

2010 SREE Conference Abstract Template

Abstract Title Page

Not included in page count.

Title: Treatment effect heterogeneity in a science professional development initiative: The case for school capacity

Author(s): Sarah Bruch, Jeffrey Grigg, Paul Hanselman

Abstract Body

Limit 5 pages single spaced.

Background/context:

Description of prior research, its intellectual context and its policy context.

Professional development policy initiatives are touted as important engines of instructional improvement and a growing body literature identifies traits of successful professional development (Birman et al., 2000; Garet et al., 2001; Gamoran et al., 2003; Penuel et al., 2007). However, the effectiveness of professional development initiatives, like all educational reforms, depends not only on the features of the intervention itself, but also on the context in which it is implemented. If schools do not possess adequate capacity to implement, then even a perfectly designed intervention will fail. Because schools differ in their capacities to implement reforms (Spillane & Thompson, 1997; Ball & Cohen, 1999), it is important to know whether and how school capacity matters for the success of interventions. An additional serious complication is that capacity is likely an integral part of selection into educational intervention. Effectiveness research must tackle school capacity and it must move beyond merely observational data. In this paper we exploit a randomized trial to investigate treatment effects on student achievement conditional on school capacity.

According to Gamoran et. al (2003), school capacity to promote high achievement lies not in structural arrangements but in the resources schools bring to bear on classroom teaching and learning. Resources are not limited to material conditions such as time, materials, and other elements that can be purchased with money; but also include human resources such as teachers' knowledge, skills, and commitments; and social resources, such as relationships of trust and shared expectations among educators upon which teachers can draw to support their endeavors. In this conception, capacities are multi-dimensional, dynamic, and potentially reciprocally related to reform efforts, particularly for teacher learning (Gamoran et al., 2003). In other words, school capacity may be necessary for the successful implementation of scaling up professional development reform, but it may also develop (or be hindered) as the result of specific reform initiatives.

The specific context of this paper is an experimental evaluation of a teacher professional development initiative in the Los Angeles Unified School District (discussed in more detail below). Two prior findings are relevant. First, although teacher participation was lower than intended, overall the treatment induced noticeable changes in teachers' instructional behavior, and the average treatment effect in the first year was detectable and negative (Borman, Gamoran, & Bowdon, 2008). Second, previous research has identified meaningful dimensions of school capacity as well as measurable variation in these dimensions across schools within the study (Bruch, Grigg, & Hanselman, 2009). These two findings suggest a unique opportunity to explore heterogeneity of established treatment effects conditional on meaningful differences in school capacity.

Investigating heterogeneity not only addresses a theoretically important dimension but also counters the "black box" critique that intervention studies fail to illuminate the processes that take place within schools to bring about the desired change in student outcomes (Cronbach et al. 1982; Howe 2004). Comparing mean differences between randomly assigned groups on a single

outcome offers an unbiased and consistent inference of the causal effect of a program, but it sheds little light on the complicated social processes in which these field trials take place and thus offer little prospect for improving the implementation of educational interventions (Stein, et al., 2008). One obvious response to this concern is to collect data on intervening conditions between intervention and outcome. However, the discussion above suggests another approach: comparing treatment effects for subpopulations of schools based on theoretically important pre-conditions. Doing so recognizes that average treatment effects may obscure underlying interactions between the intervention at hand and the necessarily complex pre-existing conditions.

Purpose / objective / research question / focus of study:

Description of what the research focused on and why.

This study focuses on how the treatment effects of a teacher professional development initiative in science differed by school capacity. In other words, we are primarily concerned with treatment effect heterogeneity. As such, this paper complements ongoing evaluation of the average treatment effects of the initiative over time.

The research question considered here is: Did existing school capacity account for heterogeneity in teacher and student outcomes? That is, do treatment effects differ for schools with low, average, or high capacity? Specifically, we consider two outcomes: teachers' reported adoption of the targeted curriculum and students' subsequent achievement scores on standardized science tests. Although our primary focus is on student outcomes, teacher behaviors are informative because they represent a necessary mechanism in the causal process that is likely influenced by school capacity.

Setting:

Description of where the research took place.

The research took place in the Los Angeles Unified School District (LAUSD), one of the largest and most diverse school districts in the United States. In addition to the city of Los Angeles, LAUSD serves other municipalities in Los Angeles County, covering an area of 710 square miles. During the 2006-2007 academic year when the randomized field trial was first deployed, over 700,000 students were enrolled in LAUSD, and the district employed over 30,000 certified teachers. The U.S. Census estimates that 49.8 million students were enrolled in grades one through twelve in 2006 (Davis & Bauman, 2008); consequently, approximately 1.5% of the United States student population was enrolled in LAUSD. Findings from LAUSD can be reasonably applied to other large urban school districts in the United States.

Population / Participants / Subjects:

Description of participants in the study: who (or what) how many, key features (or characteristics).

Eighty elementary schools were selected for the study in a block randomized design drawing from each of LAUSD's eight local districts (see "Research Design," below). The intervention and the study targeted students and teachers in the fourth and fifth grades, amounting to over 500 teachers and approximately 9,000 students. The students attending these schools are racially

diverse and comparable to LAUSD's overall student population; 70% are Hispanic, 9% are African-American, 14% are non-Hispanic White, and the remaining 6% represent other ethnicities. Nearly 80% of the students qualify for free or reduced price lunch, 40% are English language learners, and 11% receive special education services. Half of the fourth grade students proficient or above in English language arts in 4th grade, 40% were proficient in 4th grade mathematics, and 25% were proficient in science in 5th grade. Teacher survey respondents also provided information about their demographic characteristics: 26% were male, 42% were non-Hispanic White, 34% held an advanced degree, 33% had been teaching for three or fewer years, and they averaged 8 years of teaching experience in 2006. There were no statistically significant differences between control and treatment schools in terms of aggregate student characteristics, mean proficiency levels, or reported teacher characteristics, as shown in Table 1.

(Please Insert Table 1 Here)

Intervention / Program / Practice:

Description of the intervention, program or practice, including details of administration and duration.

The intervention being evaluated is a five-day teacher professional development program on “immersion units” for teaching in elementary science. These units cover segments of the California science standards for fourth and fifth grade and address features of classroom inquiry as defined by the National Science Education Standards (National Research Council, 1996; Olson & Loucks-Horsley, 2000). The immersion curricula were previously available throughout LAUSD, but simply providing materials may not be sufficient to change teaching practice (Gamoran et al., 2003); this study therefore was specifically designed to evaluate the impact of the teacher training initiative.

Research Design:

Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

The study is a randomized field trial employing a block design. Prior to the deployment of the professional development program, each of the eight LAUSD local district superintendents was asked to nominate 20 or more schools which he or she deemed capable of participating in the trial. The superintendents nominated 191 schools from which 80 schools were selected to participate using a stratified random sample to draw ten schools from each local district. These 80 schools are not statistically distinguishable from the 191 schools from which they were drawn. Five schools from each local district were then randomly selected for the intervention, leaving a total sample of 40 schools assigned the intervention condition and 40 schools as comparisons. The intervention was first introduced to fourth grade teachers in 2006-2007 and to fifth grade teachers in 2007-2008, so the 2008-2009 academic year analyses will represent the third year of implementation for fourth grade teachers and the second year of implementation for fifth grade teachers.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.

This study employs data of two primary types: teacher surveys and student test scores. In Fall 2006—at the commencement of the study—we asked the fourth and fifth grade teachers in all 80 schools to complete an extensive survey on their attitudes and practices. A total of 90% of eligible teachers completed the survey; 77 schools were represented. We used the survey responses to identify five dimensions of school capacity: Principal Leadership, Professional Development Climate, Administrative Support for Teaching, Teachers Leading Their Own Learning, and Collaborative Learning in Science. Schools above the median on four or five of these measures were classified as “high capacity” schools, schools above the median on zero or only one of these measures were classified as “low capacity” schools, and the remaining schools comprised the middle group of schools. In addition, we draw self-reported teacher data on immersion curriculum adoption from a follow-up survey administered in the third year of the study.

Student achievement is measured in two ways in LAUSD. First, the district uses a “periodic assessment” as a formative measure of performance. This assessment, which is aligned with the district instructional guide, consists of modules covering life science, earth science, and physical science and is administered in grades 4-8 (Scruggs, 2004). Second, the California Standards Test (CST) includes science in the fifth grade. This assessment is norm-referenced and aligned with state standards (California Department of Education, 2004). To examine effects for fourth grade students we will rely on the periodic assessments; for the fifth grade students we can use both the periodic assessments and the CST.

Our analytic strategy relies on the exogenous assignment of treatment to estimate causal impacts within each of the three school capacity groups as defined by baseline measures. The degree of correspondence in treatment effects across high, medium, and low capacity schools indicates the extent of capacity-related heterogeneity. Ultimately, all analyses will account for clustering of teachers (for the implementation outcomes) and students (for the achievement outcomes) in schools with multilevel models. However, individual student achievement data is not currently available. Therefore, the results from preliminary models reported in this abstract rely on aggregate school-level analyses.* Once individual-level student data become available in November 2009, we will employ multilevel models of students or teachers nested within schools to estimate the potential treatment effects.

Findings / Results:

Description of main findings with specific details.

Findings suggest that capacity conditions the success of the intervention. For teacher adoption, capacity trumped the intervention. The targeted professional development only boosted reported curriculum use in low capacity schools, because adoption was universally high among teachers in high capacity schools. For student achievement, capacity moderated the treatment effects. Preliminary results are presented Figures 1-3, which graph predicted average achievement by treatment condition and school capacity level. (Please Insert Figures 1-3 Here) In grade 4, treatment effects in year 1 (2006-2007) were negative, particularly in medium capacity schools. By year 3 (2008-2009), treatment effects were improved in all schools, leading to a positive

* The main methodological implication is that the preliminary results are under-powered. They should be thought of as suggestive, and for this reason we do not conduct tests of statistical significance here.

effects for high and low capacity schools, and much less negative for medium capacity schools. In grade 5, year 2 effects (2008-2009) are modestly negative in the high and medium capacity schools and negligible in low capacity schools. Two patterns in these figures are notable with respect to capacity heterogeneity. First, in fourth grade where two time points are available, the impacts of the treatment become more positive over time across all types of schools, suggesting that adaptation to the treatment occurred across the board. Second, treatment effects differ across groups, with low capacity schools (and high capacity in grade 4) responding more favorably to the treatment than medium capacity schools.

Together, the implementation and achievement results suggest that the treatment was experienced differently throughout the population of schools. For high capacity schools, treatment did not induce greater curriculum adoption, but achievement results suggest that the treatment led to improving instructional practice in line with the ambitious implementation model of incorporating meaningful immersion practices into the science classroom. For low capacity schools, treatment induced greater use of the curriculum, but it is not clear from the achievement results whether these teacher behaviors translated into more effective teaching. Finally, medium capacity schools were most negatively impacted by the intervention, consistent with the overall findings that this intervention was not beneficial on average (Borman, Gamoran, Bowdon, 2008).

Conclusions:

Description of conclusions and recommendations based on findings and overall study.

School capacity stands as a gatekeeper between professional development implementation and ultimate changes in instructional practice. Therefore, variations in the resources that schools bring to bear in enacting training are a key for understanding not only the success or failure of this intervention, but also the complex relationships between professional development and capacity for producing instructional change. This paper suggests significant variation in the effects of a professional development initiative at scale, both for teachers and students. In general, higher capacity is required both for program take-up and to support the successful translation of offered training into effective classroom instruction.

There are clear policy implications of the demonstrated importance and variability of school capacity. These results implore more attention to be paid to the school pre-conditions underlying educational interventions, particularly given that the average school in our study did not have the capacity to successfully respond to this intensive professional development initiative. There are two clear implications for educational evaluation. One is to direct attention to rigorous causal evaluation of school capacity building, especially given that capacity can trump the interventions more commonly subjected to experimental testing. Indeed, the current study's design does not allow us to make any casual claims about what works in that arena. The other implication is that more effectiveness evaluations should explicitly consider school capacity as an important mediating dimension. We have demonstrated here that school capacity as conceptualized in the school organizations tradition is a meaningful tool for opening up the "black box" of a randomized professional development evaluation and our methodology would be relatively easy to replicate.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

- Ball, D. L., & Cohen, D. K. (1999). Instruction, capacity, and improvement. *CPRE Research Report Series RR-43*. Ann Arbor, MI: Consortium for Policy Research in Education.
- Birman, B. F., Desimone, L., Porter, A. C., Garet, M. S. (2000). Designing professional development that works. *Educational Leadership*, 57 (8), 28-33.
- Borman, G. D., Gamoran, A., & Bowdon, J. (2008). A Randomized Trial of Teacher Development in Elementary Science: First-Year Achievement Effects. *Journal of Research on Educational Effectiveness*, 1(4), 237 - 264.
- Bruch, S., Grigg, J., and Hanselman, P. (April, 2009). "Up to the Classroom Door: Effects of a School-Randomized Trial." Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- California State Department of Education. (2004). *Key elements of testing*. Sacramento: Author.
- Cronbach, L. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- Davis, J. W., & Bauman, K. J. (2008). *School Enrollment in the United States: 2006*. (P20-559). Washington, DC: Retrieved from <http://www.census.gov/prod/2008pubs/p20-559.pdf>.
- Gamoran, A., Anderson, C. W., Quiroz, P. A., Secada, W. G., Williams, T., & Ashmann, S. (2003). *Transforming teaching in math and science: How schools and districts can support change*. New York: Teachers College Press.
- Garet, M., Porter, A., Desimone, L, Birman, Be., & Yoon, K. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915-945.
- Howe, K. (2004). A Critique of Experimentalism. *Qualitative Inquiry*, 10, 42-61.
- National Research Council. (1996). *National science education standards*. Washington, DC: National Academies Press.
- Olson, S., & Loucks-Horsley, S. (Eds.). (2000). *Inquiry and the National Science Education Standards: A guide for teaching and learning*. Washington, DC: National Academies Press.

Penuel, W. R., Fishman, B., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, 44(4), 921–958.

Scruggs, L. (2004). *Science assessments for grades 4–8* (Los Angeles Unified School District Reference Guide). Los Angeles: Los Angeles Unified School District.

Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis*, 19, 185-203.

Stein, M.L., Berends, M., Fuchs, D., McMaster, K., Sáenz, L., Yen, L., Fuchs, L.S., and Compton, D.L. (2008). Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. *Education Evaluation and Policy Analysis* 30(4): 368-388.

Appendix B. Tables and Figures

Not included in page count.

Table 1: Descriptive Statistics

Teacher Characteristics	Treatment (n = 40)	Comparison (n = 37)	t	Pr > t
Years of Teaching Experience	8.36	7.92	-0.53	0.596
Inexperienced (Fewer Than 3 Years)	0.32	0.36	0.75	0.458
Hold Advanced Degree	0.33	0.35	0.41	0.686
Male	0.26	0.25	-0.13	0.897
Nonwhite	0.58	0.60	0.23	0.818
Teach 5th grade	0.53	0.54	0.56	0.580
Science Lead Teacher	0.29	0.30	0.40	0.688

School and Student Characteristics (2006)	Treatment (n = 40)	Comparison (n = 40)	t	Pr > t
Failed to make AYP	0.40	0.35	-0.46	0.649
Proficient or Above in 4th Grade Math	0.52	0.50	-0.57	0.571
Proficient or Above in 4th Grade ELA	0.42	0.43	-0.31	0.755
Proficient of Above in 5th Grade Science	0.24	0.25	0.17	0.868
Free/Reduced Lunch Eligible	0.77	0.78	0.21	0.834
Proportion Hispanic	0.70	0.68	-0.30	0.768
Proportion African-American	0.09	0.09	-0.06	0.949
Proportion White	0.14	0.13	-0.32	0.750
Limited English Proficient	0.36	0.44	1.47	0.146
Receiving Special Education Services	0.11	0.10	-0.65	0.519

Note: Teacher characteristics drawn from System-Wide Change survey. School and students composition statistics from publicly available administrative data.

Figure 1

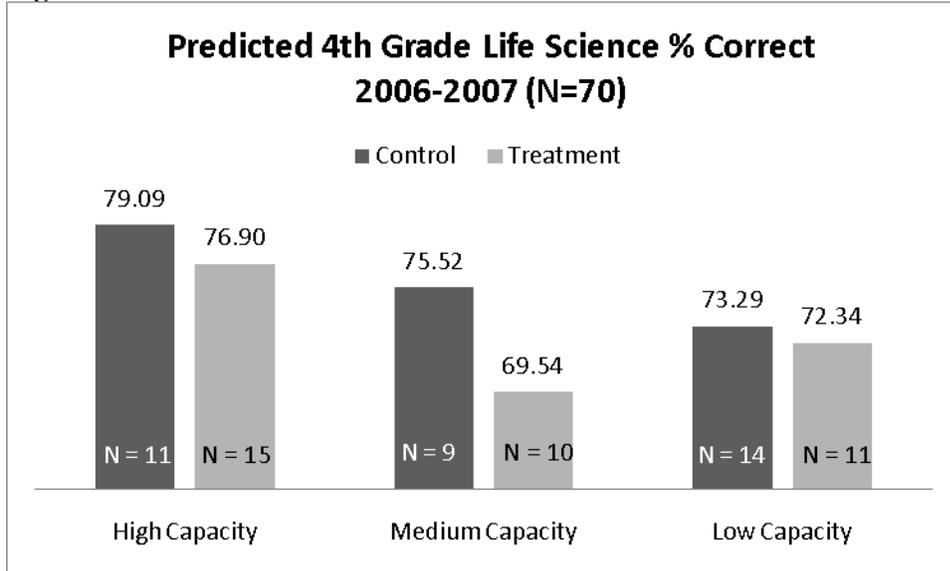


Figure 2

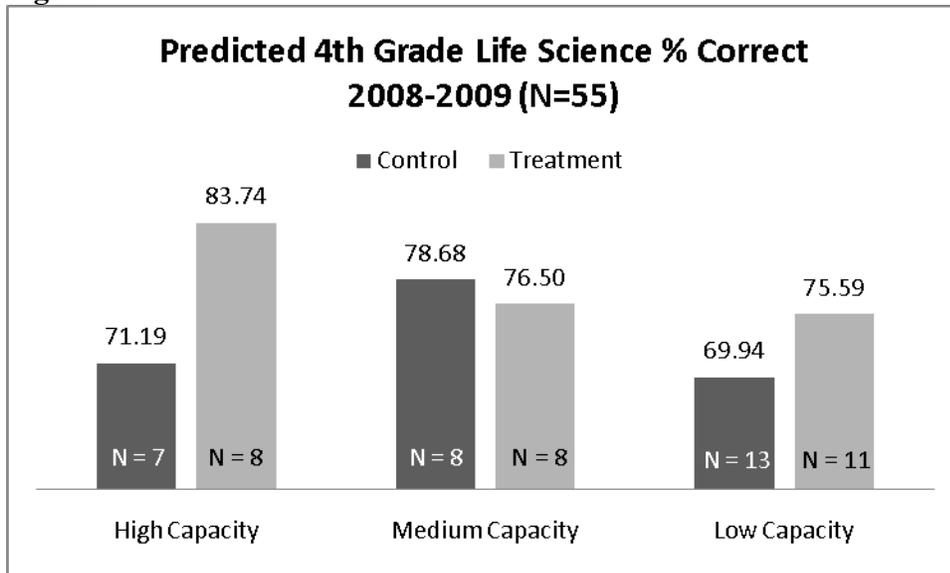


Figure 3

