Abstract Title Page

Title: The Effects of Research-Based Curriculum Materials and Curriculum-Based Professional Development on High School Science Achievement: Results of a Cluster-Randomized Trial

Authors and Affiliations:

Joseph Taylor

Susan Kowalski

Stephen Getty

Christopher Wilson

Janet Carlson

(all Biological Sciences Curriculum Study, Colorado Springs, CO 80919)

Abstract Body

Background / Context: Effective instructional materials can be valuable interventions to improve student interest and achievement in science (National Research Council [NRC], 2007); yet, analyses indicate that many science instructional materials and curricula are fragmented, lack coherence, and are not carefully articulated through a sequence of grade levels (AAAS, 2001; Schmidt et al., 2001). In order to improve student achievement in science, school districts need evidence about the efficacy of instructional materials so they can make sound decisions about their science programs. In addition, science education researchers and curriculum developers can benefit from a better understanding of the characteristics of instructional materials and PD that promote student achievement. This paper presents the results of a recently completed study that examines the connections among research-based materials, PD, and student achievement.

The findings from the TIMSS analysis (Schmidt, et al, 1997), the research syntheses, How People Learn (Bransford, Brown, & Cocking, 1999) and Knowing What Students Know (Pellegrino, Chudowsky, & Glaser, 2001), and the Framework for the Next Generation Science Standards (NRC, 2012) provide clear and compelling guidance for the development of effective instructional materials. Specifically, instructional materials should 1) address core concepts in science in a coherent way, as well as make connections between core ideas and across disciplines; 2) provide students opportunities to express and confront their prior conceptions; 3) help students to be metacognitively aware of their own learning; and 4) provide opportunities and scaffolding to enable students to engage in key science practices (argumentation, explanation development, questioning, etc.). We developed the materials in this study to align with these criteria.

Furthermore, a number of research reports indicate that well-designed, standards-based materials supported by professional development focused on the implementation of the materials can have a significant impact on teaching and learning (Briars & Resnick, 2001; Russell, 1998; Schneider & Krajcik, 2002; Taylor et al., 2003). For example, a large-scale study by Cohen and Hill (2002) found that mathematics teachers who participated in sustained professional development *based on the curriculum they were learning to teach* were much more likely to adopt effective teaching practices than those who engaged in other kinds of professional development. More recently, Lara-Alecio and colleagues (2012) found that effective instructional materials along with teacher participation in professional development were, in turn, associated with higher achievement for students. We do not have sufficient space in this proposal to fully describe the materials or the professional development. We will include a complete description of both in the full version of the paper.

Purpose / Objective / Research Question / Focus of Study: The purpose of this study is to examine the efficacy of an intervention that consists of research-based multidisciplinary science curriculum materials for high school students and curriculum-based professional development (PD) for teachers using the materials. The outcome variables are student achievement and teachers' use of reform-based classroom practices. We consider the intervention a bundled intervention because we regard classroom practice (instruction) as critical to the effectiveness of the curriculum materials and comprehensive PD as necessary to promote classroom practices that are complimentary to the goals of the curriculum. Thus, we see the role of classroom practice as one of partial mediation. That is, in addition to having a direct effect on student achievement, we hypothesize that the intervention results in more reform-based classroom instruction that in turn improves student achievement (see Figure 1).

Insert figure one about here

Setting: This study took place in traditional high schools the state of Washington. Approximately half of the schools were in rural settings in central Washington, the other half were in suburban settings in western Washington. Each treatment group had both rural and suburban schools.

Population / Participants / Subjects: The study sample included nearly 4000 ninth and tenth grade students nested within 18 high schools. The teacher sample within these 18 schools included 54 teachers.

Intervention / Program / Practice: Ninth and tenth grade science teachers in nine treatment schools received curriculum materials [program name removed in blinded version] and seven days of curriculum-based professional development for each of two years. Teachers in nine comparison schools continued to use extant instructional materials and receive extant professional development (i.e., business-as-usual). Comparison teachers received, on average, just two days of extant PD per year. Consequently, as noted below, the treatment condition differed from the comparison condition by the quantity of PD, as well as the quality of the PD and instructional materials used in the classroom.

Research Design: In order that we might have high confidence in making causal claims about the instructional materials and PD, we used a *cluster-randomized trial* design (Raudenbush et al., 2002) where schools were randomly assigned to treatment conditions. Neither matching nor blocking was used prior to random assignment as schools joined the study too close to its onset for an accurate stratification to be established prior to assignment.

Data Collection and Analysis:

The Outcome Measures

Measure of Classroom Instruction. The instrument used to measure the primary teacher outcome, reform-based instruction, was the *Reform Teaching Observation Protocol* [RTOP] (Sawada et al., 2002). Most teachers in this study were observed approximately once each month for a maximum of eight observations. A small number of teachers were observed only seven times during the school year. The dependent variable for classroom instruction was teachers' mean RTOP score across the seven or eight observations. We contracted two external researchers to visit classrooms and score instruction using the RTOP to help eliminate the potential for researcher bias. Inter-rater reliability was calculated using the intra-class correlation coefficient statistic – a measure that takes into account both the absolute agreement between raters and the correlation of their scores. Across all shared observations, the ICC for the two raters was highly satisfactory: $\rho = 0.96$ (mixed effects, absolute agreement, average measures).

Measure of Student Achievement. We used the Washington state High School Proficiency Exam (HSPE) as our measure of student achievement. The science test is administered in eighth grade and in tenth grade. We used students' eighth grade science scale scores as well their eighth grade mathematics and seventh grade writing scale scores as baseline covariates in our analytic model.

Analysis Techniques

Confirmatory Analysis: Main Effect of Treatment: For the main effect of treatment, we estimated a two-level hierarchical linear model within STATA12 to examine the statistical significance of the treatment effect. Level 1 modeled students' tenth grade science scale scores as a function of a variety of grand mean-centered or effect coded covariates. These included students' eighth grade science scale scores, seventh grade writing scale scores, eighth grade math scale scores, and their demographic characteristics (free and reduced-price lunch status [FRL], English Language Learner status [ELL], special education status [SPED], race/ethnicity, and gender). Level 2 modeled school-mean tenth grade science scale scores as a function of an effect coded treatment variable, the school mean eighth grade science scale score, the school mean eighth grade math scale score, and the percent of students in the school who are FRL eligible (all grand mean centered).

Exploratory Analysis: Teacher Practice as a Mediator (Indirect Effect on Achievement): Our mediation analyses examined the effect of the intervention on student learning as mediated by classroom practice. This analysis tests our hypothesized path of influence of the intervention. Mediation exists when: 1) the treatment variable directly affects the mediator (path *a*)***;*** 2) the mediator directly affects the outcome variable when controlling for treatment (path *b*), 3) the treatment variable may or may not directly affect the outcome variable (path *c′*), and 4) a significant *a*b* product exists (Mackinnon, 2008).

To test path *a*, we used a two-level hierarchical linear model to examine the statistical significance of the treatment on teacher practice. At level one, RTOP scores were modeled unconditionally. At level two, the school-mean RTOP score was modeled as a function of the treatment variable and a school-level random effect.

To test path *b* and *c′*, we used a three-level hierarchical linear model for a "3-2-1" mediation analysis. That is, the treatment assignment was at level 3 (school), the mediator was at level 2 (teacher), and the student outcome of interest was at level 1 (Pituch et al., 2010). Level 1 modeled students' tenth grade science scale scores as a function of their eighth grade science scale scores, eighth grade math scale scores, seventh grade writing scale scores, and student demographics (all independent variables were effect coded or grand mean centered). Level 2 modeled the mean 10^{th} grade science scale scores by teacher as a function of teachers' RTOP scores (grand mean centered). Level 3 modeled the school mean 10^{th} grade science scale scores as a function of the treatment variable (effect coded) and the school mean eighth grade science scale score, the school mean eighth grade math scale score, and the percent of students in the school who are FRL eligible (all grand mean centered).

To statistically test for the significance of mediation, we first calculated the *ab* product (the product of the coefficient for the treatment variable in path *a* times the coefficient for the RTOP variable in path *b*), divided by the standard error of the *ab* product (Sobel 1986). Because the *ab* product can follow a non-normal distribution, particularly for multilevel models, we used the computer program PRODCLIN (MacKinnon et al., 2007) to ensure more accurate Type I error rates.

Findings / Results:

Confirmatory Analysis: Main Effect of Treatment

We found the treatment coefficient to be statistically significant (γ_{01} = 3.68, SE = 1.75, p = 0.035). The Hedges' *g* effect size for this treatment effect was *g* = .09 with a 95% confidence interval of [.01<--->.17]. Bloom et al, (2008) reported that the expected normative gain for science students from eighth to tenth grade as being approximately 0.41 (Hedges' *g*). Thus, this effect size is equivalent to about 4 months of instructional time when compared to the expected normative gain for science from eight to tenth grade [(.09/.41) x 18 school months]. The intervention (instructional materials plus PD) appears to have a positive influence on student science achievement as measured by a state standardized science assessment.

Exploratory Analysis: Test of Teacher Practice as a Mediator (Indirect Effect on Achievement
For path *a*, the effect of treatment on classroom practice, we found the treatment effect to be statistically significant (γ_{01} = 16.7, SE = 3.1, p < .001). The Hedges' g effect size was *g* = 1.85. 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 reform-based instruction that would be easy for most science educators to observe. For path *b*, the effect of mediator on outcome, we found the RTOP coefficient, β_{01} = 0.13 (SE = .07, p = .07). Thus, teachers' RTOP scores do not account for significant variation in mean student science achievement at α = .05, but the coefficient does *approach* significance. In addition, the effect of the treatment estimated in the three-level model that included the mediator (*c'*) was just 1.56 (SE = 2.21, p = 0.49) compared to 3.68 (SE = 1.75, p = 0.035) from the confirmatory main effect analysis. These results are consistent with a mediation effect but formal tests of the *ab* product are recommended and reported below.

From our mediation analyses, we computed an *ab* product of 2.18 with a confidence interval of = -0.12 to 4.84 (p = 0.64). Thus, this study was unable to detect (at the α = .05 level) whether classroom instruction as measured by the RTOP mediates the relationship between the treatment (instructional materials plus PD) and student science achievement. It may be that classroom instruction does not strongly mediate the relationship between the intervention and student achievement. Alternatively, we note that the relationship between teacher RTOP scores and student achievement is positive and the *ab* product approaches significance. It may be that this study simply lacks sufficient power to detect the mediation effect. Further research with a larger pool of schools would clarify this result.

Conclusions: In this study, we empirically tested the efficacy of research-based instructional materials and PD. We briefly described the materials as attending to coherence of core ideas within and across science disciplines, providing opportunities for students to express and confront their prior conceptions, promoting student metacognition, and providing opportunities and scaffolding for students to engage in key science practices. We found that research-based science instructional materials with supporting curriculum-based PD have a strong positive effect on classroom instruction and a modest but noteworthy effect on student achievement. We have evidence that classroom practice mediates the treatment effect but this result is inconclusive at the .05 significance level (p=.064).
These results are important because they add to the growing body of evidence that research-based instructional materials supported by curriculum-based PD promote improved student achievement (Lynch et al., 2005; Lee et al., 2008). As teachers grow and learn from instructional materials and curriculum-based PD, their teaching improves and their students' achievement improves. Thus, the research-based instructional materials plus PD used in this study and the results of our research showcase one path toward improved scientific literacy in the United States. As such, this study holds important ramifications for teachers, school districts, curriculum developers, and professional development providers. Perhaps greater emphasis on the identification of high-quality research-based instructional materials and supporting teachers with PD can ultimately begin to transform science teaching and learning in the United States.

.

Appendices.

Appendix A. References

American Association for the Advancement of Science (AAAS). (2001). *Atlas of science literacy.* Washington, DC: Project 2061 and the National Science Teachers Association.

Bloom, H., Zhu, P., Jacob, R., Raudenbush, S., Martinez, A., & Lin, F. (2008). Empirical issues in the design of group-randomized studies to measure the effects of interventions for children. *MDRC Working Papers on Research Methodology*.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Briars, D., & Resnick, L. (January 2001). *Standards, assessments - and what else? The essential elements of standards-based school improvement.* Paper presented at the Local Systemic Change PI Meeting, National Science Foundation, Washington, DC.

Cohen, D. K., & Hill, H. C. (2002). *Learning policy: When state education reform works.* New Haven, CT: Yale University Press.

Committee on Science, Engineering, and Public Policy (COSEPUP). (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, D.C.: National Academies Press.

Darling-Hammond, L. (1997). *Doing what matters most: Investing in quality teaching*. New York: National Commission on Teaching and America's Future.

International Association for the Evaluation of Educational Achievement (IEA). (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study (TIMSS)*. Center for the Study of Testing, Evaluation and Educational Policy. Chestnut Hill, MA: Boston College.

Lara-Alecio, R., Tong, F., Irby, B. J., Guerrero, C., Huerta, M., & Fan, Y. (2012). The effect of an instructional intervention on middle school English learners' science and English reading achievement. *Journal of Research in Science Teaching.* doi: 10.1002/tea.21031

Lee, O., Deaktor, R., Enders, C., & Lambert, J. (2008). Impact of a multiyear professional development intervention on science achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching, 45*(6), 726-747.

Lynch, S., Kuipers, J., Pyke, C., & Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results form a planning grant. *Journal of Research in Science Teaching*, *42*: 921-946.

National Research Council. (1997). *Learning from TIMSS results of the third international mathematics and science study*. Washington, DC: National Academy Press.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of assessment*. Washington, DC: National Academy Press.

Pituch, K. A., Murphy, D. L., & Tate, R. L. (2010). Three level models for indirect effects in school- and class-randomized experiments in education. *The Journal of Experimental Education, 78*(1), 60-95.

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

Russell, S. J. (1998). Mathematics curriculum implementation: Not a beginning, not an end. *Hands On!, 21*(1), 6-29.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science & Mathematics, 102*(6), 245–253.

Schmidt, W. H. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco, CA: Jossey-Bass.

Schmidt, W. H., McKnight, C. C., & Raizen, S.A. (1997). *Splintered vision: An investigation of U.S. science and mathematics education.* Boston: Kluwer Academic.

Schneider, R.M., & Krajcik, J. (2002). Supporting science teacher learning: The role of educative curriculum materials. *Journal of Science Teacher Education,* 13, 221-245.

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. *Sociological Methodology*, *16,* 159-186.

Taylor, J. A., Powell, J., Van Dusen, D. R., Pearson, B., Bess, K., & Schindler, B. (2003, June). *Rethinking the continuing education of science teachers: An example of transformative, curriculum-based professional development.* NSTA Monograph Series: Exemplifying the More Emphasis Conditions in the National Science Education Standards.

Appendix B. Tables and Figures

Figure 1. Hypothesized Causal Pathways