structured as a NIC, and an increase in the number of NBCTs, which is one of this NIC's primary aims. Because of the 4 years required to achieve full certification, the study could not examine full certification as an outcome; 4 years of data were not available for the study. Instead, as a proxy for full certification, the study examined persistence in earning the National Board certification, operationalized as the submission of at least one certification component by the end of an academic year. The study found that, after the Network had 2 years to accelerate the recruitment and support of National Board certification candidates, it experienced a significant increase in the number of educators persisting in earning the certification by the end of their first year of pursuit. The full- and limited-time series models used by the study estimate that anywhere from

11% (n = 234) to 20% (n = 433) of persisters at the Network sites are attributable to participation in the Network. The results for the first year of outcomes were positive, but they did not reach the level of statistical significance.

The current study is different from prior research on NICs in education in two ways. First, the study examined the outcomes of a NIC implemented on a greater geographic scale than NICs in education tend to be. The National Board's NIC is based in multiple states and districts representing diverse regions of the country, while NICs in education are typically more tightly geographically bound. Second, the current study's research methodology, including propensity score matching to create a comparison group and the CITS approach to analyze results, is more sophisticated than the descriptive methodology used in prior research, allowing the current study to examine the effectiveness of the National Board's NIC with more rigor than have other studies of NICs.

The results of the study are encouraging, suggesting that a NIC could be an effective approach to addressing complex educational challenges and for doing so on a large scale—a reassuring conclusion for the stakeholders participating in NICs and their funders. However, although methodologically more sophisticated than prior research, the current study is far from the rigor required for making causal conclusions. While we used a statistically matched comparison group, the characteristics available for creating this group were a limited number of educational variables in select cities—hardly enough information for capturing the variability and complexity that is bound to exist in the sites that comprise entire states. Any future studies that use a similar approach, particularly the ones examining NICs on a large scale similar to the NIC in the current study, would benefit from the use of additional data, perhaps the Census data, to improve the counterfactual condition and strengthen the study's conclusions.

Although the completeness of data used in a matching study, as we recommend above, is desirable, such a study will approach but not reach the level of rigor appropriate for making causal conclusions. A study capable of causal conclusions will require random assignment of participants. In the current study, the Network sites were recruited by the National Board on the basis of their willingness to conduct the work and preexisting propensity toward certifying their teaching workforce, which eliminated the possibility of random assignment. In fact, the selection of NIC members for the National Board's Network or any other NIC is meant to draw together like-minded and motivated members, which is one of NICs' strengths. This feature, however, is likely to be an obstacle for any study that requires a random selection of participants. We believe that, as the existing NICs mature and new ones emerge, perhaps with different configurations, so will the opportunities to study their effectiveness more rigorously. In the meantime, matching studies like the one described in this working paper or other studies capable of assessing the effectiveness of NICs are needed to ensure that the intent behind them—to improve our ability to innovate effectively and scale practical knowledge efficiency—is indeed being realized.

References

Baron, K. (2017). *The promise of social relationships in building strong networked improvement communities*. Retrieved from https://www.carnegiefoundation.org/blog/the-promise-of-social-relationships-in-building-strong-networked-improvement-communities/

Bell, M., Jopling, M., Cordingley, P., Firth, A., King, E., & Mitchell, H. (2006). *What is the impact on pupils of networks that include at least three schools? What additional benefits are there for practitioners, organizations and the communities they serve?* Retrieved from http://www.curee.co.uk/files/shared/curee-nlc-systematic-review-impactofnetworkd.pdf

Bloom, H. S. (2003). Using 'short' interrupted time-series analysis to measure the impacts of whole-school reforms: with applications to a study of accelerated schools. *Evaluation Review*, *27*(1), 3–49.

Bryk, A., Gomez, L., & Grunow, A. (2011). *Getting ideas into action: Building networked improvement communities in education. Carnegie Perspectives*. Retrieved from https://www.carnegiefoundation.org/wp-content/uploads/2014/09/bryk-gomez_building-nics-education.pdf

Bryk, A., Gomez, L., Grunow, A., & LeMahieu, P. (2015). *Learning to improve: How America's schools can get better at getting better.* Cambridge, MA: Harvard Education Publishing.

Cantrell, S., Fullerton, J., Kane, T. J., & Staiger, D. O. (2008). *National Board certification and teacher effectiveness: evidence from a random assignment experiment.* National Bureau of Economic Research Working Paper Series14608. Retrieved from http://www.nber.org/papers/w14608

Cavaluzzo, L. (2004). *Is National Board certification an effective signal of teacher quality?* National Science Foundation No. REC-0107014. Alexandria, VA: CNA Corporation.

Clotfelter, C., Ladd, H., & Vigdor, J. (2007). *How and why do teacher credentials matter for student achievement?* Retrieved from http://www.caldercenter.org/sites/default/files/1001058_Teacher_Credentials.pdf

Coburn, C., Penuel, W., & Geil, K. (2013). *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts.* New York, NY: William T. Grant Foundation. Retrieved from http://www.spencer.org/sites/default/files/pdfs/Research-Practice-Partnerships-at-the-District-Level.pdf

Cowan, J., & Goldhaber, D. (2015). National Board certification and teacher effectiveness: Evidence from Washington state. *Journal of Research on Educational Effectiveness, 9*(3), 233–258.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching. R*eview of Economics and Statistics, 89*(1), 134–150.

Hannan, M., Russell, J., Takahashi, S., & Park, S. (2015). Using improvement science to better support beginning teachers: The case of the building a teaching effectiveness network. *Journal of Teacher Education, 1*(15), 1–15.

Manzeske, D., Park, S., Feng, L., Borman, T., Gnedko-Berry, N., West, B., & Deng, E. (2017). *Effects of National Board certified teachers on student achievement and behavioral outcomes: Studies conducted in two states.* Retrieved from https://eric.ed.gov/?id=ED572969

Martin, G., & Gobstein, H. (2016). *Organizing a network for collective action.* Retrieved from https://www.carnegiefoundation.org/blog/organizing-a-network-for-collective-action/

National Board. (2016). *What teachers should know and be able to do*. Retrieved from http://accomplishedteacher.org/wp-content/uploads/2016/12/NBPTS-What-Teachers-Should-Know-and-Be-Able-to-Do-.pdf

Proger, A., Bhatt, M., Cirks, V., & Gurke, D. (2017). *Establishing and sustaining networked improvement communities: Lessons from Michigan and Minnesota* (REL 2017–264). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/REL_2017264.pdf

Rosenbaum, P., & Rubin D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Publishing.

Appendix: Tables and Figures

Propensity Score Matching

Table A1: Characteristics used in the propensity score matching process

Educator characteristics and trends in National Board certification	
Total NBCTs	The sum of all NBCTs in each city
Certification age	The average age of candidates in each city
Years of teaching experience	The average number of years of teaching experience of candidates in each city
Gender	The percentage of candidates that are female in each city
Master's degree or higher	The average percentage of candidates with master's degrees or higher in each city
STEM	The average percentage of candidates with STEM specialization in each city
Teacher non-White	The average percent of candidates who are non-White in each city
School/student characteristics	
Total NBCTs at school	The average number of NBCTs per school in the city
Totals students	The city-level sum of students served by candidates
Percent FRL[a]	The city level percent FRL, based on the city-wide sum of the FRL students served by candidates divided by the sum of all students served by candidate

[a] FRL (free or reduced-price lunch) is calculated by summing up the students eligible for FRL to the city level before taking the percentage. We believe that the alternative approach, to average the school-level FRL percentage, would decrease the usefulness of the FRL data, as it would weight every school equally, reducing the relative importance of the largest schools, precisely the schools where we expect the largest percentages of students eligible for FRL. As it turns out, the contribution of this decision to the overall findings is of marginal significance at best.

Table A2: Standardized bias reduction between Network and non-Network cities for the final matching model

Variable	Sample	Treated	Control	Bias	% Bias reduction
Total NBCTs 2009	Unmatched	4.2192	2.1013	28.4	22.2
	Matched	4.2192	2.5714	22.1	
Total students at school 2009	Unmatched	2058.3	1374.7	23.4	5.2
	Matched	2058.3	1410.4	22.2	
Total NBCTs at school 2009	Unmatched	1.4114	1.2448	29.8	46.4
	Matched	1.4114	1.3221	16	
Percent FRL 2009	Unmatched	8.712	8.0706	14.9	64.6
	Matched	8.712	8.485	5.3	
Certified age 2009	Unmatched	39.574	40.066	-7	94.3
	Matched	39.574	39.546	0.4	
STEM 2009	Unmatched	0.18865	0.16367	8.5	59.7
	Matched	0.18865	0.17857	3.4	
Female 2009	Unmatched	0.84605	0.89138	-18.5	58.8
	Matched	0.84605	0.86473	-7.6	
Master's degree or higher 2009	Unmatched	0.82961	0.69329	41.4	98.7
	Matched	0.82961	0.83133	-0.5	
Years of teaching experience 2009	Unmatched	14.164	15.646	-29.4	93.5
	Matched	14.164	14.261	-1.9	
Teachers, non-White, 2009	Unmatched	0.16636	0.16014	2.2	-146.3
	Matched	0.16636	0.15105	5.4	
Total NBCTs 2010	Unmatched	3.5222	1.8652	26.4	21.2
	Matched	3.5222	2.2167	20.8	
Total students at school 2010	Unmatched	1921.9	1311.3	23.1	-2.4
	Matched	1921.9	1296.6	23.6	
Total NBCTs at school 2010	Unmatched	1.3673	1.2194	27.5	40.5
	Matched	1.3673	1.2792	16.3	
Percent FRL 2010	Unmatched	8.6295	8.0123	14.2	68.1
	Matched	8.6295	8.4329	4.5	
Certified age 2010	Unmatched	40.169	39.735	6.3	66.5
	Matched	40.169	40.024	2.1	
STEM 2010	Unmatched	0.16639	0.1515	5.3	30.9
	Matched	0.16639	0.1561	3.7	
Female 2010	Unmatched	0.85129	0.89181	-16.4	77.2
	Matched	0.85129	0.86054	-3.7	

Variable	Sample	Treated	Control	Bias	% Bias reduction
Master's degree or higher 2010	Unmatched	0.83872	0.69802	43.7	87.8
	Matched	0.83872	0.82152	5.3	
Years of teaching experience 2010	Unmatched	13.905	14.665	-14.7	99.4
	Matched	13.905	13.901	0.1	
Teachers, non-White, 2010	Unmatched	0.15017	0.17266	-8.2	60
	Matched	0.15017	0.14118	3.3	
Total NBCTs 2011	Unmatched	2.8867	2.1327	14	10.2
	Matched	2.8867	2.2094	12.6	
Total students at school 2011	Unmatched	1741.8	1293.7	17.8	-1.3
	Matched	1741.8	1288	18	
Total NBCTs at school 2011	Unmatched	1.3201	1.2036	23.8	51.9
	Matched	1.3201	1.264	11.5	
Percent FRL 2011	Unmatched	8.5991	8.0907	11.9	40
	Matched	8.5991	8.2939	7.1	
Certified age 2011	Unmatched	39.156	39.548	-5.5	57
	Matched	39.156	39.325	-2.4	
STEM 2011	Unmatched	0.1768	0.14958	9.5	45.3
	Matched	0.1768	0.16191	5.2	
Female 2011	Unmatched	0.85342	0.89117	-14.8	62.3
	Matched	0.85342	0.86765	-5.6	
Master's degree or higher 2011	Unmatched	0.85344	0.72405	41.5	97.4
	Matched	0.85344	0.85011	1.1	
Years of teaching experience 2011	Unmatched	12.482	13.712	-25.2	85.2
	Matched	12.482	12.664	-3.7	
Teachers, non-White, 2011	Unmatched	0.2253	0.23369	-2.6	-135.4
	Matched	0.2253	0.20557	6.1	
Total NBCTs 2012	Unmatched	2.6256	1.7477	15	18.9
	Matched	2.6256	1.9138	12.1	
Total students at school 2012	Unmatched	1680.8	1211.5	17.6	8.4
	Matched	1680.8	1250.8	16.2	
Total NBCTs at school 2012	Unmatched	1.992	1.7796	3.6	55.7
	Matched	1.992	1.8978	1.6	
Percent FRL 2012	Unmatched	11.089	10.393	6.5	75.3
	Matched	11.089	11.261	-1.6	
Certified age 2012	Unmatched	39.691	39.446	3.5	10.3
	Matched	39.691	39.91	-3.1	

Variable	Sample	Treated	Control	Bias	% Bias reduction
STEM 2012	Unmatched	0.19647	0.1507	15.3	52.6
	Matched	0.19647	0.17478	7.3	
Female 2012	Unmatched	0.84775	0.90181	-21.2	53.4
	Matched	0.84775	0.87292	-9.9	
Master’s degree or higher 2012	Unmatched	0.86348	0.73349	41.6	93.3
	Matched	0.86348	0.87221	-2.8	
Years of teaching experience 2012	Unmatched	12.41	13.019	-11.6	49.9
	Matched	12.41	12.715	-5.8	
Teachers, non-White, 2012	Unmatched	0.31635	0.27985	10	-16
	Matched	0.31635	0.27403	11.6	
Total NBCTs 2013	Unmatched	3.367	1.3587	30.4	23.3
	Matched	3.367	1.8276	23.3	
Total students at school 2013	Unmatched	1805.8	1156.1	25.3	18.6
	Matched	1805.8	1277.1	20.6	
Total NBCTs at school 2013	Unmatched	1.892	1.6691	5.2	65.2
	Matched	1.892	1.8143	1.8	
Percent FRL 2013	Unmatched	12.801	12.758	0.3	-1368.3
	Matched	12.801	13.43	-4.1	
Certified age 2013	Unmatched	39.226	39.372	-2	-38.4
	Matched	39.226	39.428	-2.8	
STEM 2013	Unmatched	0.18379	0.15682	9	47.2
	Matched	0.18379	0.16955	4.7	
Female 2013	Unmatched	0.8541	0.88632	-12.3	53.1
	Matched	0.8541	0.86923	-5.8	
Master’s degree or higher 2013	Unmatched	0.88266	0.73789	46.9	99.3
	Matched	0.88266	0.88367	-0.3	
Years of teaching experience 2013	Unmatched	11.367	12.569	-22.6	81.8
	Matched	11.367	11.586	-4.1	
Teachers, non-White, 2013	Unmatched	0.34357	0.29985	11.7	23.8
	Matched	0.34357	0.31026	8.9	

Table A3: Overall standardized bias reduction between Network and non-Network cities for the final matching model

Sample	Standardized % bias
Unmatched	17.3
Matched	7.8

Complete Results of the CITS Analysis

Table A4: Results of CITS analysis with covariate controls for linear trend

Variable	Full-time series: since 2000	Limited-time series: since 2009
Treated	2.955*** (1.048)	0.898 (1.718)
Treated*YR1 Post (2014–2015)	**0.587 (0.778)**	**0.896 (1.200)**
Treated*YR2 Post (2015–2016)	**0.607*** (0.182)**	**1.102* (0.556)**
Linear year trend	-0.137* (0.0755)	-0.183 (0.122)
Linear year trend*Post	0.234** (0.0940)	0.281 (0.205)
Educator characteristic: certified age	-0.00137 (0.00347)	0.00441 (0.00478)
Educator characteristic: female	0.269*** (0.0830)	0.102 (0.139)
Educator characteristic: non-White	0.0636 (0.123)	0.152 (0.211)
Educator characteristic: STEM field	-0.477*** (0.114)	-0.486** (0.184)
Educator characteristic: master’s degree or higher	-0.188 (0.178)	-0.478*** (0.135)
School characteristic: student enrollment	0.00191*** (0.000325)	0.00141*** (0.000469)
School characteristic: number of candidates/NBCTs at the same school	0.00401 (0.0302)	0.0128 (0.0283)
School characteristic: Percent FRL-eligible	-0.00266 (0.00329)	0.00338 (0.00418)
City fixed effects	Y	Y
Observations	7,588	3,119
R-squared	0.844	0.921

Table A5: NBCT and candidate numbers for the Network cities and matched non-Network cities by year

Year	NBCTs and National Board certification candidates in Network cities	NBCTs and National Board certification candidates in matched non-Network cities
2000-01	253	1,086
2001-02	473	1,290
2002-03	529	1,241
2003-04	614	1,137
2004-05	664	1,016
2005-06	827	1,077
2006-07	975	1,244
2007-08	1,404	1,425
2008-09	1,773	1,345
2009-10	1,877	1,044
2010-11	1,545	900
2011-12	1,181	897
2012-13	1,120	777
2013-14	1,436	742
2014-15	2,097	656
2015-16	2,112	705

Note. Prior to 2014–15, educators could become fully certified within a year. Starting with the new certification, in 2014–15, it will take educators up to 4 years to become fully certified. Therefore, the table shows the number of NBCTs in the cells corresponding to years 2000–01 through 2013–14, and the number of National Board candidates in the cells corresponding to 2014–15 and 2015–16.