Contact: Jessaca Spybrook
Jessaca.spybrook@wmich.edu

Title: Improving the Design of STEM Impact Studies: Considerations for Statistical Power

Conference Section: Research Methods

Discussant: Ben Kelcey
Ben.kelcey@gmail.com

Symposium Justification:

The purpose of this symposium is to present recent work aimed at improving the design and power analyses of impact studies of STEM interventions. It is often the case that these impact studies utilize a cluster randomized trial (CRT) in which students are nested within teachers nested within schools. The outcomes of interest may be at the student level, such as science achievement, or at the teacher level, such as teacher pedagogical content knowledge or content knowledge. In order for these CRTs to yield rigorous evidence of whether an intervention has a treatment effect at the student and/or teacher level, it must be designed with adequate power to detect a meaningful effect at either/both levels. The first paper in this symposium presents effect sizes from a meta-analysis of experimental and quasi-experimental studies of interventions with preservice and inservice science teachers, in the U.S. and internationally. The second paper provides results from secondary analyses with 8 existing educational datasets that have at least one teacher outcome in math or science and reports estimates of the intraclass correlation (ICC), and the percent of variance explained by covariate(s) (R^2). The empirically estimated design parameters (effect sizes, ICCs, R^2s) from these two papers will help researchers planning CRTs of STEM interventions conduct more accurate power analyses for teacher-level outcomes, which is critical given that lack of empirical estimates of design parameters for teacher-level outcomes. The third paper considers the design of CRTs that seek to impact both teacher outcomes and student outcomes in one study, such as an impact study that evaluates a STEM professional development program. This paper integrates the findings from papers 1 and 2 as well as recent literature on design parameters for student outcomes to examine the alignment of the power analyses when a study seeks to examine the impact of a STEM intervention on both teacher and student effects.

Title: Investigating Science Teacher Effect Sizes for A Priori Power Analyses

Presenting Author: Kowalski, S. M., BSCS

Authors: Kowalski, S. M., BSCS, Taylor, J. A., BSCS, Askinas, K. M., BSCS, Anderson, D., WMU, Maddix, W., WMU, Wang, Q., WMU & Zhang, Q., WMU

Purpose

As part of the larger project to examine a priori power analysis parameters for teacher outcomes from studies of science education interventions, we conducted a meta-analysis of experimental and quasi-experimental studies of interventions with preservice and inservice science teachers, in the U.S. and internationally. The results will inform study designs for primary researchers.

Research Question:

What effect sizes exist in the research literature on education interventions for science teachers and how do those effect sizes vary within and across designs or intervention characteristics?

Methods

Eligiblility

All eligible studies were

- Published in 2001 or later
- Studied the effect of an intervention on preservice or inservice science teachers of primary or secondary students in U.S. and international contexts
- Published in English
- Included outcome measures for preservice or inservice science teachers
- Included baseline equivalence measures

We selected 2001 because it coincides with the passage of the *No Child Left Behind* act that played an important role in transforming the types of research conducted in educational settings in the U.S.. We wished to identify studies designed to meet that call for causal effects research. We included only studies published in English as we lacked capacity for translation.

Identifying Studies

We identified studies by searching education databases (e.g., ERIC, Dissertation Abstracts), web-based databases (Google Scholar, Yahoo), general education, policy journals, evaluation journals, science education journals, conference programs, research reports from IES and National Science Foundation funded studies, reports from national databases such as the WWC and the Best Evidence Encyclopedia, white papers, and reports from educational publishers' websites.

Screening

We double-screened all abstracts and the PI resolved conflicts. We identified 250 potentially eligible studies based on abstract screening. We identified 164 unique studies for coding after screening full texts.

Coding Procedures

Using a database, we coded information related to research design, sample size and characteristics, treatment and comparison group characteristics, the nature of the intervention and of dependent variables, and effect size data. We calculated standardized mean difference effect sizes. Studies reported findings in many formats. We converted all statistics to d-family effect size indices (see Lipsey & Wilson, 2001).

Interrater Agreement

Coders jointly coded studies and met to discuss codes. During discussions, coders resolved any inconsistencies. They continued to jointly code one study at a time until they reached complete agreement (we initially aimed for 90% agreement, but surpassed that goal by the final two jointly coded studies). The remainder of the studies were divided among the coders.

Statistical Independence

When two studies included overlapping participant samples, we linked the two studies with a single Study ID and used all linked manuscripts to code the single study.

We coded all effect sizes for each study. We accounted for the lack of statistical independence of effect sizes (multiple effects within a single study) using Robust Variance Estimation (RVE) meta-regression (Hedges, Tipton, & Johnson, 2010).

Analysis

Statistical adjustments. We Winsorized sample sizes and effect sizes to ensure that extremely large values did not overly influence regression coefficients. We also adjusted sample sizes for studies that had cluster-level assignment (Higgins et al., 2011). Discrete predictor variables account for study design, sample characteristics, and intervention characteristics. We grand mean centered all variables. We gave more weight to studies whose effect size estimates have greater precision, where precision is primarily a function of study sample size. However, in our approach to weighting, we include a within-study as well as a between-study component to the variance. The within-study component is the average variance across effect sizes within the study, and the between-study component is calculated following Hedges, Tipton, and Johnson (2010).

Analytic model. Our RVE meta-regression equation is:

$$ES_{ij} = \beta_0 + \beta_1 RCT + \beta_2 Pub + \sum_{k=3}^{6} \beta_k Sci_{ij} + \sum_{m=7}^{8} \beta_l DV_{ij} \sum_{l=9}^{13} \beta_m TxComp_{ij}$$
$$+ \beta_{14} AltTx + \beta_{15} TxDif + \beta_{16} AsessDev + \beta_{17} DurHrs + e_{ij}$$

Table 1 defines the variables.

Results

We coded 789 posttest effect sizes across the 164 studies. Table 1 summarizes the results of the regression using RVE (using the program Robumeta in R; Fisher, Tipton, & Hou, 2017). In general, we found a grand mean effect size across interventions of 0.42. Randomized studies tend to have lower effect sizes that quasi-experiments, biology interventions tend to have larger effect sizes than those in other science disciplines, effects tend to be larger when the outcome measured is a practice measure, studies that compare one treatment to another alternative treatment tend to have smaller effects than those that compare a treatment to either business-as-usual or no treatment, and interventions that include curriculum materials for a teacher to use also tend to produce larger effect sizes than interventions that do not include curriculum as a treatment component. We are currently developing a calculator (R-based shiny app) that will allow users to estimate an a priori effect size for a planned intervention. The estimated effect sizes will allow authors to better estimate the number of participants needed in their studies when little other information about an effect size from prior research is available. One surprising finding is that publication status (peer reviewed or not) did not substantially impact the effect size ($\beta = -.08$). That said, our funnel plot (Figure 1) is not symmetrical indicating that we may not be adequately capturing publication bias in our Pub regression coefficient. Studies with low effect sizes may never have been shared in any publicly available format, peer reviewed or otherwise.

Conclusions

We found that effect sizes of teacher outcomes in science education intervention studies are highly variable. In some ways, the variation is expected. One might expect that randomized studies have smaller effect sizes than quasi-experiments. On the other hand, when controlling for randomization and other variables, publication status was *not* highly correlated with effect size. This finding contradicts our own previous work (Authors, 2017) and that of others (Polanin, Tanner-Smith, & Henessy, 2016). We also note that none of the coefficients are statistically significant predictors of effect sizes. Thus, the coefficients pertain specifically to our sample of studies. But given the wide-ranging nature of our search and the inclusion of such a wide array of experiments and quasi-experiments, we assert that the coefficients can provide valuable information to science education researchers as they design their own studies.

Tables and Figures

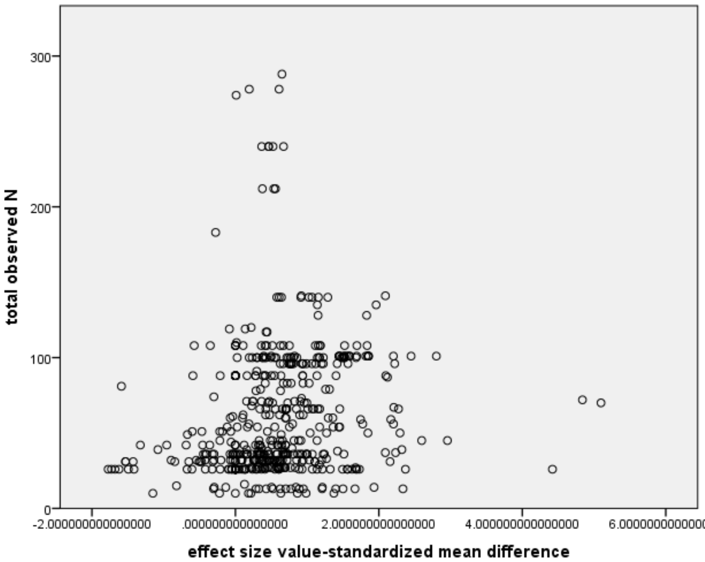
Table 1. Predictor variables in meta-regression model and their meaning.

Variable	**Meaning (1 = yes; 0 = no)**	**Reference group**
RCT	Was the sample randomly assigned to treatment or comparison?	NA
Pub	Was the study published in a peer-reviewed journal?	NA
Chem	Was the science discipline chemistry?	biology
Phys	Was the science discipline physics?	
PhysSci	Was the science discipline physical science?	
EaMul	Was the science discipline Earth/Space or Multidisciplinary?	
ContDV	Was the dependent variable a knowledge outcome?	attitudes
PracDV	Was the dependent variable a classroom practice outcome?	
CurTx	Did the intervention include a curriculum component?	NA
SciTx	Did the intervention include the opportunity to learn science concepts?	NA
MentTx	Did the intervention include a mentorship componenet?	NA
AdMtlTx	Did the intervention include additional materials (not curriculum)?	NA
CollTx	Did the intervention included the opportunity for teachers to collaborate?	NA
AltTx	Did the comparison group receive a focused treatment (rather than BaU)?	NA
TxDif	How many components did the treatment have that the comparison did not?	NA
AssessDev	Was the author the assessment developer?	NA
DurHrs	What was the duration of the study in hours?	NA

Table 2. Regression estimates from Robust Variance Estimation (RVE) meta-regression.

β Coeff	Variable Name	Estimate	SE	t-value	dfs	p-value	95% CI L	95% CI U
0	Intercept	0.41678	0.17091	2.439	17.89	0.0254	0.05756	0.776
1	RCT	-0.32095	0.22741	-1.411	19.89	0.1736	-0.79547	0.1536
2	Pub	-0.08982	0.30095	-0.298	17.06	0.769	-0.7246	0.545
3	Chem	-0.52391	0.43765	-1.197	15.47	0.2493	-1.45426	0.4064
4	Phys	-0.5077	0.41712	-1.217	15.8	0.2414	-1.39285	0.3775
5	PhySci	-0.83323	0.61133	-1.363	4.71	0.2344	-2.43434	0.7679
6	EaMul	-0.36742	0.43449	-0.846	16.16	0.4101	-1.28776	0.5529
7	ContDV	0.12558	0.24075	0.522	18.85	0.608	-0.37858	0.6297
8	PracDV	0.47867	0.36555	1.309	12.12	0.2146	-0.31687	1.2742
9	CurTx	0.27147	0.50537	0.537	7.43	0.6069	-0.90964	1.4526
10	SciTx	0.17696	0.35721	0.495	9.65	0.6314	-0.62294	0.9769
11	MentTx	-0.50857	0.30678	-1.658	10.33	0.1274	-1.18921	0.1721
12	AdMtlTx	0.15388	0.28809	0.534	11.95	0.6031	-0.47412	0.7819
13	CollTx	-0.0906	0.21077	-0.43	18.16	0.6724	-0.53313	0.3519
14	AltTx	-0.34305	0.28009	-1.225	9.83	0.2492	-0.96857	0.2825
15	TxDif	-0.01724	0.10345	-0.167	18.76	0.8694	-0.23396	0.1995
16	AssessDev	0.18133	0.25423	0.713	23.56	0.4827	-0.34389	0.7066
17	DurHrs	0.00248	0.00431	0.575	13.82	0.5746	-0.00677	0.0117

Figure 1. Funnel plot of effect sizes and sample sizes (shows some bias in favor of larger effect size studies).



References

Studies WITHOUT asterisks were included in the meta-analysis. When multiple research reports refer to the same study sample, we coded the multiple reports as a single study.

Abd-El-Khalick, F. ;Akerson, V. (2009). The Influence of Metacognitive Training on Preservice Elementary Teachers' Conceptions of Nature of Science International Journal of Science Education, 31(16), 2161-2184.

Akcay , Nilufer Okur;Doymus , Kemal;Simsek , Umit;Okumus , Seda (2012). The effect of cooperative learning model on academic achievement in physics Energy Education Science and Technology Part B-Social and Educational Studies, 4, 1915-1924.

Akcay, B. (2007). The influence of the history of science course on pre-service science teachers' understanding of the nature of science concepts (Doctoral dissertation, The University of Iowa). Dissertation Abstracts International, , .

Alayyar , Ghaida M.;Fisser , Petra;Voogt , Joke (2012). Developing Technological Pedagogical Content Knowledge in Pre-Service Science Teachers: Support from Blended Learning Australasian Journal of Educational Technology, 28, 1298-1316.

Alkan , Fatma (2013). The Effect of Alternative Assessment Techniques on Chemistry Competency Perceptions and Chemistry Success of Prospective Science Teachers Journal of Baltic Science Education, 12, 774-783.

Anderson, J., & Barnett, M. (2011). Using Video Games to Support Pre-Service Elementary Teachers Learning of Basic Physics Principles Journal of Science Education and Technology, 20, 347-362.

Anderson, Janice L. ;Barnett, Mike (2013). Learning physics with digital game simulations in middle school science Journal of Science Education and Technology, 22(6), 914-926.

Atasoy , Sengul (2013). Effect of Writing-to-Learn Strategy on Undergraduates' Conceptual Understanding of Electrostatics Asia-Pacific Education Researcher, 22, 593-602.

AYDENIZ,M., PABUCCU, A., CETIN, P. S., & KAYA, E. (2012). Argumentation and students' conceptual understanding of properties and behaviors of gases. International Journal of Science and Mathematics Education, 10, 1303-1324.

Aydin , Serap Oz;Sahin , Selin;Sicaker , Aysun;Laborda , JC Ozdamli , F Maasoglu , Y. (2014). The effect of protein synthesis game in the class on the students' understanding of protein synthesis subject. Paper presented at the meeting of the , .

Aydin , Suleyman (2011). Effect of Cooperative Learning and Traditional Methods on Students' Achievements and Identifications of Laboratory Equipments in Science-Technology Laboratory Course Educational Research and Reviews, 6, 636-644.

Aydogdu , Cemil (2012). The Effect of Problem Based Learning Strategy in Electrolysis and Battery Subject Teaching Hacettepe Universitesi Egitim Fakultesi Dergisi-Hacettepe University Journal of Education, , 48-59.

Ayyildiz , Yildizay;Tarhan , Leman (2013). Case study applications in chemistry lesson: gases, liquids, and solids Chemistry Education Research and Practice, 14, 408-420.

Baser , Mustafa;Durmus , Soner (2010). The Effectiveness of Computer Supported versus Real Laboratory Inquiry Learning Environments on the Understanding of Direct Current Electricity among Pre-Service Elementary School Teachers EURASIA Journal of Mathematics, Science & Technology Education, 6, 47-61.

Batiza , Ann Finney;Gruhl , Mary;Zhang , Bo;Harrington , Tom;Roberts , Marisa;LaFlamme , Donna;Haasch , Mary Anne;Knopp , Jonathan;Vogt , Gina;Goodsell , David;Hagedorn , Eric;Marcey , David;Hoelzer , Mark;Nelson , Dave (2013). The Effects of the SUN Project on Teacher Knowledge and Self-Efficacy regarding Biological Energy Transfer Are Significant and Long-Lasting: Results of a Randomized Controlled Trial CBE - Life Sciences Education, 12, 287-305.

Bell, R. L. ;Matkins, J. J. ;Gansneder, B. M. (2011). Impacts of Contextual and Explicit Instruction on Preservice Elementary Teachers' Understandings of the Nature of Science Journal of Research in Science Teaching, 48(4), 414-436.

Benli , Esra;Sarikaya , Mustafa;Baskan , GA Ozdamli , F Kanbul , S Ozcan , D. (2012). The investigation of the effect of problem based learning to the academic achievement and the permanence of knowledge of prospective science teacher: the problem of the boiler stone. Paper presented at the meeting of the , .

Bhattacharyya , Sumita;Volk , Trudi;Lumpe (2009). The Influence of an Extensive Inquiry-Based Field Experience on Pre-Service Elementary Student Teachers' Science Teaching Beliefs Journal of Science Teacher Education, 20, 199-218.

Bhattacharyya, S. (2003). Influence of an extensive inquiry-based field experience on pre-service elementary student teachers' science teaching beliefs (Doctoral dissertation, Southern Illinois University at Carbondale). Dissertation Abstracts International, , .

Bilgin , I.; Senocak, E.; Sozbilir, M. (2009). The effects of problem-based learning instruction on university students' performance of conceptual and quantitative problems in gas concepts Eurasia Journal of Mathematics, Science and Technology Education, 5, 153-164.

Bilgin , Ibrahim (2006). Promoting Pre-Service Elementary Students' Understanding of Chemical Equilibrium through Discussions in Small Groups International Journal of Science and Mathematics Education, 4, 467-484.

Bilgin , Ibrahim;Karakuyu , Yunus;Ay , Yusuf (2015). The Effects of Project Based Learning on Undergraduate Students' Achievement and Self-Efficacy Beliefs Towards Science Teaching Eurasia Journal of Mathematics Science and Technology Education, 11, 469-477.

Bilgin, İ., & Geban, Ö. (2006). The Effect of Cooperative Learning Approach Based on Conceptual Change Condition on Students' Understanding of Chemical Equilibrium Concepts. Journal of Science Education and Technology, 15(1), 31-46.

Bradshaw, T. J. (2012). Impact of inquiry based distance learning and availability of classroom materials on physical science content knowledge of teachers and students in central Appalachia (Doctoral dissertation, University of Kentucky). Dissertation Abstracts International, PhD, .

Bradshaw, T. J., Cunningham, J. D., Osborn , J. L., & Bradley, K. D. (). Impact of Newton's Universe Materials on Middle School Teacher and Student Understanding of Physical

Science Unpublished manuscript, University of Kentucky & Western Carolina University, .

Buschang , Rebecca E. (2012).

Cakmakci , Gultekin;Aydogdu , Cemil (2011). Designing and Evaluating an Evidence-Informed Instruction in Chemical Kinetics Chemistry Education Research and Practice, 12, 15-28.

Caliskan , Serap;Selcuk , Gamze Sezgin;Erol , Mustafa (2010). Instruction of Problem Solving Strategies: Effects on Physics Achievement and Self-Efficacy Beliefs Journal of Baltic Science Education, 9, 20-34.

Capps , Daniel K.;Crawford , Barbara A. (2013). Inquiry-Based Professional Development: What does it take to support teachers in learning about inquiry and nature of science? International Journal of Science Education, 35, 1947-1978.

Cavalluzzo , Linda;Barrow , Lisa;Henderson , Stephen (2014).

Cayci , Baris (2012). The Effect of Technology Based Course Material Use on the Science Teaching Self-Efficacy Beliefs of Pre-service Elementary Teachers New Educational Review, 27, 221-231.

Celikler , Dilek;Aksan , Zeynep;Baskan , GA Ozdamli , F Kanbul , S Ozcan , D. (2012). The effect of the use of worksheets about aqueous solution reactions on pre-service elementary science teachers' academic success. Paper presented at the meeting of the , .

Celikler , Dilek;Aksan , Zeynep;Yalin , HI Adiloglu , F Boz , H Karatas , S Ozdamli , F. (2011). The effect of computer assisted instruction in teaching ionic compounds on pre-service elemantary science teachers' academic achievement and permanent learning. Paper presented at the meeting of the , .

Coban , Gul Unal (2013). The Effects of Inquiry Supported by Argument Maps on Science Process Skills and Epistemological Views of Prospective Science Teachers Journal of Baltic Science Education, 12, 271-288.

Cone, N. T. (2006). The effects of community-based service-learning on preservice elementary teachers' self-efficacy beliefs about equitable science teaching and learning (Doctoral dissertation, University of South Florida). Dissertation Abstracts International, , .

Cotabish, A., Dailey, D., Hughes, G. D., & Robinson, A. (2011). The Effects of a STEM Professional Development Intervention on Elementary Teachers’ Science Process Skills Research in the Schools, 18(2), 16-25.

Cotabish, Alicia ;Dailey, Debbie ;Robinson, Ann ;Hughes, Gail (2013). The Effects of a STEM Intervention on Elementary Students' Science Knowledge and Skills School Science and Mathematics, 113(5), 215-226.

Demirbag , Mehmet;Gunel , Murat (2014). Integrating Argument-Based Science Inquiry with Modal Representations: Impact on Science Achievement, Argumentation, and Writing Skills Kuram Ve Uygulamada Egitim Bilimleri, 14, 386-391.

Demircioglu , Tuba;Ucar , Sedat (2015). Investigating the Effect of Argument-Driven Inquiry in Laboratory Instruction Educational Sciences: Theory and Practice, 15, 267-283.

Diaconu , Dana Viorica;Radigan , Judy;Suskavcevic , Milijana;Nichol , Carolyn (2012). A Multi-Year Study of the Impact of the Rice Model Teacher Professional Development on Elementary Science Teachers International Journal of Science Education, 34, 855-877.

Diamond, B. S. (2013). Effect of a curricular and professional development intervention on elementary teachers' science content knowledge (Doctoral dissertation, University of Miami). Dissertation Abstracts International, , .

Diamond, Brandon S. ;Maerten-Rivera, Jaime ;Rohrer, Rose Elizabeth ;Lee, Okhee (2014). Effectiveness of a curricular and professional development intervention at improving elementary teachers' science content knowledge and student achievement outcomes: Year 1 results Journal of Research in Science Teaching, 51(5), 635-658.

Dogru , Mustafa (2008). The Application of Problem Solving Method on Science Teacher Trainees on the Solution of the Environmental Problems International Journal of Environmental and Science Education, 3, 9-18.

Dogru, Mustafa ;Kalender, Suna (2007). Applying the Subject "Cell" through Constructivist Approach during Science Lessons and the Teacher's View Journal of Environmental Science Education, 2(1), 3-13.

Ebrahim , Ali H. (2012). The Self-Efficacy of Preservice Elementary Teachers in Kuwaiti Science Programs Education, 133, 67-76.

Edgar, D. W. (2007). Structured communication: Effects on teaching efficacy of student teachers and student teacher-cooperating teacher relationships (Doctoral dissertation, Texas A&M University). Dissertation Abstracts International, , .

Ekmekci , Adem;Gulacar , Ozcan (2015). A Case Study for Comparing the Effectiveness of a Computer Simulation and a Hands-on Activity on Learning Electric Circuits EURASIA Journal of Mathematics, Science & Technology Education, 11, 765-775.

Enderle , Patrick;Dentzau , Michael;Roseler , Katrina;Southerl, , Sherry;Granger , Ellen;Hughes , Roxanne;Golden , Barry;Saka , Yavuz (2014). Examining the Influence of RETs on Science Teacher Beliefs and Practice Science Education, 98, 1077-1108.

Eren , C. D.;Akinoglu , O. (2013). Effect of problem-based learning (PBL) on critical thinking disposition in science education Journal of Environmental Protection and Ecology, 14, 1353-1361.

Evren , Aysegul;Sulun , Yusuf;Uzunboylu , H. , (2010). The effect of teaching animal physiology through V-Diagrams" on students' success , retention level Innovation and Creativity in Education, 2(2), 4285-4292.

Ewers, T. G. (2001). TEACHER-DIRECTED VERSUS LEARNING CYCLES METHODS: EFFECTS ONSCIENCE PROCESS SKILLS MASTERY AND TEACHER EFFICACY AMONG ELEMENTARY EDUCATION STUDENTS (Doctoral dissertation, University o f Idaho). Dissertation Abstracts International, PhD, . (University Microfilms No. 3022333)

Fies, C. H. (2005). Classroom response systems: What do they add to an active learning environment? (Doctoral dissertation, The University of Texas at Austin). Dissertation Abstracts International, , .

Fies, C. H. (2005). Classroom response systems: What do they add to an active learning environment? (Doctoral dissertation, The University of Texas at Austin). Dissertation Abstracts International, , .

*Fisher, Z., Tipton, E., & Hou, Z. (2017). Robumeta statistical software package in R. https://github.com/zackfisher/robumeta

Fishman , Barry;Konstantopoulos , Spyros;Kubitskey , Beth W.;Vath , Richard;Park , Gina;Johnson , Heather;Edelson , Daniel C. (2013). Comparing the Impact of Online and Face-to-Face Professional Development in the Context of Curriculum Implementation Journal of Teacher Education, 64, 426-438.

Gatling, A. P. (2010). Investigating the impact of field verses university-based science methods on preservice teachers' belief and abilities to design inquiry -based science instruction for diverse learners (Doctoral dissertation, Boston College). Dissertation Abstracts International, , .

Glogger-Frey , Inga;Fleischer , Corinna;Grueny , Lisa;Kappich , Julian;Renkl , Alexander (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction Learning and Instruction, 39, 72-87.

Goldenberg , Lauren B.;Culp , Katherine McMillan;Clements , Margaret;Pasquale , Marian;,erson , Alice (2014). Online Professional Development for High-School Biology Teachers: Effects on Teachers' and Students' Knowledge Journal of Technology and Teacher Education, 22, 287-309.

Grimberg, Bruna Irene ;Gummer, Edith (2013). Teaching Science from Cultural Points of Intersection Journal of Research in Science Teaching, 50(1), 12-32.

Gurbuz , Hasan;Kisoglu , Mustafa;Erkol , Mehmet;Alas , Ali;Kahraman , Sakip;Uzunboylu , H. (2010). The effect of PowerPoint presentations prepared and presented by prospective teachers on biology achievement and attitudes toward biology Innovation and Creativity in Education, 2, 3043-3047.

Hanegan, Nikki ;Friden, Kelley ;Nelson, C. Riley (2009). Authentic and Simulated Professional Development: Teachers Reflect What is Modeled School Science and Mathematics, 109(2), 79-94.

Harlen, W. ;Doubler, S. J. (2004). Can teachers learn through enquiry on-line? Studying professional development in science delivered on-line and on-campus International Journal of Science Education, 26(10), 1247-1267.

Harlen, W., & Altobello, C. (2003). An Investigation of "Try Science" Studied On-line and Face-to-face. Cambridge, MA: TERC - CESSE.

Hartshorne , Richard (2005). Effects of Integrating Hypermedia into Elementary Science Professional Development on Science Content Knowledge Journal of Science Education and Technology, 14, 415-424.

Hartshorne , Richard (2008). Effects of Hypermedia-Infused Professional Development on Attitudes toward Teaching Science Journal of Educational Computing Research, 38, 333-351.

Heller , Joan I. (2012). Effects of Making Sense of SCIENCE[TM] Professional Development on the Achievement of Middle School Students, Including English Language Learners. Final Report. NCEE 2012-4002. : .

Heller , Joan I.;Shinohara , Mayumi;Miratrix , Luke;Hesketh , Sophia Rabe;Daehler , Kirsten R. (2010). Learning Science for Teaching: Effects of Professional Development on Elementary Teachers, Classrooms, and Students , , .

Heller, Joan I. ;Daehler, Kirsten R. ;Wong, Nicole ;Shinohara, Mayumi ;Miratrix, Luke W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science Journal of Research in Science Teaching, 49(3), 333-362.

Henrichs , Lotte F.;Leseman , Paul P. M. (2014). Early Science Instruction and Academic Language Development Can Go Hand in Hand. The Promising Effects of a Low-Intensity Teacher-Focused Intervention International Journal of Science Education, 36, 2978-2995.

Herman, J., Osmundson,E., Dai, Y., Ringstaff, C., & Timms, M. (2015). Investigating the dynamics of formative assessment: relationships between teacher knowledge, assessment practice and learning. Assessment in Education, 22(3), 344-367.

*Higgins, J.P.T., & Green, S. (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Hrepic , Zdeslav;Reed , RH Berque , DA Prey , JC (2009). Deploying Tablet PCs and DyKnow Software in Teaching an Inquiry-Based Physical Science Course Unpublished manuscript. .

Hrepic, Zdeslav (2008). Deploying Tablet PCs and DyKnow software in teaching inquiry-based physical science course
In (Ed.), The Impact of Tablet PCs and Pen-based Technology on Education: Evidence and Outcomes
(pp. 55-62). : Purdue University Press.

Hypolite, Christine Collins (2003). The impact of whole-plant instruction of preservice teachers' understanding of plant science principles (Doctoral dissertation, Louisiana State University , Agricultural & Mechanical College). Dissertation Abstracts International, , .

Izzet Kurbanoglu , N.;Taskesenligil , Y.;Sozbilir , M. (2006). Programmed instruction revisited: A study on teaching stereochemistry Chemistry Education Research and Practice, 7, 13-21.

Jackson, D. C. (2005). The effect of concept mapping on preservice elementary teachers' knowledge of science inquiry teaching (Doctoral dissertation, Syracuse University). Dissertation Abstracts International, , .

Johnson , Carla C.;Fargo , Jamison D. (2010). Urban School Reform Enabled by Transformative Professional Development: Impact on Teacher Change and Student Learning of Science Urban Education, 45, 4-29.

Jung, M. L. (2004). Out-of-school science-teaching preparation for elementary teacher-education students (Doctoral dissertation, Wayne State University). Dissertation Abstracts International, , .

Karpudewan , Mageswary;Ismail , Zurida;Roth , Wolff-Michael (2012). Fostering Pre-service Teachers' Self-Determined Environmental Motivation Through Green Chemistry Experiments Journal of Science Teacher Education, 23, 673-696.

Karpudewan, Mageswary; Ismail, Zurida; Roth, Wolff-Michael (2012). Ensuring sustainability of tomorrow through green chemistry integrated with sustainable development concepts (SDCs) Chemistry Education Research and Practice, 13, 120-127.

Karpudewan, Mageswary; Ismail, Zurida; Roth, Wolff-Michael (2012). Promoting pro-environmental attitudes and reported behaviors of Malaysian pre-service teachers using green chemistry experiments Environmental Education Research, 18(3), 375-389.

Kleickmann , T.;Trobst , S.;Jonen , A.;Vehmeyer , J.;Moller , K. (2016). The Effects of Expert Scaffolding in Elementary Science Professional Development on Teachers' Beliefs and Motivations, Instructional Practices, and Student Achievement Journal of educational psychology, 108(1), 21-42.

Kocak , Canan (2013). The Effects of Process-Based Teaching Model on Student Teachers' Logical/intuitive Thinking Skills and Academic Performances Journal of Baltic Science Education, 12, 640-651.

Krall, Rebecca McNall, . Straley, Joseph PShafer, Sally A. & Osborn , Jeffrey L (2009). Impact of inquiry based distance learning and availability of classroom materials on physical science content knowledge of teachers and students in central Appalachia Journal of Science Education and Technology, 18, 173-186.

Lee , Eunmi;Brown , Michelle N.;Luft , Julie A.;Roehrig , Gillian H. (2007). Assessing Beginning Secondary Science Teachers' PCK: Pilot Year Results School Science and Mathematics, 107, 52-.

Lee , Luann Christensen (2011). Under One Big Sky: Elementary Pre-Service Teachers Use Inquiry to Learn about the Moon, Construct Knowledge, and Teach Elementary Students around the World via the Internet (Doctoral dissertation, Ball State University). Dissertation Abstracts International, , .

Liang , Ling L.;Richardson , Greer M. (2009). Enhancing Prospective Teachers' Science Teaching Efficacy Beliefs through Scaffolded, Student-Directed Inquiry Journal of Elementary Science Education, 21, 51-66.

Liang, L. L. ;Gabel, D. L. (2005). Effectiveness of a constructivist approach to science instruction for prospective elementary teachers International Journal of Science Education, 27(10), 1143-1162 .

Lin, H. S. ;Chen, C. C. (2002). Promoting preservice chemistry teachers' understanding about the nature of science through history Journal of Research in Science Teaching, 39(9), 773-792.

Logerwell , Mollianne G. (2009). The Effects of a Summer Science Camp Teaching Experience on Preservice Elementary Teachers' Science Teaching Efficacy, Science Content

Knowledge, and Understanding of the Nature of Science (Doctoral dissertation, George Mason University, 2009). Dissertation Abstracts International, PhD, .

Lusk , M. G.;Bickmore , B. R.;Christiansen , E. H.;Sudweeks , R. R. (2006). Use of a mentored creative writing project to improve the geology education of preservice elementary teachers Journal of Geoscience Education, 54, 31-40.

Maerten-Rivera , Jaime Lynn;Huggins-Manley , Anne Corinne;Adamson , Karen;Lee , Okhee;Llosa , Lorena (2015). Development and Validation of a Measure of Elementary Teachers' Science Content Knowledge in Two Multiyear Teacher Professional Development Intervention Projects Journal of Research in Science Teaching, 52, 371-396.

Martin-Dunlop , Catherine;Hodum , Peter (2009). Scientist-Science Educator Collaborations: Do They Improve Students' Understanding of the Nature of Science? Journal of College Science Teaching, 39, 66-75.

Mazzie, D. D. (2008). The effects of professional development related to classroom assessment on student achievement in science (Doctoral dissertation, University of South Carolina). Dissertation Abstracts International, PhD, . (University Microfilms No. 3321420)

McGinnis, J. Randy,Kramer, Steve, Shama, Gilli, Graeber, Anna O. , Parker, Carolyn A., Watanabe, Ted (2002). Undergraduates' Attitudes and Beliefs about Subject Matter and Pedagogy Measured Periodically in a Reform-Based Mathematics and Science Teacher Preparation Program Journal of Research in Science Teaching, 39(8), 713-737.

Michalsky , Tova;Kramarski , Bracha (2015). Prompting Reflections for Integrating Self-Regulation Into Teacher Technology Education Teachers College Record, 117, 5.

Mintzes , Joel J.;Marcum , Bev;Messerschmidt-Yates , Christl;Mark (2013). Enhancing Self-Efficacy in Elementary Science Teaching with Professional Learning Communities Journal of Science Teacher Education, 24, 1201-1218.

Minuskin , Sondra (2009). Effects of Professional Development on the Knowledge and Classroom Practices of Elementary School Science Teachers (Doctoral dissertation, Nova Southeastern University). Dissertation Abstracts International, , . (University Microfilms No. 3385691)

Moseley , Christine;Utley , Juliana (2006). The Effect of an Integrated Science and Mathematics Content-Based Course on Science and Mathematics Teaching Efficacy of Preservice Elementary Teachers Journal of Elementary Science Education, 18, 1-12.

Moseley , Christine;Utley , Juliana (2008). An Exploratory Study of Preservice Teachers' Beliefs About the Environment Journal of Environmental Education, 39, 15-30.

Mutch-Jones , Karen;Puttick , Gillian;Minner , Daphne (2012). Lesson Study for Accessible Science: Building Expertise to Improve Practice in Inclusive Science Classrooms Journal of Research in Science Teaching, 49, 1012-1034.

Mutlu , Mehmet (2013). Effect of using Roundhouse Diagrams on Preservice Teachers' Understanding of Ecosystem Journal of Baltic Science Education, 12, 205-218.

Nadelson, L. S. (2007). Preservice teachers' understanding of evolution, the nature of science, and situations of chance (Doctoral dissertation, University of Nevada , Las Vegas). Dissertation Abstracts International, , .

Nugent , Gwen;Tol, , Michael D.;Levy , Richard;Kunz , Gina;Harwood , David;Green , Denise;Kitts , Kathy (2012). The Impact of an Inquiry-Based Geoscience Field Course on Pre-Service Teachers Journal of Science Teacher Education, 23, 503-529.

Ogan-Bekiroglu , Feral;Arslan , Arzu;Laborda , JG (2014). Examination of the Effects of Model-Based Inquiry on Students' Outcomes: Scientific Process Skills and Conceptual Knowledge. Paper presented at the meeting of the , .

Oskay , Ozge Ozyalcin;Dincol , Sinem (2011). The Effects of Internet-Assisted Chemistry Applications on Prospective Chemistry Teachers' Cognitive Structure. Paper presented at the meeting of the , .

Osmundson , Ellen;Dai , Yunyun;Herman , Joan (2011). Year 3 ASK/FOSS Efficacy Study. CRESST Report 782 , , 782.

Ozdilek , Zehra (2015). Teaching the properties of chromium's oxidation states with a case study method Chemistry Education Research and Practice, 16, 39-52.

Pekbay , Canay;Yilmaz , Serkan (2015). The Effect of Explicit-Reflective and Historical Approach on Preservice Elementary Teachers' Views of Nature of Science International Journal of Progressive Education, 11, 113-131.

Polanin, J. R., Tanner-Smith, E. E., & Henessy, E. A. (2016). Estimating the Difference Between Published and Unpublished Effect Sizes: A Meta-Review. *Review of Educational Research, 86*(1), 207 – 236.

Qing , Zhou; Jing , Guo; & Yan , Wang (2010). Promoting preservice teachers' critical thinking skills by inquiry-based chemical experiment Procedia Social and Behavioral Sciences, 2, 4597-4603.

Qing, Zhou;Jing, Guo ; Yazhuan, Liu; Ting, Wang ;Junping, Ma (2010). Promoting preservice teachers' critical thinking Disposition by inquiry-based chemical experiment Procedia Social and Behavioral Sciences, 9, 1429-1436.

Reinfried , S. (2006). Conceptual change in physical geography and environmental sciences through mental model building: The example of groundwater International Research in Geographical and Environmental Education, 15, 41-61.

Rhoton, Jack ;McLean, James E (2008). Developing Teacher Leaders in Science: Catalysts for Improved Science Teaching and Student Learning Science Educator, 17(2), 45-56.

Riegle-Crumb , Catherine;Morton , Karisma;Moore , Chelsea;Chimonidou , Antonia;Labrake , Cynthia;Kopp , Sacha (2015). Do Inquiring Minds Have Positive Attitudes? The Science Education of Preservice Elementary Teachers Science Education, 99, 819-836.

Roth, K. J. ;Garnier, H. E. ;Chen, C. ;Lemmens, M. ;Schwille, K. ;Wickler, N. I. Z. (2011). Videobased Lesson Analysis: Effective Science PD for Teacher and Student Learning Journal of Research in Science Teaching, 48(2), 117-148.

Ruhf, R. J. (2006). Analyzing the effects of inquiry-based instruction on the learning of atmospheric science among pre-service teacher education students (Doctoral dissertation, Western Michigan University). Dissertation Abstracts International, , .

Sampson, Victor ;Enderle, Patrick ;Grooms, Jonathon ;Witte, Shelbie (2013). Writing to learn by learning to write during the school science laboratory: Helping middle and high school students develop argumentative writing skills as they learn core ideas Science Education, 97(5), 643-670.

Sang , Guoyuan;Valcke , Martin;van Braak , Johan;Zhu , Chang;Tondeur , Jo;Yu , Kailian (2012). Challenging science teachers' beliefs and practices through a video-case-based intervention in China's primary schools Asia-Pacific Journal of Teacher Education, 40, 363-378.

Sasser, S. K. (2014). Effect of structure in problem based learning on science teaching efficacy beliefs and science content knowledge of elementary preservice teachers (Doctoral dissertation, Southern Illinois University at Carbondale). Dissertation Abstracts International, , .

Sato , M.;Wei , R. C.;Darling-Hammond , L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification American Educational Research Journal, 45, 669-700.

Scarlett, T., III (2008). An exploratory study of the impact of two versions of inquiry-based science program professional development on teachers' perceptions of their Pedagogical Content Knowledge (Doctoral dissertation, University of Hawai'i at Manoa). Dissertation Abstracts International, , .

Schaal , Steffen;Matt , Monica;Gruebmeyer , Sonja;Baskan , GA Ozdamli , F Kanbul , S Ozcan , D. (2012). Mobile Learning and biodiversity - bridging the gap between outdoor and inquiry learning in pre-service science teacher education. Paper presented at the meeting of the , .

Selçuk , G. S.; Çalışkan , S. (2010). A small-scale study comparing the impacts of problem-based learning and traditional methods on student satisfaction in the introductory physics course Procedia - Social and Behavioral Sciences, 2(2), 809-813.

Selcuk , Gamze Sezgin;Caliskan , Serap;Uzunboylu , H. (2010). A small-scale study comparing the impacts of problem-based learning and traditional methods on student satisfaction in the introductory physics course Innovation and Creativity in Education, 2, 809-813.

Selçuk, G. S. (2010). The effects of problem-based learning on pre-service teachers' achievement, approaches and attitudes towards learning physics International Journal of Physical Sciences, 5, 711-723.

Selcuk, G. S. ;Sahin, M. ;Acikgoz, K. U. (2011). The Effects of Learning Strategy Instruction on Achievement, Attitude, and Achievement Motivation in a Physics Course Research in Science Education, 41(1), 39-62.

Sendur , Gulten;Toprak , Mustafa (2013). The role of conceptual change texts to improve students' understanding of alkenes Chemistry Education Research and Practice, 14, 431-449.

Senocak , Erdal;Taskesenligil , Yavuz;Sozbilir , Mustafa (2007). A Study on Teaching Gases to Prospective Primary Science Teachers through Problem-Based Learning Research in Science Education, 37, 279-290.

Sever , Songul;Oguz-Unver , Ayse;Yurumezoglu , Kemal (2013). The Effective Presentation of Inquiry-Based Classroom Experiments Using Teaching Strategies that Employ Video and Demonstration Methods Australasian Journal of Educational Technology, 29, 450-463.

Simsek , Umit (2012). Effects of two cooperative learning strategies on achievement in chemistry in undergraduate classes Energy Education Science and Technology Part B- Social and Educational Studies, 4, 901-912.

Sodervik, Ilona ; Mikkila-Erdmann, Mirjamaija ; Vilppu, Henna (2014). Promoting the Understanding of Photosynthesis among Elementary School Student Teachers through Text Design Journal of Science Teacher Education, 25, 581-600.

Staples, K. A. (2002). The effect of a nontraditional undergraduate science course on teacher and student performance in elementary science teaching (Doctoral dissertation, The University of Alabama.). Dissertation Abstracts International, , .

Stoddart , Trish;Bravo , Marco;Mosqueda , Eduardo;Solis , Jorge (2013). Restructuring Pre-Service Teacher Education to Respond to Increasing Student Diversity Research in Higher Education Journal, 19, .

Tasdelen , Ugur;Koseoglu , Fitnat (2008). Learner-friendly textbooks: Chemistry texts based on a constructivist view of learning Asia Pacific Education Review, 9, 136-147.

Taslidere , Erdal (2015). A Study Investigating the Effect of Treatment Developed by Integrating the 5E and Simulation on Pre-Service Science Teachers' Achievement in Photoelectric Effect EURASIA Journal of Mathematics, Science & Technology Education, 11, 777-792.

Taslidere, Erdal ;Eryilmaz, Ali (2012). The Relative Effectiveness of Integrated Reading Study Strategy and Conceptual Physics Approach Research in Science Education, 42(2), 181-199.

Tasoglu , Aslihan Kartal;Bakac , Mustafa;Uzunboylu , H. (2010). The effects of problem based learning and traditional teaching methods on students' academic achievements, conceptual developments and scientific process skills according to their graduated high school types Innovation and Creativity in Education, 2, 2409-2413.

Tatar , Erdal (2011). The effect of guided inquiry and open inquiry methods on teacher candidates' science process skills Energy Education Science and Technology Part B- Social and Educational Studies, 3, 669-680.

Temel , Senar (2014). The effects of problem-based learning on pre-service teachers' critical thinking dispositions and perceptions of problem-solving ability South African Journal of Education, 34, 769-769.

Tessier , Jack (2010). An Inquiry-Based Biology Laboratory Improves Preservice Elementary Teachers' Attitudes about Science Journal of College Science Teaching, 39, 84-90.

Thompson, R. (2003). The effect of a laboratory-based, in-context, constructivist teaching approach on preservice teachers' science knowledge and teaching efficacy (Doctoral dissertation, University of North Texas). Dissertation Abstracts International, Ed.D, . (University Microfilms No. 3095382)

Tosun , Cemal;Taskesenligil , Yavuz (2013). The Effect of Problem-Based Learning on Undergraduate Students' Learning about Solutions and Their Physical Properties and Scientific Processing Skills Chemistry Education Research and Practice, 14, 36-50.

Trimmell, M. D. (2015). The effects of stem-rich clinical professional development on elementary teachers' sense of self-efficacy in teaching science (Doctoral dissertation, California State University , Long Beach). Dissertation Abstracts International, , .

Trumper, R. (2006). Teaching future teachers basic astronomy concepts - Seasonal changes - At a time of reform in science education Journal of Research in Science Teaching, 43(9), 879-906.

Trundle, K. C. ;Atwood, R. K. ;Christopher, J. E. (2002). Preservice elementary teachers' conceptions of moon phases before and after instruction Journal of Research in Science Teaching, 39(7), 633-658 .

Turacoglu , Ilker;Alpat , Senol;Ellez , A. Murat (2013). Effects of Jigsaw on Teaching Chemical Nomenclature Egitim Ve Bilim-Education and Science, 38, 256-272.

Tural , G. (2013). The functioning of context-based physics instruction in higher education Asia-Pacific Forum on Science Learning and Teaching, 14(1), 1-23.

Tural , Guner (2014). The Functioning of Context-Based Physics Instruction in Higher Education Asia-Pacific Forum on Science Learning and Teaching, 14, .

Tuysuz , Cengiz;Demirel , Ozan Emre;Yildirim , Bilal;Odabasi , HF (2013). Investigating the effects of argumentation, problem and laboratory based instruction approaches on pre-service teachers' achievement concerning the concept of 'acid and base'. Paper presented at the meeting of the , .

Ucar , Sedat;Trundle , Kathy Cabe (2011). Conducting Guided Inquiry in Science Classes Using Authentic, Archived, Web-Based Data Computers & Education, 57, 1571-1582.

Ucar, S. ;Trundle, K. C. ;Krissek, L. (2011). Inquiry-based Instruction with Archived, Online Data: An Intervention Study with Preservice Teachers Research in Science Education, 41(2), 261-282 .

Ultay, N., Durukana, G. U., & Ultayb, E. (2015). Evaluation of the effectiveness of conceptual change texts in the REACT strategy Chemistry Education Research and Practice, 16, 22-38.

Unal , Ibrahim;Okur , Nilufer;Kapucu , Serkan;Uzunboylu , H. (2010). The effect of using animations on pre-service science teachers' science achievement Innovation and Creativity in Education, 2, 5357-5361.

van Aalderen-Smeets , S,ra I.;van der Molen , Juliette H. (2015). Improving Primary Teachers' Attitudes toward Science by Attitude-Focused Professional Development Journal of Research in Science Teaching, 52, 710-734.

Walker , ,rew;Recker , Mimi;Ye , Lei;Robertshaw , M. Brooke;Sellers , Linda;Leary , Heather (2012). Comparing technology-related teacher professional development designs: a multilevel study of teacher and student impacts Etr&D-Educational Technology Research and Development, 60, 421-444.

Weakley , Kevin Douglas (2010). The Effects of an Inquiry-Based Earth Science Course on the Spatial Thinking of Pre-Service Elementary Teacher Education Students (Doctoral dissertation, Western Michigan University). Dissertation Abstracts International, , . (University Microfilms No. 3410420)

Willden , Jennifer L.;Crowther , David T.;Gubanich , Alan A.;Cannon , John R. (2002). A Quantitative Comparison of Instruction Format of Undergraduate Introductory Level Content Biology Courses: Traditional Lecture Approach vs. Inquiry Based for Education Majors. Paper presented at the meeting of the Annual International conference for the Association for the Education of Teachers in Science, Charlotte, NC.

Willden, J. W. (2002). Introductory biology instruction method for elementary education majors: Inquiry vs. traditional approach (Doctoral dissertation, University of Nevada , Reno). Dissertation Abstracts International, , .

Ye, L. (2013). Integrating technology, curriculum, and online resources: A multilevel model study of impacts on science teachers and students (Doctoral dissertation, Utah State University). Dissertation Abstracts International, , .

Yildiz , Ali (2012). Prospective Teachers' Comprehension Levels of Special Relativity Theory and the Effect of Writing for Learning on Achievement Australian Journal of Teacher Education, 37, .

Yildiz , Ali; Buyukkasap , Erdogan (2011). The Level of Understanding of the Photoelectric Phenomenon in Prospective Teachers and the Effects of 'Writing with Learning' on their Success Rates Educational Sciences: Theory & Practice, 11(4), 2268-2274.

Yoon , Heojeong;Woo , Ae Ja;Treagust , David;Ch,rasegaran , A. L. (2014). The Efficacy of Problem-based Learning in an Analytical Laboratory Course for Pre-service Chemistry Teachers International Journal of Science Education, 36, 79-102.

Zacharia , C. Z. (2007). Comparing and Combining Real and Virtual Experimentation: An Effort to Enhance Students' Conceptual Understanding of Electric Circuits Journal of Computer Assisted Learning, 23, 120-132.

Zacharia, Z. (2003). Beliefs, attitudes, and intentions of science teachers regarding the educational use of computer simulations and inquiry-based experiments in physics Journal of Research in Science Teaching, 40(8), 792-823.

Zacharia, Z. C. (2002). The effects of an interactive computer-based simulation prior to performing a laboratory inquiry-based experiment on science teachers' conceptual understanding of physics (Doctoral dissertation, Columbia University). Dissertation Abstracts International, PhD, . (University Microfilms No. 3039662)

Zhang, D. (2007). EVALUATION OF THE INTEGRATED EXPERIENTIAL LEARNING PROGRAM OF SCIENCE IN CHINA (Doctoral dissertation, UTAH STATE UNIVERSITY). Dissertation Abstracts International, M. S., . (University Microfilms No. 1454880)

Zhang, Danhui ;Campbell, Todd (2012). An Exploration of the Potential Impact of the Integrated Experiential Learning Curriculum in Beijing, China International Journal of Science Education, 34(7), 1093-1123.

Title: Estimates of IntraClass Correlation and Outcome-Covariate Correlations for Teacher Outcomes in Evaluations of Math and Science Interventions

Presenting Author: Fatih Unlu, RAND

Authors: Fatih Unlu, RAND; Joe Taylor, BSCS; Jessaca Spybrook, WMU; Carl Westine, UNCC; and Brent Anderson, RAND

Background:

The past fifteen years have seen a dramatic increase in the use of cluster randomized trials (CRTs) to evaluate educational interventions. In a CRT, entire clusters (such as schools) are randomly assigned to the study conditions. In order to yield rigorous evidence on program effectiveness, a CRT must have a strong design, implementation, and analysis. This paper focuses on one of these components, the design of CRTs, and specifically improving the power analyses. The research on power analyses highlights the importance of using accurate, relevant, and reliable design parameters noting that the power analysis is only as good as the design parameters. This paper focuses on two specific design parameters: the percent of variance in the outcome measures between clusters, the intraclass correlation (ICC), and the percent of variance explained by covariate(s) (R^2).

Prior work on design parameter estimates suggests that there is substantial variation in these parameters across outcome domains, measures within a domain, subject matters, and grade-levels (e.g., Bloom, Richburg-Hayes, & Black, 2007; Hedges & Hedberg, 2013; Phelps, Kelcey, Jones, & Liu, 2016; and Spybrook, Westine, & Taylor, 2016). While the literature covers student achievement quite well, only a few studies provide estimates for teacher outcomes such as use of research based instructional practices, content knowledge, and pedagogical content knowledge. The evidence is sparser for teacher outcomes in math and science than reading. This is a critical limitation for educational evaluations because teacher outcomes are at least as important as student outcomes. For example, the interventions in the early development stage may choose to initially focus on teacher-level outcomes rather than student-level outcomes because the former are more proximal to the intervention. Therefore, examining the former would give an early indication of the potential effects on the latter and may greatly inform and enhance the intervention development.

Purpose:

Education researchers designing studies to evaluate interventions with teacher outcomes in math and science often have to borrow design parameter estimates from other measures and subject—setting up a high likelihood for inappropriately powered studies, inconclusive results, and stagnation of knowledge in science and math education research. This paper addresses this limitation by conducting secondary analyses with 8 existing educational datasets that have at least one teacher outcome in math or science and reporting estimates of ICC and R^2 values. One of these datasets is from the Measures of Effective Teaching (MET) study and the other seven are from relatively large and federally funded studies[1]. We have selected these data sets because they provide us the unique opportunity to track outcomes from both national and regional samples of individual teachers over time (i.e., for baseline-outcome R^2 analysis), and contain a grouping variable for estimating cluster effects (i.e., for ICCs). Further, these data sets include data on other teacher characteristics (e.g., content knowledge, years of experience) for use in

[1] One study is funded by NSF and the others are funded by the Institute of Education Sciences.

other R^2 analyses. The combined samples of these datasets include over 6,000 teachers in more than 1,000 schools.

We currently have access to all datasets and have completed preliminary analyses of two. We will complete the analyses of the remaining datasets by the time of the conference and include the corresponding results in our presentation should our proposal be accepted.

Method:

The design parameters are estimated using 2-level hierarchical linear models (HLMs) with teachers nested within schools.[2] First, we estimate an unconditional HLM specified as:

Level 1 (teacher): $Y_{ij} = \alpha_{0j} + e_{ij},\ e_{ij} \sim N(0, \sigma^2)$ [1]

Level 2 (school): $\alpha_{0j} = \gamma_{00} + \lambda_{0j},\ \lambda_{0j} \sim N(0, \upsilon^2)$ [2]

where Y_{ijk} is the outcome (e.g., instructional practices) for teacher i in school j; e_{ij} is the error associated with teachers and σ^2 is the teacher-level variance; and λ_{0j} is the error associated with schools and υ^2 is the school-level variance. The ICC coefficient is given by:

$$ICC_{L2} = \frac{\upsilon^2}{\sigma^2 + \upsilon^2} \quad [3]$$

To calculate the variance explained by covariates, the unconditional model is modified to include teacher and school-level covariates. We examine three covariate sets: (1) baseline version of the outcome (i.e., pretest) only; (2) teacher and school demographic characteristics only; and (3) full covariate set including the pretest and all other teacher and school-level baseline characteristics. We define two R^2 terms that capture the proportion of outcome variance at the teacher and school level explained by the covariates:

$$R^2_{L1} = \frac{\sigma^2 - \sigma^2_{|X}}{\sigma^2} \text{ and } R^2_{L2} = \frac{\upsilon^2 - \upsilon^2_{|X,W}}{\upsilon^2} \quad [4]$$

where **X** and **W** are vectors containing teacher and school-level covariates and $\sigma^2_{|X}$ and $\upsilon^2_{|X,W}$ are conditional versions of the teacher and school-level variance terms. We also report standard errors of the ICC and R2 estimates using formulae provided by Siddiqui et al. (1996) and Olkin and Finn (1995).

[2] We also estimate 3-level models – teachers nested within schools nested within districts – with datasets that have sufficiently large number of schools per district.

Results:

Table 1 shows the ICC and R2 estimates from the preliminary analyses. The estimates vary substantially across studies and across outcomes within each study. This variation has implications for statistical power, minimum detectable effect sizes, and sample size requirements. Specifically, the ICC values in the first dataset are considerably larger than those in the second dataset; therefore, using the former would lead to larger sample size requirements than the latter. One factor underlying this variation could be the different levels of heterogeneity of teachers and schools across the two datasets.

The observed variation across datasets highlights the importance of calculating design parameters with samples from different settings and locations and providing contextual and descriptive information about the samples so that the users of the estimated design parameters could choose the values for their designs accordingly. The completed version of the paper will provide this information, examine factors that may explain the variation in the design parameter estimates, and assess the generalizability of the parameter estimates using the approach described in Jacob, Zhu, and Bloom (2010).

Table 1. Estimated Values of ICC and R^2 Coefficients

Measure	Subject	Dataset	ICC		Level 1 (Teacher) R2		Level 2 (School) R2	
			Estimate	Std Error	Estimate	Std Error	Estimate	Std Error
Instructional Practices	Science	Science Teachers Learning through Lesson Analysis1	0.57	0.09	0.01	0.002	0	-
Content Knowledge	Science		0.3	0.12	0.58	0.042	0.46	0.057
Pedalogical Content Knowledge	Science		0.43	0.11	0.08	0.013	0	-
Mathematical Quality of Instruction	Math	Measures of Effective Teaching	0.03	0.03	0.15	0.009	0.97	0.000
Quality of Science Teaching	Science		0.11	0.32	N.A.	N.A.	N.A	N.A

Notes: [1] For all outcomes, the sample includes 133 teachers in 77 schools. R^2 values are from conditional models that control for baseline measure of the outcome. [2] The sample includes 767 teachers in 225 schools for the math outcome and 45 teachers in 33 schools for the science outcome. R^2 values are from conditional models that control for baseline measure of the outcome.

References

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using Covariates to Improve Precision for Studies That Randomize Schools to Evaluate Educational Interventions. Educational Evaluation and Policy Analysis, 29(1), 30–59.

Hedges, L. V., & Hedberg, E. C. (2013). Intraclass Correlations and Covariate Outcome Correlations for Planning Two- and Three-Level Cluster-Randomized Experiments in Education. Evaluation Review, 37(6), 445–489.

Jacob, R., Zhu, P., & Bloom, H. (2010). New Empirical Evidence for the Design of Group Randomized Trials in Education. Journal of Research on Educational Effectiveness, 3(2), 157–198.

Olkin, I., & Finn, J. D. (1995). Correlations redux. Psychological Bulletin, 118(1), 155–164.

Phelps, G., Kelcey, B., Jones, N., & Liu, S. (2016). Informing Estimates of Program Effects for Studies of Mathematics Professional Development Using Teacher Content Knowledge Outcomes. Evaluation Review, 40(5), 383–409.

Spybrook, J., Westine, C. D., & Taylor, J. A. (2016). Design Parameters for Impact Research in Science Education. AERA Open, 2(1), 23-32.

Title: An Investigation of Design and Statistical Power for Planning Cluster Randomized Trials with Student and Teacher Outcomes

Presenting Author: Qi Zhang, Western Michigan University

Authors: Qi Zhang and Jessaca Spybrook, Western Michigan University; Fatih Unlu, RAND

Background: Cluster randomized trials (CRTs) are commonly conducted to assess the efficacy of educational interventions, which are often implemented at the school level. CRTs with schools as the natural unit of random assignment often have teachers nested within schools and students nested within teachers. Further, it is often the case that the outcomes of interest are at both the student- and the teacher-level. For instance, CRTs designed to evaluate the efficacy of teacher professional development (PD) programs are interested in determining the effect of PD programs on teacher content knowledge and teacher practice, as well as their impact on student achievement. Since design parameters used to calculate the statistical power, such as the intra-class correlation coefficients (ICCs) and the outcome-covariate correlations (R^2 coefficients) tend to be different for the teacher and the student outcomes, a single power analysis is generally not sufficient to assess whether the study is adequately powered to detect effects for teachers and students. Therefore, researchers need to conduct two power analyses—one for teacher outcomes and one for student outcomes—and determine sample size requirements according to both analyses, as a design that is adequately powered to detect a meaningful impact at the student level is not necessarily powered to detect a meaningful impact at the teacher level and vice versa.

Purpose: This paper examines design considerations for studies that seek to evaluate the effectiveness of educational interventions for both teachers and students within one study. Specifically, this paper incorporates new empirical work presented in this symposium on design parameters for planning CRTs with student and teacher outcomes to provide insights for CRT designs of K-12 science educational interventions. The goal of this paper is to estimate the statistical power of these studies, examine the alignment of the power analyses when a study seeks to examine both teacher and student effects, and suggest considerations that can be incorporated into future planning to maximize the efficiency of the study design.

Method: We calculated statistical power, represented by the minimum detectable effect size (MDES) for CRTs that examine interventions that seek to improve student science achievement, as well improvements in science teacher content knowledge, teacher pedagogical content knowledge, and teacher practice. We assumed school as the unit of random assignment and allowed the total number of schools to vary from 25 to 65 schools. For the teacher level outcomes, we assume 5 teachers nested within each school, which is common within a particular grade in elementary schools. Assuming 25 students nested within each teacher, we set the number of students per school to 125 for the design to assess student outcomes. We compared MDES for studies with student and teacher outcomes based on best estimates of design parameters.

The power calculations for student outcomes are based on a 2-level Hierarchical Linear Model (HLM) that nested students within schools to estimate impacts. We ignored the teacher-level for this calculation, which is not expected to influence the calculated MDES (Zu, Jacob, Bloom, & Xu, 2012). Table 1 shows the design parameters used for this calculation, which are based on those reported by Westine, Spybrook, & Taylor (2014) for grade 5-11 science achievement outcomes. For the ICC and the school-level R^2, we used an upper and lower bound. We used a single value for the individual-level R^2. Different combinations of these parameter values lead to a range for the MDES estimates, which are discussed below.

The MDES calculations for teacher outcomes are based on a 2-level HLM that nests teachers within schools. The design parameters for these calculations were based on plausible values from the empirical analyses presented in paper 2 of this symposium for teacher practice and content knowledge outcomes.

Table 1. Empirical estimates used for the MDES calculations*.

	School-level ICC	**School-level (Level-2) R^2**	**Individual-level (Level-1) R^2**
Student Outcome	0.20, 0.26	0.50, 0.80	0.40
Teacher Outcome	0.15, 0.30	0.10, 0.20	0.25

*Calculations were based on these additional assumptions: two-tailed test, alpha = 0.05, equal allocation at all levels.

Equation (1) below shows the formula for calculating MDES for a 2-level CRT and we operationalized the formula in the program PowerUp! (Dong & Maynard, 2013).

$$MDES_{2LCRT} = M_{J-3}\sqrt{\frac{\rho(1-R_{L2}^2)}{P(1-P)J} + \frac{(1-\rho)(1-R_{L1}^2)}{P(1-P)Jn}} \qquad \text{Equation (1)}$$

Where n is the number of individuals (teachers or students) per cluster and J is the number of clusters. M is the multiplier for two-tailed test with $J - 3$ degrees of freedom. ρ is the ICC, which is the proportion of variance in outcome that is between clusters. R_{L1}^2 and R_{L2}^2 are the proportion of variance explained by level-1 and level-2 covariates, respectively. P is the proportion of level-2 units randomized to treatment.

Results: Figure 1 shows the MDES calculated based on student achievement outcome (dashed line) and teacher outcomes (solid line). As the number of schools approach 45, the range of MDES for student outcomes (0.18–0.31) is similar to the effect sizes for educational interventions noted by Hill, Bloom, Black, and Lipsey (2008) (0.14-0.24). For the same number of schools, note that the MDESs for teacher content knowledge outcomes was higher (0.43-0.52), due to the small number of teachers per school, the larger value of ICC, and the smaller school-level R^2. However, the range for teacher content knowledge outcome was consistent with the median effect size (0.52) observed for science teacher interventions in paper 1 of this symposium. As mentioned in paper 2 of this symposium, the final version of this paper will incorporate results from our analysis of a large number of additional existing datasets. Further, we will optimize our calculations with a larger set of parameters, including varying number of teachers and baseline data collection activities.

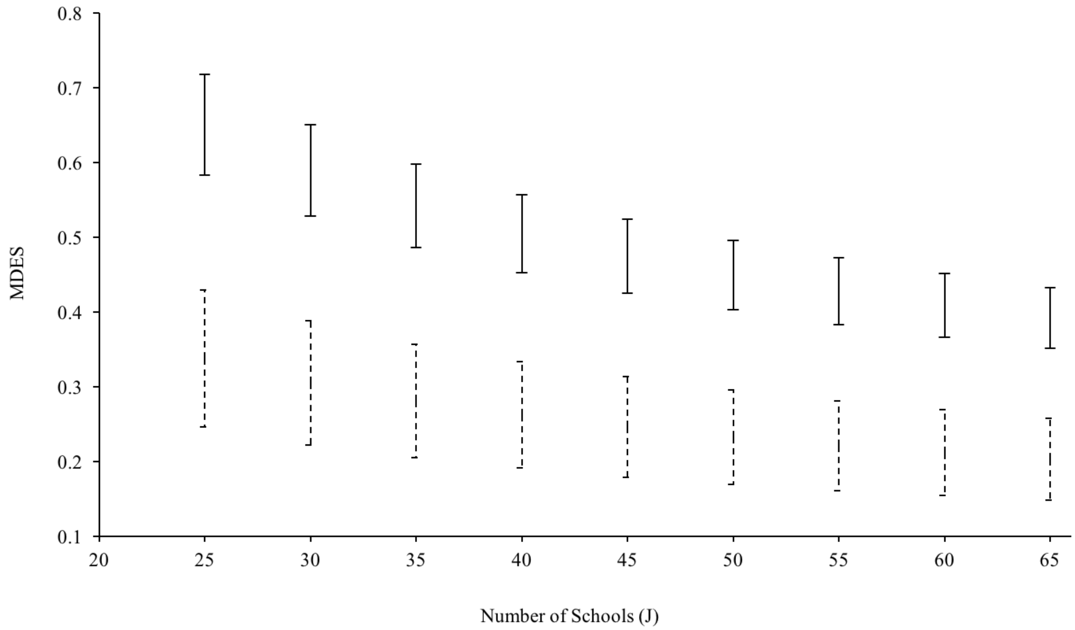


Figure 1. Calculated MDES based student achievement outcome (dashed line) and teacher content knowledge outcome (solid line), with varying number of schools.

Conclusion: Our result suggested studies that include at least 45 schools, 5 teachers per school, and 25 students per teacher may be able to detect a meaningful effect of the intervention for both students and teachers. This was possible because the larger effect size for teachers observed in the meta-analysis study (paper 1 of the symposium) compared to the effect size for students. Studies with less than 45 schools may be able to detect meaningful effects for students, but they may have some difficulties to achieve the same for teachers. This study will further extrapolate the results of power calculations for other teacher outcomes, as well as the implication of these results in planning CRTs that examine intervention effects both student and teacher.

References

Dong, N**.,** & Maynard, R. A. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, *6*(1), 24-67.

Hill, C. J.; Bloom, H. S.; Black, A. R.; & Lipsey, M. (2008). Empirical Benchmark for Interpreting Effect Sizes in Research. *Child Development Perspectives*, 2(3), 172-177.

Westine, C. D.; Spybrook, J.; & Taylor, J. A. (2013). An Empirical Investigation of Variance Design Parameters for Planning Cluster-Randomized Trials of Science Achievement. *Evaluation Review*, 37(6), 490-519.

Zhu, P.; Jacob, R.; Bloom, H.; & Xu, Z. (2012). Designing and Analyzing Studies that Randomized Schools to Estimate Intervention Effects on Student Achievement Outcomes without Classroom-level Information. *Educational Evaluation and Policy Analysis*, 34(1), 45-68.