AMERICAN EVALUATION ASSOCIATION

PATHS TO THE FUTURE OF

EVALUATION 2019

NOVEMBER 11 – 16

MINNEAPOLIS, MN

Evaluating Research Quality and Impacts: the Standards of Evidence Review

Evelyn Gordon, Horizon Research, Inc.

November 13, 2019

AMERICAN EVALUATION ASSOCIATION

PATHS TO THE FUTURE OF EVALUATION 2019

horizon RESEARCH, INC.

Session Plan

- History and moving towards the future
- Purpose of a Standards of Evidence for Empirical Research (SoE) review
- What types of manuscripts are reviewable
- Process and criteria for the review
- Ways to use review findings for formative and summative evaluation

History of the Standards of Evidence for Empirical Research Review Tool

- Developed by Horizon Research, Inc. (HRI), with Education Development Center, to review what is known about key topics in mathematics and science teaching and learning (EHR-0445398)
- Revised by HRI to summarize contributions to STEM education literature by projects funded under the REESE program (NSF DACS10CL617)
- Ongoing use and refinement in evaluation work

This material is based upon work supported by the National Science Foundation under Grant No. EHR-0445398 and Contract No. NSF DACS10CL617. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Moving toward the future

- In the past, evaluation of education interventions generally focused on the quality and impacts of the interventions
- Several National Science Foundation programs have moved toward requiring grant awardees to conduct their own research…
  often looking at quality and impacts
- Present and future evaluations
  - Smaller focus on quality and impacts to avoid duplicating project work
  - Larger focus on evaluating research

Purpose

- Characterize contributions of a publication to the field's knowledge base
  - What is known from the findings?
  - What confidence can be placed in the findings?

- Formative evaluation: provide feedback on manuscripts prior to submission
- Summative evaluation: describe the evaluation client's research contributions

Reviewable Manuscripts

- To be reviewable, manuscripts must report results of a study that
  - Addressed an identifiable question or issue
  - Systematically gathered or obtained data
  - Analyzed the data to address the question/issue

- The review is applicable to a broad range of research designs and methodologies

Unreviewable Manuscripts

- A SoE review is not appropriate for some research-related products, for example
  - Descriptive reports of a program
  - Theoretical articles
  - Guidance for practitioners
  - Opinion pieces
- These types of products cannot be meaningfully reviewed using the SoE, even though they may offer valuable contributions to the field

SoE Review Process

- Read the manuscript and identify
  - Research questions
  - Results
  - Claims made – immediate findings related to research questions; broader conclusions, generalizations, implications
- Assess if documentation criteria are met
- Judge if validity criteria are met
- Determine an overall rating about the quality of the empirical evidence to support findings related to each research question and broader claims

Review Criteria Categories

- Documentation for the overall manuscript and each research question
- Validity considerations for each research question
  - Avoidance of bias in research design
  - Avoidance of bias in conducting research
  - Appropriateness of data collection methods
  - Appropriateness of analysis
  - Appropriateness of reporting
  - Consideration of alternative explanations
- Overall rating of strength of empirical evidence for each research question
- Appropriateness of generalizations, conclusions, and implications

Example Study: Introduction

AEA's annual conference offers a variety of session types, as is common at professional conferences. However, there is limited research available to indicate which types of sessions are most worthwhile for conference goers (Faux, 2015). *Other Conference* attendees found single-presenter sessions more informative than multi-paper sessions (Bogus, 2014), and *Another Conference* participants who talked to poster presenters recalled more about the poster than about sessions in which they were part of a passive audience (Faux, 2015). This paper provides results of a (fictive) study to investigate the effectiveness of single-presenter sessions, multi-paper sessions, and workshop sessions.

Example Study: Methods

Forty attendees of the *Evaluation 2018* conference were recruited to complete a brief, online questionnaire following each session they attended on two days of the conference. The questionnaire included 8 closed-ended questions, 4 addressing the session quality and 4 addressing impacts; one open-ended question; and a link to indicate which session the questionnaire was rating. Participants rated the quality and impact questions using a 5-point Likert type scale. A hierarchical analysis was conducted, with ratings nested in participants, to compare two outcomes for the three session categories: a session quality composite and a session impact composite.

Example Study: Results

Single-presenter sessions were rated as higher quality than either multi-paper sessions or workshops (statistical test result and effect size), and workshop sessions were rated as having a higher impact than other session types (statistical test result and effect size). In a follow-up analysis with session time added to the model, the only significant result found was that after-lunch sessions were rated as lower in both quality and impact than sessions at any other time of day. Responses to the open-ended question suggest that participants often found after-lunch sessions boring, confusing, or both.

Example Study: Conclusions and Implications

Our results suggest that there may be differences in quality and impact of sessions associated with the session type. However, additional research is needed to distinguish effects of session type and session time, as well as to investigate whether our results generalize to other professional conferences. An implication for conference planners is that they should consider including a siesta session following lunch.

Pause for thought (and questions)

- What are some strengths of this manuscript and the study it describes?
- What are some weaknesses of this manuscript and the study it describes?

- No study is perfect
- Manuscripts must balance journal requirements, including space, and other considerations

Documentation for the overall manuscript

Intended contribution/research questions
Theoretical background
Current knowledge
Constructs as they are operationalized
Researcher disclosure
Intended generalizability
Directions for future research

Documentation for the example

Intended contribution/research questions
Theoretical background
Current knowledge
Constructs as they are operationalized
Researcher disclosure
Intended generalizability
Directions for future research

Documentation for each research question

Units of Study	Research site, participants, and event
Design	Sampling/assignment strategy, design type
Collection of data and instrumentation	Methods, where/when/how data were gathered
Analysis	Strategy and results
Findings	Empirical support, limitations

Documentation for the example

Units of Study	Research site, participants, and event
Design	Sampling/assignment strategy, design type
Collection of data and instrumentation	Methods, where/when/how data were gathered
Analysis	Strategy and results
Findings	Empirical support, limitations

Validity for each research question

Avoiding bias in design	Sample bias, unfair comparisons
Avoiding bias in conducting research	Non-response bias, attrition bias, missing data, contamination, investigator bias
Appropriate data collection methods	Methods and instruments appropriate for research question, triangulation
Appropriate and systematic analysis	Unit of analysis, methods of analysis, sample suitable for planned analysis
Appropriate reporting of results	Null and discrepant evidence indicated, effect size
Considering alternative explanations	Alternative explanations considered through the design, analytic strategy, discussion, and/or in recommendations for future research

Validity for each research question

Avoiding bias in design	Sample bias, unfair comparisons
Avoiding bias in conducting research	Non-response bias, attrition bias, missing data, contamination, investigator bias
Appropriate data collection methods	Methods and instruments appropriate for research question, triangulation
Appropriate and systematic analysis	Unit of analysis, methods of analysis, sample suitable for planned analysis
Appropriate reporting of results	Null and discrepant evidence indicated, effect size
Considering alternative explanations	Alternative explanations considered through the design, analytic strategy, discussion, and/or in recommendations for future research

Overall rating for each research question

- Overall numeric rating to indicate strength
  - **Level 1**: does not meet standards because the design does not align with the stated problem, analysis does not align with design, the findings are not supported by evidence, or there is insufficient documentation to rate a Level 2 or Level 3
  - **Level 2**: Meets standards with reservations
  - **Level 3**: Meets standards (strengths generally outweigh limitations)
- Narrative to justify rating, indicate strengths, and indicate limitations that were not outweighed by the design and analysis

Conclusions, Generalizations, and Implications

Conclusions aligned with the study's findings	Logical case made, discrepant findings acknowledged/explained
Generalizations stated with appropriate caveats or bounds	Sensitive to the sample or context of the study, context adequately described to provide confidence for any generalizations made, caveats or bounds of generalization stated
Implications aligned with findings and sensitive to limitations	Implications logically derived from findings and sensitive to important limitations

For More Information

- http://www.horizon-research.com/standards-of-evidence-codebook
- http://www.mspkmd.net/papers/research_support_tool.pdf
- http://www.mspkmd.net/papers/consumer_guide.pdf

Questions? Comments?

Thank you!