Abstract Title Page

Not included in page count.

Title: The impact of curriculum-based professional development on science instruction: Results from a cluster-randomized trial

Author(s): Joseph Taylor, Susan Kowalski, Stephen Getty, Christopher Wilson, Janet Carlson

Abstract Body

Limit 4 pages single spaced.

Background / Context:

The research described in this abstract is part of a larger, IES-funded study titled: *Measuring the Efficacy and Student Achievement of Research-based Instructional Materials in High School Multidisciplinary Science* (Award # *R305K060142).* The larger study seeks to use a cluster-randomized trial design, with schools as the unit of assignment, to make causal inferences about the effect of treatment on both students and teachers. In the context of this study, the treatment is defined as teacher and student use of a comprehensive, year-long, program of instructional materials, as well as a seven-day professional development (PD) program for treatment group teachers that is directly focused on use of the instructional materials. The comparison group continues to use extant instructional materials and receive extant professional development (i.e., business-as-usual). Outcome measures for all students include science achievement test scores. Outcome measures for all teachers include measures of classroom instruction.

The treatment is a combination of instructional materials and professional development because the developers hypothesized that the quality of classroom instruction and materials implementation is as critical, if not more, to effects on student achievement than simply having the instructional materials in classrooms. Further, the larger study is funded as an "efficacy trial" and therefore seeks to study the effects of the instructional materials under more ideal conditions and with high standards of internal validity. As such, an on-going professional development program is needed to encourage high fidelity use of the instructional materials and thus allow for study of student and teacher effects under more ideal conditions.

The developers' hypothesis regarding the critical role of classroom instruction is in fact one of *mediation*. That is, the developers hypothesize that classroom instruction mediates the relationship between the treatment and student achievement (see Figure 1).

Insert Figure 1 about here

The portion of the larger study that is described in this abstract is the portion that examines the effect of treatment on the teacher outcome of classroom instruction, or path *"a"* in Figure 1. The developers' hypothesis that the professional development program, focused on use of the instructional materials, will improve classroom instruction is based on prior studies conducted by the developers. In these studies, the effects of instructional materials-based professional development on teacher outcomes were promising (*citations removed in blinded version*). The data reported in this abstract were collected during the 2009-10 academic year.

Purpose / Objective / Research Question / Focus of Study:

The research described in this abstract addresses the following research question associated with path "a" in Figure 1:

1. What is the mean difference in teacher outcome (i.e., instruction) across the treatment groups?
    a. What is the effect size (practical significance)?
    b. Is the difference statistically significant at the $\alpha = .05$ level?
2. If practically or statistically significant differences in instruction exist across treatment groups, to what extent can the differences be attributed to the treatment (instructional materials and PD)?

Setting:
The research reported here takes place in both suburban and rural high schools in the state of Washington. In particular, the suburban schools are clustered near Seattle/Tacoma and the rural schools are clustered near Yakima.

Population / Participants / Subjects:
The participants in the research reported here are 53 ninth-grade science teachers distributed across 18 high schools, nine schools in suburban settings, and nine in rural settings. Twenty-six teachers were in treatment schools, twenty-seven in comparison (business-as-usual) schools. Each of the two treatment groups includes rural and suburban schools. All high schools in the study are 'traditional" high schools. That is, the sample of high schools does not include any non-traditional high schools such as vocational/technical, magnet, or correctional.

Intervention / Program / Practice:
The year-long professional development program includes seven days of instruction distributed across four events. The first event is a four-day summer institute. The summer institute is followed by three follow-up events distributed throughout the school year. The professional development program focuses on introducing teachers to the physical and philosophical components of the instructional materials as well as strengthening their content background and use of key instructional strategies essential to effective, high-fidelity use of the materials.

Research Design:
The design is a cluster-randomized trial (Raudenbush 1997) where schools were randomly assigned to treatment conditions. Neither matching nor blocking was used prior to random assignment. Treatment assignments were determined using a random number generator (even = treatment, odd= comparison). The design can also be thought of as a ***Post-Test Only Control Group*** design (Shadish, Cook et al. 2002) as it was not possible to obtain a pre-intervention measure of classroom instruction.

Data Collection and Analysis:
The instrument used to measure instruction in both treatment groups was the Reform Teaching Observation Protocol (Piburn, Sawada et al. 2000). The Reform Teaching Observation Protocol (RTOP) includes 25 rating scale items. Each scale varies from a score of "0" – never occurred to a score of "4" – very descriptive. The maximum RTOP score for a given classroom observation is 100. The subscales of the RTOP include:

1. Lesson Design and Implementation
2. Content
    a. Propositional Knowledge
    b. Procedural Knowledge
3. Classroom Culture
    a. Communicative Interactions
    b. Student/Teacher Relationships

Lesson Design and Implementation includes five items. Content and Classroom Culture consist of 10 items each. As a whole, the protocol addresses teacher attention to students' prior knowledge, student engagement in a learning community, and teacher's use of inquiry to promote an atmosphere of problem solving and student generated ideas. The face validity of the

RTOP was established based in part on the National Council of Teachers of Mathematics' *Professional Standards for Teaching Mathematics* (NCTM, 1991) as well as the National Research Council's *National Science Education Standards* (National Research Council (NRC) 1996). Validation studies of the RTOP suggest that it can have strong psychometric properties. Specifically, construct validity statistics are promising (see Table 1).

Insert Table 1 around here

Most teachers in this study were observed approximately once a month for a total of eight observations. A small number of teachers were observed only seven times during the school year. The dependent variable for teachers is their mean RTOP score across the eight (or in some cases, seven) observations. Two observers were contracted to visit classrooms and score the instruction using the RTOP. Some teachers' mean RTOP scores were based on observations from both observers. Therefore, planned redundancy was built into the observation schedule. Specifically, 10% of the total number of observations was conducted by both observers at the same time. Cronbach's alpha was computed as an expression of inter-rater reliability between the observers and its value was deemed acceptable ($\alpha = 0.94$)

Findings / Results:

Upon inspecting the descriptive statistics, a noteworthy difference was observed in school-level mean RTOP scores across treatments (see Table 2).

Insert Table 2 about here

To examine the statistical significance of this treatment effect, a two-level hierarchical model was tested. Level one of this model was unconditional, where the average RTOP score for teacher *i* in school *j* was modeled as a function of the school-mean RTOP score (β_{0j}) and the random effect (r_{ij}) for teacher *i*. At level two, the school-mean RTOP score was modeled as a function of the grand mean of school mean RTOP scores (γ_{00}), an effect coded treatment effect (γ_{01}), and the school-level random effect (μ_{0j}). The results of this analysis are in Tables 3 and 4.

Level 1: $RTOP_{ij} = \beta_{0j} + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(TREAT)_j + \mu_{0j}$

Insert Tables 3 and 4 about here

It is clear from Table 3 that the treatment effect is statistically significant at the $\alpha = .05$ significance level. The sample of teachers in this study is not an equal probability sample so generalization of findings is quite limited. That caveat acknowledged, to inform the extent to which this treatment effect might apply to a larger, similar population of science teachers, we computed a 95% confidence interval around the treatment effect (γ_{01}) using the following expression suggested by Raudenbush and Bryk (2002);

$$\gamma_{01} \pm 1.96\, [\mathrm{Var}(\gamma_{01})]^{1/2}$$

This yields an interval of [10.7<----->23.5]. In a random sampling context, an interpretation of this interval is that we can be 95% confident that the true treatment effect (difference in RTOP school means across treatment groups) in the larger population is between 10.7 and 23.5 points. In addition to statistical significance, we computed an effect size for this treatment effect using Hedges' g for means, corrected for small sample size (see below).

$$\hat{g} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)SD_1^2 + (n_2-1)SD_2^2}{N_{total}-2}}} \times \left(1 - \frac{3}{4(n_1+n_2)-9}\right)$$

From this analysis, Hedges' g = 2.45. The approximate interpretation for this value is that the mean of school-level RTOP means in the treatment group is 2.45 pooled standard deviations larger than the equivalent in the comparison group. There are few studies of this type to which comparisons can be made but it is defensible to say that this effect size has *practical or substantive* significance given that this difference on the RTOP measure would translate to a difference in classroom instruction that would be easy for most science educators to observe. However, the statistically significant χ^2 value in Table 4 for the random school effect (μ_{0j}) suggests that the model is somewhat underspecified and that the addition of school-level variables as covariates could increase the model's explanatory power.

The research team suspected that teaching experience may influence teacher's RTOP scores. Thus, we aggregated this variable across teachers in each school to create a school-level mean teaching experience covariate as shown in the model below.

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{TREAT})_j + \gamma_{02}(\text{SCHMEANEXP})_j + \mu_{0j}$

The HLM results with this covariate included at level two are shown in Tables 5 and 6. Although the main effect of teaching experience was not significant at $\alpha = .05$, including this variable in the model did improve its explanatory power. Using variance estimates from Tables 4 and 6 in the following expression:

$[\tau_{\mu 0j \text{ (compact model)}} - \tau_{\mu 0j \text{ (augmented model)}}] / \tau_{\mu 0j \text{ (compact model)}} = [22.6 – 17.3] / 22.6 = .23$,

we observed that by adding teaching experience to the model we have reduced the amount of unexplained variance in school-mean RTOP scores by 23%. Again, the statistically significant χ^2 value in Table 6 suggests that the model is still somewhat underspecified and that including additional school-level variables as covariates may increase the model's explanatory power.

Conclusions:

Research Question 1. The data from this analysis suggest that the PD treatment was more effective in fostering reform-oriented science instruction, on average, than was the extant PD experienced by the business-as-usual comparison group. This difference was both statistically and practically significant. Applying this result to our hypothesis of mediation, we now have confidence that one of the causal paths (path a) that are necessary to argue mediation is trustworthy. Further study of path b is necessary to understand whether instruction is serving as a mediator of the treatment effect. That said, there is evidence in the literature suggesting that the possibility of a significant b path is quite real. For example, Hedges and Hedberg (2007) found that in school-level interventions, a considerable amount of the variance in outcomes was attributable to teacher and /or classroom effects.

Research Question 2. Threats to internal validity that are noteworthy include limitations in our confidence that the post-intervention differences in RTOP scores were not pre-existing (i.e., not attributable to the treatment). Unfortunately, we did not have a baseline RTOP measure that could have served as a covariate in the main effect analysis of treatment. Use of such a covariate would have likely provided a more precise estimate of the treatment effect. Further, because the comparison group received business-as-usual PD, this experience was highly variable across teachers. The research team has only cursory knowledge of the nature and duration of extant PD experienced by the comparison group. As such, there is limited clarity in the PD experiences to which the treatment is being compared.

In the context of an efficacy trial, external validity (i.e., generalizability) of findings is not paramount. However, it should be noted again that our sampling approach was not random. Therefore, we are cautious not to suggest that our treatment effect estimates would generalize far beyond our sample of rural and suburban schools in Washington state.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60-87.

Hulleman, C. S., & Cordray, D. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*(1), 88-109.

National Council of Teachers of Mathematics (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.

National Research Council (NRC) (1996). *National science education standards*. Washington DC: National Academy Press.

Piburn, M., Sawada, D., Falconer, K., Turley, J., Benford, R., Bloom, I. (2000). *Reformed teaching observation protocol (RTOP)*. (ACEPT Report no. IN-003).

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods 2*(2), 173-185.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi experimental designs for generalized causal inference.* Boston: Houghton Mifflin Company.

Appendix B. Tables and Figures
Not included in page count.

Figure 1. Hypothesized Causal Pathways

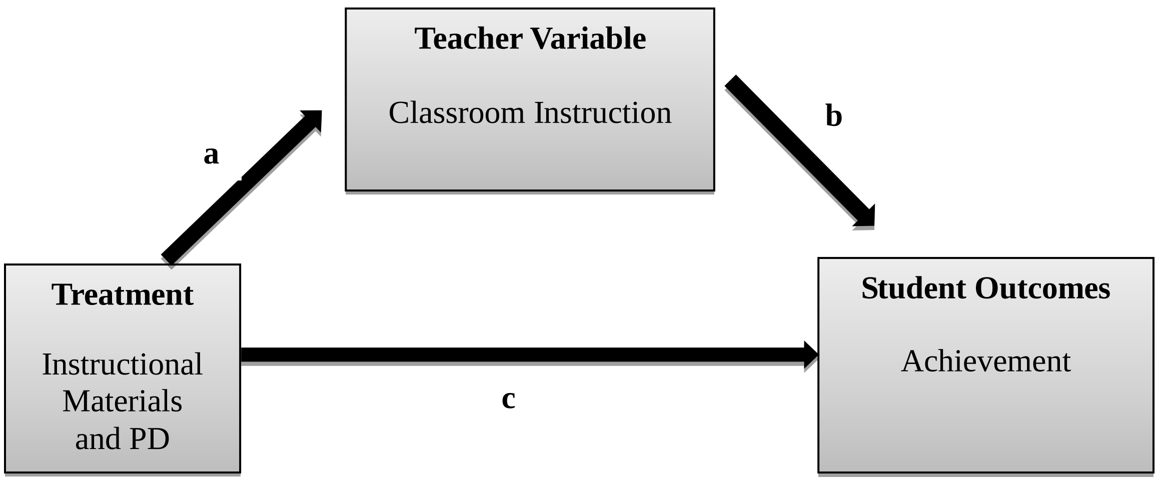

Table 1. Subscales as Predictors of the RTOP Total Score

Subscale	R^2 of Subscale as Predictor of Total Score
1	**0.956**
2a	**0.769**
2b	**0.971**
3a	**0.967**
3b	**0.941**

Table 2. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
School Mean RTOP Score (Comparison)	9	46.5	61.5	53.9	4.9
School Mean RTOP Score (Treatment)	9	62.0	86.1	72.0	8.7

Table 3. Estimation of Fixed Effects

Fixed Effect	**Coefficient**	**Standard Error**	**t-ratio**	**Approximate df**	**p-value**
Intercept (γ_{00})	62.9	1.6	39.3	16	0.000
Treatment (γ_{01})	17.2	3.2	5.4	16	0.000

Table 4. Estimation of Variance Components

Random Effect	**Standard Deviation**	**Variance Component**	**Approximate df**	χ^2	**p-value**
μ_{0j}	4.8	22.6	16	34.3	0.005
r_{ij}	7.7	58.5			

Table 5. Estimation of Fixed Effects (Augmented Model)

Fixed Effect	**Coefficient**	**Standard Error**	**t-ratio**	**Approximate df**	**p-value**
Intercept (γ_{00})	63.1	1.5	42.2	15	0.000
Treatment (γ_{01})	15.4	3.1	5.0	15	0.000
SchoolMeanExp (γ_{02})	0.51	0.27	1.9	15	0.072

Table 6. Estimation of Variance Components (Augmented Model)

Random Effect	**Standard Deviation**	**Variance Component**	**Approximate df**	χ^2	**p-value**
μ_{0j}	4.2	17.3	15	28.0	0.021
r_{ij}	7.6	57.7			