Routledge
Taylor & Francis Group

School Effectiveness and School Improvement
An International Journal of Research, Policy and Practice

ISSN: 0924-3453 (Print) 1744-5124 (Online) Journal homepage: http://www.tandfonline.com/loi/nses20

Research-informed practice improvements: exploring linkages between school district use of research evidence and educational outcomes over time

Carolyn J. Heinrich & Annalee Good

To cite this article: Carolyn J. Heinrich & Annalee Good (2018): Research-informed practice improvements: exploring linkages between school district use of research evidence and educational outcomes over time, School Effectiveness and School Improvement, DOI: 10.1080/09243453.2018.1445116

To link to this article: https://doi.org/10.1080/09243453.2018.1445116

SCHOOL EFFECTIVENESS AND SCHOOL IMPROVEMENT, 2018
https://doi.org/10.1080/09243453.2018.1445116

Routledge
Taylor & Francis Group

ARTICLE

Research-informed practice improvements: exploring linkages between school district use of research evidence and educational outcomes over time

Carolyn J. Heinrich [a] and Annalee Good[b]

[a]Peabody College of Education and Human Development, Department of Leadership, Policy and Organizations, Vanderbilt University, Nashville, TN, USA; [b]Wisconsin Center for Education Research, University of Wisconsin-Madison, Madison, WI, USA

ABSTRACT

The field of education is progressively building capacity and tools for producing rigorous research evidence to use in improving educational practice and outcomes. The knowledge base is lacking, however, in explicating the dynamics of research-based decision making and exploring connections between the work of research–practice partnerships and changes in educational outcomes. Drawing on experiences and data collected over a decade in mixed-methods research, we describe a case example of how research evidence developed in a multidistrict collaboration was used by one district partner to inform ongoing improvements in publicly funded tutoring programs. We employ both qualitative and quantitative analysis in exploring probable linkages between research-based changes made by the district partner and academic outcomes over time.

ARTICLE HISTORY
Received 6 June 2017
Accepted 20 February 2018

KEYWORDS
Research partnerships; evidence-based policymaking; tutoring; student achievement

Introduction

Education policy in the United States has been a central focus of federal efforts to build stronger evidence of policy and program effectiveness, including through large-scale initiatives such as Race to the Top and Investing in Innovation (Haskins & Margolis, 2014). A key aim of these efforts has been to develop tools for more rigorously identifying evidence-based interventions (Whitehurst, 2012), while also fostering partnerships to increase the production of high-quality research evidence that is relevant to those in a position to apply it in practice (Easton, 2010). Although there are many potential benefits to research-based decision making, especially in the context of established research–practice partnerships, such as more timely information for decision making and increased practitioner capacity for using research evidence (Owen & Larson, 2017), the knowledge base is still limited in explicating how research generated in such collaborations is used in practice to improve education outcomes (Coburn, Penuel, & Geil, 2013). Coburn and Penuel (2016) point out that the current body of research is more likely to describe challenges in sustaining research partnerships in the dynamic

CONTACT Carolyn J. Heinrich carolyn.j.heinrich@vanderbilt.edu

and politicized contexts of education systems, rather than how to overcome them, and they further note that there is little research that examines both these dynamics *and* how they relate to outcomes.

The research we present here seeks to fill this gap in part. We first aim to explicate some of the key mechanisms through which a long-term research–practice partnership was structured, supported, and sustained – as it expanded from a single school district to six school districts over time – drawing on the growing literature on evidence-based policy making through research–practice partnerships. A second core objective is to document, through rich description of mixed-methods research conducted in the multi-site collaboration, the specific ways in which the research evidence generated was used to inform district policy and program changes over time. Lastly, focusing on the longest running district partner, we explore potential linkages between the evidence-informed changes made by this district in a specific program intervention to changes in the academic outcomes of students participating in that intervention over time.

The long-term research–practice partnership in which this research is situated originated in the context of the US No Child Left Behind (NCLB) Act, which expanded assessment and accountability activities and mandated increasing levels of interventions and sanctions for schools not making adequate yearly progress, including the provision of free supplemental educational services (out-of-school tutoring). The burden of implementing these services and new accountability regimes fell primarily to local educational agencies, which were required to contract with private providers to offer choice in tutoring services and to hold them accountable for increasing student achievement (Heinrich, 2010). In 2006, Milwaukee Public Schools (MPS) reached out to the Wisconsin Center for Evaluation Research (WCER) at the University of Wisconsin-Madison to request assistance in evaluating its tutoring providers, initiating a research relationship between MPS and researchers at the University of Wisconsin-Madison that continues today (see Figure 1, which describes the timing and activities of the research collaboration).

Through networks with other school districts cultivated at WCER, this research collaboration, centered on evaluating the implementation and effects of supplemental educational services, subsequently expanded to include five other large, urban school districts – Austin Independent School District (ISD), Chicago Public Schools, Dallas ISD, Los Angeles Unified School District, and Minneapolis Public Schools – beginning in 2009.[1] Indeed, the impetus for growing the research alliance into a network of district relationships arose naturally, as we found that these other large urban school districts were encountering similar challenges in implementing NCLB provisions and could gain considerably from exchanging information and ideas around the application of research to practice (see Figure 1). In addressing important questions of "what works where, when, and under what conditions" (Coburn et al., 2013; p. 10; see also Bryk, Gomez, & Grunow, 2011), the expansion of the partnership efforts proved critical, enabling the research investigation to go both deep into local district settings and across sites to more usefully explore the potentially causal versus mediating factors and relationships in a given context, such as those driven by local implementation decisions versus those commonly experienced across districts (e.g., due to federal policy design). The expansion, as well as the purposeful structures of communication that were developed with it (e.g., annual research briefings in the districts), allowed key stakeholders from each district site (district program directors, research and evaluation staff, and

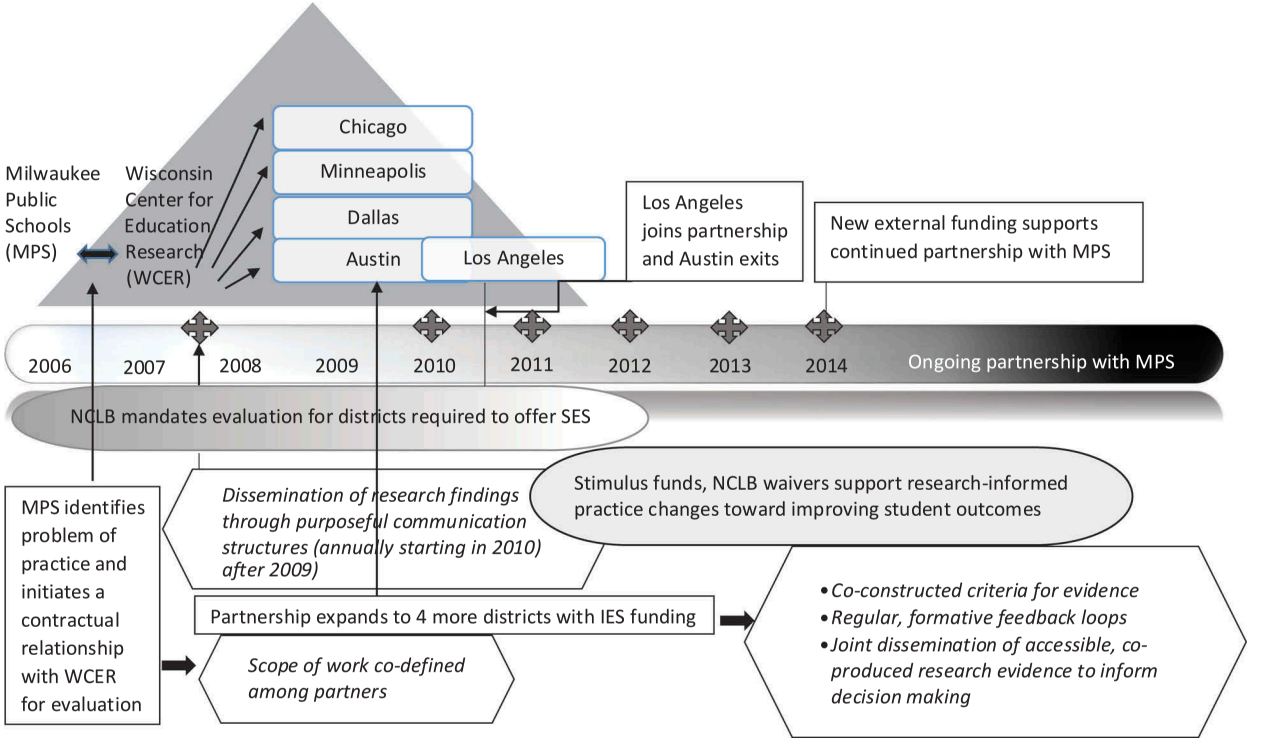

Figure 1. Actors and timeline of collaborative activities in the research–practice partnership.

tutoring program vendors) to interact with one another around these "problems of practice" via our research partnership.

The research undertaken in the context of this multi-site partnership employed a longitudinal, mixed-methods design, in which quantitative and qualitative methods were tightly integrated in developing the research plan, collecting and analyzing data, and disseminating the research evidence through multiple phases of program development, evaluation, revamping, and more evaluation. In defining what constitutes research evidence with our practice partners, we followed the perspective of Gueron and Rolston (2013), who made clear that experimental and quasi-experimental evidence is "not the only type of evidence (or even the most important) that would matter to policymakers" (p. 426). They conveyed an appreciation for "multiple techniques to address a broader range of questions – including those raised by practitioners and managers, and by researchers who used different methods to diagnose the problems and understand why people behave as they do and how social programs work in practice" (p. 426). Accordingly, in addition to drawing on quantitative data and econometric methods commonly used to assess student outcomes over time, we collected qualitative data in classroom observations and interviews with program administrators and staff across the six research sites (over time) to identify collaboratively, that is, in joint discussions with our district partners, configurations in program management and contextual factors in the instructional settings that could aid in understanding local challenges to policy implementation and inform potential strategies for improving program outcomes. Although in the absence of experimental data we do not make any causal claims in our analyses, we point to patterns in the quantitative data, corroborated by qualitative findings, that suggest possible linkages between evidence-informed changes made by the districts and improvements in student outcomes.

We begin by providing additional background and information on the motivation, scope, and aims of the research, as well as the nature, mechanisms, and progression of the research–practice partnership developed to achieve these aims. We next introduce our research samples and data, describing first the data collected in all six research sites, and then discussing the data and methods employed in the analysis (specific to our originating partner, MPS) to investigate the link between evidence-informed program changes and student outcomes. We continue with a description of how research evidence generated in this collaboration was used by the school district partners to implement policy and practice changes to their tutoring interventions, and we follow with our empirical analysis of changes in academic achievement among students participating in MPS tutoring interventions (following the research-informed improvements), drawing on qualitative data to inform the interpretation of these results. We conclude with a brief discussion of the implications of the findings for future efforts to promote evidence-based policy making and program and practice improvements in Kindergarten to Grade 12 (K–12) education.

Aims and supporting structures and activities of the research collaboration

Defining the research scope and questions

The Elementary and Secondary Education Act (ESEA) of 1965 encouraged the provision of supplemental instruction as a means to increase the quality and effectiveness of

instruction for low-income students. School districts across the United States accordingly spend millions of Title I dollars annually on tutoring for economically and academically disadvantaged students. Under NCLB, the market for tutoring services expanded rapidly, as increasing numbers of schools were *mandated* to offer supplemental instruction using their Title I funds, and the law specified a larger role for private sector providers and parental choice in improving the quality of tutoring offered. However, NCLB fell short in providing resources to school districts to evaluate tutoring providers and generate information that could be used by parents to make informed choices, and it also limited the ability of state and local agencies to regulate the market and provider practices (Heinrich et al., 2014). Hence, the research collaboration with MPS was initiated with a focus on a particular "problem of practice" that was pressing for the district, and it sought to produce answers to questions posed by the district (concerning tutoring provider performance) through research (Coburn et al., 2013, pp. 2–3). Coburn et al. (2013) identify a focus on research and program innovations driven by local (e.g., district) needs as one of the hallmarks of successful research–practice partnerships. Similarly, Dagenais et al. (2012) distinguish the production of this type of research evidence – "local research-based information" that is both "produced locally and intended for local use" – from "general research-based information" (p. 286) that is typically accessed in the published academic literature.

Figure 1, as noted above, depicts the actions and actors that engaged in this research–practice cycle, detailing the structures and activities (shown in italics) and the progression of the collaboration (with key steps identified along the timeline). Agreeing on a central purpose of the research – developing measures of tutoring provider performance and understanding factors contributing to their effectiveness – was only a first step in co-defining the scope and direction of the collaborative work. As Tseng (2012) explains, to cultivate partnerships that are effective and sustainable, the research and practice partners need to open a "two-way street" of "practice to research and back" (p. 22). In this approach to joint partnership work, local policy and practice inform the line of research investigation, and the findings of the research in turn are integrated into local program development and implementation, with the research cycle continuing iteratively and learning ongoing on the part of both the researchers and practitioners. In the context of the multi-site partnership (represented in the triangle in Figure 1), the collaboration occurred not only between researchers and practitioners but also among practitioners across sites, facilitated by partnership infrastructure and a flexible orientation to the nature of both the research and district programming. The research–practice partnership also created opportunities for continuous appraisal of stakeholder interests, needs, and expectations for the partnership efforts.

Infrastructure for supporting and sustaining the partnership

To facilitate this multi-pronged approach to creating research evidence for instrumental uses, that is, for application directly to problems of "practice" identified by our partners (Tseng, 2012), we developed infrastructure for regularly sharing information about the research activities and for convening discussions of the research findings with our district partners and other stakeholders, such as parents and community-based organizations. The more formal, annual interactions (identified on the timeline in Figure 1)

were timed strategically around the school year calendar to coincide with the programmatic activities and decision making involved in implementing the tutoring programs. For example, prior to the start of each school year, we conducted in-person district research briefings to discuss the research findings and their implications for potential program modifications that could be made by district staff engaged in program implementation. At the same time, maintaining channels for more informal but ongoing exchange and an open line to hearing our practitioner partners' (and stakeholders') questions and concerns was critical to communicating research findings that could have immediate implications for practice or to adapting our research to be responsive to our district partner needs. In an example of the former, we promptly reported patterns found in administrative data on provider invoicing that suggested billing for services was inconsistent with recorded hours of tutoring provided to students. Alternatively, when a district administrator learned that a contract with an online service provider was abruptly terminated, she requested a summary of the peer-reviewed (not industry-generated) research evidence on online tutoring programs and assistance with developing research-relevant language for a new request for proposals.

The involvement of five other urban school districts facing similar challenges in implementing supplemental educational services under NCLB also substantially enriched both the research investigation and the research–practice exchanges. For instance, many of the tutoring providers were operating in the national market for tutoring services, and thus, we were able to observe the same provider offering services in different school districts, often at varying hourly rates and/or under different arrangements, for example, on-site at schools versus off-site, differing student–teacher ratios and instructor qualifications (see Appendix 1, Table A1). This is just one example of how the multi-site structure of the partnership was critical to generating stronger evidence for research-informed improvements by any one district – the cross-district variation not only provided grist for our empirical mill in analyses of the relationship of district and provider characteristics to tutoring hours and student outcomes, but it also spurred discussion among the district partners of policy levers and strategies they were applying (within these market constraints) to increase the hours of tutoring offered to students and the effectiveness of the services. To support these cross-district exchanges of information, we held joint webinars with the district partners (typically before the end of the school year or the start of the next school year), so that they could draw on the latest study findings in these discussions and use them to inform program planning and modifications for the coming school year. Head (2016) describes these types of activities as "knowledge-brokering," in that they go beyond a simple report of the findings to creating mechanisms for "dialogue and co-production of insights in new contexts" (p. 8) among the end users of this knowledge.

The nature and accessibility of the evidence produced was likewise critical to promoting its use by the practice partners in an iterative cycle with the researchers. One of the striking findings from interviews conducted by Nelson, Leffler, and Hansen (2009) with education stakeholders such as congressional staff, deputy state commissioners of education, legislators, school board trustees, district personnel, principals, and teachers was how infrequently they mentioned using any *research* evidence, expressing a lack of confidence in its veracity and applicability. Instead, these important potential users of research evidence were more likely to draw on other sources of information such as local

data and personal experiences and communications. Of perhaps even greater concern, in a comprehensive review of existing research on the use of research evidence by school practitioners, Dagenais et al. (2012) found very little use of research findings, whether academic or "local" in nature. Additionally, Finnegan, Daly, and Che (2013) point to enormous variation in what is seen as "legitimate" sources of evidence, depending on the perspective of the practitioner and his or her position in the education system. Recognizing these potential barriers to receptivity and confidence in research findings, we sought to incorporate multiple types of data into the production of the research evidence, both qualitative and quantitative, including data collected through observations of instructional settings, interviews with key personnel, and administrative data sources.

In addition, the cross-district webinars were an important source of information "co-produced" by the research partners, which we in turn integrated into our mixed-methods analysis of the study data. And in MPS, for a number of years we jointly supported an "embedded" researcher who worked on-site in the district administration building and was available to readily interface with program administrative and research support staff on data issues or questions needing immediate attention. This shared approach to creating and drawing on the data for research insights appeared to increase practitioners' trust of the data and their view of it as not only "local" but also relevant to their work (Finnegan et al., 2013). In a study of data-based decision making in Dutch secondary education, Schildkamp, Poortman, Luyten, and Ebbeler (2017) likewise found that collaboration and trust (among teachers) were key elements ungirding their use of data.

Lastly, in preparing the research findings for dissemination and discussion, we were also attentive to how the research partners and stakeholders preferred to "consume" and use the research evidence. For example, Nelson et al. (2009) found in their research that policy makers and practitioners often prefer a brief (e.g., 1–2-page) concise summary of findings and other important knowledge for decision making, conveyed in nontechnical language and written with a given audience or stakeholder in mind. Dagenais et al. (2012) also found that developing research products tailored to practitioners' needs was key to ensuring their application of the findings. We have correspondingly developed these types of research products and made them publicly available, including research and policy briefs written for program administrators and other school-level staff, as well as briefs for parents and community organizations on how to use the information generated in the research. For example, Chicago Public Schools asked for assistance in integrating research evidence on tutoring provider performance into booklets for parents that could help inform their choices of tutoring providers for their children, a practice that was adopted by other districts through our collaboration and that was even codified into law in Texas when Dallas ISD encountered resistance in implementing this strategy.[2] Some of the district partners also posted the research briefs and reports on their websites to make them available to parents and other interested parties, and we responded by creating direct channels of communication for the public (i.e., a toll-free study information line) to be available for questions and to ensure that the results (and their limitations) were understood and used with appropriate caution by the various research stakeholders.

We were also continually aware, however, of the constraints faced by our district partners in applying the research evidence to effect changes in practice. As Yohalem and Tseng (2015) point out, even if a given research study or body of work is directly relevant to the practitioner or policymaker's issue at hand, there may be other forces driving the decision making that are "beyond the potential influence of social science" (p. 118). Some of these forces identified in Nelson et al.'s (2009) interviews included political outlook, public sentiment, legal and economic considerations, media pressure, and constituent concerns, suggesting that our expectations for evidence-based policy making should be tempered by these realities and moderated to aspirations for "evidence-informed policy" (Head, 2016). As Carol Weiss (1979) explained, research is less likely to influence the outcome of policy or program decision making directly, but rather to help inform practitioners' choices and understanding indirectly, in combination with other sources of information on which they regularly rely.

Studies of research–practice efforts also point to problems of information and role "overload" on the part of practitioners (Yohalem & Tseng, 2015) that limit their capacity for engaging in and using research. As we have seen in our decade of interactions with school district personnel, those tasked with research-related responsibilities are often balancing them with a heavy load of program and administrative obligations that typically take precedence over engagement in research. Furthermore, their ability to commit time to research on a given program or intervention will be shaped not only by time requirements of other obligations, but also by how those obligations are prioritized by other forces (i.e., political, economic/budgetary, etc.). In our research–practice partnership focused on the implementation of supplemental educational services under NCLB, the fact that it was federally mandated, consumed a significant portion of school district Title I funds, and could benefit from research support in its implementation made it a higher priority for a research relationship. Once the research relationship was established and the value of research for districts made clear, external (federal) pressures and obligations were no longer a major factor motivating district engagement in the partnership (especially following the initiation of NCLB waivers starting in 2012, as shown in Figure 1).

Research samples, data, and methods

Multi-site research samples and data

As described above and shown in Figure 1, this research collaboration involved six school district partners, with data collection taking place over the 2007–2008 to 2012–2013 school years for most of the districts, and continuing through the present in MPS. Across the districts and years, our sample frame for the quantitative analysis in each district was defined by the students in K–12 schools who were eligible under NCLB (and district policies) for free tutoring services. The school districts by law offered these opportunities first to low-income students (those eligible for free or reduced-price lunch), and they also consistently prioritized students who lagged further behind in their academic performance (as measured by standardized tests). Appendix 1, Table A2 presents basic information on the student samples for this research over the years of the multi-site collaboration on the implementation and effects of supplemental educational services.

The qualitative data that we drew on in the multi-site component of the research (2009–2012) included: observations of full tutoring sessions (n = 162) using a classroom observation instrument designed for this study; semi-structured interviews with tutoring provider administrators and tutoring staff (n = 153) about instructional programs and formats, staff professional background and training, curricula and assessments, challenges in implementation; and adaptations for special student needs; semi-structured interviews with district and state administrators (n = 34) regarding program implementation; focus groups with parents (n = 183) or caregivers of students who were eligible to receive tutoring; and document analysis of curriculum and staff training manuals, assessments used, and policy documents on federal, state, or district policies (see Appendix 1, Table A3 for these numbers by site). In addition, in MPS, we collected data in an additional 42 observations, seven provider interviews, four district and state administrator interviews, and from 19 focus groups participants in the 2012–2013 through 2014–2015 (post-NCLB waiver) school years.

The quantitative data used in our analyses were assembled from the school districts each year and included information from individual student records, that is, gender, race/ethnicity, free and reduced-price lunch eligibility, English proficiency, student special needs, absences, current and prior tutoring program attendance, as well as their standardized test scores (see Appendix 1, Table A4 for additional details on these measures and Table A2 for sample sizes). We analyzed the qualitative and quantitative data jointly each year (as described below) to understand the implementation and effects of free tutoring on student outcomes. For example, we employed both criterion sampling and extreme case sampling in the qualitative sample selection over time (Palinkas et al., 2015), beginning with samples that represented different types of tutoring providers, for example, school-based, community, in-home, and online, and drawing on relationships or patterns of effects identified in the quantitative analysis to refine the sample selection each year and purposively select tutoring providers identified as generating higher (or lower) than average effect sizes on student achievement. We employed a tightly integrated mixed-methods approach to the research (Burch & Heinrich, 2015), in which insights from the qualitative analysis also informed empirical exploration of relationships in the data, such as refining the construction of measures of tutoring hours to account for how the districts were tracking and recording tutoring hours and thereby more accurately capture instructional hours. Information gathered in the yearly research briefings to disseminate the findings to the districts and guide their modifications to tutoring interventions also, in turn, informed the ongoing qualitative and quantitative data analyses.

Research sample specific to evaluating MPS' post-waiver tutoring program reforms

Prior to receiving waivers from NCLB (up until 2012, as shown in Figure 1), school districts were not permitted to make policy changes that constrained tutoring provider practices, such as restricting their hourly rates or imposing minimum qualifications for tutors or instructional programs. The analysis we undertake to explore linkages between the evidence-informed changes made by MPS to its tutoring program and program outcomes thus necessarily occurs in the post-waiver phase of this research, that is, the

2012–2013 to 2014–2015 school years, when MPS utilized the knowledge created in the multi-site research–practice partnership to implement policy and practice changes for improving program (student) outcomes. MPS was the first of the partner districts to move forward with post-waiver tutoring program changes, creating the Tutoring for You (or T4U) program for elementary students in need of supplemental instructional opportunities, and we received new external funding in 2014 to continue evaluating these program reform efforts in the district (see Figure 1).

The sample frame for the analysis of T4U program effects includes students eligible for T4U in MPS. The eligibility criteria for MPS tutoring programs were clearly defined and consistently based on the following factors: (a) students attending low-achieving schools (defined by test scores and proficiency rates), (b) grade level (2nd or 3rd grade), and (c) student math and reading proficiency levels. In estimating the effects of T4U on student educational outcomes (and comparing them to estimates of the effects of supplemental educational services under NCLB), we use students who were eligible for these programs but did not receive tutoring as an internal comparison group, which reduces the likelihood of bias associated with unobserved differences between participating and nonparticipating students (Cook, Shadish, & Wong, 2008). Importantly, Hallberg, Cook, Steiner, and Clark (2016) demonstrate that particularly in circumstances such as this, where the pretest score is an important factor for selection into treatment (free-tutoring), controlling for the pretest or baseline measures of the outcome (student standardized test scores) typically removes all or most of the bias associated with non-random selection into treatment.[3] Table 1 presents descriptive statistics on MPS students eligible for and receiving free tutoring under NCLB (2010–2012) and the post-NCLB, T4U phase (2012–2015), and we further discuss our models for estimating program effects below.

Qualitative methods

The validity of the qualitative instrumentation employed in this research was ensured in the development process, whereas its structure and content were based on well-tested, existing observation instruments for supplemental instruction, existing literature on best practices for tutoring, and the theory of action in the supplemental services policy. Reliability trainings were conducted regularly with the qualitative researchers throughout the project to ensure consistency in observation ratings and coding and analysis of all qualitative data.

In the analysis of the qualitative data, we used an iterative and constant comparative method to develop and refine our understanding of patterns in tutoring practices and program changes across districts and providers. Throughout the analytic process, we examined potential patterns in the instructional setting, program management, and policy implementation that could aid in understanding the local challenges in policy implementation and potential strategies to address them. Analytic codes were developed from these patterns and in response to the research questions, and then reapplied via qualitative analytic software to interview, observation, and archival data in order to establish findings. Data analysis occurred both concurrent to and after the data collection process, taking into consideration findings from the quantitative analysis, as well as research insights and information that emerged in district research briefings, cross-district webinars, and other forums where the research evidence was reviewed and discussed.

Table 1. MPS sample of students eligible for and receiving tutoring under NCLB and T4U.

Student characteristics	Eligible School year 2010–2011	NCLB participants	Eligible School year 2011–2012	NCLB participants	Eligible School year 2012–2013	T4U participants	Eligible School year 2013–2014	T4U participants	Eligible School year 2014–2015	T4U participants
Number of students	16,439	2,592	20,905	4,010	13,159	2,615	4,919	960	2,234	871
Asian	0.036	0.026	0.042	0.040	0.030	0.028	0.042	0.032	0.024	0.014
Black	0.678	0.758	0.688	0.750	0.588	0.614	0.554	0.548	0.736	0.796
Hispanic	0.208	0.176	0.198	0.172	0.294	0.276	0.326	0.334	0.196	0.152
White	0.072	0.038	0.066	0.036	0.082	0.074	0.070	0.070	0.040	0.036
Other race	0.006	0.002	0.006	0.002	0.008	0.008	0.008	0.016	0.004	0.002
Female	0.458	0.496	0.464	0.518	0.480	0.492	0.496	0.532	0.476	0.494
English language learner	0.112	0.142	0.106	0.148	0.120	0.172	0.186	0.172	0.104	0.074
Free lunch eligible	0.882	0.984	0.894	0.986	0.902	0.940	0.922	0.928	0.964	0.980
Student with disabilities	0.224	0.260	0.242	0.244	0.218	0.210	0.194	0.166	0.182	0.138
MAP math score	−0.170	−0.576	−0.242	−0.492	−0.070	−0.410	−0.094	−0.078	−0.264	−0.152
MAP reading score	−0.176	−0.554	−0.234	−0.454	−0.082	−0.438	−0.120	−0.154	−0.274	−0.132

Quantitative methods

We employed quasi-experimental methods – value-added models with school and student fixed effects – to control for school and student characteristics that we expect were related to whether eligible students received free tutoring in estimating average tutoring program effects. Prior research has found a high degree of consistency in estimates produced by alternative value-added and fixed-effects model specifications (Heinrich et al., 2014; Heinrich & Nisar, 2013; Shanley, 2016), and thus, we present only the following specification:

$$A_{jst} - A_{jst-1} = \alpha T_{jt} + \beta X_{jt-1} + \pi_s + \mu_{gt} + E_{jst} \quad (1)$$

where A_{jst} is the achievement of student j attending school s in year t; T_{jt} is an indicator function if the student j attended tutoring in year t; X_{jt-1} are student characteristics; π_s are school fixed effects; μ_{gt} are grade by year fixed effects, and E_{jst} is the random error term. The purpose of taking into account A_{jst-1} (baseline test scores) in the main estimating equation is that students with different abilities may be more or less likely to participate in tutoring, and accounting for pretest measures has been shown to be a reliable way to address this self-selection into treatment, as discussed above (Bifulco, 2012; Cook et al., 2008; Hallberg et al., 2016). Identification in this specification comes from the average gain in student achievement after controlling for student characteristics (and their underlying achievement trajectories, as reflected in prior test scores), and school and grade year effects. A common alternative specification includes the pretest score (Y_{t-1}) on the right-hand side of the model (with other conditioning variables, X) as a predictor of student achievement (measured by the test score after tutoring, Y_1):

$$A_{jst} = A_{jst-1+} \alpha T_{jt} + \beta X_{jt-1} + \pi_s + \mu_{gt} + E_{jst.} \quad (2)$$

In all cases, our estimates from this "level on previous level" model specification fell within 95% confidence intervals of the original estimates (results available from the authors). Importantly, because there is still potential for unobserved, nonrandom differences between students who received free tutoring and those who were eligible but did not enroll in the programs, we interpret our estimated effects as associations between program participation and student achievement, rather than as causal program impacts.

Research-informed policy and practice changes in MPS tutoring programs

Given the tightly integrated, mixed-methods approach that we applied in our research, the instructive examples of evidence-informed policy and programming changes that we describe in this section draw on both qualitative and quantitative analysis conducted over time in the context of the research–practice partnership described in Figure 1.

Tutoring hourly rates and dosage

Under NCLB, state and local educational agencies were explicitly discouraged from taking any actions that might limit the supply of tutoring providers or range of choices available to parents, and they likewise could not specify or constrain hourly rates

charged by providers. Accordingly, a large number of diverse organizations – national and local, for-profit and nonprofit, online and on-site providers with widely varying hourly rates, service costs, tutor qualifications, tutoring session length, instructional strategies, and curricula – entered the market to offer supplemental educational services (as seen in Appendix 1, Table A1). Although prior research showed little consistency in the relationships among provider hourly rates charged for tutoring services and attributes such as their student–teacher ratios, number of sessions offered, and student attendance rates (Heinrich, 2010), one very basic lever shown to increase tutoring effectiveness was the intensity (or number of hours) of tutoring provided (Heinrich et al., 2014; Zimmer, Hamilton, & Christina, 2010). Some studies suggested that reaching a minimum threshold of tutoring hours (i.e., approximately 30–40 hr) could be critical to producing measurable effects on students' achievement (Jones, 2009; Lauer et al., 2006), while other research (Deke, Dragoset, Bogen, & Gill, 2012) was inconclusive on this issue.

Through our multi-site research collaboration, we were able to draw on the cross-site and within-district variation in hours of tutoring provided, driven in part by district policy and administrative decisions (including funding allocations per student), to better understand the relationship of hours of tutoring to student achievement. Initially, only in Chicago Public Schools (CPS) were students routinely reaching thresholds of 35 or more hours of tutoring each school year, largely because of lower rates charged per hour by tutoring providers there, and we observed a positive, statistically significant relationship between hours of tutoring and student achievement in CPS (see Appendix 1, Table A5). In addition, in both Dallas ISD in 2009–2010 and Minneapolis Public Schools in 2010–2011, we were able to take advantage of natural "policy experiments," in which limited-time policy or program changes directly increased the number of hours of tutoring that students received in those districts only in those years. In Dallas ISD, the district used federal stimulus funds in 2009–2010 to increase the allotted district expenditure per student and thereby boost the number of hours of tutoring students received. Average hours of tutoring increased from approximately 22 hr in 2008–2009 to 35 hr in 2009–2010, and positive effects of tutoring on student achievement were observed only in this school year (as tutoring hours fell by half again in the subsequent school years). In Minneapolis, the district introduced a new program in 2010–2011 for a subsample of tutoring participants that compelled providers to deliver at least 40 hr of tutoring. Students in this trial program received an average of more than 30 hr of tutoring (only for that year), and consistent with what we observed in Dallas ISD, we found positive program effects on the achievement of these students in that year. In interviews across all our district sites, district-level staff lamented how NCLB provisions constrained their ability to require more hours of tutoring from providers. Interestingly, some tutors also expressed frustration at not being able to fully serve their students:

> I think the biggest challenge is when the students run out of hours or time. They want to continue coming to tutoring. The biggest challenge, I think, is turning those students away that have completed their hours. Because they want to be there. They want to get that extra help. They want the tutoring, and when they run out of hours, you have to say, you know, "We can only have you for so many hours." And to see that look on that child's face when they can't come to tutoring anymore, that to me is the biggest challenge. (Interview with a tutor in MPS, 2010)

Following federal waivers from NCLB provisions that had precluded school districts from controlling hourly tutoring rates or setting tutoring hours requirements, MPS used this research evidence – accumulated through the district research briefings and cross-district webinars (i.e., the purposeful communications depicted in Figure 1) – to guide their decision to establish a maximum hourly rate of $35 per hour for tutoring providers starting in 2012–2013, the first school year of the T4U program. Under supplemental educational services (before waivers from NCLB), tutoring providers in MPS had charged anywhere from $55 to $108 per hour; however, MPS knew from the briefings that some of these same providers were charging less than half these rates in other districts such as CPS (allowing students to receive more tutoring hours). This provided additional justification for MPS to establish a maximum rate that was one half to one third of what some tutoring providers had been charging in the district. As a result, MPS saw a steady increase in the hours of tutoring students received, from 21.7 hr (on average) under NCLB in 2011–2012 to 33.7 hr under T4U in 2014–2015. Table 2 summarizes this and other policy and program changes made by MPS (described below) in launching the T4U program, based on evidence generated in the research–practice partnership.

In addition, our observations of 146 tutoring sessions across four districts from 2009–2012 showed consistent differences between the advertised time of tutoring sessions and the actual instructional time. Irrespective of format, students tended to receive less instructional time than what was advertised by providers, although tutoring completed in the student's home most closely matched advertised time (less than 3 min difference on average), whereas that provided in school and community settings was often considerably less than average advertised time (19–25 min less on average). Our field observations suggested much of this time was lost to administrative necessities such as coordinating snacks or transportation home for students. Drawing on this more nuanced understanding of the importance of time in tutoring sessions, MPS established a minimum threshold of 40 hr of tutoring per student under T4U, a maximum tutoring session length of 1 hr, and a maximum 5:1 student-to-tutor ratio, while also increasing district monitoring of tutoring sessions with on-site (school) coordinator reviews and observations. The district stipulated that results of these reviews and desktop audits could be shared with the public and the state educational agency. An MPS district administrator described some of these important, research-based changes in an interview in 2012:

> Some of our providers offered as little as 9 hr [of tutoring], and some offered, you know, about 27 hr. And according to your research … the nominal 40 hr is required, but we couldn't require these private companies to offer services, you know, more than the DPI application. The other thing that it will allow us to do is to set the number of hours per day and per week that students can be tutored. And before the waiver, some of the providers tutored our students two nights, three nights at 2 hr each tutoring session.

The importance of these changes was echoed in an interview with administrators in the state Department of Public Instruction in 2012: "it seems like Milwaukee … it does seem like they use a lot of the research that you had gotten in terms of the per-pupil amount, seems to be a little more reasonable." By the 2013–2014 school year, the median number of hours of tutoring received had reached 40 hr, and the tutor-to-

Table 2. MPS Tutoring 4 You (T4U) evidence-based program redesign.

2011–2012 UNDER NCLB	2012–2013 TO 2014–2015 UNDER T4U
Hours of service	**Hours of service**
9–26	>40
Class size	**Class size**
1:1–10:1+	1:1–5:1
Grade Level	**Grade Level**
K–5/K–12	K5–K–12 (only Grade 2 in 2014–2015)
Service Delivery Model	**Service Delivery Model**
In home, in community, online, school	School, online (synchronous)
Math, or literacy; can combine in one session	Either math or literacy for the year
Hours per session	**Hours per session**
2 hr – 2 × 3 per week	1 hr – 2 × 3 per week
Enrollment	**Enrollment**
Providers canvas schools and community	Schools conduct process
Providers service a variety of schools	Providers assigned to 1–2 schools (reduced to 2 total providers in 2014–2015)
	No provider marketing – MPS identifies & recruits
	Parent/guardian must complete the registration
Hourly rate	**Hourly rate**
$55.00–$108.00 per child in session 1:1–10:1+	$35 per hour rate per session 1:1 to 5:1
English Language Learners	**English Language Learners**
Services provided in native language and/or English	Services in native language if low LAU* level and English
Materials not in native language	Materials provided in native language if low LAU level
Special Education	**Special Education**
Services lacked differentiation	Services appropriate to individual education plan
Tutor Qualifications	**Tutor Qualifications**
No Department Public Instruction (DPI) guidelines	Certified Teacher in subject are tutored
	Match elementary tutor to Math or Literacy area
Target Population	**Target Population**
Schools identified in need of improvement; free lunch – eligible; all students at site, priority to minimal and basic	Focus schools to address gaps
	Minimal/basic level of proficiency
	No services for proficient or advanced students
Monitoring	**Monitoring**
DPI, Principal, Coordinators, Supplemental Educational Services Office	Fidelity of implementation "walks"
	Site coordinator, Desktop audits
	Provider site visits
	Tutoring observations, file review, review of findings meeting.
	On-site reviews may be announced or unannounced – findings may be shared with the public and/or the Wisconsin Department of
	Public Instruction
	Parent Survey
Curriculum	**Curriculum**
DPI-approved curriculum	District-aligned supplemental reading or math curriculum
Site Coordinator	**Site Coordinator**
MPS not mandatory	Position exists at school level but no additional compensation (2012–2013)
Termination	**Termination**
Terminate for cause	Terminate for failure to achieve desired student progress listed in learning plan
Right to Cure	
Individual student services	

*LAU refers to levels of English language proficiency: **Lau Code A –** Pre-functional level limited English proficiency; **Lau Code B –** Beginning level limited English proficiency; **Lau Code C –** Intermediate and Advanced level limited English proficiency; **Lau Code D –** Exited from English language support services; **Lau Code E –** English proficient, no language service required.

student ratio in all observations of sessions was no larger than 1:3. In addition, the difference between advertised and instructional time (as recorded in observations during the 2013–2014 school year) had been reduced to less than 5 min.

Curriculum and instruction

The observations of tutoring sessions across the districts also revealed considerable *intra*-provider variation in both instruction and curriculum materials, observed in a variety of formal (website or provider materials) and informal sources (tutors' own resources or students' work from day school) used in sessions. As a result, the "in-use" curriculum was sometimes inconsistent with the formal curriculum, which is problematic given that conflicting day school and tutoring instructional strategies can negatively affect a student's day school instruction and hinder tutor efforts to meet students' instructional needs (Good, Burch, Stewart, Acosta, & Heinrich, 2014). Furthermore, although 18 out of 25 providers in our qualitative sample across five districts advertised they could serve students with disabilities, our field observations and review of tutor training materials suggested that with few exceptions, tutors did not have specific training or experience in supporting students with disabilities. Similarly, 21 of 25 providers advertised services for English learners, but our observations often found tutoring staff without targeted training or experience in serving English learners. In addition, some tutoring providers did not have access to school records or staff with knowledge about students' needs to appropriately adapt instruction. Accordingly, we found across all of our district research partners that tutoring services were less effective for students who were English learners or had special needs. MPS subsequently required all T4U providers to use a curriculum directly aligned to that of the district day school, with drill-down goals for students with special needs. T4U providers are required to provide appropriate services for students with special needs and English learners, and MPS requires information sharing between the schools and tutors on the needs of students with disabilities. Additionally, when sufficient staff are available, T4U tutors are required to be certified teachers. For example, the majority (8 of 10) tutors observed in T4U sessions in 2013–2014 were certified teachers or specialists.

Procurement and management of tutoring providers

Finally, MPS made two additional program design changes in the first 2 years of T4U that aimed to better regulate the pool of tutoring providers, reduce program management challenges and improve student outcomes. First, 13 of the 14 tutoring providers returned to the school district in 2012–2013 to offer T4U services (despite the dramatically reduced hourly rates paid for tutoring). Under NCLB, we had observed in MPS and across all research sites (and through our focus groups with parents) how tutoring providers marketed their services heavily to parents and used incentives to attract students to their programs. On the basis of these findings, our partnering districts attempted to implement policies, within the constraints of NCLB, that limited providers' use of incentives. Following the waiver from NCLB, MPS took these efforts a step further, and rather than allowing providers to compete for student "market shares," MPS assigned each provider to one to two schools in the district, contributing to a more even distribution of students across the providers (who were prohibited from conducting marketing campaigns and directly enrolling students in their programs). In fact, some providers expressed in interviews a preference for this system, as it allowed them to focus more on programming and staffing than on recruitment. Second, MPS began contracting with only two tutoring providers via a competitive process in

2013–2014: one that delivers services in online tutoring sessions and the other that operates solely in face-to-face formats. MPS selected two providers that our research showed had positive effects on reading test scores and were among the most effective in increasing students' math test scores in 2012–2013. In effect, nearly every change that MPS made in developing the T4U program (see Table 2) was based on evidence generated through the years of collaborative research, dialogue, and co-production of insights in the partnership.

Patterns in student outcomes following evidence-based program changes

An important objective of this paper was to move beyond solely descriptive documentation of the mechanisms and potential benefits of research–practice partnerships, such as building district capacity for data use, to explore the potential linkages between the work of the partnership and changes in students' educational outcomes over time. Here, we present the results of the value-added models (described above) that were estimated to examine associations specifically between student participation in T4U and changes in their math and reading achievement, as well as to compare them with the estimated effects of participation in supplemental educational services (before evidence-based program changes were made).

In the 1st year of T4U (2012–2013), we estimated both average program effects (as shown in Equation [1]) and effect sizes for each of the 13 providers offering tutoring in math and reading (adding provider dummies to this equation). While in the prior school year, only one tutoring provider had been identified as effective in increasing student's math achievement and one other as effective in increasing student's reading achievement, in 2012–2013, all but two of the 13 tutoring providers were effective in increasing student achievement in at least one of these subjects (see Figures 2 and 3 that present provider-specific effects graphically). In the subsequent school years (2013–2014 and 2014–2015), when only two providers were delivering T4U services, we saw additional gains in students math and reading achievement; see Figures 4 and 5, which compare average gains in student achievement across all four school years (under NCLB in 2011–2012 and 2012–2015 for T4U). Standard error bars on the effect estimates shown in Figures 2 and 3 indicate that while there was no statistically significant effect of tutoring on student reading or math achievement during 2011–2012, all estimated effects of T4U on student math and reading achievement were positive and statistically significant in the subsequent years.

Furthermore, additional fixed-effects models were estimated to examine whether the increased number of hours of tutoring students received under T4U was an important mechanism (as suggested by the research–practice partnership evidence) for increasing student achievement. The findings showed (as reported to MPS) that for each additional hour of tutoring received, students participating in T4U increased their performance on the math MAP test by 0.01323 units (or 0.0016 standard deviations per additional hour of tutoring) and 0.05037 units (or 0.003 standard deviations) on the reading MAP test.[4]

These increases in student test scores were also accompanied by changes in the instructional settings observed in the T4U program. Although we cannot make general claims about all T4U sessions based on data from our sample of observed sessions in 2012–2013 and 2013–2014, we did find contrasts in instructional quality under T4U compared to observations of supplemental educational services from 2009–2012 (under

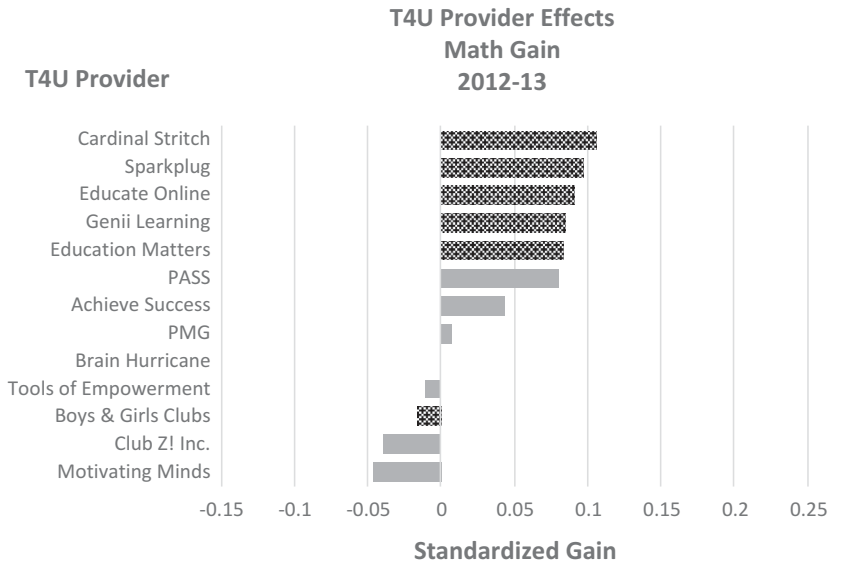

Figure 2. Estimated effects of T4U providers on student math achievement, 2012–2013.
Note: Patterned bars indicate statistically significant effects at α = 0.05,

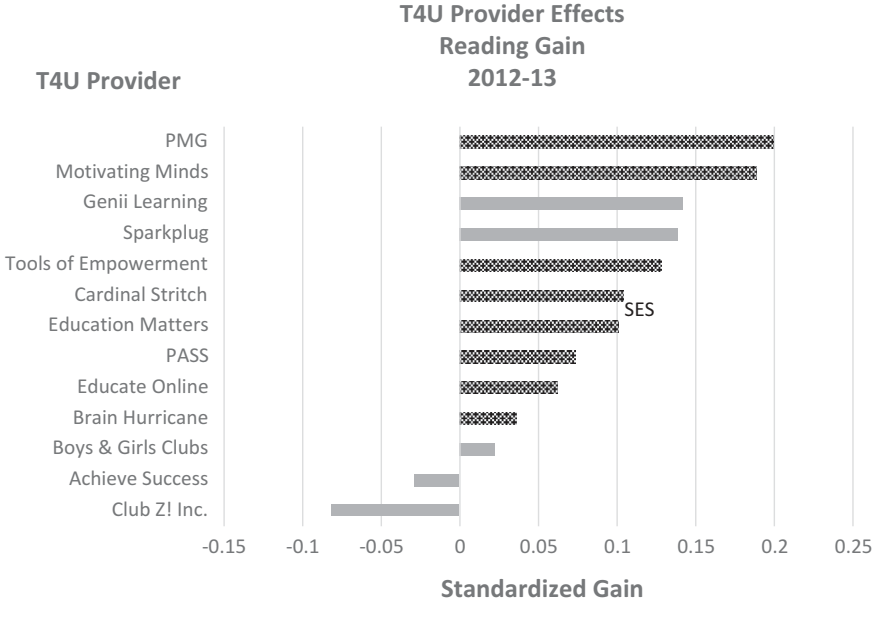

Figure 3. Estimated effects of T4U providers on student reading achievement, 2012–2013.
Note: Patterned bars indicate statistically significant effects at α = 0.05.

NCLB). For example, where the observed tutor-to-student ratio in MPS ranged as high as 1:9 under NCLB, the ratio of all observed sessions in 2013–2014 under T4U were consistently equal to or less than 1:3. We also observed improvements in the actual

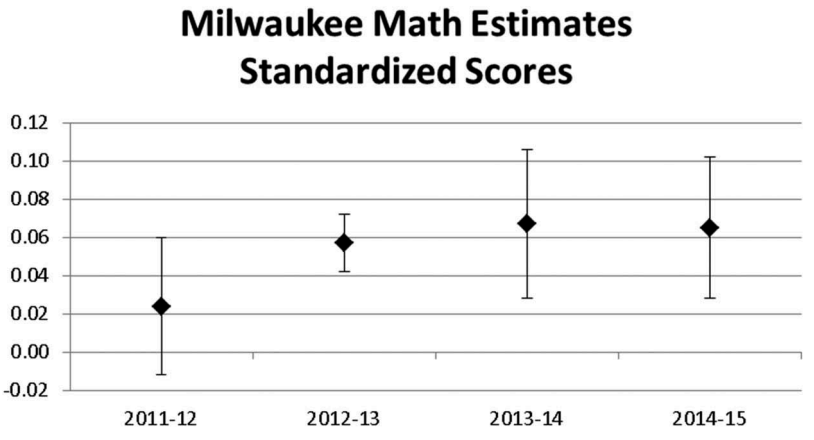

Figure 4. Estimated T4U effects on student math achievement, 2011–2015.

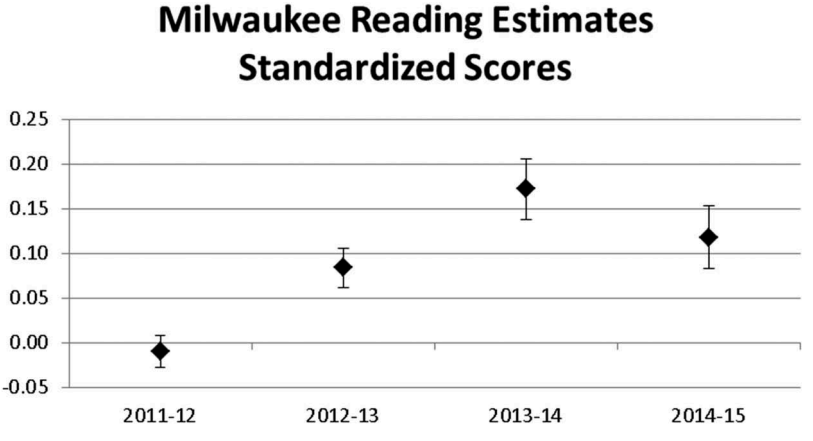

Figure 5. Estimated T4U effects on student reading achievement, 2011–2015.

instructional time in each session. Compared to an average difference of 20 min for school-based tutoring formats under NCLB in the five district sites, the difference between advertised and instructional time was less than 5 min in all but one observation of T4U tutoring in 2013–2014.

In addition, 8 of 10 tutors in T4U sessions observed in 2013–2014 were certified or retired teachers or specialists. This is compared to a broad array of experience and training in the NCLB context, where there were no requirements on tutor qualifications. And overall, consistently positive student engagement was observed across both online and in-person tutoring formats under T4U. For example, we saw in the observations of tutoring more instances where students were asked to actively apply skills and knowledge by an engaged tutor in small-group settings. This was in contrast to supplemental educational services in which we commonly observed large-group "homework help" sessions or worksheet drills, where even the tutors themselves made comments such as, "I am not a big fan of [the provider's] materials. I like to pick and pull – I'm not a drill-

and-kill person." Lastly, parents and family members across two focus groups in the spring of 2014 reflected more positively on the T4U experience and believed tutoring was improving their students' skills, as well as their confidence level in school.

Conclusions

The research collaboration with MPS and other school districts described here did not begin with a "grand plan" of how to make a long-term research relationship in support of evidence-informed policy-making work, nor was the infrastructure or large-scale funding in place from the start to "grow" it. The research relationship began with a small-scale, district-initiated and funded study and then built structures for collaboration over time, including the trust that undergirds them. The external funding that came in to support and expand the research to other sites and over time was critical for deepening the research relationships and strengthening the feedback loops between the researchers and practitioners. We also learned along the way the importance of a flexible (and long-term) orientation to the nature and focus of the research investigation, allowing for local policy and program priorities to help shape the research agenda and scope of the work over time. Some of the partnership infrastructure that was developed, such as end-of-year/summer research briefings, promoted the iterative research–practice cycle, with learning ongoing on the part of both the researchers and practitioners.

Not surprisingly, seeing visible program successes in both implementation and outcomes as the partnership progressed bolstered the interest of district personnel in continuing the research and in constructing systematic approaches to engaging with it. The ongoing collaboration with MPS facilitated many opportunities to apply findings and to assess the effects of the changes made to policy and practice. In due course, we were able to observe patterns in associations between the research-informed policy and program changes implemented and the end goal of increasing student achievement. The district's steady engagement in the research process and its regular use of the co-produced findings to guide program improvement were in itself an important transformation.

However, in the absence of a randomized experimental design, it is not possible to assert a causal relationship between changes made by the district in response to the research evidence and the subsequent improvements observed in student reading and math outcomes (in a specific, post-NCLB intervention) over time. Future research should strive to develop empirical measures of the use of research evidence, which we expect would require the types of infrastructure developed in this partnership for regularly sharing data, policy documentation, and practitioner insights. Ideally, a study designed for causal identification of the effects of the use of research evidence on student outcomes would incorporate both a randomized design and rigorous, empirical measures of evidence use. On the basis of our own experiences, we see this as a lofty research goal, and it is important for academics setting out to establish strong and enduring research collaborations with education stakeholders to have realistic expectations of the time and resources involved in developing and sustaining a long-term research relationship, especially given these types of investments may not always be "rewarded" in academia. Moreover, academic partners need to be open to refining their own research interests or agenda in order to come to a productive middle ground that generates research evidence that is responsive, applicable,

and will be valuable to all partners in the collaboration. This also entails, as we have shown, flexibility in the research design and openness to developing new approaches and infrastructure for sharing research findings and maintaining "two-way" channels for communicating and co-producing evidence over time.

In turn, policymakers and practitioners likewise face challenges in making longer term commitments to research collaboration, especially given the time demands of adding research responsibilities to an already onerous load of regular programmatic and administrative responsibilities. They may also be required to take some "political" risks in pursuing research collaborations with academics, where there will be expectations for publicly disseminating the findings, even if outcomes are not favorable. Risks and expectations (for positive outcomes and benefits for stakeholders) may grow with greater investments of time and resources over the course of a partnership, raising the question of whether the stakeholders are open to learning and continuing with the partnership, even in the face of disappointing results. In this regard, seeing unexpected or negative results as an opportunity to develop and implement new research-informed program or practice changes, as we experienced in our research–practice partnership with MPS and other districts, can go a long way toward identifying pathways to improved program (and student) outcomes.

Notes

1. The expansion of this research to a multi-site evaluation of the implementation and impacts of supplemental educational services was made possible by a 3-year grant from the Institute of Education Sciences. Other external funding sources have since supported the continuation of the research partnership with MPS.
2. Texas House Bill 753 was developed by a member of our research team and a Texas state representative to ensure that parents have full access to existing, rigorous information on the effectiveness of tutoring providers. The bill was signed into law on 14 June 2013, and became effective on 1 September 2013 (see http://openstates.org/tx/bills/83/HB753/).
3. In prior research (Heinrich et al., 2014), we confirmed that whether controlling for one or two pretest waves of standardized test scores, we obtained consistent estimates of the effects of tutoring programs on student achievement.
4. Results are available from the authors upon request.

Acknowledgments

We thank the funder of this research, the Institute of Education Sciences, PR/Award number: R305A090301, Education Policy, Finance and Systems Research Program, Goal 3. In addition, we express our most sincere thanks to the staff in the study school districts for their time and support for this project. We would also like to thank staff at the Wisconsin Center for Education Research, the University of Wisconsin-Madison, and the University of Texas at Austin Center for Health and Social Policy for their help and support of this work.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Institute of Education Sciences under Grant number [R305A090301].

Notes on contributors

Carolyn J. Heinrich (PhD, University of Chicago) is a Professor of Public Policy and Education in the Department of Leadership, Policy, and Organizations at the Peabody College and a Professor of Economics in the College of Arts and Sciences at Vanderbilt University. Her research focuses on education, workforce development, social welfare policy, program evaluation, and public management and performance management.

Annalee Good (PhD, University of Wisconsin-Madison) is a researcher at the Wisconsin Center for Education Research, co-director of the Wisconsin Evaluation Collaborative, and Director of the WCER Evaluation Clinic. Her current projects include studies of digital tools in K–12 supplemental education, evaluation of academic tutoring partnerships, and the challenges of instruction and assessment for advanced learners.

ORCID

Carolyn J. Heinrich http://orcid.org/0000-0002-7731-9138

References

Bifulco, R. (2012). Can nonexperimental estimates replicate estimates based on random assignment in evaluations of school choice? A within-study comparison. *Journal of Policy Analysis and Management*, *31*(3), 729–751. doi:10.1002/pam.20637

Bryk, A., Gomez, L., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. In M. T. Hallinan (Ed.), *Frontiers in sociology of education* (pp. 127–162). Dordrecht: Springer.

Burch, P., & Heinrich, C. J. (2015). *Mixed methods for policy research and program evaluation*. Thousand Oaks, CA: SAGE.

Coburn, C. E., & Penuel, W. R. (2016). Research–practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher*, *45*(1), 48–54. doi:10.3102/0013189X16631750

Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013). *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts*. New York, NY: William T. Grant Foundation.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, *27*(4), 724–750. doi:10.1002/pam.20375

Dagenais, C., Lysenko, L., Abrami, P. C., Bernard, R. M., Ramde, J., & Janosz, M. (2012). Use of research-based information by school practitioners and determinants of use: A review of empirical research. *Evidence & Policy*, *8*(3), 285–309. doi:10.1332/174426412X654031

Deke, J., Dragoset, L., Bogen, K., & Gill, B. (2012). *Impacts of Title I supplemental educational services on student achievement* (NCEE 2012-4053). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Easton, J. Q. (2010, April-May). *Out of the tower, into the schools: How new IES goals will reshape researcher roles*. Presidential address presented at the Annual Meeting of the American Educational Research Association, Denver, CO.

Finnegan, K. S., Daly, A. J., & Che, J. (2013). Systemwide reform in districts under pressure: The role of social networks in defining, acquiring, using, and diffusing research evidence. *Journal of Educational Administration*, *51*(4), 476–497. doi:10.1108/09578231311325668

Good, A. G., Burch, P. E., Stewart, M. S., Acosta, R., & Heinrich, C. (2014). Instruction matters: Lessons from a mixed-method evaluation of out-of-school time tutoring under No Child Left Behind. *Teachers College Record*, *116*(3), 1–34.

Gueron, J. M., & Rolston, H. (2013). *Fighting for reliable evidence*. New York, NY: Russell Sage Foundation.

Hallberg, K., Cook, T. D., Steiner, P. M., & Clark, M. H. (2016). Pretest measures of the study outcome and the elimination of selection bias: Evidence from three within study comparisons. *Prevention Science*. Advance online publication. doi:10.1007/s11121-016-0732-6

Haskins, R., & Margolis, G. (2014). *Show me the evidence: Obama's fight for rigor and results in social policy*. Washington, DC: Brookings Institution Press.

Head, B. W. (2016). Toward more "evidence-informed" policy making? *Public Administration Review*, *76*(3), 472–484. doi:10.1111/puar.12475

Heinrich, C. J. (2010). Third-party governance under No Child Left Behind: Accountability and performance management. *Journal of Public Administration Research and Theory*, *20*(Suppl 1), i59–i80. doi:10.1093/jopart/mup035

Heinrich, C. J., Burch, P., Good, A., Acosta, R., Cheng, H., Dillender, M., … Stewart, M. S. (2014). Improving the implementation and effectiveness of out-of-school-time tutoring. *Journal of Policy Analysis and Management*, *33*(2), 471–494. doi:10.1002/pam.21745

Heinrich, C. J., & Nisar, H. (2013). The efficacy of private sector providers in improving public educational outcomes. *American Educational Research Journal*, *50*(5), 856–894. doi:10.3102/0002831213486334

Jones, C. (2009). *The 2009 Supplemental Educational Services program: Year 4 summative evaluation*. Chicago, IL: Chicago Public Schools Office of Extended Learning Opportunities and Office Research, Evaluation, and Accountability.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*(2), 275–313. doi:10.3102/00346543076002275

Nelson, S. R., Leffler, J. C., & Hansen, B. A. (2009). *Toward a research agenda for understanding and improving the use of research evidence*. Portland, OR: Northwest Regional Educational Laboratory.

Owen, J. W., & Larson, A. M. (Eds.). (2017). *Researcher–policymaker partnerships: Strategies for launching and sustaining successful collaborations*. Abingdon: Taylor and Francis.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*(5), 533–544. doi:10.1007/s10488-013-0528-y

Schildkamp, K., Poortman, C., Luyten, H., & Ebbeler, J. (2017). Factors promoting and hindering data-based decision making in schools. *School Effectiveness and School Improvement*, *28*(2), 242–258. doi:10.1080/09243453.2016.1256901

Shanley, L. (2016). Evaluating longitudinal mathematics achievement growth: Modeling and measurement considerations for assessing academic progress. *Educational Researcher*, *45*(6), 347–357. doi:10.3102/0013189X16662461

Tseng, V. (2012). The uses of research and evidence in policy and practice. *Social Policy Report*, *26*(2), 1–16.

Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review*, *39*(5), 426–431. doi:10.2307/3109916

Whitehurst, G. J. (2012). The value of experiments in education. *Education Finance and Policy*, *7*(2), 107–123. doi:10.1162/EDFP_a_00058

Yohalem, N., & Tseng, V. (2015). Commentary: Moving from practice to research, and back. *Applied Developmental Science*, *19*(2), 117–120. doi:10.1080/10888691.2014.983033

Zimmer, R., Hamilton, L., & Christina, R. (2010). After-school tutoring in the context of No Child Left Behind: Effectiveness of two programs in the Pittsburgh Public Schools (PPS). *Economics of Education Review*, *29*(1), 18–28. doi:10.1016/j.econedurev.2009.02.005

Appendix 1 Tables providing additional study details

Table A1. Tutoring providers operating in district partner sites (2009–2013).

Provider	Austin	Chicago	Dallas	Los Angeles	Milwaukee	Minneapolis	FORMAT
A+ Tutoring Services		X (10–12)				X (09–13)	School, home, community
A+ Markem			X (11–13)				community
Aavanza				X (12–13)			Digital (mixed)
Academic Advantage				X (12–13)			Digital
Academic Coaches	X (11–13)						School
A.I.M. High		X (09–13)					Home
AISD Supplemental Educational Services Tutors	X (09–10)						School, home
A Better Grade					X (09–10)		Home
Apex Academics	X (10–11)						School
ATS Project Success						X (09–13)	Digital (not live)
BabbageNet		X (09–12)					Digital (not live)
Black Star Project		X (09–12)					School
Brain Hurricane					X (12–13)		School
COMEDI	X (09–10)						School
Confidence Music			X (11–13)				School
Educate Online		X (09–12)	X (09–12)		X (09–13)		Digital (live)
Focus First	X (10–12)		X (11–12)				School
Group Excellence	X (10–11)		X (09–13)				School, community
Learning Disabilities Association						X (09–12)	Home
Learning Exchange					X (09–10)		School
Learn It Systems	X (10–11)			X (12–13)			Digital (live)
Mainstream Development					X (09–12)		School
MIGIZI Native Academy						X (09–13)	Community, home
Newton Learning	X (12–13)						School
Orion's Mind		X (09–13)	X (10–11)				School
Rocket Tutoring			X (10–11)				School
Salem Educational Initiative						X (09–12)	Home, community
Supplemental Educational Services Texas Tutors			X (09–10)				School
Sparkplug					X (09–13)		School
Step Ahead					X (09–12)		School, home
TutorCo						X (09–12)	Digital (live)
Wisconsin Education Matters					X (10–12)		Home

Table A2. Number of students eligible, registered, and attending free tutoring by school year and district.

	Eligible	Registered	Attending
2007–2008 School year			
Milwaukee	8,284	3,704	2,194
Minneapolis	9,217	1,345	1,124
Chicago	166,386	46,856	37,095
Dallas	24,031	6,179	4,632
Austin	6,297	489	212
2008–2009 School year			
Milwaukee	11,992	4,267	2,689
Minneapolis	10,618	2,567	1,412
Chicago	212,504	69,073	56,921
Dallas	36,770	8,619	4,911
Austin	7,330	2,761	1,929
2009–2010 School year			
Milwaukee	26,798	6,933	4,998
Minneapolis	16,484	4,910	3,533
Chicago	135,418	65,531	33,465
Dallas	30,916	10,950	10,637
Austin	4,470	2,986	1,306
2010–2011 School year			
Milwaukee	16,439	3,271	2,592
Minneapolis	16,985	3,028	2,245
Chicago	101,930	34,838	30,306
Dallas	35,026	10,685	8,661
Austin	1,805	n.a.	1,069
2011–2012 School year			
Milwaukee	20,905	5,239	4,221
Minneapolis	16,055	3,284	2,576
Chicago	98,203.	33,348	11,357
Dallas	39,091	10,862	7,941
Los Angelesa	326,117	63,603	26,430

*Los Angeles Unified School District joined the research collaboration in 2011, and Austin was no longer required to offer supplemental educational services after 2011.

Table A3. Number of observations of full tutoring sessions and number of participants in interviews/focus groups across districts in 2009–2010, 2010–2011, and 2011–2012 school years.

	Observations	Provider interviews	State/district admin. interviews	Parent focus group
Austin	21	25	6	13
Chicago	20	24	4	16
Dallas	25	21	8	45
Los Angeles	29	11	2	15
Milwaukee	35	33	6	33
Minneapolis	32	39	8	61
TOTAL	**162**	**153**	**34**	**183**

Table A4. Description of variables available for empirical analysis across the study sites.

Control Variables	Description	Site-Specific Details
Eligibility and Registration	Eligible for out-of-school Tutoring (OST)	
	Registered for OST Tutoring with district	
	Received Tutoring (non-zero hours)	
Student Identification and Enrollment Information	Student ID	
	District Code	
	District Assigned Local Identification Number	
Student Demographic Information	Student Gender	
	Student Ethnicity (Black, White, Hispanic, Asian, or Other)	
	Student Age	
	Limited English Proficiency/English Language Learner (ELL) Indicator	
	Economic Disadvantage Status (Indicated by Free or Reduced Lunch)	
	Enrolled in Special Education Program during School Year	
	Retained in the Same Grade as the Prior School Year	
	Period for which attendance is recorded	
Attendance and Absence Information	Percent of Days Absent in Prior School Year	
	Number of Days Absent During the Reporting Period	
Basic Treatment Measures	**Description**	**Site-Specific Details**
Treatment Information	Attended Any Tutoring in Prior School Year	
	Hours of Tutoring Received (Invoiced)	
	Tutoring Provider	
Primary Outcome Measures	**Description**	**Site-Specific Details**
Reading Measures	Change in Reading Scores (Standardized with District Average Test Scores)	*Dallas, Austin*: TAKS, STAAR,
	Change in Reading Scores (Standardized with Average Test Scores of students eligible for supplemental educational services)	*Chicago*: ISAT, ITBS *Los Angeles*: Standardized Testing and Reporting (STAR) *Milwaukee*: WKCE *Minneapolis*: MCA-II
Math Measures	Change in Math Scores (Standardized with District Test Score Averages)	*Dallas, Austin*: TAKS, STARR
	Change in Math Scores (Standardized with Average Test Scores of students eligible for supplemental educational services)	*Chicago*: ISAT, ITBS *Los Angeles*: Standardized Testing and Reporting (STAR) *Milwaukee*: WKCE *Minneapolis*: MCA-II, MTELL

Table A5. Estimated average impacts of attending tutoring by school district and year.

	2008–2009		2009–2010		2010–2011		2011–2012	
School District	# of Students with gain scores	Effect size	# of Students with gain scores	Effect size	# of Students with gain scores	Effect size	# of Students with gain scores	Effect size
Reading Achievement								
Chicago	61,171	**0.043**	63,506	**0.094**	205,187	**0.075**	68,541	**0.042**
Minneapolis	2,862	–0.202	1,602	–0.202	5,025	**0.144**	4,247	–0.037
Milwaukee	4,697	–0.079	1,841	–0.079	2,826	0.021	3,668	–0.020
Dallas	9,294	–0.109	14,106	**0.111**	13,428	0.016	14,670	0.011
Los Angeles	n.a.	n.a.	n.a.	n.a.	32,453	–0.0115	44,383	**0.041**
Math Achievement								
Chicago	61,464	**0.046**	63,773	**0.053**	204,094	**0.064**	68,411	**0.045**
Minneapolis	1,400	–0.011	789	–0.011	5,045	**0.191**	4,298	0.050
Milwaukee	4,772	–0.048	1,870	–0.048	2,831	–0.043	3,663	0.031
Dallas	9,294	–0.076	13,807	**0.127**	13,333	0.016	14,361	0.054
Los Angeles	n.a.	n.a.	n.a.	n.a.	31,990	–0.0115	43,607	**0.061**

Note: Boldface indicates statistically significant effect size at $\alpha = 0.05$.