Learning as a Community: Maximizing the Impact of Research Syntheses in STEM Education

5 August 2014

Washington, DC

Presider: Joseph A Taylor, BSCS

BSCS

Panel Participants

Discussants

- Joseph Taylor - BSCS
- Christopher Wilson - BSCS

Panelists

- Alina Martinez – Abt Associates
- Erin Furtak – University of Colorado at Boulder
- Susan Kowalski - BSCS

Panel Focus Questions

- How do you view the role of research syntheses in advancing STEM education or other education research fields?
- In what ways are common practices the STEM education community facilitating and/or inhibiting the impact of research syntheses?
- What would you recommend education researchers do to maximize the validity, usefulness, and impact of research syntheses?

Agenda in Broad Strokes

- Goals and Introductions – 10 min
- Each panelist will speak for 10-15 minutes – 45 minutes
- Breakout groups by panelist - 20 minutes (please elect someone as a notetaker and reporter)
- Reports from the breakout groups – 20 minutes total (~6 minutes per group plus cross talk)
- Comments from the discussants plus large group discussion of general impressions – 25 minutes

Role of Research Syntheses in Education Research:
Insights from CADRE's Compendia of STEM Research Instruments

Alina Martinez
DR K-12 PI Meeting 2014

BOLD THINKERS DRIVING REAL-WORLD IMPACT

Community for Advancing Discovery Research in Education

NSF Grant No. DRL-0822241

Role of Syntheses in Advancing Education Research

- Potential to advance theory, practice, and methodology
- Facilitate the accumulation of knowledge that is being generated by individual (or team) efforts
- Contribute over and above the contributions of individual studies

Necessary Conditions May Include

- **Interest** on the part of the investigators that will conduct the synthesis
- **Opportunity** in the form of available resources
- **Accumulation** of a body of work
- **Access** to the information
- **Relevance/Utility** of the final product to a larger group

Systematic Reporting is Key

- Individual studies build foundation for syntheses
- Systematic reporting facilitates
  - Understanding of work
  - Better replication
  - Syntheses

CADRE’s Work on Instrumentation

- CADRE (Community for Advancing Discovery Research in Education)
- Purpose was to pull together information on available instruments
  - What are the instruments, constructs, and methods being used to study teacher outcomes?
  - What are the instruments, constructs, and methods being used to study student outcomes?
- Included multiple cohorts of NSF-funded DR K-12 grants
- Involved three phases of work.
  1. Review of project materials
  2. Search for instrument-specific information (reliability and validity evidence, development and piloting, accessibility of the instrument, administration, and variables measured)
  3. Fine-grained analysis of constructs measured and psychometric evidence
- Resulted in collection of instruments commonly used for gathering information about educational innovations

Conditions for this Work

- **Interest** – Interest in identifying areas where there was a need to develop measures, and where measures existed.
- **Opportunity** –CADRE’s charge includes looking across the work of individual DR K-12 projects
- **Accumulation** - Extant, named instruments as opposed to new instruments
- **Access** – Relied on information shared with CADRE or information that is publicly available.
- **Relevance/Utility** – Two compendia were produced.
  - Instruments to assess teacher practices, PCK, and content knowledge
  - Instruments to measure students’ content knowledge, reasoning skills, and psychological attributes.

Role of Community in Syntheses

- Individual investigation can be done in isolation; synthesis requires contributions of a community
- Accumulation, opportunity, and access are key areas where the community may facilitate or inhibit research syntheses;
  - **Accumulation -** Researchers in the community conduct the work that can be synthesized.
  - **Opportunity -** Funding and researchers' attention may need to be redirected from individual studies.
  - **Access -** Information researchers make available is critical.
- Additional roles that the community plays include dissemination and use.

CADRE Instruments Compendia

- Psychometric reporting practices limited the syntheses, as well as the utility of the aggregated work
  - Instruments that measure teacher constructs
    - 36% were missing information on reliability
    - 50% were missing information on validity
  - Instruments that measure student outcomes
    - 37% were missing information on reliability
    - 40% were missing information on validity

Maximizing Potential of Research Syntheses

- Funders may want to support novel research, while a synthesis may seem to be investigating what we already know (or assume we know)
- Consider what we individually and as a community do that affects research syntheses
- Syntheses require access to detailed information, while investigators may prefer to protect their intellectual property
- Make relevant details publically available
  - Report methodological detail
  - In this case psychometric information on the tools
- Could we reach consensus on what should be reported?

This work was conducted as part of the Community for Advancing Discovery Research in Education (CADRE). This material is based on work supported by the National Science Foundation under Grant No. DRL-0822241. Its contents are solely the responsibility of the authors and do not represent the official views of NSF.

Website: cadrek12.org

Contact: Alina Martinez
alina_martinez@abtassoc.com

Challenges in Developing Classroom Assessments Linked to Multidimensional Learning Progressions

Erin Marie Furtak

School of Education, University of Colorado at Boulder

NARST 2014, Pittsburgh

University of Colorado
Boulder

How do you view the role of research syntheses in advancing science education?

Review of Educational Research
September 2012, Vol. 82, No. 3, pp. 300–329
DOI: 10.3102/0034654312457206

Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis

Erin Marie Furtak
University of Colorado at Boulder

Tina Seidel
TUM School of Education

Heidi Iverson
University of Colorado Denver

Derek C. Briggs
University of Colorado at Boulder

Although previous meta-analyses have indicated a connection between inquiry-based teaching and improved student learning, the type of instruction characterized as inquiry based has varied greatly, and few have focused on the extent to which activities are led by the teacher or student. This meta-analysis introduces a framework for inquiry-based teaching that distinguishes between cognitive features of the activity and degree of guidance given to students. This framework is used to code 37 experimental and quasi-experimental studies published between 1996 and 2006, a decade during which inquiry was the main focus of science education reform. The overall

How do you view the role of research syntheses in advancing science education?

- *By better operationalizing the instructional approaches we are investigating, and then relating them to student learning*
  - *Model of inquiry*

TABLE 2
Codes and subcategories

Domain of inquiry	Description
Procedural	Asking scientifically oriented questions
	Experimental design
	Executing scientific procedures
	Recording data
	Representing data
	Hands-on
Epistemic	Nature of science
	Drawing conclusions based on evidence
	Generating and revising theories
Conceptual	Drawing on/connecting to prior knowledge
	Eliciting students' ideas/mental models
	Providing conceptually oriented feedback
Social	Participating in class discussions
	Arguing/debating scientific ideas
	Presentations
	Working collaboratively

How do you view the role of research syntheses in advancing science education?

- *By better operationalizing the instructional approaches we are investigating, and then relating them to student learning*
  - *Teacher role*

Teacher-led reform	←→	Traditional
Student-led reform	←→	Traditional
Teacher-led	←→	Student-led

TABLE 4
Mean effect size by model of inquiry contrasted

Contrast	*N* studies	*N* papers	Min	Max	*SD*	Mean	Median
EC	3	2	–.04	0.63	.38	.19	–.01
S	8	3	–.30	1.05	.43	.11	.09
PECS	2	1	.24	0.25	.01	.24	.24
PES	6	5	.05	1.74	.61	.72	.72
E	3	3	.55	0.92	.19	.75	.79

Note: Overall mean effect size = .50 across the 37 studies. Table does not provide mean effect size for studies that did not explicitly study guidance or for which there was only one study in a category. P = procedural; E = epistemic; C = conceptual; S = social.

TABLE 5
Effect sizes by guidance contrasted in study

Guidance	*N* studies	Min	Max	*SD*	Mean	Median
Student led versus teacher led	6	–.04	0.04	.03	.01	.01
Traditional versus student-led reform	5	–.30	0.96	.45	.25	.19
Traditional versus teacher-led reform	10	–.01	1.74	.57	.65	.60

Note. Overall mean effect size = .50 across the 37 studies. Table does not provide mean effect size for studies that did not explicitly study guidance or for which there was only one study in a category.

In what ways is the science education community facilitating and/or inhibiting the impact of research syntheses?

- *Insufficient descriptions of teaching interventions – methods sections often did not describe in much detail*
- *Insufficient data*
    - *Many studies did not include N's, means, SD's necessary for inclusion*
- *Arguments over terminology*
    - *e.g. Klahr & Nigam, 2004; Kirschner, Sweller & Clark, 2006*
    - *the field could benefit by focusing on smaller elements like scientific practices and the role of teacher guidance, rather than terms like 'inquiry,' 'discovery,' or 'hands-on.'*

What would you recommend science education researchers do to maximize the validity, usefulness, and impact of research syntheses?

- *Develop a standard for what types of information should be provided for interventions in teaching studies (e.g. not just duration, but details about instructional approaches, teacher and student role, materials used, etc.)*
- *Develop a standard for the data provided (e.g. tables must report N's, means, SD's)*

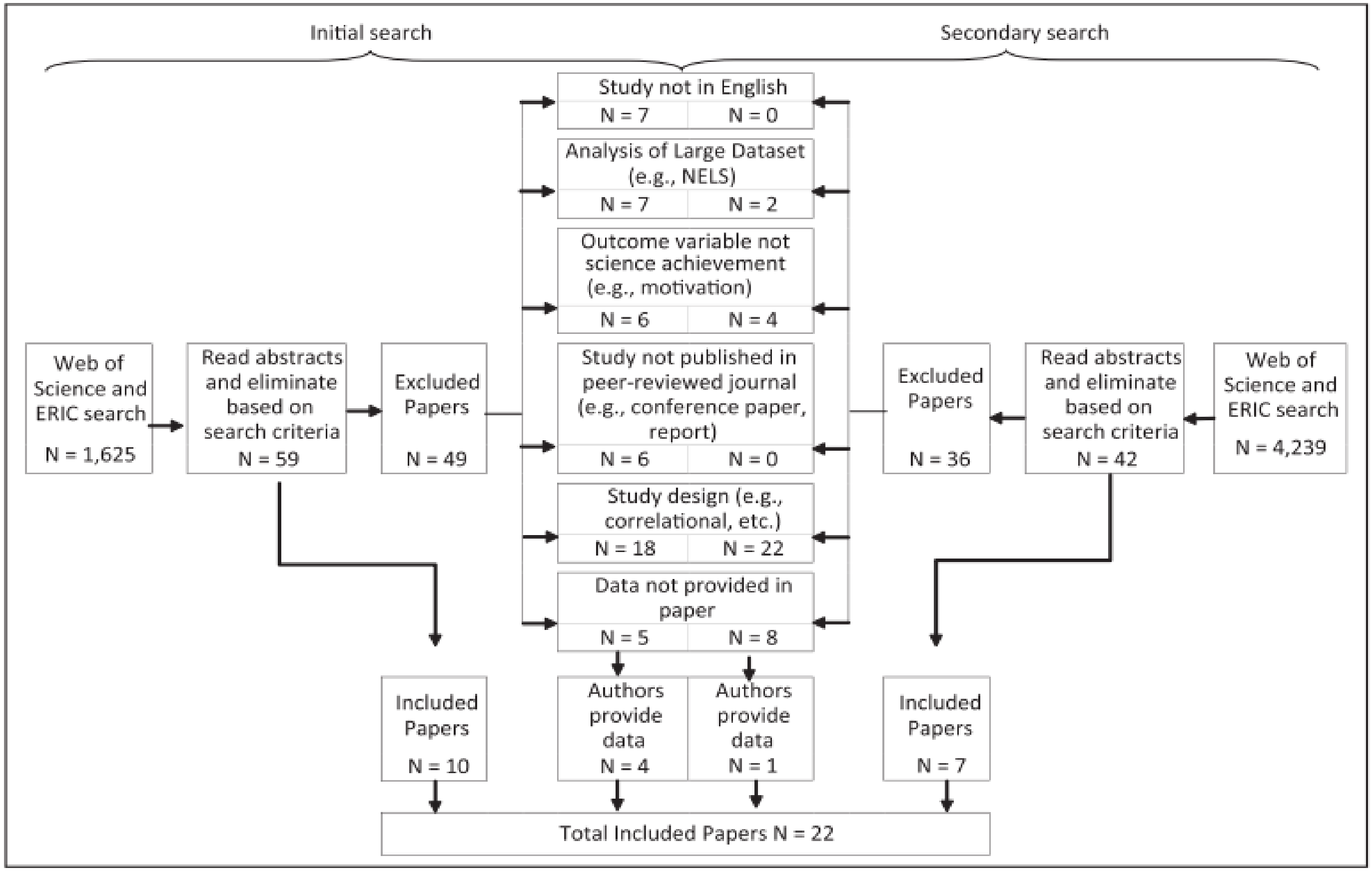


FIGURE 2. *Study selection flowchart.*

Reporting Practices in Science Ed

The Current State of the Field

Susan Kowalski - BSCS

BSCS

How do you View the Role of Research Syntheses in Advancing Science Education or other Education Research Fields?

Short answer…

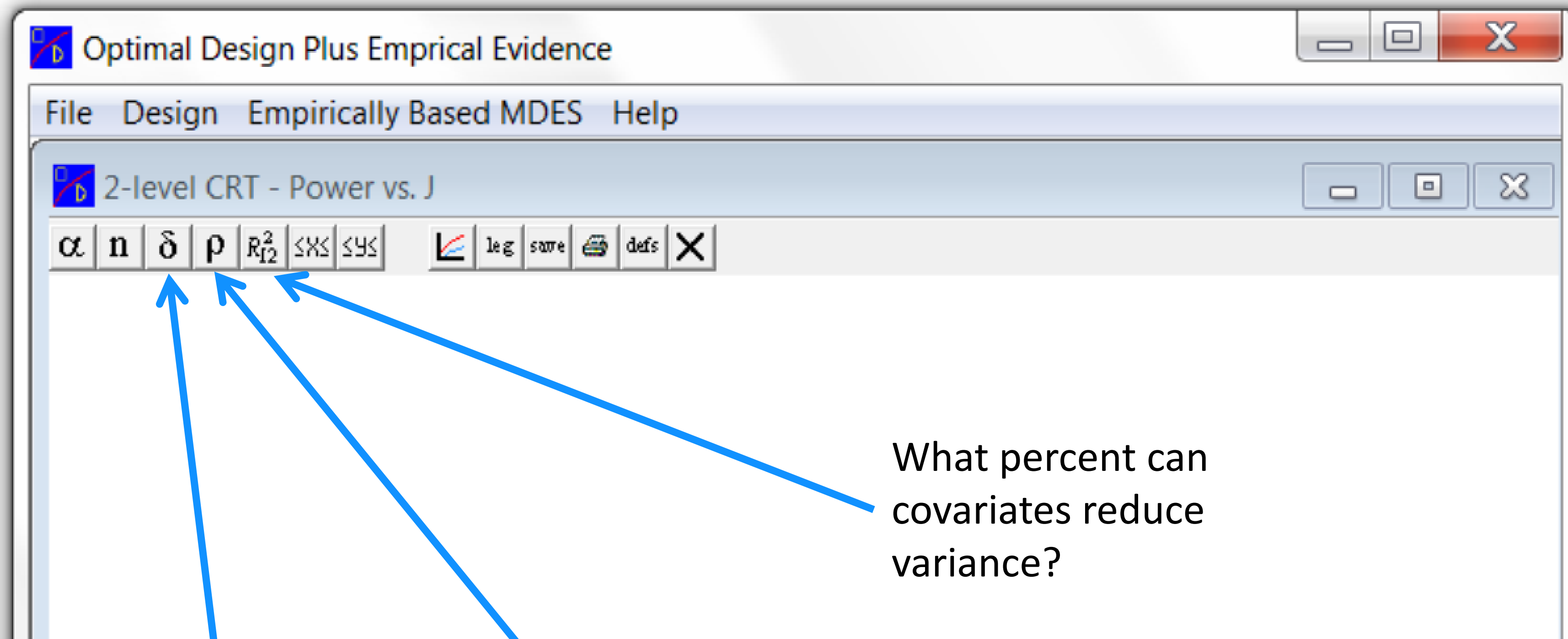
- Help researchers design group randomized (or cluster randomized) trials

- We are helping answer the question,

"How many schools (or teachers or districts) do I need to have an adequately powered study?"

Optimal Design Plus

- Freely available power analysis software
- Funded by WT Grant Foundation
- Requires that researchers supply good power analysis estimates



What percent can covariates reduce variance?

What is the anticipated effect size?

How similar are the groups or clusters to one another?

Meta-analysis Currently in Progress

- BSCS and Western Michigan University
  - – Joe Taylor (BSCS)
  - – Jessaca Spybrook (WMU) are co-PIs
- Funded by NSF PRIME, Grant # DRL 1118555
- Part of a larger effort to identify power analysis parameters for the design of CRTs

Inclusion Criteria

- Experimental or quasi-experimental studies
  - An identifiable intervention
  - Comparison of at least two groups
  - Student science achievement or attitudes/motivation outcomes
- Published between 2001 – 2011
- Preschool – Grade 12
- US students
- Studies obtained from 13 education research journals (more to come)

Journals (more to come)

- American Educational Research Journal
- Science Education
- Journal of Research in Science Teaching
- Research in Science Education
- International Journal of Science Education
- Journal for Science Teacher Education
- Education Evaluation and Policy Analysis
- Educational Researcher
- Journal of Educational Research
- Science & Education
- Journal of Science Education and Technology
- Journal of Research on Educational Effectiveness
- Electronic Journal of Science Education

Will Report Effect Sizes by…

- Type of intervention
- Science discipline
- Grade band
- Proximity of Outcome Measure to Intervention

End Product

- Researchers designing CRTs will be able to look up summary effect sizes that may be similar to their proposed intervention
- Use the summary effect size (along with ICCs and R-squared values) to power a study
- Promote responsible use of taxpayer $$

In what ways is the science education community facilitating and/or inhibiting the impact of research syntheses?

Current State of Reporting Practices in Science Education

- Authors rarely report effect sizes
- 52% of studies required author query to obtain enough information to calculate an effect size
- Studies lacked one or more of the following:
  - Number of participants in each treatment group (25%)
  - Standard deviations by treatment group (25%)
  - Means, covariate-adjusted means, or regression coefficients (58%)

Investigating Equity of Interventions

- Need descriptive statistics on the outcome measure and demographics *by treatment condition*
- Most authors report study-wide demographics only

What is normally reported…

“Approximately 71% of the students were European American, 18% were Latino/a, 4% were African American, and 7% were from other ethnic backgrounds. Ten percent of these students indicated that they spoke a language other than English at home. The students ranged in age from 15 to 17 years (M = 15.77, SD = 0.62). Forty-two percent of these students were in 10th grade, 54% were in 11th grade, and 4% were in 12th grade.”

Data Needed to Investigate the Equity of an Intervention—Example

Treatment Group	Number	Mean Pretest	Mean Posttest	SD
White Students—Treatment				
White Students—Comparison				
African American Students—Treatment				
African American Students—Comparison				
Students eligible for FRL—Treatment				
Students eligible for FRL—Comparison				
Etc.				

Added Difficulties

- Sometimes journal editors specifically ask authors to take descriptive information out
- Authors often no longer have access to data so author query is fruitless
  - retirement
  - moved institutions
  - data discarded after a specified period of time
- Several authors have not responded to author queries

What would you recommend science education researchers do to maximize the validity, usefulness, and impact of research syntheses?

Making it Easier to Learn from Each Other
AERA Reporting Practices (2006)

- Index of magnitude of quantitative relation between variables (treatment effect; regression coefficient; odds ratio)
- Indication of uncertainty in the index (SE or confidence interval)
- Exact test statistic and exact significance level
- Qualitative interpretation of the index and the effect describing its meaningfulness

Making it Easier to Learn from Each Other

- Report all descriptives *for each treatment and comparison condition*
  - *Adjusted* posttest scores if available
  - Individual-level (kid-level) standard deviations
  - Numbers of individuals
  - Demographic characteristics of each treatment and comparison group
- Calculate effect sizes
- Report confidence intervals around effect sizes
- Report non-significant findings

When Reading the Work of Others

- Don’t dismiss small-scale studies of interventions because of non-significant p-values
- Always look for effect sizes and calculate them yourself if necessary (and possible)
  - David Wilson’s Effect Size Calculator, George Mason University
  - http://cebcp.org/practical-meta-analysis-effect-size-calculator/standardized-mean-difference-d/
- Compare the effect size to that of similar interventions in your field

Small group discussions

- 20 minutes
- Focus mainly on question 3
  - *What would you recommend education researchers do to maximize the validity, usefulness, and impact of research syntheses?*
- Capture ideas
- Elect a spokesperson to report back to the large group (6-7 min max)

Discussant Summary and Impressions

Elaboration and Standardization

- More comprehensive statistical and/or psychometric reporting
- Fuller intervention descriptions
- Report potential moderators of effects- e.g., study artifacts

- Ideas for getting the word out
  - Point folks to AERA guidelines?
  - Are these sufficient?
  - STEM-specific and/or DRK-12 Guidelines?

On Uniqueness

- Robert Slavin (2008) What works? Issues in synthesizing educational program evaluations. Educational Researcher. 31(1), 5-14.
- Where are we now?
- Issues:
  - Funding
  - Systems
  - Dissemination
  - Numbers of researchers

On Uniqueness

- So we do causal effects research, we just do it badly:
  - Methodology
  - Reporting
  - What we study
- Where to?
  - Replication
  - Economic links
  - Focus across research traditions

Value of standardization to our field?

- Standardized reporting practices:
  - inform the focus and design of new, unique studies that generate new knowledge
  - facilitate ***replication*** to increase our confidence in what we think we know

Both are important

- 1960s - Schwab (1964) and Kuhn (1962)
  - Fluid Inquiry/Revolutionary Science: new studies based new ways of conceptualizing problems or phenomena
  - Stable Inquiry/Normal Science: focused work on a set of studies within a stable paradigm
- Contemporary
  - *A Framework for K-12 Science Education*
  - Proposed R&D agenda for NGSS notes the importance of large-scale replication studies of NGSS-aligned programs or practices

Others thinking about replication

Bauernfeind, R. H. (1968). The Need for Replication in Educational Research**.** *The Phi Delta Kappan* 50 (2), 126-128.

***Large Scale Replication Research: Three Examples and the Issues They Raise* – John Ioannidis – SREE Spring 2014**

- Based on Ioannidis, J. (2005). Why most published research findings are false. *PLoS Med, 2*(8), e124.
- **RISK of NOT Replicating: Drawing false conclusions from single studies of a program or practice**
- Specifically, the risks are greatest when:
  - There are few studies on a given question
  - The studies are small (low power)
  - The effect sizes are small (i.e., publication biases, confidence intervals approach zero)
  - ***The designs, definitions, and analytic conventions are less standardized***
  - There are conflicts of interest
  - Only one research team is pursuing a question or set of questions

Self-Reflection as a Community?

- Are we overemphasizing uniqueness in new studies? Is replication overlooked in the process?
- Where can we find the evidence?
  - What value to do we place on replication when advising graduate students and new doctorates?
  - How valuable do journal editors find replication studies?
  - What do expressed funding priorities suggest about the value of replication?
- If our community decides that lack of replication is a problem, and value systems must change, then the success of a movement toward replication relies heavily on standardized reporting practices.

Lingering questions?

- For panelists?
- For discussants?

Thank you!

- Session materials will be posted to bscs.org/sessions