

What Have We Learned  
About RCTs, Gold  
Standards, and Credible  
Evidence?

Stewart I. Donaldson, Ph.D.  
Claremont Graduate University  
stewart.donaldson@cgu.edu

AES Conference  
Canberra, Australia  
September 2, 2008

---

---

---

---

---

---

---

---

What Counts as  
Credible Evidence  
in **APPLIED  
RESEARCH and  
EVALUATION  
PRACTICE?**

Stewart I. Donaldson  
Christina A. Christie  
Melvin M. Mark  
Editors

---

---

---

---

---

---

---

---

Overview

- Contemporary Evaluation Practice
- Evidence-based Practice
- Debates about What Counts as Evidence
- Experimental Routes to Credible Evidence
- Non-Experimental Approaches
- Moving Beyond the Debates

---

---

---

---

---

---

---

---

## Contemporary Evaluation Practice

- Booming
- Global
- Diverse Contexts
- Many More Evaluands
- Multidisciplinary
- Many New Approaches & Methods
- More than Traditional Social Science Research Methods

---

---

---

---

---

---

---

---

## Evidence-Based Practice

- Highly Valued
- Global
- Multidisciplinary
- Many Applications

---

---

---

---

---

---

---

---

## Sample of Applications

- Evidence-based Medicine
- Evidence-based Mental Health
- Evidence-based Management
- Evidence-based Decision Making
- Evidence-based Education
- Evidence-based Coaching

---

---

---

---

---

---

---

---

## Sample of Applications

- Evidence-based Social Services
- Evidence-based Policing
- Evidence-based Conservation
- Evidence-based Dentistry
- Evidence-based Policy
- Evidence-based Thinking about Health Care

---

---

---

---

---

---

---

---

## Sample of Applications

- Evidence-based Occupational Therapy
- Evidence-based Prevention Science
- Evidence-based Dermatology
- Evidence-based Gambling Treatment
- Evidence-based Sex Education
- Evidence-based Needle Exchange Programs
- Evidence-based Prices
- Evidence-based Education Help Desk

---

---

---

---

---

---

---

---

## Evidence-based Paradox

- Strong Agreement that Credible Evidence is Highly Desirable
- Strong Disagreements about what is Credible Evidence

---

---

---

---

---

---

---

---

## Sample of The Debates

- Qualitative-Quantitative Debate
- Visions for the Desire Future of Evaluation Practice
- AEA Statement vs. Not AEA Statement
- EES Statement
- The Lipsey vs. Scriven Debate
- What Counts a Credible Evidence?

---

---

---

---

---

---

---

---

## Experimental Design: Gold Standard?

- Random Assignment
- Experimental Control
- Ruling Out Threats to Validity



---

---

---

---

---

---

---

---

## Supreme Courts of Credible Evidence

- What Works Clearinghouse
- Campbell Collaboration
- Cochrane Collaboration



---

---

---

---

---

---

---

---

## Experimental Approaches

- Henry: When Getting it Right Matters
- Bickman & Reich: RCTs - A Gold Standard with a Feet of Clay
- Gersten & Hitchcock – The What Works Clearinghouse
- Julnes & Rog - Methods for Producing Actionable Evidence

---

---

---

---

---

---

---

---

## Non-Experimental Approaches

- Scriven: Demythologizing Causation and Evidence
- Greene: Evidence as “Proof” and Evidence as “Inkling”
- Rallis: Reasoning With Rigor and Probity: Ethical Premises for Credible Evidence
- Mathison: Seeing Is Believing: The Credibility of Image- Based Research and Evaluation
- Schwandt: Toward a Practical Theory of Evidence for Evaluation

---

---

---

---

---

---

---

---

## Challenges of the Gold Standard

- AEA Statement vs. Not AEA Statement
- Theoretical
- Practical
- Methodological
- Ethical
- Ideological
- Political
- Scriven’s Summative Conclusion

---

---

---

---

---

---

---

---

## AEA Statement vs. Not AEA Statement

- AEA Opposition to Priority on RCTs
  - “Privileging RCTs: Back to the Dark Ages”
  - “Priority Manifests Fundamental Misunderstandings Causality and Evaluation”
- AEA Members Opposition to AEA Statement
  - “Lack of Input from Key AEA Members”
  - “Unjustified, Speciously Argued, Does Not Represent Norms or Many AEA Members Views”
- AEA Compared to the Flat Earth Society

---

---

---

---

---

---

---

---

## Diverse Prescriptive Theories of Evaluation Practice

- Social experimentation
- Science of valuing
- Results oriented management
- Utilization focused evaluation
- Empowerment evaluation
- Realist evaluation
- Theory-driven evaluation
- Inclusive evaluation
- Fourth generation evaluation

---

---

---

---

---

---

---

---

## RCTs Not Practical/Feasible

- Often Impossible to Implement Well
- Not Cost Effective
- Very Limited Range of Applications
- Chapter Authors Provide Evidence to the Contrary

---

---

---

---

---

---

---

---

## RCT Ethical Issues

- Unethical to Withhold Treatment from Control Groups
- Why Evaluate if Treatment is Better?
- Delay Treatment
- Non Evidence-Based Programs are Unethical

---

---

---

---

---

---

---

---

## Methodological Challenges

- Zero Blind vs. Double Blind – Experimenter Effects
- Allegiance Effects
- Unmasked Assignment
- Misguided Arguments about Causality
- External Validity Concerns
  
- Chapter Authors Claim Recent Methodological Developments to Overcome Some Challenges Noted in the Past

---

---

---

---

---

---

---

---

## Political Concerns

- The RCT Gang has hijacked the term “evidence-based” for political and financial gain
- “Evidence” and especially “scientific or rigorous evidence” have become code for RCTs
- Focusing evaluation around these particular ideas about “scientific evidence,” allows social inquiry to become a tool for institutional control and to advance policy in particular directions

---

---

---

---

---

---

---

---

## Political Concerns

- It is epistemological politics, not the relative merits of RCTs, that underlie federal directives on methodology choice
- The demand for evidence advances a "master epistemology." The very dangerous claim is that a single epistemology governs all science
- Privileging the interests of the elite in evaluation is radically undemocratic

---

---

---

---

---

---

---

---

## Ideological Differences: Paradigm Wars

- "The positivist can't believe their luck, they've lost all the arguments of the last 30 years and they've still won the war!"
- "The world view underlying the current demand for evidence is generously speaking a form of conservative post-positivism, but in many ways is more like a kind of neo-positivism."

---

---

---

---

---

---

---

---

## Ideological Differences: Paradigm Wars

- Many of us thought we'd seen the last of this obsolete way of thinking about the causes and meanings of human activity, as it was a consensual casualty of the great quantitative-qualitative debate in the latter part of the 20<sup>th</sup> century.
- Human action is not like activity in the physical world.
- Social knowledge is interpreted, contextual, dynamic or even transient, social or communal, and quite complicated. Privilege and honor complexity.

---

---

---

---

---

---

---

---



## Ideological Differences: Paradigm Wars

- Evidence-based evaluation concentrates evaluation resources around one small question, does the program work?, and uses but one methodology, despite a considerable richness of options. The result is but one small answer.
- So what kind of evidence is needed? Not evidence that claims purchase on the truth with but a small answer to a small question, neat and tidy as it may be.

---

---

---

---

---

---

---

---

## Scriven, 2009

To insist we use RCTs is simply bigotry ... not pragmatic and not logical. In short, it is a dogmatic approach that is an affront to scientific method.

---

---

---

---

---

---

---

---

## So What Kind of Evidence is Needed? (Greene, 2009)

### Evidence:

- that provides a window into the messy complexity of human experience
- that accounts for history, culture, and context
- that respects differences in perspective and values
- about experience in addition to consequences
- about the responsibilities of government not just responsibilities of its citizens
- with the potential for democratic inclusion and legitimization of multiple voices - evidence not as proof but as inkling

---

---

---

---

---

---

---

---



---

---

---

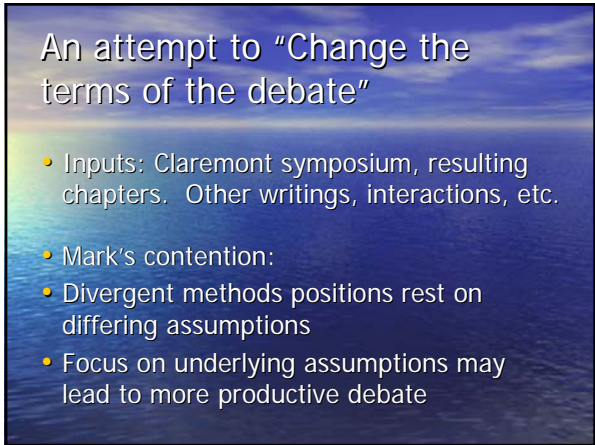
---

---

---

---

---



---

---

---

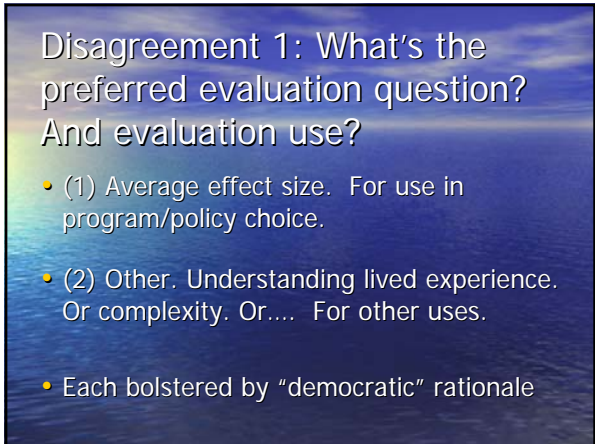
---

---

---

---

---



---

---

---

---

---

---

---

---

## Alternative debate topics

- Value of estimating the effect of a given program?
- Value, relative to addressing other questions?
- Who decides the above, and how?
- If program's average effects are of interest, what ancillary methods are needed?

---

---

---

---

---

---

---

---

## Gold Standards *in context*

- "Unfortunately, too many people like to do their statistical work [or their evaluation/applied research planning] as they say their prayers – merely substitute in a formula found in a highly respected book written a long time ago." *Hotelling et al. (1948)*

---

---

---

---

---

---

---

---

## *So What Counts as Credible Evidence? A Contingency Perspective*

*It depends on:*

*Question(s) of Interest*

*The Context(s)*

*Assumptions of Evaluators & Stakeholders*

*Theory of Practice*

*Practical, Time, & Resource Constraints*

---

---

---

---

---

---

---

---

## CDC: Gathering Credible Evidence

- **Definition:** Compiling information that stakeholders perceive as trustworthy and relevant for answering their questions. Such evidence can be experimental or observational, qualitative or quantitative, or it can include a mixture of methods. Adequate data might be available and easily accessed, or it might need to be defined and new data collected. Whether a body of evidence is credible to stakeholders might depend on such factors as how the questions were posed, sources of information, conditions of data collection, reliability of measurement, validity of interpretations, and quality control procedures.

---

---

---

---

---

---

---

---

*From Experimenting Society to Evidence-based Global Society?*

*From "RCTs" as the Gold Standard to "Methodological Appropriateness" as the Platinum Standard*

---

---

---

---

---

---

---

---

*"The issue of what constitutes credible evidence isn't about to get resolved. And it isn't going away. This book explains why. The diverse perspectives presented are balanced, insightful, and critical for making up one's own mind about what counts as credible evidence."*

---

---

---

---

---

---

---

---

*And, in the end, everyone must take a position. You simply can't engage in or use research and evaluation without deciding what counts as credible evidence. So read this book carefully, take a position, and enter the fray."*  
**(Patton, 2009)**

---

---

---

---

---

---

---

---