ESERA 2019 Pre-Conference Workshop

1. **Title of Workshop**: Designing Adequately Powered Impact Studies for Science Education Interventions

2. **Contact Information**:

Joe Taylor

University of Colorado, Colorado Springs

Jtaylo18@uccs.edu

303.548.6648

3. **Requested maximum number of participants**: 30 participants
4. **Short description of relevant areas of expertise for each workshop facilitator**:

   Joseph Taylor, Ph.D., is former Director of Research and Development at BSCS and now faculty in statistics at the University of Colorado, Colorado Springs. His primary role includes conducting methodological research and consulting on the design of impact studies, including the use of *a priori* power analyses. Taylor was principal investigator of a NSF-funded meta-analysis of effect sizes, intraclass correlations, and covariate correlations for science education effectiveness trials. His work on estimating these three parameters has already improved impact study design in science education. In addition, Taylor was co-principal investigator and primary designer of several large-scale impact studies in science education including the NSF–funded effectiveness trial (DRK-12 Award # 0918277) and IES (US Dept. of Ed.) funded award (#R305K060142, 2006-12). All three studies were published, making both substantive and methodological contributions to the literature.

5. **Workshop abstract**:
Researchers often wish to be able to make causal links between an intervention and student or teacher outcomes. When researchers wish to make causal claims, they generally use either quasi-experimental or experimental designs comparing a treatment group to a comparison group. But how many participants does a researcher need for a given study? Accurately estimating the number of participants to detect an effect has huge implications for both the cost of a study and the responsible use of researcher dollars (often provided by taxpayers). Too many participants results in excessive costs, and too few participants can result in a failure to detect an important intervention effect. Accurately estimating the number of participants for a study is a critical and a sophisticated decision-making process, particularly when studying interventions in classroom contexts where students are nested within teacher, and teachers are nested within schools. To appropriately power a study, researchers need to know how to find and use a variety of power analysis parameters, including but not limited to an estimate of the expected effect size, an estimate of how similar clusters of participants are to one another (for example, how similar one class of students is to another class), and how much a researcher can expect to explain differences between groups using covariates such as pretest scores and demographic characteristics.

In this workshop, we will focus on how to design and calculate the statistical power for effectiveness studies of science interventions. The workshop will focus on a variety of effectiveness trial designs, including those with and without nesting at multiple levels. The principles learned will apply to both experiments and quasi-experiments and therefore cover a wide variety of effectiveness trial designs. Participants will learn how to use the *Optimal Design Plus* Software, a *freely available* and user-friendly program for calculating the power for effectiveness trials. We will also share our recent research findings related to empirical estimates of design parameters specifically for effectiveness trials related to science outcomes. The workshop will combine short presentations of information with multiple opportunities for discussion and hands-on practice using the *Optimal Design Plus* Software. The target audience for this workshop is science education researchers interested in planning and conducting effectiveness trials. It will be of particular interest to any researcher who would like to submit a proposal of an effectiveness trial for federal funding. Funding agencies are demanding increasingly sophisticated arguments that any effectiveness trial will include a sufficient number of participants in order to detect a treatment effect, should that effect exist. Participants will leave the workshop with the skills to make those nuanced arguments.

6. **Workshop description**:

Proposed Goals of the workshop:

At the conclusion of this workshop, participants will be able to define, describe, and identify statistical power and key elements of power analysis. In addition, participants will be able to map power analysis parameters to study designs, including those with and without nesting, as well as nesting at multiple levels (students nested within teachers, and teachers nested within schools). After helping participants build a foundation of understanding, participants will consider the benefits and limitations of various study designs, particularly from a power perspective. Last, participants will apply what they have learned to identify the appropriate power analysis parameters to be used in any given study; use the Optimal Design Plus Software to conduct a power analysis; interpret the power curves for a variety of designs from Optimal Design Plus Software; and finally, incorporate the results of a power analysis into a narrative format that would be compelling to reviewers.

Schedule and Activities:

Below is the proposed structure for the half day workshop assuming we meet from 9-12:30.

9:00-9:15 *Welcome and Introductions*: Participants will share who they are, a brief statement of what they know about power analysis, and what they hope to gain from the workshop.

9:15-9:45 *Session 1:* Discussion of Person-Level Experimental and Quasi-Experimental Trials for Science Intervention Studies: When are they useful and why?

Format: Whole Group and Small Group Discussion

Activities: Using examples from science education research literature, presenters will engage participants in a discussion of person-level causal effects studies, their basic statistical properties, and their affordances and limitations in science education research.

9:45-10:30 Session 2: Discussion of Cluster Randomized Trials for Science Intervention Studies

Format: Whole Group and Small Group Discussion

Activities: Using examples from science education research literature, presenters will engage participants in a discussion of cluster level causal effects studies in science education (e.g., students or teachers nested within schools), their basic statistical properties, and their affordances and limitations in science education research. Participants will also engage in identifying studies that used a nested design but did not account for nesting in analyses.

10:30 - 10:45 Break

10:45-11:45 Session 3: The Two-level Effectiveness Trial for Science Intervention Studies

Format: Short presentation on software use followed by small group work with *Optimal Design Plus*

Activities: Participants will work in small groups as they apply what they have learned about two-level effectiveness trials, including their basic statistical properties, strategic use of covariates, and power analysis parameters to conduct a power analysis using *OD Plus.*

11:45-12:15 Session 3: The Three-level Effectiveness Trial for Science Intervention Studies

Format: Short presentation followed by small group work with *OD Plus*

Activities: Presenters will share the nuances associated with moving from two-level to three-level designs and the ramifications for power analyses. Following the presentation, participants will apply what they have learned about three-level effectiveness trials, their basic statistical properties, and use of covariates to conduct a power analysis with *OD Plus*.

12:15-12:30 Session 5: Wrap up, Questions, Evaluation

Format: Discussion

Activities: Summary of designs and material, evaluation

Role of Workshop Participants and Facilitators:

The workshop facilitators will primarily play the role of facilitating guided discussions to help participants become familiar with a variety of statistical issues associated with power analysis. Prior to the workshop, facilitators will identify a variety of studies to use as examples in discussions. The facilitators will also actively assist participants during the small group work and hands-on activities. Participants will be engaged in reflecting on a variety of study designs, offering their ideas about these designs, as well as using information gained from discussions to conduct power analyses using *Optimal Design Plus* software.

Short Review of Relevant Literature:

Federal funding agencies have been placing increasing emphasis on the role of effectiveness trials in identifying interventions that enhance science teaching and learning. For example, since 2002 the US Dept. of Education's *Institute of Education Sciences* (IES) has funded more than 40 trials seeking the causal effects of math or science interventions. But effectiveness trials are generally very expensive, and funders want to know that taxpayer dollars will be well spent. Funders want coherent arguments that a study has neither too many nor too few participants. The latter case is also particularly costly, because an expensive intervention trial does not have the policy impact it could have had if additional participants had been included.

Calculating the statistical power for effectiveness trials in schools is complex because of the nested structure of the data: e.g. students nested in schools nested in districts. As a result, in the past 15 years there has been quite a bit of attention focused on how to calculate the statistical power for effectiveness trials where randomization occurs at the cluster (e.g., school or district) level (e.g. Bloom, 2005; Donner & Klar, 2000; Konstantopolous, 2008; Murray, 1998; Raudenbush, 1997; Raudenbush & Liu, 2000; Raudenbush, Martinez, & Spybrook, 2007; Schochet, 2005). Several important concepts have emerged from this body of work.

First, it is critical to correctly identify the unit of assignment to treatment groups (i.e., treatment or comparison). For example, is it a 2-level trial, such as students nested within schools and assignment at the school level, or is it a 3-level CRT with students nested within classrooms nested within schools and treatment assigned at the school level? Correctly identifying the design is critical since the parameters required for a power analysis differ depending on the design. Design identification will be an important part of the workshop.

A second key concept that has emerged from this research is the importance of good design parameters to use in the power analyses, noting that the power analysis is only as accurate as the design parameters. That is, if any of the design parameters are inaccurate, then too many or too few schools may be recruited resulting in higher costs or an underpowered trial. For example, suppose that a researcher planning a 2-level effectiveness trial with students nested within schools assumes that the intraclass correlation (ICC), or the between school variance is 0.05. A power analysis suggests that with 28 schools the study is powered at 0.80 to detect a difference of 0.20 standard deviations between treatment and control schools. However, suppose that in reality the ICC was 0.18. In this case, the 28 schools would not be sufficient to detect a difference of 0.20 standard deviations and the study would be underpowered.

Several research teams have conducted empirical investigations of design parameters for *reading* and *mathematics* outcomes including ICCs, covariate-outcome correlations, and effect sizes. Hedges and Hedberg (2007) used national longitudinal studies to determine ICC and R^2 values using schools as the cluster-level. Bloom, Richburg-Hayes, and Black (2007) used district reading and mathematics assessment data to calculate and compile ICC and R^2 values. Taylor received a grant from the US National Science Foundation in 2011 to conduct the first empirical investigation of design parameters for effectiveness trials of science outcomes. Findings from this work will be incorporated in the proposed workshop as information that participants can use when conducting their own power analyses.

Materials or Artifacts from the Workshop:

All workshop materials including PowerPoint slides and hand-on exercises will be provided for participants.