

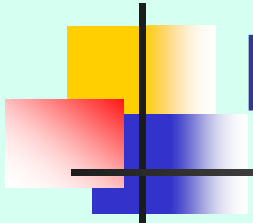


Alternatives and Trade-offs in Generating and Evaluating Evidence: Perspectives from Education

**Madhabi Chatterji
Teachers College, Columbia University**

January 8, 2009

**Institute of Medicine of The National Academies
Committee on Evidence Frameworks for Obesity
Prevention and Decision-making**



Purpose

- **ISSUE 1**
Identifying “what works” in:
 - Education
 - Public health
 - Obesity prevention.....

when **complex, field-based** programs, policies and interventions are the object of inquiry

- What research methods and designs are most likely to yield the **best evidence** on the extent to which interventions are **effective**?
- Are randomized experiments or methods that approximate “true” experiments, the very best by themselves?



Purpose..

- **ISSUE 2**

Making the evidence usable-“actionable”

Informing individual/community action, moving towards social betterment and change

- Sources of evidence?
- Types and forms of evidence?
- Quality of the evidence
- Criteria for appraising the evidence?

Focus of discussion: Research-based evidence



Object of Inquiry

- What are “**Complex, Field-based Interventions**”?

Example in Education:

A **reading program** delivered to a **target population** (say, **grade 5 students**) by **teachers** in particular **classrooms**, nested in **schools**, resourced by **school/district** leaders providing materials, training, staff support and infrastructure.

- **Social Elements**
- **Human Elements**
- **Multi-component “Treatment”-- Intervention Content/Materials, Delivery Processes, and Operations**
- **Organizational Dynamics**
- **Real-time (non-laboratory), multivariate settings**



Object of Inquiry

- What are “**Complex, Field-based Interventions**”?

Example in Mental Health:

Therapy (e.g., anti-depressant drug with counseling supports) delivered to a target patient population by a team of doctors and health professionals, in particular clinics, situated in different hospitals

- **Social Elements**
- **Human Elements**
- **Multi-component “Treatment”-- Medication, Dosages, Supporting Regimens/Routines and Operations**
- **Organizational Dynamics**
- **Real-time (non-laboratory), multivariate settings**



Typical Formulation

Effectiveness/Impact Evaluation Research Designs in Education

- *Population= Grade 5 Students*
- *Intervention (treatment)= X*
- *Expected Outcome=Y*
(E.g., better reading/language arts achievement in students who receive the intervention compared to those who do not)

Causal Link to be Tested:

$X \rightarrow Y$



Typical Formulation

Effectiveness/Impact Research Designs in Mental Health

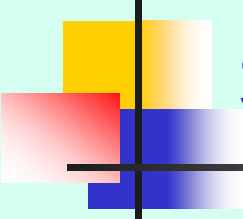
Population = Mentally Depressed Patients

Therapy = X

Expected Outcome = Y (normally-functioning patients with reduced/ no symptoms of depression following treatment)

Causal Link to be Tested

X → Y



Traditional Research Designs for Study Cause/Effect Questions

Comparative Experiments

- **Randomized Controlled Trials (RCT)**

(RCT Designs--also called “randomized clinical trials” in medicine or “true experiments” in social sciences)

- **Quasi-experimental Designs**

(e.g., matched groups, time series, and a range of other design options—next best thing to RCTs)



Policy Context

Call for Evidence-based Education

- Policy mandate to encourage professionals to use interventions and practices that are supported with solid research evidence
- Dichotomous distinction: “Strong” versus “Weak” evidence tied to use of comparative experimental designs
- Endorsement of a hierarchy of research designs with respect to evidence quality
 - RCTs=“strong evidence”
 - Quasi-experiments=“acceptable”
 - Other research designs =“weak” unacceptable evidence

What Works Clearinghouse standards in education (www.w-w-c.org)
- Debate continuing among educational researchers, evaluators and professionals



Main Arguments

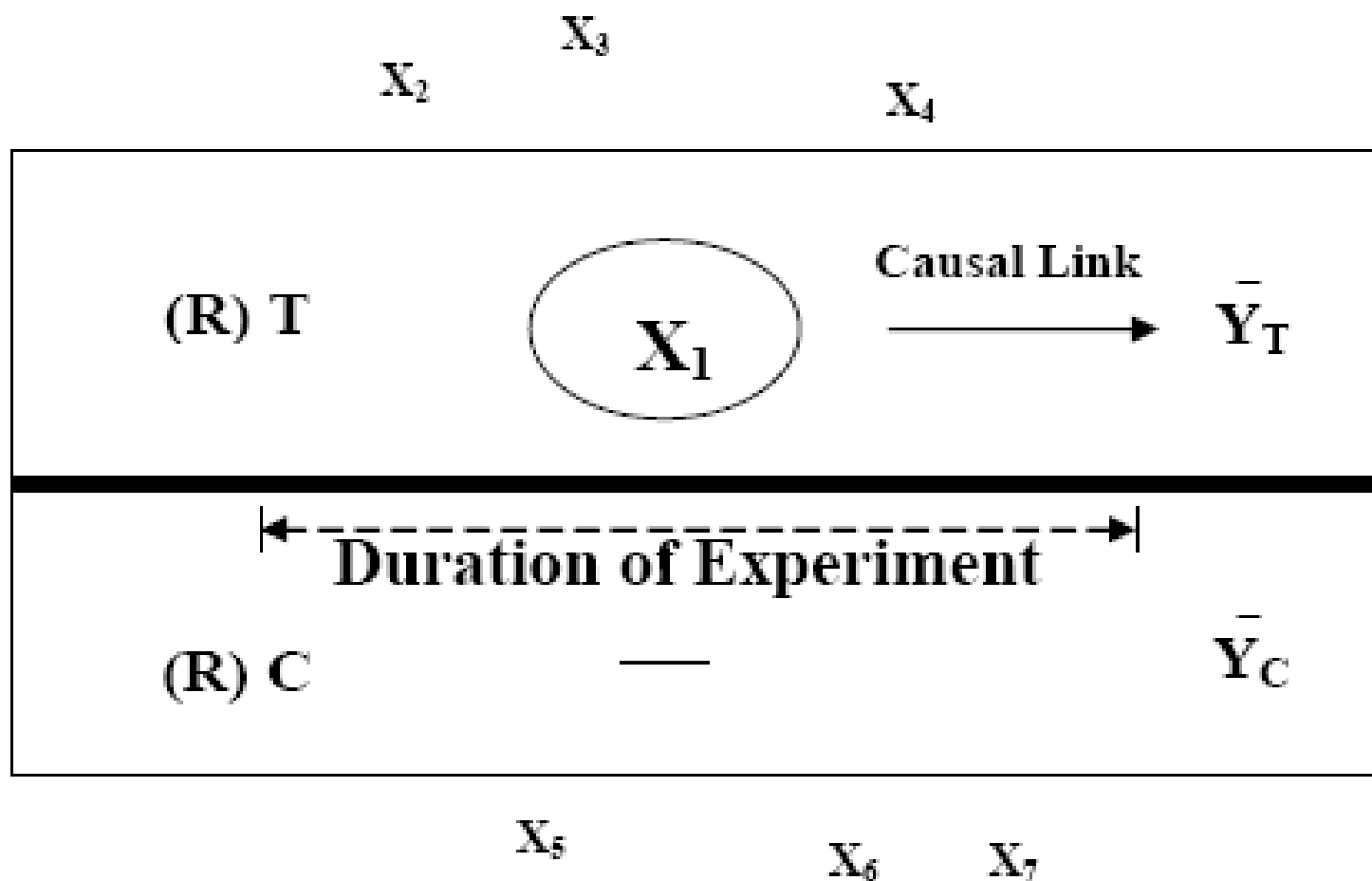
.Impossible to get “strong” evidence, as defined, under actual field conditions as even RCTs get severely compromised; often executed poorly!

-When X is a Complex Social Intervention **NEITHER** Randomized Controlled Trials (RCTs) **NOR** Quasi-experimental Designs are adequate by themselves

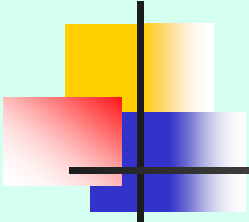
To properly gauge and understand meaning of effects, generalizability or sustainability in “action” environments:

- Need for systemic, multi-phase, mixed-method designs
- Can aspire towards the best **grade of evidence** on effects by scaffolding causal designs with other methods
- Need broader and different criteria and methods for evidence appraisal and synthesis

Figure I: Theoretical Assumptions of RCTs



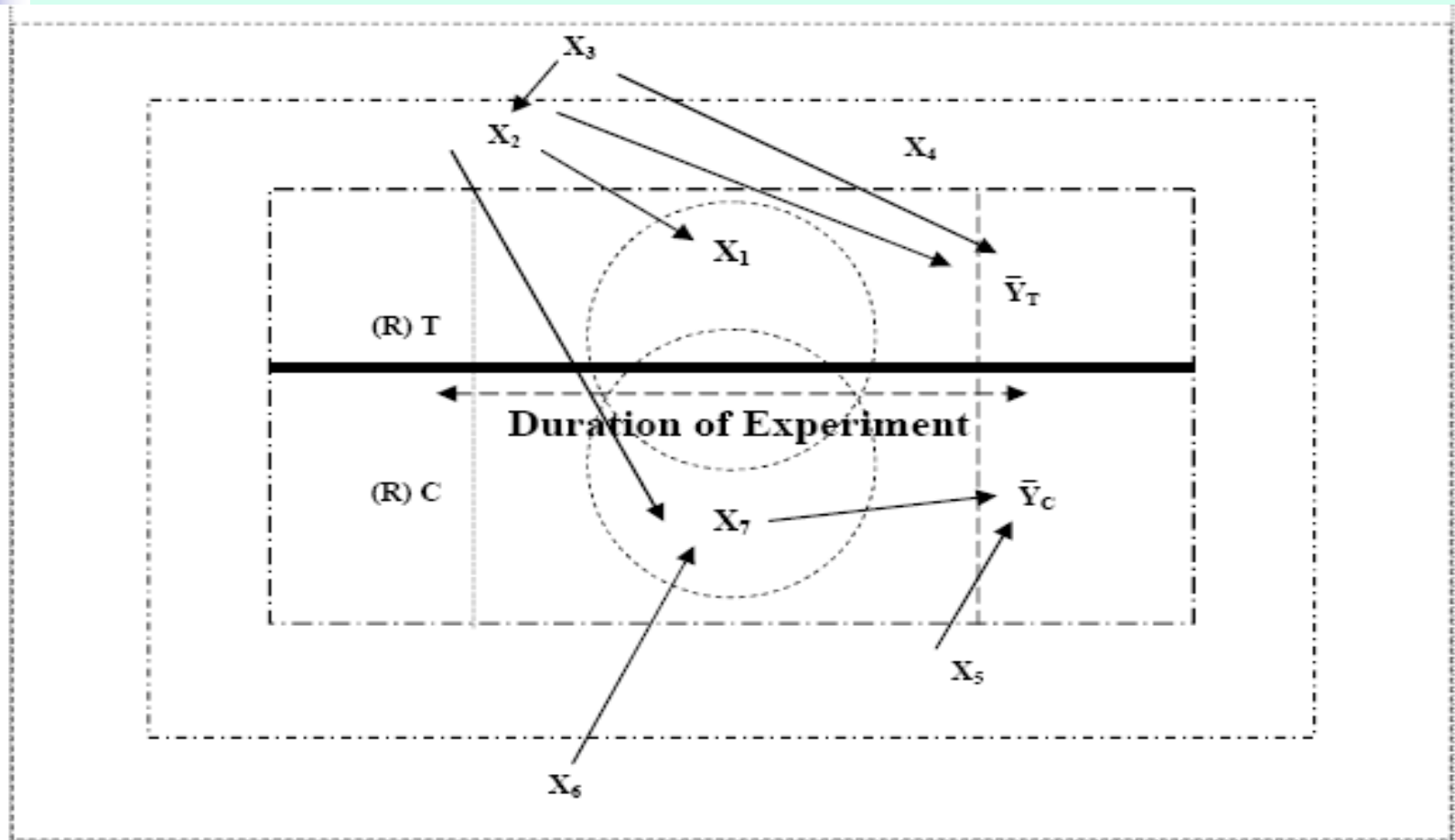
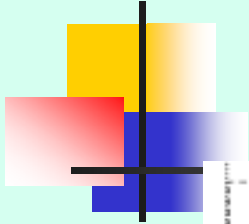
Merits of RCTs When Assumptions are Met....



Random assignment to Treatment (T) groups and Control (C) groups results in:

- equivalent groups as determined by chance factors (same on pre-existing factors that could influence Y)
- independence of cases
- expected mean outcome in both groups is the same before treatment
- optimal conditions for statistical significance tests for testing causal effects of X and Y
- All other variables held constant or controlled to permit an assessment of “net effect” of X

Figure II: Typical Field Conditions for Experiments with Complex Social Interventions





Trade-offs and Some Alternative Design Solutions

1. Broaden research questions to gather more comprehensive data that will better inform evidence-based practice:

“What type of therapy, with what type of client, produces what kind of effect”? After Gordon Paul, as cited by Glass, 2000, p. 8)

“ WHAT is it about this kind of intervention that works, for WHOM, in what CIRCUMSTANCES, in what respects, and WHY?” (Pawson et al., 2005)

PRIMARY EVIDENCE- CAUSAL LINKS

COMPLEMENTARY EVIDENCE-TO DESCRIBE/DOCUMENT INTERVENTION AND ECOLOGY WHEN LINKS ARE TESTED



Trade-offs: Some Alternative Design Solutions

2. **Realistic and More Accurate Definition of Intervention (not BINARY, 1/0)**
 - **Identify, observe, document/measure relevant factors that define intervention and influence delivery dynamics**
 - **Use systemic “logic modeling” as a pre-condition to designing impact studies**
Model pertinent variable pathways based on logical plan of operation— “program theory-driven evaluation designs” (Weiss, 2003; Chen, 2003; Donaldson, 2007)
 - **Monitor and gather evidence on implementation fidelity/stability**
 - **Delay testing for effects until intervention is stable**
 - **Use descriptive, qualitative, and quantitative methods**



Trade-offs and Alternative Design Solutions

3. Document barriers to isolating “Net” Effects of X on Y under field conditions (lowers internal validity)

Dedicate resources to study intervention environment with qualitative and descriptive methods first

Identify factors likely to contribute to “gross” effects (Rossi, Freeman, & Lipsey, 2003); interpret “gross” effects more meaningfully accounting for contributing factors (multiple causation)

Model and test for intervention effects mediated or moderated by relevant environmental factors , X_2 , X_3 , X_4 ...after ecology is better understood

Use covariates or defensible methods to make meaningful comparisons when effects can be tested



Trade-offs and Some Alternative Design Solutions

- **3. Weigh Internal validity and External Validity Issues with User Context in Mind**

Environment adds complexity but gives ecological validity—generates useful evidence for “action” environments

Sample to maximize generalizability of effects over persons, times, settings, conditions (external validity), and track sample size and composition

Use designs with hierarchical analytic models (attention to nesting and cluster sampling)

Use long-term designs to test for change and sustainability of effects



Trade-offs and Some Alternative Design Solutions

4. When controlled experiments cannot be mounted, other quantitative designs can and should be considered

Causal modeling can be reasonably applied to non-experimental (observational) data, large survey databases—don't rule them out!

Approximating randomized experiments with propensity score matching, instrumental variables, regression discontinuity designs—other methods now available.



Summing Up...

Better studies examining effects of complex, field-based programs, policies and interventions, improve the **Grade of Evidence** by using :

- Broader, more nuanced guiding questions
- Multiple methods and types of evidence
 - Primary evidence
 - Complementary evidence
- Multi-phase designs
 - Exploratory studies
 - Confirmatory studies
- Systemic Designs
 - Test multiple causal pathways
 - Document or model intervention, environmental, placebo conditions



References

- Chatterji, M. (2004). Evidence of “What Works”: An argument for extended-term mixed method (ETMM) evaluation designs. *Educational Researcher*, 33(9), 1-13. Reprinted: *Educational Researcher*, 34(6), 13-24.
- Chatterji, M. (2007). Grades of evidence: Variability in quality of findings from effectiveness studies of complex field interventions. *American Journal of Evaluation*, 28 (3), 231-239.
- Chatterji, M. (2008). Synthesizing evidence from impact evaluations in education to inform action: Comments on Slavin. *Educational Researcher*, 37 (1) 23-26.