Evidence-Based Practice: Where do we Stand?

Thomas D. Cook

Northwestern University

Tampa, March, 2007

Some Recent History

- Tony Blair's Electoral Campaign
- George W. Bush's First Electoral Campaign
- Federal Agencies--e.g., FDA, IES, Justice and their lists of effective practices
- Professional Organizations - lists of evidence-based effective practices from which selections are to be made for improving local practice

What is Evidence-based?

- My neighbor tells me that…is also evidence
- Best scientific evidence for a given type of knowledge - e.g. only experiments for cause
- Adequate and permissible for given type-which quasi-experiments should be added
- Organizationally certified as acceptable, but organizations vary in standards,even battling each other--e.g. school violence domain

Some Realities

- Battles over aspects of standards of evidence--in education today, for example
- Ancillary standards need to be applied and so always multiple--e.g., measurement, "exact" replication, transfer to other pops.settings, etc.
- Standards setting is a technical and political process, subject to historical change
- But today relatively well settled in many domains relevant to both generalization and cause
- Debates about how far down preference hierarchy to go; not about best or even second-best methods

Four major Types of Evidence re:

- Entities - have I validly measured the constructs X and Y, and changes in them? Indicators of need, of causal levers and of outcomes of interest
- Co-variation--how certain am I that X and Y truly co-vary?
- Causation--how certain am I that X causes change in Y and that the size of the relationship is Z ?
- Generalization--does X generally causes change in Y; can I indentify the conditions under which X causes change in Y; will X "work" for me?

Purposes of this Talk

- Analyze our capacity to draw strong conclusions about these four matters--how good are our relevant theories about…?
- Analyze our actual research practices to see how well we do in these four areas-to identify where we are dong well and where we need to improve

Conclusions

- We are generally doing well with theories of entitivity, covariation and causation, but not with theories of causal generalization
- Practice in measurement, statistical analysis and causal work is generally OK but can be improved in a few important areas we identify
- Practice about generalizing causal knowledge is lamentable; efficacy study results are often over-generalized. How to make evidence-based mean effectiveness-based rather than efficacy-based?

Is it X or Y?: Construct Validity

- Need a substantive theory of X, specifying its "nature', antecedents and consequences--we validate a theory and not a measure
- Need multiple measures for convergent val., hopefully measured in different ways
- Need cognate measures for discriminant val.-- anxiety is not depression
- Need validation against a criterion not subject to the same suspected biases.
- NCLB: respective role of state tests and NAEP.

Practice Limitation: 1

- Measurement is context-dependent
- NCLB again--mismatch of state and NAEP trend results--why? How high the stakes?
- So which is valid if each is imperfect?
- Golden rule a: All indicators are corruptible.
- Golden rule b: Multi-measurement needed under conditions chosen to vary major sources of anticipated bias--all is impossible

Practice Limitation: 2

- Measuring gaps and changes in gaps
- The Black-White ach gap revisited.
- As raw scores; in standard deviation form; as logs--what's the appropriate metric, especially over long time periods?
- Changing the cutoff point for sanctions

Covariation: are X and Y related?

- Theory of stat sig testing well worked out
- Practice shortfall #1--capitalizing on chance with multiple tests, but improving
- Shortfall #2-- failure to control for clustering, but improving thanks to HLM
- Shortfall #3--size of relationships by developer vs independent researcher, only recent sensitivity
- Shortfall #4. Sample sizes and power- progress
- But our standards for inferring covariation are arbitrary anyway--Fisher and .05 level.

Causation: Does X cause Y?

- Growing institutionalization of RCT--preferred for funding, publishing, training
- Upsurge of interest in cluster-based RCTs a
- New interest in violated contamination assumption
- Huge interest in RDD in theory and practice
- Growing interest in abbreviated ITS as design
- Also in propensity scores for analysis
- Hierarchy at top getting accepted in general--RCT, RDD matching--but not quite everywhere yet

Problems are:

- No consensus on what is good enough causal study
- None of these methods are necessary for causal knowledge--as in history
- RCT not a routine gold standard in practice
- Disaffected practitioners of old methods, esp in education, feel their identify is denied
- Sad reality: Better studies needed because we are in the game of detecting modest effects

Causal Dilemmas now most Evident in Educational Research

- The epistemological shift 30 years ago
- Made reputations since using other methods
- Enough limitations to RCT to make it reasonable to resist gold standard rhetoric
- BUT frequency of RCTs in school prevention work
- Frequency of RCTs in early education work with achievement as the dependent variable
- Role of funding agencies, and other institutions

Within-Study Comparisons

- Nail in the coffin to the resisters
- RCTs vs RDD studies--3 of them
- RCTs vs a priori group matching studies--2
- RCTs versus workhorse design with or without pretest on same scale as outcome-- disappointing however analyzed.

How general is X-Y causal Link?

No general theory of causal generalization
- Two problems--of representation and of extrapolation
- What do most RCTs represent--one version of intervention, one version of outcome, one setting, one time, and one population
- Efficacy trials--contrast with effectiveness
- Developer presence; atypical fidelity?

Extrapolation

- Identify the set of conditions under which will work either thru meta-analysis or identifying causal mediating mechanism
- Meta-analysis of specific programs rare
- Definitively identifying necessary and sufficient conditions (crucial mediators) also very difficult.
- We have a practical problem of extrapolation from effiacy trials to conditions of application of general interest
- And to my local interests

What we need to worry about

- Developer role
- Program specifics that limit implementation -- case of Success For All
- Fidelity vs adaptation dynamics
- Mismatch of outcome heterogeneity in research and often broader in applications
- Scale effects--SFA and gaps
- To date we have evidence based efficacy policy masquerading as effectiveness policy

New Frontier: Evidence-based Effectiveness Policy

- How to do better effectiveness-based research since this is the evidence we need
- We are good at doing efficacy-based research, but is this the evidence we need?
- Example of class size reductions in Calif.

Result: Lists of Recommended Programs for Local Choice

- Growth of prescribing program selection from a limited list--smorgorsbord model to facilitate local tailoring
- Example of violence prevention in schools
- Multiple lists with some similar and some unique criteria
- Some programs on all lists; some on but one
- Agency pressure to include programs they funded

Selecting from the List

- How do I know what best suits my project profile and how this will fit with what I already have?
- How do I navigate between fidelity to the program and adaptations to it to suit my circumstances?
- How do I navigate between what is on the list and what I prefer that is not on the list?
- Where's the booze?
- When can I retire?
- Can Tom Cook help?

Short Answer: No

- These dilemmas are becoming salient again, given the evidence-based rhetoric of today
- We never solved them before in diffusion of innovations literature.
- They constitute the new frontier for research on evidence-based policy.
- But they could only become the frontier once we learned to measure better, to do analyses of association better, and to do causal studies better

Long Answer: Maybe

But we need new theories of causal generalization

We need to do qualitative studies of implementation of efficacy studies

We need more attention, not on whether X causes Y, but on the conditions under which X causes Y.

Our problems are the products of our progress.