

**National Research Council Randomized Field
Trials (RFTs) in Education:
Implications for Research & Practice
Washington, DC**

September 24, 2003

Anthony E. Kelly
George Mason University

Implications for Educational Researchers

- If educational researchers believe in the RFT model and that it can be used successfully in educational contexts, and if they believe that they can draw “causal claims” from its use and that program evaluation satisfies their career ambitions, then:
- They need to consider, afresh, **threats to external and internal validity** (i.e., trading off random selection and random assignment)

Implications 2

- They need to consider the **ethical implications** of withholding a “promising method” over a “business-as-usual” “control” implementation
- They need to consider the **political implications** of such withholding or the possible rejection of “experimental” treatments by some parents.

Implications 3

- They need to consider the potential Federal **political influence** on what problems they study and what method (or combination of methods they use): NCLB; and that their reports may be dismissed (National Reading Panel; NCLB). Ultimately, that the **peer evaluation** aspect of “the scientific method” (NRC, 2002) may be trumped by politics.

Implications 4

- They need to consider that they may **lose the freedom** to determine which measures of “student achievement” (learning?) they can use. Standardized tests (typically produced at the state level) are becoming the de facto standard of learning – yet, these measures have been faulted in many ways (Knowing what students know, NRC). Yet, alignment of the tests with the “State Standards” may not have been demonstrated (or is open to contention). IRT; Rule Space . . .

Implications 5

- They may have to choose to work in the “context of **verification**” rather than in the “context of discovery.” (Schickore & Steinle, 2002). Thus, in areas in which little is known (e.g., how to teach and how students learn statistics), exploratory or descriptive work may have to be eschewed.
- They may feel pressured to accept that the qualitative methods in which they have been primarily trained play, at most, a **supportive role** to RFTs. Many may find themselves at a funding disadvantage.
- The **career implications** of this subservience of methods for new graduates or mid-career professionals is unknown.

Implications 6

- Many who actually conducts RFTs may find (as did their senior professors – many who were trained in experimental psychology, e.g., Ann Brown, Allan Collins) – as did Cook et al., (2000) that RFTs are notoriously difficult to conduct in practice and that the promised “scientific claims with warrant” are more **elusive** than the textbook version of RFTs promises.
- Where do RFTs live in Sloane and Gorard’s (2003) **model-building structure**?
- Where to RFTs live in Bannan-Ritland’s (2003) **research-as-design structure**?

“In theory there is no difference between theory and practice; but, in practice, there is.” Yogi Berra
Cook et al., 2000.

- Designing the “intended” but living with the “achieved” experimental design: “the Chicago Comer effort did not exist in a vacuum” (p. 592)
- weak statistical power (19 schools total), yet HLM analyses were considered
- selective attrition (“selective attrition vitiates the randomized experiment . . . because the treatment is clearly confounded with principal turnover.”)
- being unwilling or unable afford the resources to describe “the control schools”;
- living with clinical confounds “How much was this a Comer effect and how much a Youth Guidance effect?”;
- accepting treatment infidelity, “By the end of 1997, the ethnographers were not willing to classify any school as faithfully following the program guidelines, although some were rated as close.”

Validity, a perennial

- Living with weak measures, “The conclusion depends on the validity of the implementation index, and the one we used meets all the usual psychometric criteria with respect to reliability and face validity. It was created after writing a paper on the program together with its designer (Anderson et al., 1991). It was checked out with both Yale and Prince George’s County Comer officials before being used.” (p. 590).
- [discussion of the implementation index results . . . Then], “The preceding results depend on single items of inevitable questionable reliability” (p. 561).
- See Mestre (2000). “Progress in research: The interplay among theory, research questions, and measurement techniques.”

The Art of the Design of Designing Research

- Having to choose among methods: “It is noteworthy that the program and control schools started at about the same position for the nonachievement outcomes suggesting that the results may well be causal. However, the program school started out behind the controls in academic achievement and then caught up with them, thus creating differences in slopes rather than eighth-grade means. Moreover, the achievement results obtained with the inferentially more powerful longitudinal design were not replicated with the weaker cross-sectional design. So causal inferences about the differences in achievement gain is especially dependent on the design used and the adequacy of the statistical control for selection.” (p. 589).

Living with a lesser form of certainty

- Living with the complex realities of assigned or convenience samples. “At [the Prince George’s County] site, middles schoolers were the target of the intervention, not younger students, the schools were located in suburban rather than in urban areas, program exposure was mostly for 2 years rather than for 6 years, the facilitators were teachers or counselors rather than social workers, and the program variant did not have the emphasis on academic achievement that came to be important in Chicago . . . We cannot be sure which of these many site differences, alone or in combination, account for the different students by site. However, all the presumptions are that the Chicago site is superior.” (p. 589).
- Living without having the luxury of allowing a “treatment” to mature (and hoping for no attrition among teachers or administrators), “Most of the effects on students are not clear until the last 2 program years (out of 6).” See above, also Cline & Mandinach (2000) “The corruption of a research design: A case study of a curriculum intervention project.”

Evidence matters. So does context, design, measurement theory, assumed causal model, social, political and historical factors, volition, motivation, resources, cost, how people learn, knowing what students know ...

- What claim would you choose to replicate in Cook et al.? What are the warrants?
- What is the underlying phenomenon or process that is assumed to be mechanistic enough (or perceptible enough) to be described using a “causal model” arising from this RFT? Is it still an RFT? Has it become a queasy experimental design?
- Even if the intervention “worked,” on what basis can we assume that anyone else would adopt it and would implement it with fidelity and with what resources and for what cost and for how long under what social conditions (i.e., how should RFTs articulate with the diffusion of innovations literature)?

“...teachers’ level of support for SFA [Success for All] did not necessarily predict the degree of fidelity with which they implemented it. Almost all teachers made adaptations to the program, in spite of the developers’ demands to closely follow the model. Teachers supported the continued implementation . . . [though] many teachers felt that the program constrained their autonomy and creativity.” Datnow and Castellano (2000, p. 775).

References

- Bannan-Ritland, B. (2003). The role of design in research: The Integrative Learning Design Framework. Educational Researcher, 32(1), 21-24.
- Cline, H. F., & Mandinach, E. (2000). The corruption of a research design: A case study of a curriculum intervention project. In A. E. Kelly & R. Lesh (Eds.). (pp. 169-190). Handbook of research design in mathematics and science education. Mahwah, NJ: Erlbaum.
- Cook, T. D., Hunt, H. D., & Murphy, R. F. (2000). Comer's School Development Program in Chicago: A theory-based evaluation. *American Educational Research Journal*, *37*(2), 535-597.
- Datnow, A., & Castellano, M. (2000). Teachers' responses to Success for All: How beliefs, experiences, and adaptations shape implementation. American Educational Research Journal, 37(3). 775-799.
- Mestre, J. (2000). Progress in research: The interplay among theory, research questions, and measurement techniques. (pp. 151-168). In A. E. Kelly & R. Lesh (Eds.). Handbook of research design in mathematics and science education. Mahwah, NJ: Erlbaum.

References

- National Research Council (2000). Knowing what students know. Washington, DC: National Academy Press.
- National Research Council (2002). Scientific research in education. Washington, DC: National Academy Press.
- Schickore, J. & Steinle, F. (Eds.). (2002). Revisiting discovery and justification, Preprint 211. Berlin: Max Planck Institute for the History of Science.
- Sloane, F. C., & Gorard, S. (2003). Exploring modeling aspects of design experiments. Educational Researcher, 32(1), 29-31.