# Evaluation Designs

*Taken from Professor Jennifer Greene's Evaluation Methods course (EPSY 471)*
*Please do not distribute or cite.*

Context for design application: Continuing controversy over whole language versus phonics instruction for beginning literacy, in primary grades (K-3).

## 1. Experimental designs

*Evaluation questions:*
Causal, explanatory, predictive
Did the program being evaluated cause the intended outcomes?
Can observed outcomes be attributed to the program's treatment?

*Key features:*
Randomization = random selection of "units" (usually people in programs) from population *and* random assignment of units to experimental and control groups
Analytically comparative
Quantitative, standardized measurement, to enable statistical analysis
Best design for cost-benefit questions

*Key technical concerns:*
Ruling out rival explanations or alternative hypotheses
Threats to internal validity (causal inference)
Challenges of external validity (generalization)

*Sample evaluation question in "whole language" evaluation context:*
How well do children in whole language primary grade instruction do on standardized measures of reading and writing achievement, compared to similar children in phonics instruction and a combined instructional program?

## 2. Quasi-experimental designs

*Evaluation questions:*
Same as above

*Key features:*
Randomization not possible or not ethical or not desirable
Experimental-Control comparison achieved via other means:
- Identification of a comparison (not control) group
- One-group comparisons, e.g., pre to post
- Statistical adjustments
- Comparisons made to existing data bases, existing literature, or existing "knowledge"

Analytically comparative
Quantitative, standardized measurement, to enable statistical analysis

Known quasi-experimental designs, Cook and Campbell… examples (and see Weiss, plus Joe R)

- One group designs
  After-only
  Before and after
  Pre-post + program theory
  Pre-post + dose-response or levels-of-use analysis
  Time series
- Two+ group designs
  After-only
  Before and after
  Cohorts as comparisons
  Multiple time series
  Regression discontinuity
  Propensity score matching
  Statistical adjustments for non-equivalence
  Existing data sets used as comparisons

*Key concerns:*
Ruling out rival explanations or alternative hypotheses
Threats to internal validity (causal inference)
Challenges of external validity (generalization)
Adequacy of comparison

*Sample evaluation question in "whole language" evaluation context:*
How well do children in whole language primary grade instruction do on standardized measures of reading and writing achievement, compared to similar children in phonics instruction and a combined instructional program?

3. **Descriptive and relational designs**

*Evaluation questions:*
Relational, descriptive, associative, ameliorative
What are the primary characteristics of participants in this program – before, during, and after their participation? Is there any relationship between these characteristics and program participation?
What is this program like? What are its important characteristics and dimensions?
What kind of people does the program serve? What profiles of participation can be drawn?

*Key features:*
Correlational, cross-sectional
Use of survey approaches, extended time series designs, administrative data monitoring, secondary data analysis
Analytically descriptive, relational, and representative
Eclectic measurement and methods

*Key concerns:*

Representativeness
Data quality, challenges of measuring outcomes
*Sample evaluation questions in "whole language" evaluation context:*
What kinds of children do whole language programs serve?
What are key features of whole language curricula and instruction? How well are they being implemented?
What kinds of teachers and schools favor whole language programs?
How well do children in whole language classes do on various measures of literacy?
Is their performance comparable to local, state, national norms for similar children?
Do some kinds of children do particularly well in whole language classes?
How satisfied are parents of children in whole language classes with their literacy learning?

4. **Holistic, case-oriented designs**

*Evaluation questions:*
Emphasis on holistic, contextual understanding
In what ways and to what extent is this program experience meaningful, of high quality, beneficial for its various participants?

*Key features:*
Contextual and local and therefore multiplistic, relativistic
Emergent, not preordinate or a priori design
Analytically holistic
Use of case study, narrative, interview/observation, ethnographic approaches
Strong preference for qualitative methods

*Key concerns:*
Challenges of attaining holistic, contextualized understanding
Discerning important dimensions of program quality
Credibility (not validity)

*Sample evaluation questions in "whole language" evaluation context:*
In what ways and to what extent do children in selected whole language programs experience these programs as meaningful, of high quality, and beneficial? How do their teachers and their parents view and understand these programs?
In what ways and to what extent do children's experiences in these programs facilitate "adequate" literacy learning, according to a variety of standards and measures?
What about these programs contributes, or not, to children's meaningful literacy development?

5. **Action- or dialogue-oriented designs**

*Evaluation questions:*
Actionable, dialogic; change-oriented, critical; democratic
How can this context be more equitable and just for all participants? What changes are needed? How can dialogue contribute to these changes?

*Key features:*
Attends to structure and history in context
Analytically critical, both contextual <u>and</u> generalizable
Eclectic design and methods

*Key concerns:*
Democratic dialogue and action
Power
Challenges of "representing" the Other
Challenges of legitimizing social inquiry

*Sample evaluation questions in "whole language" evaluation context:*
In what ways and to what extent does whole language instruction serve to equalize literacy achievement among children grouped by class, race, gender and other demographic markers of differential achievement in society?
What changes are needed in whole language instruction to further advance democratic aims of education?

6. **Mixed methods designs**

   *Evaluation questions:*
   Any of the above, most likely a mix of the above

   *Key features:*
   Intentional mixing of data and methods at technical, methodological, and/or assumptive (philosophical) levels
   Requires thoughtful attention to purposes and rationales for mixing and then to mixed methods designs that well enact these purposes
   A good match to an evaluation with multiple questions and one or more high-stakes constructs of interest

   *Key concerns:*
   Mixed methods 'theory' still emerging
   Criteria for judging quality and warrant of inferences inadequately developed
   Many evaluators and stakeholders unfamiliar with the domain of mixed methods

   *Sample questions in "whole language" evaluation context:*
   Any of the above, most likely a mix of them