Symposium Title:
Informing Practice through Research:
Lessons from four research-practice partnerships

Symposium justification:

This symposium features four papers that describe how research, conducted in close collaboration with program partners, informed decisions about programming. In order to produce findings that are useful to practitioners, we aimed at being responsive to the changing needs of developing programs, as well as our broader research aims. Aligned with this year's conference theme, each author will highlight instances where the research process and findings informed program and policy decisions. We use the dimensions of research-practice partnership proposed by Henrick, Cobb, Penuel, Jackson, and Clark (2017) to reflect on our process.

The first paper documents how conversations between researchers and program staff helped generate a theory of action and thresholds for fidelity of implementation for an in-school and after-school middle grades science program. By clarifying causal mechanisms in advance of the first year of implementation and sharing implementation data at the end of the year, this process helped develop practitioner capacity to engage in the research process by elaborating hypotheses about program functioning and developing systems to monitor implementation.

The second paper details a collaboration between internal researchers and program staff to iteratively develop a daily log used by reading coaches to report on the reading content and coaching methods used in their work with K-2 teachers. Throughout the program's pilot year, researchers distributed monthly reports on these logs to a variety of stakeholders, including district leadership, program leadership, and the coaches themselves. These reports and other descriptive analyses supported practitioners is several ways: they built a shared understanding of the work, grounded leadership's guidance to coaches, and allowed all stakeholders to reflect on the work.

The third study discusses a collaboration between a researcher and a network improvement community (NIC) of youth development organizations focused on improving students' social emotional learning skills. The author will share both strategies used to improve the reliability of measurement and data collection procedures, rigorous analytic procedures used to identify sites associated with large positive changes in students' socio-emotional outcomes, and lessons learned through this close collaboration about how to maximize the relevance of results.

The fourth paper describes how evidence from a mixed-method evaluation of a district-wide initiative aimed at improving outcomes for Black and Latino young men informed decisions about the program's structure and implementation over time. It also describes post-evaluation fora and reflection exercises designed to identify lessons from this study that can inform educational improvement efforts more broadly.

Combining these four papers into a symposium is a unique opportunity to examine and compare multiple examples of research-practice partnerships. Across all of these papers, authors will share how research findings were generated and presented to maximize their relevance to informing practitioners' ongoing implementation decisions. The discussant, Kylie Klein, will provide comments on each paper and the symposium as a whole.

Getting Started: Specifying the theory of action and fidelity of implementation measures to guide program improvement

By Cheri Fancsali

Background

Design2Learn (D2L), funded through the US DOE's Investing in Innovation fund, is an afterschool model for middle school students. The model is aimed at increasing students' interest, engagement, and academic performance in science. Students participate in weekly science instruction after school, facilitated by a three-person team (one certified teacher; two informal educators) and designed to connect science learning goals to real-world activities. Students enroll in Design2Learn in sixth grade, and continue the program through eighth grade.

The program combines three research-based strategies:

1) **collaborative teaching**, which brings together certified teachers and afterschool educators to jointly deliver instruction, and has been found to predict positive student performance (Moorehead and Grillo, 2013; Zito, 2011). D2L builds on this powerful approach by teaming in-school teachers with afterschool educators who have expertise in positive youth development practices (Krishnamurti, Ballard, and Noam, 2014).

2) **curricular bridging,** which coordinates in-school and afterschool instruction to create seamless student learning experiences (Noam, 2003). Curricular bridging is associated with increased student performance and shows promise as a strategy for engaging students by connecting their "divergent worlds," making learning "more meaningful and relevant to [their] life experience" (Noam, 2003; Noam et, al., 2002).

3) **design-based learning**, in which students identify a design problem; consider options and constraints; and plan, model, test, and iterate solutions, making higher-order thinking skills more tangible and visible. Research shows that design-based activities are highly motivating, can help students develop deep conceptual understanding, and support the development of self-guided inquiry skills critical to success in science (Kimmel, 2006; Kolodner, 2000; Sadler, 2007).

Research Questions

This study is an implementation and impact study of the Design2Learn model. The confirmatory impact research question is: What is the impact of the opportunity to participate in Design2Learn for at least two years on 8th grade students' interest and engagement in science? The study also explores questions related to program implementation, such as fidelity to the model and implementation facilitators and barriers.

Setting and Study Participants

This study is located in 32 high-needs afterschool programs (i.e., serving high proportions of students in poverty and with low prior achievement) in NYC. We recruited 8 sites in 2016 (cohort 1), and an additional 14 sites in 2017 (cohort 2). Within each program, we recruited two after-school educators and one in-school science teacher who deliver afterschool programming to approximately 25 students.

Research Design

This study is a cluster randomized controlled trial with program-level randomization. Programs were randomly assigned to participate in D2L (treatment) or business as usual STEM programming (control). Educator teams in the treatment sites participate in the D2L professional development and coaching support.

Data Collection/Analysis

We measured fidelity of implementation (FOI) for three key components of the program: attendance at the professional development sessions, participation in site-based strategic planning meetings, and delivery of and participation in on-site instructional coaching and technical assistance. Professional development and strategic planning were measured through attendance rosters collected by the PD facilitators. Instructional coaching was measured through reports completed by coaches each day. From educators, we collected end of the year surveys to measure changes in practice, and successes and barriers to implementation. We also conducted observations of treatment and control program activities and interviewed program staff to determine the treatment/control contrast. To measure impact on students, we collected pre/post interest and engagement surveys, attendance (in the afterschool and in school), and science standardized test data.

We will use a multilevel model to estimate the impact of Design2Learn on student outcomes.

Findings

Prior to implementation, we worked with the program developers to refine and revise the intervention's theory of action (TOA). Through a series of discussions, we identified the key inputs and activities; expected changes in educator practices, knowledge and attitudes as a result of those activities; and hypothesized short and long-term student outcomes. The discussions helped elicit the expected causal mechanisms by which the intervention was intended to work. Next, we worked with the developers to identify the level at which implementation (inputs and activities) needed to occur to reasonably expect an impact on the hypothesized outcomes. The developers reported that this process helped them to identify and prioritize the activities that were most likely to result in the intended outcomes. For example, hypothesizing that school administrator attendance at the strategic planning sessions would support curricular bridging, the developers emphasized this as an expectation and included it as a criteria for FOI. Findings from implementation data supported this hypothesis.

After the first year of implementation, the researchers and developers revisited the TOA and fidelity thresholds, using implementation data to determine if adjustments were necessary. For example, educator survey data indicated a need for additional instructional materials (such as lesson plans and design activities) as they planned year 2. The program developers adjusted the PD and coaching to provide these supports, and FOI thresholds were adjusted to incorporate instructional materials as part of the on-site coaching.

In this presentation, we will describe the process we used to generate the TOA and FOI indicators (including parallel implementation indicators from the control sites to assess the treatment/control contrast) and use them to guide implementation. Ultimately, this process will allow us to investigate the relationship between implementation and impacts. For example, if we find no impact on hypothesized outcomes, the TOA and FOI data will provide insight into whether this is due to poor implementation, cross-over effects, or a flawed TOA.

Conclusion

This work points to the value of researchers and program developers investing effort early in a project's life cycle to specify the intervention's TOA and criteria for determining adequate delivery and implementation of the intervention. Doing so supports continuous reflection and allows for evidence-based adaptations to the intervention. Further, it ensures that the evaluation collects data that is necessary to determine which components work and which do not, informing the interpretation of impact findings, as well as future replication and scale-up of the model.

References

Kimmel, H., Carpinelli, J. Burr-Alexander, L., & Rockland, R. (2006, June). Bringing engineering into K–12 schools: A problem looking for solutions? *Proceedings of the American Society for Engineering Education Annual Conference and Exposition. Chicago.*

Krishnamurthi, A., Ballard, M. & Noam, G. (2014, July). Examining the impact of afterschool STEM programs. A paper commissioned by the Noyce Foundation.

Moorehead, T. & Grillo, K. (2013). Celebrating the Reality of Inclusive STEM Education: Co-Teaching in Science and Mathematics. *TEACHING Exceptional Children, v45 n4 50-57*

Noam, G. (2003). Learning with excitement: Bridging school and after-school worlds andproject-based learning. *New Directions for Youth Development, 97: 121-138*

Noam, G., Biancarosa, G. & Dechausay, N. (2002). Learning to Bridge–Bridging to Learn: A Model and Action Plan to Increase Engagement Between Schools and Afterschool Programs in Boston. *Program in Afterschool Education and Research (PAER), Harvard University. A Report Commissioned by Boston's After-School for All Partnership Learning Goal Research September, 2002*

Zito, M. (2011). Is working together worth it? Examining the Relationship Between the Quality of Teacher Collaboration, Instruction, and Student Achievement. Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of Doctor of Education, May 2011.

Coaching Logs:

A tool for informing, grounding, and reflecting on pilot year implementation.

By Rachel Cole, Lauren B. Goldenberg, Andrew Fletcher, & David Braslow

Introduction: One way to support teachers in understanding how children acquire reading skills is providing job-embedded professional learning through coaching. Coaching can help ensure that new skills are transferred to classroom practice (Joyce & Showers, 1980, 1981). Time spent with literacy coaches can lead to gains in student achievement (Bean et al, 2008; Elish-Piper & L'Allier, 2011).

Understanding how coaches spend time in schools is essential. Researchers and practitioners co-developed a log for measuring coach time use. Researchers conducted regular discussions of patterns in the data and observed the influence on the practice of program leadership, coaches, and other stakeholders, and on themselves.

With 70% of students not proficient on the 3rd grade reading assessments, district leaders invested in a major early literacy initiative. In the pilot year 2016-17, the district hired over 100 teachers, trained them in reading content and the craft of coaching, and placed them in 107 schools. Coaches were charged to work with teachers on research-based reading instruction; they were not to work directly with students except for modeling instructional practices. Coaches worked with approximately 1,700 teachers who taught about 30,000 students. Program leadership collaborated with researchers in order to use research and data for decision-making.

Research Design: Longitudinal descriptive study.

Data Collection and Analysis: In an effort to unpack the black box of how coaches spent their time, researchers and program leaders co-developed, field tested, and established the content validity of a daily digital coaching log through collaborative, iterative design process. The log tracked coaches' time use class period and included the teachers worked with, reading content, and coaching moves. Coaches completed the log on a daily basis from November-June and researchers shared monthly reports and other analyses. Periodically, we shared initiative-wide data with coaches; at these times leadership shared comments and directions about coach time use. Finally, we offered a workshop at the end of the year for coaches.

Whenever sharing data, the research team went through a systematic process of capturing reactions and questions, and reflecting on these data after meetings. Through our close collaboration and frequent observations of coach professional learning sessions, we observed how these reactions developed into recommendations and policy clarifications to coaches.

Findings:

Informing high level stakeholders: **Sharing reports and other analyses with principal supervisors and central office policymakers gave these leaders a more detailed understanding of coaching activity; this understanding informed and undergirded leaders' messaging to other stakeholders.** For example, when a district leader spoke with a group of supervisors, she clarified that initiative's model was for coaches to spend their time with teachers improving their instruction, and raised a concern that some school administrators were not allowing coaches full access to teachers. Citing log data as well as anecdotal evidence, she asked the principal supervisors to intervene with principals where this occurred. Similarly, while debriefing a monthly log report with a supervisor, researchers raised a concern about a number of schools where coaches reported working with teachers primarily around comprehension, to the exclusion of other reading competencies. The principal supervisor agreed that a broad spectrum of reading competencies were important for teachers and indicated she would

speak to principals about this matter. Finally, log reports and analysis were available to satisfy key data requests on an ad-hoc basis. For example, when a district leader was asked by a local education advocate about rumors that coaches were not working with teachers of self-contained special education classes, researchers were able to provide data that showed this was not a widespread issue. In this case, evidence influenced practice primarily by avoiding a potential distraction.

Empirical grounding for coach guidance: **The data provided an empirical grounding for leadership instructions to coaches about their work.** Leadership's empirically grounded instructions allowed coaches to receive general instructions (e.g. *time with teachers is important*) as well as instructions specific to themselves (e.g. *since my log report tells me I spend substantially less of my time with teachers than the initiative-wide average, I particularly need to spend more time with teachers*). Log data also informed program leadership by helping identify coaches who were not spending time as the program encouraged. For example, the average percent of time coaches spent with teachers was 34%, but at the coach level this varied from 12% to 55%. In cases where this number was low, leadership reached out to coaches to understand why this was the case, and to ameliorate it if possible--in the event of a recalcitrant principal, for example.

Reflection and goal setting: **Reviewing log data provided an opportunity for shared reflection.** At an end-of-year workshop, researchers presented initiative-wide trends and asked coaches to: review their individual reports; reflect on trends they observed; discuss patterns as a group; and set goals for the following year. This process lead to a deeper understanding of the ways that initiative goals played in particular school contexts. For example, discussion of log data patterns lead to a more nuanced understanding of key trade-offs, e.g.:

- Should coaches work with stronger teachers in order to develop model classrooms, or ones who need more instructional support?
- Should motivated teachers or weaker teachers received more support?
- How does the size and stability of the staff impact the decision?

These reflections strengthened coaches' understanding of the work in ways that we hope will translate into time-use decisions better aligned to program priorities in the 2017-18 school year.

Conclusions:

The evidence shows that coaching logs can be a helpful tool for multiple stakeholders in coaching initiatives. Features that made the tool particularly helpful were its scope and geographical dispersion; integration of log data with other data sources (e.g., human resources data, teacher evaluation, student scheduling and demographic data); and reporting for multiple audiences. Log data are limited in some important ways: they do not assess the quality of the coaching; time use is not captured in a precise manner; it is based on self-reports. Once log data are integrated with student reading performance data, they will help provide insights into program impacts.

References:

Bean, Rita, and William Isler. "The School Board Wants to Know: Why Literacy Coaching?." *Literacy Coaching Clearinghouse* (2008).

Elish-Piper, Laurie, and Susan K. L'Allier. "Examining the relationship between literacy coaching and student reading gains in grades K–3." *The Elementary School Journal* 112.1 (2011): 83-106.

Joyce, Bruce, and Beverly Showers. "Improving inservice training: The messages of research." *Educational leadership* 37.5 (1980): 379-385.

Joyce, Bruce R., and Beverly Showers. "Transfer of training: the contribution of" coaching"." *Journal of Education* (1981): 163-172.

Putting research to use:

Identifying positive deviants for a Network Improvement Community

By Lisa Merrill

Background:

Youth development organizations seek to improve Social Emotional Learning (SEL) because it has consequences for the students' academic achievement, quality of relationships, and career success (Nagaoka et al., 2015; Zins et al., 2007). These organizations offer a variety of programming and serve students after school hours and/or during the school day. Some of these organizations are more successful than others at strengthening student SEL, but they have are limited opportunities for them to learn from one another and share promising practices which can hinder their ability to improve.

In New York City, however, 48 youth development organizations joined a network to learn faster, share practices and improve their students' outcomes. The network follows a Network Improvement Community model (Dolle et al., 2007), and its members have committed to using the same student self-report survey instrument that measures SEL and to attending network-wide meetings every two months. A small network staff and their research partner work together to provide members with the common data collection tool, support in data administration and data analyses. The network intends to use data and research as well as practitioner expertise to help programs identify, learn, and spread promising practices across the network.

At the beginning of the research-practice partnership, the network's survey data quality was not conducive to identifying program effects or promising practices. They did not track response rates and the instrument needed improvement. After one year of collaborative work, the research partner and network staff educated members about the importance of creating a reliable and valid survey instrument and ensuring high response rates for every program as pre-conditions for comparing results across programs. One year later, in 2016-17, over half of the members were able to collect survey data with high response rates.

This paper describes how researchers identified sites that are associated with large positive changes in students' SEL and then helped the network investigated promising practices at those sites.

Purpose:

The purpose of this research is to illustrate how a positive deviate analysis can be used in a Network Improvement Community context. To conduct the analysis researchers created clusters of similar sites with high quality data and then identified positive deviants within those groups. The network then built on this knowledge using practitioner expertise to identify specific promising practices and plan to test their promising practices in the field next year.

Setting:

The setting are youth development organizations that are members of the network and serve NYC public school students enrolled in grades 6-12 and or are school aged and involved in alternative, adult education.

Sample:

The sample includes students who took the pre survey in the fall of 2016 and the post survey in the spring of 2017 who enrolled in a member program at a site with at least a 66% response rate. The data include 2,303 students in 48 sites. The students are 48 percent female, 75 percent Black or Hispanic, 27 of students were in middle school, 68 percent in high school, 3 percent in college, and 2 percent in alternative adult education programs.

Intervention / Program:

The sites have very different program models. Some of the sites provide computer science education during after school for a semester, some provide academic preparation for competitive high schools over multiple years, and others provide school-based mentoring for an entire school year.

Research Design: Cluster analysis and random effects multi-level model with shrinkage estimates. The outcome of interest are post SEL score constructs controlling for student pre-scores.

Data Collection:

The analytic data set includes student background characteristics (gender, race, and school level), program and site identifiers, and pre and post responses to the survey (2016-2017 school year) including measures of Growth Mindset, Academic Self-Efficacy, Problem Solving, Sense of Belonging, Interpersonal Skills, Self-Advocacy, and Academic Behaviors. Each of these constructs have an Alpha above .70 and are positively correlated with one another and concurrent academic outcomes.

Data analysis and findings:

Researchers applied PROC FASTCLLUS in SAS to the site level data to create groups of similar sites because it performs well with data containing poorly separated clusters (SAS Institute, 2008). We selected the number of clusters based on the R-square value, Cubic Clustering criterion (CCC) and the Pseudo F. The four cluster divided the sites into two middle school clusters—one with sites with high baseline scores and one with lower baseline scores; and two high school, college, and adult education clusters—one with sites with high baseline scores and one with lower baseline scores.

Researchers then built a multi-level model predicting students' post-SEL score controlling for students' pre-scores and background characteristics, with separate models estimated for each cluster. The model employed PROC MIXED with a random statement that nested students in sites and estimated shrinkage estimates for each site. Positive deviants were defined as sites with positive and statistically significant shrinkage estimates. Ten positive deviant sites were identified across the four clusters.

The researchers then discussed how these findings could be used to begin identifying promising practices at the bright spot sites. Researchers provided a rubric and guidance on a structured interview with site personnel to identify possible promising practices. The possible promising practices were presented to all of the network members and will hopefully be tested in the upcoming school year.

Conclusions:

The research pursued in this study is a testament to a long-term research-practice partnership. We could not identify positive deviant sites without first educating and supporting members in college high-quality data. Using results of the cluster analysis within weeks of their completion to interview practitioners was facilitated by strong relationships between practitioners and researchers.

The next step in the partnership is for other sites to implement these potential promising practices and to test their efficacy across multiple contexts. The research-practice partnership will collaborate to teach members how to develop and measure outcomes when implementing these new promising practices.

Works Cited

Nagaoka, J., Farrington, C. A., Ehrlich, S. B., & Heath, R. D. (2015). Foundations for Young Adult Success: A Developmental Framework. Concept Paper for Research and Practice. *University of Chicago Consortium on Chicago School Research*, Chicago

Dolle, J. R., Gomez, L. M., Russell, J. L., & Bryk, A. S. (2013). More than a network: Building professional communities for educational improvement. *National Society for the Study of Education Yearbook*, *112*(2), 443-463.

SAS Institute Inc. 2008. SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc

Zins, J. E., Bloodworth, M. R., Weissberg, R. P., & Walberg, H. J. (2007). The scientific base linking social and emotional learning to school success. *Journal of Educational and Psychological Consultation*, *17*(2-3), 191-210.

Informing Program Implementation through Research:

Lessons from an Evaluation of a College Readiness Program

By Adriana Villavicencio

Context

In New York City, high school graduation rates continue to rise for males of color (by more than 10 percentage in the last decade), but very few of these students are leaving high school with the requisite skills to succeed in college. Among students scheduled to graduate in 2010, only 9 and 11 percent of Black and Latino males respectively graduated college ready[1] (Villavicencio, Battacharya, & Guidry, 2013). In response, the NYC Department of Education (DOE) launched the Expanded Success Initiative (ESI), designed to increase college and career readiness among Black and Latino male students.

Research Questions

Our evaluation of ESI was designed to answer two broad sets of research questions: 1) What services and programs did ESI schools provide to their staff and students as a result of this initiative? What challenges did schools face in implementing these programs, and how did they attempt to address those challenges? 2) Did ESI impact students' academic and/or non-academic outcomes?

Setting & Population

ESI was implemented in 40 NYC public high schools with 1) student enrollment that included at least 35 percent Black and Latino males, with at least 60 percent of students qualifying for free and/or reduced priced lunch, 2) a four-year graduation rate above 65 percent, and 3) an "A" or "B" on the latest high school Progress Report, grades assigned to schools by the district to provide a snapshot of the school's performance during the previous year. Our study also utilized a set of comparison schools. Using a multistep statistical process, we identified 80 schools which were most similar to ESI schools with respect to both student demographics and student achievement.

Intervention

Each high school participating in ESI received $250,000 over three years to create or expand existing supports for Black and Latino male students in three areas: academics, youth development, and college/career culture. In addition, a team at the DOE provided ongoing professional development, enrichment opportunities for their students, and regular meetings with staff designated as ESI point people in their schools.

Research Design

Beginning in the summer of 2012 and extending though the spring of 2016, our evaluation examined ESI over four school years, following one cohort of 9th graders through their scheduled graduation (and three additional cohorts through the 2015-2016 school year). Our ESI evaluation consists of 1) an implementation study, which examined the services and supports that were planned and implemented under ESI and 2) an impact study designed to determine whether students who were exposed to ESI-related interventions and supports achieved better outcomes than they would have if their school had not been involved in ESI.

Data Collection and Analysis

[1] "College readiness" refers to NYS Education Department's Aspirational Performance Measure (i.e., receiving a score of 80 or higher on a Mathematics Regents examination and a score of 75 or higher on an English Regents examination).

The implementation study relied largely on field visits to each ESI school, which included 60-minute semi-structured interviews and focus groups with school leaders, ESI point people, and 3-5 teachers in each school. Each year, we also administered a highly structured questionnaire about the details of ESI programming to assess the fidelity and intensity of ESI in each school. For the impact study, we received administrative records (e.g., grades, graduation rates) from the NYC DOE. We assessed non-academic outcomes (e.g., academic self-perception, postsecondary goals) via a student survey designed and conducted by the Research Alliance. For each survey construct, we measured the impact of ESI on students' responses using regression analyses, with participation in ESI being the main explanatory variable. To estimate ESI's effects on non-survey data, we used a Comparative Interrupted Time Series to estimate the impacts of ESI by comparing student performance in ESI schools with the performance of a) students in the same schools prior to the implementation of ESI and with b) students in comparison schools during the same time period.

Findings: Informing Practice though Research

Many of our research findings prompted program improvement. For example, our first-year interviews revealed that educators in ESI schools desired more opportunities to learn how other schools were implementing ESI. As a response, the program team created monthly meetings for ESI point people in each school to share best practices. In addition, our internal memos to the program team on the fidelity and intensity of ESI at each school enabled the team to provide specific feedback to individual schools about how to improve implementation. It also helped the team decide which professional development opportunities they should provide to all ESI schools. After our second year, our preliminary findings saw little gains on academic outcomes. As a result, the program team provided targeted professional development in increasing academic rigor across the core subject matters and shared with schools more fine-grained data reports on the academic progress of their students disaggregated by race and gender.

Findings from the last year of the evaluation—which show impacts on students' socioemotional outcomes, but none on other key outcomes—are now being communicated to the broader community of researchers and districts engaged in efforts to improve outcomes for Black and Latino males, specifically through a multi-district convening held in the fall of 2017 and an online platform for research-practice partnerships to share strategies, instruments, and research findings. Both have provided us with the opportunity to share lessons about this work, highlight the particular strengths of ESI, and raise critical questions that could help inform other initiatives around the country.

Conclusions

Our evaluation of ESI revealed important lessons about how research can support school-based interventions. In addition to highlighting how our findings informed practice, this paper will also identify which factors supported evidence use, including: 1) an iterative process of incorporating feedback loops between the research team and the program partners, 2) multiple forms of communication with program partners including regular briefings and internal memos, and 3) a strong relationship between the research group and the program team committed to both the rigor of the research and the goals of the program.