From Knowledge to Results: What is Missing in Educator Professional Development Design?

The purpose of this symposium is to examine the effects of three educator training programs—two for teachers and one for principals—that have been subjected to rigorous empirical investigation via randomized control trials. The range of outcomes studied includes leader and teacher perceptions, teacher practices, educator turnover and—in all three—student achievement. While not a comprehensive meta-analysis, these three studies represent well-executed randomized control trials that document remarkably similar inabilities of contrasting professional development designs to effect anticipated outcomes. The results are consistent: the three different professional development programs did not impact teachers or students in expected ways.

In the two studies examining teacher professional development, teachers reported knowing more as a result of the professional development in which they engaged. In one of the studies, teachers also reported that their practices had changed. Teacher practices were examined in both studies, as was student achievement. Despite support that teachers’ knowledge increased, actual classroom practices were not affected in significant ways, nor did student achievement increase. One study examined possible explanations for the null findings and reported that teacher will and resource availability appeared to be weak explanations for the outcomes. Organizational barriers, difficulty of ambitious practice, and weak instructional press were strong factors that may have contributed to static teacher practices and a resultant lack of student achievement change.

The study examining principal professional development found increases in treatment school leaders’ reported sense of efficacy, change leadership, and knowledge of curriculum, instruction and assessment, but no change in direct involvement in curriculum, instruction and assessment. Simultaneously, teachers in treatment schools reported no changes in their principals’ practices nor did they report changing their own practices. Similar to the other two studies, student achievement was not affected.

In sum, increasing educators’ knowledge base does not necessarily translate into more distal outcomes such as changed practices or improved student outcomes. Change can be challenging and complicated and may require myriad supports that are not always part of intervention designs, including more external facilitation and support, a greater press for knowledge implementation, more on-the-ground sustained practice, and more collaborative work with teachers to support change.

These results are important to consider because each year in the United States nearly 50 million dollars are spent on educator professional development. Since 2002, the federal government has likewise invested in experimental analyses of promising, well-designed interventions, most of which have failed to achieve hypothesized outcomes. While many studies simply catalog the results, we have organized this symposium to think deeply about the underlying design issues and mechanisms that may explain these failures in order to inform the design of future interventions and experiments studying their causal impacts.
Knowledge is not enough: An inductive analysis of experimental evidence examining the relation between educator professional development and impacts on participant behavior

Roger D. Goddard
Yvonne L. Goddard
The Ohio State University

Background/Context

Knowledge alone has no inertia in the physical world beyond the realm it occupies in our synapses. Without talented people who act on it day by day, knowledge remains contemplative at best. In other words, knowledge alone is insufficient. In the absence of action, the best of ideas possess only potential energy. Without change in practice, change in the human condition—and thus knowledge itself—is unlikely.

Change in practice in response to new knowledge, however, is a formidable undertaking. In a study of school districts serving largely poor students, The New Teacher Project estimated that 6-9% of annual education operating budgets were spent on professional development (PD) (TNTP, 2015). If this is true of public schools writ large, then the U.S. investment in educator PD may approach $50 million dollars annually. Unfortunately, the TNTP report indicates the record of improvement based on professional learning initiatives known largely as “professional development” is questionable at best. Thus, we need to know more about the circumstances under which PD does and does not produce the outcomes such investments expect.

Purpose

The purpose of our work was to examine the probable link between the professional development design provided to school principals and the outcomes of that PD recently observed through a randomized control trial funded by the US Institute of Education Sciences1. Our goal was to reach conclusions about why the PD as designed and implemented impacted certain outcomes but not others.

Setting, Participants and Intervention Design

In the study on which we report, half of a group of principals from largely poor, rural northern Michigan elementary schools were randomly assigned to participate in the Balanced Leadership Program® - a research-based professional development program created by McREL International. The treatment was intensive with a total of 20 training days delivered to principals by highly-qualified instructors over two years. As reported elsewhere, the program’s content validity and fidelity of implementation were strong (Jacob et al., 2015; Miller et al., in press; Schroeder et al., 2012). Distinguishing features of the training included principal cohort groups led through a well-developed, research-based curriculum focused on principal responsibilities that predict student achievement (Marzano, Waters, & McNulty, 2005). By design, teachers were not directly involved in the training. Additionally, treatment group principals were not held

1 Award number R305A080696
accountable for enacting their learning in practice beyond conversations about their experiences during the regularly-scheduled training meetings. Although the opportunity to communicate with trainers between sessions was offered, anecdotal evidence suggests that such interactions were infrequent. The PD was mainly classroom training for principals and did not include onsite implementation assistance. Finally, the modal district had only one participating school and there was no formal design for school district implementation support. Principals were thus largely left on their own to manage their implementation.

**Findings and Discussion**

We analyzed impacts of the professional development described on school climate, principal leadership, teacher practice, educator turnover, and student achievement based on data gathered from teachers, principals and the state department of education. A critical finding is that treatment principals reported higher on average than control principals across measures of school leadership, climate and teacher practice; however, the teachers in treatment schools did not report higher than teachers in control schools on parallel outcomes (see Table 1). Therefore, it is not surprising that, per Table 2, multilevel models demonstrated no statistically significant differences in student achievement between treatment and control schools. Finally, Table 3 shows that treatment school principals were 16% less likely than control principals to leave their position over the course of the study and teachers in the treatment schools demonstrated smaller but significant attenuation in turnover probability (5%) compared to their control school counterparts.

The findings above established a perplexing question regarding how a principal training program could cause principals in treatment schools to report growth in leadership, climate and teacher practice that teachers did not confirm while simultaneously reducing turnover for both principals and teachers. To identify possible explanations, we disaggregated treatment effects based on the principal self-administered questionnaire (SAQ) data into principal leadership component outcome domains and report these in Table 4. Here a striking pattern emerges regarding the instructional leadership areas in which treatment principals’ did and did not demonstrate growth. The largest effects were those for principals’ efficacy for instructional leadership (.62 SD) and principal reports of their knowledge of curriculum, instruction and assessment (.32 SD). Conversely, the smallest of the treatment effects was for principals’ reports of direct involvement in curriculum, instruction and assessment with teachers (.05 SD).

Thus, the PD appears to have caused treatment principals to feel more confident in and knowledgeable about instructional leadership but by their own self-reports, not more directly involved with teachers in curriculum, instruction and assessment. Moreover, in the absence of knowledge enactment, it is probable that teachers in treatment schools would not have experiences leading them to report differently on measures of leadership, climate or teacher practice, which is precisely what we observed. A series of principal interviews in five treatment schools over the three-year study identified a lack of both teacher and central office support as implementation barriers. From this perspective, the consistency in turnover findings also makes sense—more knowledgeable and efficacious principals are less likely to quit; and, a reduction in treatment principal turnover alone may aide in teacher retention.
In sum, it seems likely that the desirable principal efficacy, knowledge and educator turnover impacts we observed were real outcomes of the training; however, changes in teacher perceptions and student learning were not achieved. Lessons for future intervention designers include that even PD focused on leadership might benefit from direct teacher involvement as their practice is a main cause of student learning. Even so, making changes based on information alone is an incomplete paradigm. Other potentially beneficial design elements for converting knowledge to practice include onsite, between session implementation support; accountability for implementing practical change based on knowledge acquired; establishing communities of practice and district implementation support that may include additional resources.
References


Table 1. Impact of PD on Principal and Teacher Reports of Leadership, Climate and Practice

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>β</th>
<th>SE</th>
<th>p-value</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Principal Report (N=88)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal leadership</td>
<td>0.18</td>
<td>0.11</td>
<td>0.10</td>
<td>0.33</td>
</tr>
<tr>
<td>Collaboration</td>
<td>0.26</td>
<td>0.14</td>
<td>0.06*</td>
<td>0.40</td>
</tr>
<tr>
<td>School climate</td>
<td>0.21</td>
<td>0.10</td>
<td>0.03**</td>
<td>0.34</td>
</tr>
<tr>
<td>Collective differentiated instruction</td>
<td>0.44</td>
<td>0.18</td>
<td>0.01**</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>Teacher Report (N=1546 teacher in 91 schools)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal leadership</td>
<td>0.05</td>
<td>0.12</td>
<td>0.67</td>
<td>0.03</td>
</tr>
<tr>
<td>Collaboration</td>
<td>0.07</td>
<td>0.06</td>
<td>0.27</td>
<td>0.06</td>
</tr>
<tr>
<td>School climate</td>
<td>0.02</td>
<td>0.05</td>
<td>0.64</td>
<td>0.02</td>
</tr>
<tr>
<td>Collective differentiated instruction</td>
<td>0.07</td>
<td>0.05</td>
<td>0.17</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Table 2. Impact of PD program on student achievement (n=126)

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>n</th>
<th>ITT β</th>
<th>SE</th>
<th>TOT β</th>
<th>SE</th>
<th>p-value</th>
<th>ITT Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>119</td>
<td>0.78</td>
<td>0.81</td>
<td>1.18</td>
<td>1.18</td>
<td>0.34</td>
<td>0.04</td>
</tr>
<tr>
<td>Grade 4</td>
<td>115</td>
<td>-0.01</td>
<td>1.03</td>
<td>-0.01</td>
<td>1.56</td>
<td>0.99</td>
<td>0.00</td>
</tr>
<tr>
<td>Grade 5</td>
<td>109</td>
<td>1.48</td>
<td>1.40</td>
<td>2.24</td>
<td>2.12</td>
<td>0.29</td>
<td>0.04</td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3</td>
<td>119</td>
<td>-0.45</td>
<td>0.78</td>
<td>-0.68</td>
<td>1.19</td>
<td>0.56</td>
<td>-0.02</td>
</tr>
<tr>
<td>Grade 4</td>
<td>115</td>
<td>-0.27</td>
<td>1.01</td>
<td>-0.40</td>
<td>1.53</td>
<td>0.79</td>
<td>-0.01</td>
</tr>
<tr>
<td>Grade 5</td>
<td>109</td>
<td>0.81</td>
<td>1.25</td>
<td>1.23</td>
<td>1.89</td>
<td>0.52</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Note. Covariates include %minority, %free and reduced lunch, school size and average 3rd grade reading and math MEAP scores from 2008. All schools did not serve all grades in each year of the study, thus there are differences in the sample sizes across grades. TOT estimates adjusted by participation rate; Effect size reflects both within- & between-level variance (i.e. $\frac{\gamma_{01}}{\sqrt{\sigma^2 + 1/\tau_{00}}}$).

Table 3. Impact of PD on Principal and Teacher Turnover

<table>
<thead>
<tr>
<th>Administrative Data</th>
<th>ITT β</th>
<th>SE</th>
<th>TOT β</th>
<th>SE</th>
<th>p-value</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher turnover (N=1764)</td>
<td>-0.05</td>
<td>0.03</td>
<td>-0.07</td>
<td>0.04</td>
<td>0.05*</td>
<td></td>
</tr>
<tr>
<td>Principal turnover (N=122)</td>
<td>-0.16</td>
<td>0.09</td>
<td>-0.23</td>
<td>0.13</td>
<td>0.10*</td>
<td></td>
</tr>
</tbody>
</table>

Note. Covariates include corresponding baseline measures, %minority, %free and reduced lunch, school size, and 3rd grade reading and math MEAP scores. Effect size reflects both within- & between-level variance (i.e. $\frac{\gamma_{01}}{\sqrt{\sigma^2 + 1/\tau_{00}}}$).

**Statistically significant at p<=.05 *Statistically significant at p<=.10 level.**
Table 4. Impact of PD on instructional leadership.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Sample Items</th>
<th>Effect Size</th>
</tr>
</thead>
</table>
| Principal Efficacy                            | • I have the capabilities required now to help my teachers improve their skills in ways that will lead to improved student learning.  
• My leadership skills are currently sufficient to improve instruction in this school in ways that will foster high levels of student learning. | .62         |
| Leading Change                                | • I have become more effective as a school leader because of professional development I received in the last 12 months.  
• I have changed the way I respond to new challenges since last year. | .53         |
| Monitor                                       | • I continually monitor the effectiveness of the instructional practices used in our school.  
• At any time, I can accurately determine how effectively our school enhances student learning. | .41         |
| Involvement in Curriculum, Instruction, & Assessment | • I am directly involved in helping teachers make instructional decisions using assessment data.  
• I am directly involved in helping teachers address instructional issues in their classrooms. | .32         |
| Principal Leadership                          | • Teachers in this school feel comfortable discussing instructional issues with me.  
• I set high standards for teaching. | .27         |
| Knowledge of Curriculum, Instruction, & Assessment | • I am very knowledgeable about effective instructional practices.  
• I am very knowledgeable about effective classroom assessment practices. | .05         |
Impacts of professional development (PD) in classroom assessment on teacher and student outcomes: Implications for PD design

Tedra Clark
McREL International
Bruce Randel
Century Analytics, Inc.

Context
This presentation addresses the implications of teacher-led professional development (PD) program for improving teacher practice and student learning. Specifically, we ask the question: Is a model of PD that relies on self-study and teacher-led learning teams with no external facilitation or monitoring sufficient for improving teaching practice and student outcomes? We ask this question in the context of a randomized experiment examining the impacts of a widely-used teacher PD program in classroom assessment.

Research suggests that high quality classroom assessment used to support ongoing instruction and learning leads to greater student self-efficacy and motivation and, in turn, improved student learning and achievement (Black & Wiliam, 1998; Brookhart, 1997; Butler & Winne, 1995; Kingston & Nash, 2011). Teachers, however, often receive insufficient training in classroom assessment in teacher preparation programs and indicate the need for more professional development (PD) in classroom assessment (DeLuca & Klinger, 2010; O’Sullivan & Chalnik, 1991; Schaffer, 1993).

To address needs for PD in classroom assessment, education agency decision-makers have several options including, but not limited to, externally facilitated workshops (of varying frequency and duration), independent study, real-time coaching, teacher learning teams, or a combination of formats. However, evidence on the impact of different program types is limited.

Objective/Intervention
In this presentation, the authors describe an impact study of a widely used PD program in classroom assessment – Classroom Assessment for Student Learning (CASL) (Stiggins, Arter, Chappuis, & Chappuis, 2004). CASL uses a format of teacher independent study combined with teacher-led learning teams. With the support of a primary textbook, DVDs, ancillary texts and a learning team facilitator handbook, CASL is designed to be self-executing, without external facilitation. The content of CASL is consistent with research on high quality teacher practices for improving student motivation and learning. Specifically, CASL emphasizes matching learning targets with assessment methods, providing descriptive feedback, and activating student involvement in learning. However, mastery of this complex process requires sustained effort and practice (Brookhart, 1997), and it is unclear whether the design of the CASL PD will provide enough support for sound implementation in the classroom.

Research Design/Questions
Conforming to the self-executing nature of the program, researchers examined the impact of CASL as implemented under real-world conditions and without any involvement of, or requirements from, the researchers. Because the goals of PD in classroom assessment include
enhanced teacher knowledge/skills and improved student learning and achievement, we designed this study to estimate the impact of CASL on three teacher outcomes and student achievement. Although CASL is applicable in all content areas, we focused on mathematics in response to regional needs expressed during study preparation.

Researchers randomly assigned 67 Colorado elementary schools to receive CASL materials (treatment schools) or continue with regularly scheduled PD (control schools). The student sample included all grades 4 and 5 mathematics classrooms and their students in each school.

Research questions were: Does CASL PD increase teacher knowledge of classroom assessment, involvement of students in assessment, and/or quality of assessment practice? Does CASL PD positively impact student outcomes (presumably as a result of increases in teacher outcomes)?

Data Collection/Analysis

The outcome measures for teachers included a test of assessment knowledge, teacher self-reports of student involvement in assessment, and teacher assessment practice as measured using assessment work samples scored on six dimensions of assessment quality – focus of goals on learning, alignment of learning goals and task, alignment of goals and assessment criteria, clarity of assessment criteria, feedback type, and feedback eliciting student involvement. The student outcome measure was mathematics scores on the Colorado statewide standardized assessment (CSAP).

All impacts of CASL were estimated using two-level models with students or teachers nested within school, and impacts estimated at the school level. Student achievement and teacher impact models included an individual-level pretest covariate and a school-level pretest covariate for each respective outcome.

Findings

Results showed positive and statistically significant impacts of CASL on two teacher outcomes. On the test of assessment knowledge, the treatment schools' adjusted average score was 2.78 points higher (SE = 0.99; p < .01) than that of control schools. Further, teachers in the treatment schools reported involving their students in assessment activities 9% more days (SE = 0.03, p < .001) than the control teachers.

In contrast, analyses failed to yield statistically significant impacts of CASL on teacher practice. Specifically, treatment and control teachers had very similar scores on assessment work samples used to measure teacher practice on six dimensions of classroom assessment (adjusted mean difference = 0.01, SE = 0.05, p > .05). Finally, no meaningful difference was found between treatment and control schools on student CSAP mathematics achievement (adjusted mean difference = 0.58, SE = 3.47, p > .05).

Conclusions

Overall, results of this study suggest that if schools purchase CASL and implement it under similar conditions as those found in this study, positive impacts may be found on the proximal outcomes of teacher assessment knowledge and student involvement in assessment. However,
our study provides no evidence that impacts on the distal outcomes of formative assessment practice or student performance on the state mathematics tests will be realized. The fact that results were positive for teacher assessment knowledge and student involvement but not for teachers’ assessment practices suggests that improving teacher assessment knowledge and increasing student involvement may be easier than improving such assessment practices as articulating clear learning objectives and assessment criteria and providing students with descriptive feedback. These findings have implications for how PD should be designed and delivered. Specifically, self-study and teacher-led learning teams may improve knowledge, but are probably not enough to support complex instructional strategies. Instead, these strategies may require sustained external facilitation and support along with ongoing practice of skills in order to produce meaningful results in the classroom.

References
Dividing by Zero:
Exploring Null Results in a Mathematics Professional Development Program

Heather C. Hill
Harvard Graduate School of Education

Robin T. Jacob
University of Michigan

Douglas Corey
Brigham Young University

Background

Since 2002, U.S. federal funding for educational research has favored the development and rigorous testing of interventions designed to improve student outcomes. However, a large proportion of the programs developed and rigorously tested in the past decade have shown null results on student outcomes. Scholars reporting on null results often explain these findings informally. In this paper, we argue for a more systematic approach to examining null results and illustrate this approach via an examination of one program’s failure to impact teaching and learning. Although identifying the causes of a null result is a little like dividing zero by zero—results are always indeterminate owing to the large number of possible contributing factors—gaining purchase on potential explanations would allow programs to adopt more effective designs.

Purpose

This paper develops a framework to test substantive explanations for null results. We argue for basing this framework in the existing literature on null results as well as the more-established scholarship on policy implementation. Over a period of forty years this latter literature has identified a number of major causes of program failure, making it potentially useful in designing a method to interpret null results from educational evaluation. We also integrate insights from the teacher professional development literature, which both complements the implementation findings and makes its lessons more education-specific.

Intervention

The intervention used in this paper, Math Solutions (www.mathsolutions.com), is a mathematics professional development program that operates nationally, and is designed to offer a full range of professional development resources to teachers, principals, and district leaders. Three goals are central in the professional development program; they include helping teachers to 1) learn more mathematics, 2) understand how children learn math, and 3) develop mathematical tasks for effective classroom instruction. To operationalize these goals, the program uses many of the best practices identified in Garet et al. (2001) and recommended by professional development organizations (e.g., NSDC, 2001). Observational evidence suggests Math Solutions is mathematically intensive, thoughtful regarding teacher learning, and carefully designed and implemented. Teachers participating in the program attend a five-day summer institute, then
spend six days during the school year engaged in further workshops, teacher team meetings, and classroom-based observation and coaching. However, a three year cluster randomized experiment found no statistically significant impacts on teachers’ instructional practice and no impact of the program on either a state standardized assessment or a project-administered math assessment.

**Research Design:**

To assess potential explanations for the null results observed in this study, we specified a set of hypotheses related to specific factors that have been identified as barriers in the scholarly literature on policy implementation and professional development, including *a lack of will, insufficient resources, organizational barriers, the difficulty of ambitious practice, weak instructional press and the fit of the intervention to teacher’s needs*. We then drew on quantitative, observational, and interview data (described below) to evaluate each. Although our analyses cannot definitively identify the reason(s) for program failure, we argue that this highly structured search for substantive explanations can both inform the field and provide the basis for the future aggregation of results across programs.

**Data Collection and Analysis:**

As part of the RCT longitudinal data were collected over a period of three years. A self-administered questionnaire (SAQ) assessed teachers’ mathematical knowledge and self-reported instructional practices at baseline and in the spring of each of the three study years. Both state and study-specific student tests were used to assess student achievement. Researchers attended 19 days of professional development led by Math Solutions staff. These days were split roughly equally between the summer and the school year. Researchers took notes on the tasks completed by teachers, the discussions teachers had with one another, and the whole-group discussions between the providers and teachers. During the first two years of the study, lessons taught by treatment group teachers were video recorded; in the third and final year of the study, lessons taught by both treatment and control group teachers were captured. At each time point, participating teachers recorded six lessons averaging roughly one hour each. Lessons were sampled in three blocks of two back-to-back taping days. Finally, at the conclusion of the project, school researchers conducted interviews with 31 of the 33 teachers who remained in the treatment group at the end of the project. Interviews were also conducted with the service providers.

**Findings / Results:**

Results are shown in Table 1. They suggest that in this example, organizational barriers, the difficulty of ambitious practice and weak instructional press are the most likely contributors to the null findings observed. In other evaluations different factors may prove to be most important. However, systematically testing competing explanations for the failure of educational evaluations can help researchers and program developers gain purchase on the most likely causes of program failure.
<table>
<thead>
<tr>
<th>Potential Explanations</th>
<th>Evidence</th>
<th>Extent of Support for Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of Will</td>
<td>Teachers indicated on a survey that they learned and had changed their practice as a result of the PD. All teachers interviewed indicated at least one change resulting from the PD. Many positive comments about PD quality.</td>
<td>Weak</td>
</tr>
<tr>
<td>Insufficient Resources (including knowledge and materials)</td>
<td>Analysis of data showed no differential effect of mathematical knowledge for teaching on instructional quality scores. Post-lesson teacher logs showed that MS curriculum was rarely used, and when it was used, it resulted in slightly lower instructional quality scores.</td>
<td>Weak</td>
</tr>
<tr>
<td>Organizational Barriers</td>
<td>When asked about possible difficulties in implementing MS like practices, more than half of the teachers volunteered restrictive district policies as a reason.</td>
<td>Strong</td>
</tr>
<tr>
<td>Difficulty of Ambitious Practice</td>
<td>For many MS practices were only superficially implemented, resulting in more conventional-looking mathematics instruction</td>
<td>Strong</td>
</tr>
<tr>
<td>Weak Instructional Press</td>
<td>Researcher observations of MS PD activities indicate that there was opportunity for participants to engage in meaningful mathematical work but often fell short with the discussion focused on lighter pedagogical points. In interviews with MS PD providers they indicated that pressing the teachers for mathematical work might disrupt their relationships with the participants and result in participants failing to return for more training.</td>
<td>Strong</td>
</tr>
<tr>
<td>Fit of Intervention to Teacher's Needs</td>
<td>Content fit well to the needs observed in some teachers' instruction; other needs, however, were not addressed by the program.</td>
<td>Moderate</td>
</tr>
</tbody>
</table>