

A View of Multiple Methods from a Review of the Evaluations of Mathematics Curricula

Jere Confrey

Washington University in St. Louis

Three parts

- Brief Summary of the NRC report “On Evaluating Curricular Effectiveness” and its use of multiple methods
- Application of model to Huntley et. al study
- Compare and contrast of multiple methods in OECE and Raudenbush proposal

***On Evaluating Curricular Effectiveness:
Judging the Quality of K-12 Mathematics Evaluations***
J. Confrey and V. Stohl, Editors

**Steering Committee for a Review of the Evaluation Data
on the Effectiveness of NSF-Supported and Commercially
Generated Mathematics Curriculum Materials**

**Mathematical Sciences Education Board
Center for Education
Division of Behavioral and Social Sciences and Education
National Research Council of the National Academies
The National Academies Press**

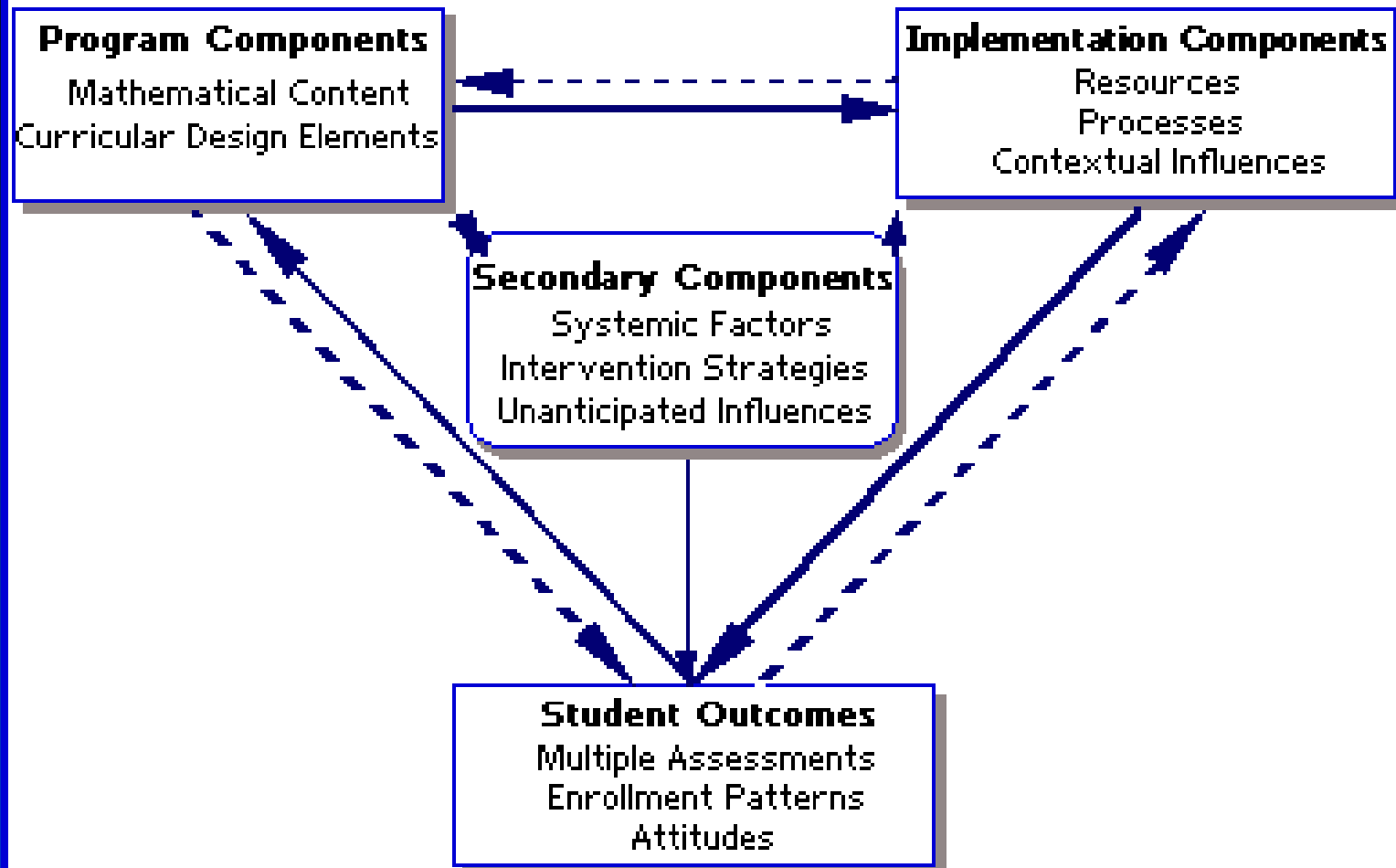
Sponsored by the National Science Foundation

Of Studies Examined

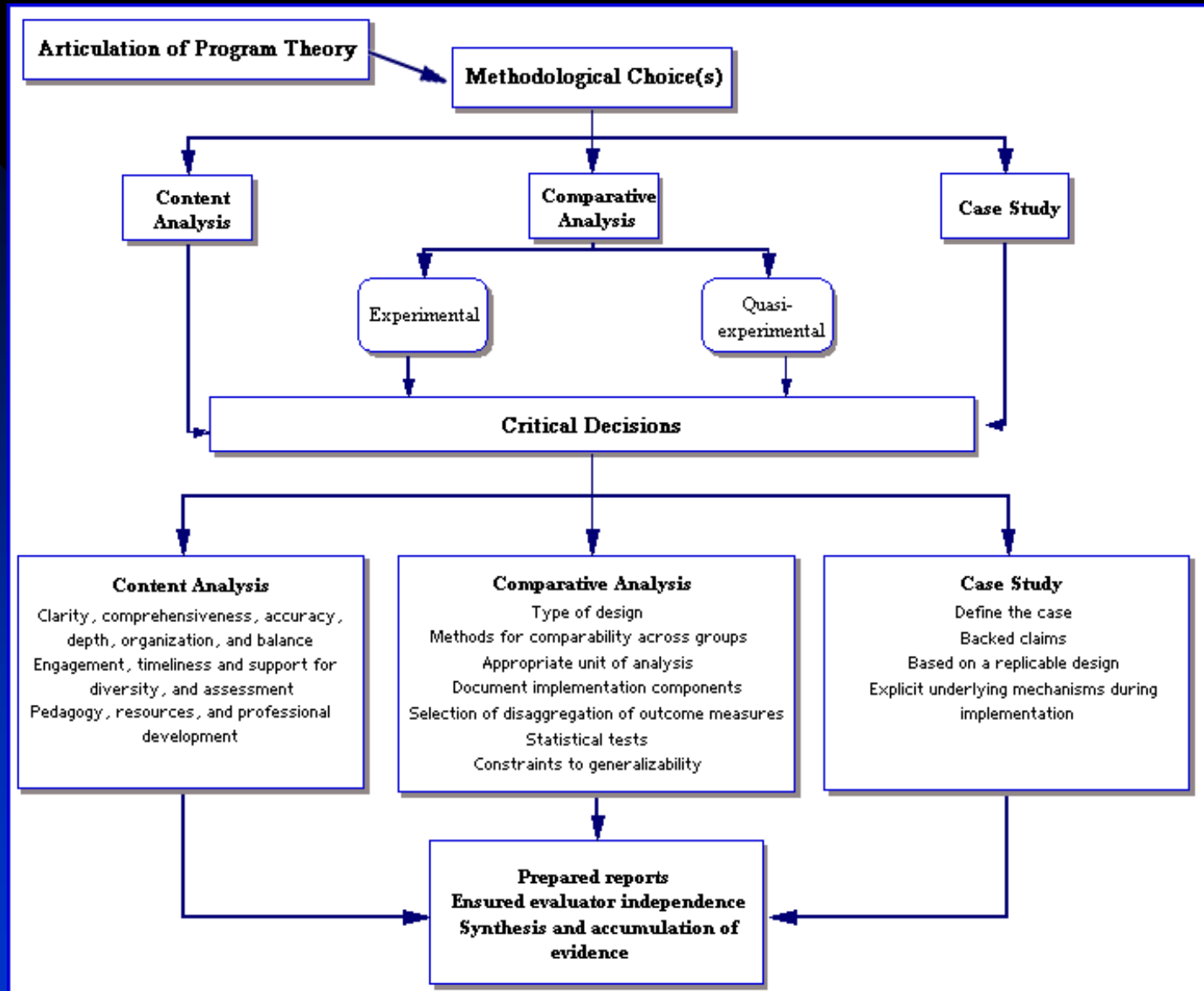
Type of Study	Met Initial Classification for the Study	Met Final Criteria for the Study
Content Analysis	36	36
Comparative	95	63
Case Study	45	32
Synthesis	16	16
TOTAL	192	147

Framework for Evaluating Curricular Effectiveness

ARTICULATION OF PROGRAM THEORY



Methodological Choices and Critical Decisions



A Scientifically Valid Study...

- Address the primary components in the framework
- Conform to methodological expectations of appropriate category of evaluation (content analysis, comparative study, case study)

Curricular Effectiveness

- ↳ "...cannot be established by a single scientifically valid study: instead, a body of studies is needed."
- ↳ "...is an integrated judgment based on interpretation of a number of scientifically valid evaluations that combine social values, empirical evidence, and theoretical rationales."

Principles

↳ General Principle

↳ "... a curricular program's effectiveness should be ascertained through the use of multiple methods of evaluation, each of which is a scientifically valid study. Periodic synthesis of the results across evaluation studies should also be conducted. "

↳ General Principle of Multiple Methods

↳ A single methodology, even when replications and variations of a study are conducted, is inadequate to establish curricular effectiveness, because some types of critical information will be lacking.

Strengths of Huntley et.al: A Comparative Study

- Clearly articulated program theory:
view of algebra
- Some implementation fidelity
through teacher interviews
- Use of multiple contrasting
outcome measures (curricular
validity of measures)

Weaknesses in Design and Analysis

- Incorrect unit of analysis
- No named comparison group
- Volunteer sample and field test sites

Value of multiple authors and perspectives

- Designed with a deep understanding of complexity and variations in practice
- Authors represented math, math education, & methodology
- Recognized the interplay of student sample, teacher quality, contexts of implementation, varied types of outcomes, school policies etc.

Why effectiveness is an integrated judgment combining values, theory and empirical evidence

- *“Our study does not provide information needed to answer the question about what mathematics is most worth learning, but it does suggest the kinds of trade-offs that might be expected when one allocates time to topics in ways that differ from allocations in the typical U.S. high school curriculum.”(Huntley et al, p. 354.)*

Why effectiveness is an integrated judgment

- *Recognizes importance of reporting variation in implementation*
- *“The overall results suggesting achievement patterns related to curricular content disguise very substantial differences in implementation and results at different sites.” (p. 355).*

Why effectiveness is an integrated judgment

- *Need to see the larger picture—transition to college etc.*
- *“There is clearly a great deal of room for improvement in the evident achievement of students in algebraic reasoning, problem solving and calculation.” (p. 357)*

Why effectiveness is an integrated judgment

- “We believe that our results do provide support for the basic reform position”
- This is clearly a warranted conclusion if one accepts the design and statistical tests, and depending on one’s values and theories, for example about technological dependency, one could disagree with the author’s conclusions.

Is it mixed methods?

- Combination of interviews and tests
- Uses different methods to explain the results (especially discrepancies across sites) within a comparative study

Does a comparative study establish effectiveness, if weaknesses are corrected?

- Need a *content analysis* to determine if the curricula is comprehensive, accurate, and complete (program theory)
- Highly desirable to have *case studies* to understand implementation better, how teacher effects, professional development, collective practice, students' class assignments may be interacting with curricular treatment (implementation)

Compare/Contrast OECE with Raudenbush's proposal

- Both support multiple methods
- Both identify importance of instructional core (regimes, curricula interplay)
- Both recognize centrality of issues of equity and equal opportunity
- Both recognize problems of precise outcomes

Compare/Contrast OECE with Raudenbush's proposal

- OECE views multiple methods as a research programme, and not as a sequential process
- OECE has a model of curricular evaluation that recognizes theory and values in all interpretations and not as a delineation “what works” after a RTF and prior descriptive/correlational studies

Raudenbush's proposal identifies specific targets for Description/correlation as prior to RFT to:

- 1) identify promising interventions (p. 26)
- 2) ensure they are capable of broad scale implementation (p. 26) and
- 3) precisely define relevant outcomes (p. 28)

“The synthesis of research from a variety of methods conducted at different scales ought to be a pre-requisite for the construction of large-scale randomized field trials” (p. 29).

OECE sees need for supplements to comparative studies:

Necessary to understand the interaction among implementation, professional development and opportunity to learn;

Understand limitations/restrictions of generalizability due to the role of context and embedded variables (states, districts, schools, classrooms)

Recognize the influence of timeline, incentives, and authority for changes and importance of understanding conceptual trajectories over time

Consider the comprehensiveness, completeness, view of assessments, engagement of tasks, use of inquiry and reasoning in the materials in the evaluation process.

Take into consideration a changing curricular focus, with new topics, new technologies and new career paths.

Current controversy

Importance of avoiding reductionism

Value of multidisciplinary perspectives

Recognize the importance of theory and values

Feasibility in our promises to the field

Recognizing contingencies and contextual variance inherent in practice

Describe likelihoods and tendencies to inform decision-making rather than to make static and uniform claims

Domain of Applied Educational Research

- “Question is whether and how to employ these [mixed] methods in service of a newly dominant research agenda that aggressively seeks to evaluate claims about the causal effects of interventions aimed to improve teaching and learning in the nation’s classrooms” p. 14

Domain of Applied Educational Research (revised statement)

- “Question is whether and how to employ these [multiple] methods in service of a newly dominant research agenda that aggressively seeks to evaluate claims about the *effectiveness of interventions in order to provide decision-makers with adequate information to make informed judgments* about how to improve teaching and learning *in their particular contexts and classrooms*”