CHAPTER 6

Narrow Questions, Narrow Answers: 
The Limited Value of Randomized Controlled Trials for Education Research

ANNETTE LAREAU*

With their emphasis on "scientific rigor" and references to a "gold standard," recent federal guidelines leave little room for debate about the importance of randomized controlled trials for education research.1 In this chapter I offer a critical assessment of this position. I suggest that randomized controlled trials, although useful in selective research situations, are not a helpful research method for the field as a whole. The most important questions in education simply cannot be answered with this approach. I argue that the medical research model has been misunderstood by advocates of randomized controlled trials in education and misapplied to the field of education research.

The limited value of randomized controlled trials in education, and in various areas of medicine, has many sources. I focus on three. One problem is that the research questions that can be asked using this method are relatively narrow and thus cannot accommodate the complexity of the educational process. The method seeks to control all features except for one important variable or "treatment" effect. Yet, as I argue below, educational institutions are multifaceted organizations; there are constant changes in many of the "independent" variables that have an impact on the "dependent" or outcome variables. A further aspect of the complexity problem is that it involves questions of meaning. Randomized controlled trials cannot address this type of "variable." Understanding organizational processes and the meanings participants assign to events is crucial to the development of a rich picture of educational institutions; the effects of educational treatments or variables depend in critical ways on what they mean to students, parents, and teachers, among others. Important research questions—such as the mechanisms through which the position of parents transmits advantages, the processes through which some principals are
more effective than others, or the micro-interactional strategies that build trust among principals and teachers—thus fall outside the methodological realm of randomized controlled trials.

Second, randomized controlled trials pose formidable problems of execution. It is common for researchers to have difficulty selecting a control group that truly acts as a control. In education, even studies blessed with talented staff and generous resources have encountered difficult problems. Researchers cannot sustain core elements of the method, including the expectation that treatment groups are delivered the treatment and control groups do not receive it. Also, randomized controlled trials are based on the assumption that the control group is deprived of the potentially valuable treatment. In a population of schoolchildren, this traditional assumption raises complex ethical concerns. Researchers promoting randomized controlled trials are guilty of a naïve hopefulness that these problems in execution will not overwhelm the research. Yet previous studies (some of which I discuss in this chapter) have not been promising in terms of the ability of researchers to execute core elements of the design.

Third, in their enthusiastic embrace of randomized controlled trials as an education research method, policy analysts have not paid sufficient attention to the crucial issue of implementing research results. Decades of research in education clearly demonstrate that it is not possible for policymakers to mandate the successful adoption of an educational policy. Rather, successful reform requires that educators “buy into” the new policy. In medicine, too, “buy-in” significantly affects the likelihood of achieving change. Even when medical research demonstrates that many lives and millions of dollars can be saved by a relatively simple innovation, the necessary change may be resisted if it runs counter to the logic and culture of the organization. Moreover, as I discuss below, in both education and medicine, promising reforms may be difficult or impossible to implement or “scale up” beyond the initial research setting.

In sum, there are formidable limits to the randomized control approach—limits which the proponents of this methodology have not sufficiently recognized. To achieve “rigor” in education research, we need to adopt a more realistic grasp of the problems that plague attempts to implement research results as well as greater openness to other approaches that offer crucial insights into social processes in organizations. For these purposes, qualitative methods, including participant observation and in-depth interviews, are likely to be the most promising.

**What Kinds of Questions Can Be Answered?**

Randomized controlled trials involve the random assignment of participants to different experiences or “treatments,” with all other components of the experience “held constant” (Mosteller and Boruch 2002). A close cousin to a randomized control trial is a study that compares individuals who received a “treatment” with individuals who sought but failed to get on the waiting list. As the materials in this section demonstrate, randomized controlled trials have been in vogue for recent years. The authorizing legislation for the importance of using “scientifically rigorous methods” (Cook and Payne 2002:177) provides evidence of the impact of this methodological preference.

The superiority of random assignment in educational research is well acknowledged in philosophy, in the sciences, and in education. A well-designed experiment is superior to a casual study, for example, in determining the cause and effect relationship of variables. In the social sciences, evidence from randomized controlled trials is almost always considered the gold standard by the publishing community. The randomized experiment requires random assignment of units to experimental conditions, and the random assignment of units to experimental conditions is an essential component of this approach. So, if experiments can be conducted, they provide a better warrant for such claims than the only alternative, a correlational study. It requires, however, that the researcher can actually randomize the assignment process. This is often more difficult than one might think, and if they do, the results can be crisp, clean, and causally compelling to the researcher, who can then claim to have demonstrated the impact of an intervention. For this reason, randomized controlled trials are often conducted in controlled environments, such as in medical studies where subjects are randomly assigned to receive a new treatment or a placebo. In educational research, randomized controlled trials are often conducted in schools or classrooms where teachers randomly assign students to receiving instruction in a particular format or curriculum. In this way, the researcher can ensure that the treatment conditions are equivalent for all students, allowing for a more direct comparison of the effects of different interventions. However, there are challenges to conducting randomized controlled trials in real-world settings, such as schools or classrooms, where teachers and students may differ in ways that could affect the results. These challenges include logistical issues, ethical considerations, and the potential for bias. Despite these challenges, randomized controlled trials remain a powerful tool for educational research, providing rigorous evidence of the effectiveness of different interventions.
"treatment" with individuals who sought the same treatment but remained on the waiting list. As the materials in the appendices to this volume clearly demonstrate, randomized controlled trials have skyrocketed in popularity in recent years. The authorizing legislation for No Child Left Behind emphasizes the importance of using "scientifically based research" methods (see Appendix A). Cook and Payne (2002:177) provide a ringing endorsement:

The superiority of random assignment for drawing inferences about the consequences of planned change attempts is routinely acknowledged in philosophy, medicine, public health, agriculture, statistics, micro-economics, psychology, criminology, prevention research, early childhood education, marketing and those parts of political science and sociology concerned with improving opinion surveys... This article does not argue that correct causal conclusions come only from experiments. It does argue, though, that experiments provide a better warrant for such conclusions than any other method. So, if experiments can be conducted in schools, they should be. Not to use them requires a very strong justification.

It is relatively rare that social scientists successfully carry out a randomized controlled trial in a high-quality, thoughtful, and well-executed study. But when they do, the results can be crisp, clear, and causally irrefutable. Indeed, when a trial is appropriately designed and conducted, and when its purpose is specifically to evaluate the effect of an intervention or treatment, the results can be stunningly appealing. One example is an evaluation of the Big Brother/Big Sister program in Philadelphia, conducted by the non-profit group Public/Private Ventures (Tierney, Grossman, and Resch 2000). The study compared those on the waiting list for a Big Brother/Big Sister with those involved in the program. The researchers found that program participants had less drug use compared to those on the waiting list. Since the problem of self-selection has been effectively handled in this design, the results are convincing. The program has a positive impact. Similarly, in a series of thoughtful studies the Moving to Opportunity and New Hope reforms used a random assignment of services for low-income families to rent an apartment with a combined quantitative and ethnographic research approach (Duncan, Huston, and Wessner 2007; Kling, Lieberman, and Katz 2007). This random-assignment study clearly showed that the housing service that placed families in less economically depressed neighborhoods improved mothers' mental health by reducing their rates of depression. The housing service also had an impact on participants' obesity, but not on other measures of physical health. The neighborhood contest did not raise adults' income, however, and the results for children's school performance were mixed: generally quite positive for girls but quite negative for boys (Kling et al. 2007). Importantly for my purposes here, the randomized controlled trial itself could not shed light on why effects differed for girls and boys. For insights
into why gender might have had an impact, the study team turned, appropriately, to ethnographic research (Clampet-Lundquist et al. 2006).

Even when conducted in tandem with qualitative investigations that are better suited to answering why and how questions, random-assignment studies have formidable built-in constraints. They must isolate one dependent variable and control all other factors except the "treatment." Examination of interaction effects (i.e., where the results are contingent upon one set of circumstances but not another) is possible, but cumbersome, as Cook and Payne (2002:152) acknowledge:

> At their most comprehensive, [random-assignment] experiments can responsibly test only a modest number of the possible interactions between treatments. So, experiments are best when a causal question involves few variables, is sharply focused, and is easily justified.

The models also presume that during the study period, conditions are static, not dynamic. Educational institutions do not conform to these basic assumptions. Schools are complex organizations with many different levels, including the classroom (and student groupings within the classroom), the school, and the district. Many of these levels are in flux, particularly in a climate of educational reform (Hubbard, Stein, and Mehan 2006; Sunderman, Kim, and Orfield 2005). Dynamics in classroom learning are shaped by many factors; it is difficult to delineate sharply one causal question.

Randomized controlled trials focus on factors that can be manipulated via policy intervention. But crucial aspects of social life that affect education cannot be manipulated this way. For example, there is an extensive body of research on the friendship networks of youth and the influence of peers on educational aspirations, homework patterns, and other aspects of schooling (Hallinan 2006). Students cannot be assigned to different friendship networks by a research study, let alone be assigned randomly to peer groups by such researchers. Similarly, social-class differences in child rearing, in parents' knowledge about educational institutions, and in the flexibility parents have for attending school events during work hours are not readily subjected to manipulation (Entwisle, Alexander, and Olson 1997; Lareau 2000, 2003). Arguably, many classic studies in education would never have been undertaken in the current research climate, since the topics these studies have addressed—peer networks, physical attractiveness, immigrant status, and students' aspirations—would not be amenable to a randomized controlled trial approach. Further, decades of research on educational achievement have shown that it is the elements least open to external manipulation (especially family background and peer networks) that are the most important influences on student learning (Hallinan 2006). In randomized controlled trials the scope of questions that can be studied is narrowed considerably, leaving unstudied precisely the factors that prior research has identified as key influences on educational outcomes.

The "narrow questions/narrow answer" approach used in clinical trials does not work for many other problems. For example, drug trials with a double-blind placebo control have yielded many important insights into treatment effectiveness. But randomized controlled trials are not the answer to many of the most important questions. For example, the critical problem is that many patients (Haynes, McDonald, and Garg 2002; and Education 2007) and residents are non-compliant (or non-adherent) with their treatments at the time(s) specified. More important, patients with psychological disorders (e.g., attention deficit disorder) or patients with manic depression (Olfson 2007) often do not take their medicine. Information and Education 2007). In both cases, the difficulty individual patients face in purchasing affordable insurance, and residents of New York City and Los Angeles—among today's most important policy questions—require much more than an experimental approach to research based on the methods of Rubin 2007).

Nor is random assignment useful to evaluate policies as a result of malfunction in several important cases:

- Natural disasters: For example, the effects of the death of the elderly, the para-military services, the social isolation of the elderly, the disfigurement of the elderly, and the health-care crisis by political organizations have been paraded through the literature (Klinenberg 1998). Similarly, the effects of multiple institutions, of work, and of education on the lives of children have been studied in detail.

Moreover, the randomized controlled trial is not well-suited to examining events in the lives of individuals. Take, for example, the extensive research that has been done on parent involvement in student learning. Parent involvement as an important influence on student learning is established. There is little doubt that teachers value parent involvement. They are not regular students, and they are not regular parents. They are not regular school volunteers. They are not regular school administrators. They are not regular school staff. They are not regular school board members. They are not regular school trustees. They are not regular school counselors. They are not regular school teachers. They are not regular school principals. They are not regular school superintendents. They are not regular school board members. They are not regular school staff. They are not regular school trustees. They are not regular school counselors. They are not regular school teachers. They are not regular school principals. They are not regular school superintendents. They are not regular school board members. They are not regular school staff. They are not regular school trustees. They are not regular school counselors. They are not regular school teachers. They are not regular school principals. They are not regular school superintendents. They are not regular school board members. They are not regular school staff. They are not regular school trustees.
The "narrow questions/narrow answers" problem is also found in medicine. For example, drug trials with a double-blind design, a placebo, and controlled conditions have yielded many important insights about which drugs are most effective. But randomized controlled trials have not provided insight into arguably the most important questions in medicine today. For example, a critical problem is that many patients are non-compliant in the taking of drugs (Haynes, McDonald, and Garg 2002; National Council on Patient Information and Education 2007). Research suggests that approximately one-half of the time patients are non-compliant (or non-adherent) in taking pills or other medical treatments at the time(s) specified. Four-compliance rates are particularly high for psychological disorders (e.g., an 80% rate of non-compliance among patients with manic depression) (Osterberg and Blaschke 2005). Doctor-patient communication problems and patients' failure to grasp the nature of their medical problem also are ubiquitous (National Council on Patient Information and Education 2007). The high cost of medical insurance for employers, the difficulty individuals with chronic health problems have purchasing affordable insurance, and the relatively large medical costs U.S. residents shoulder compared to their Western European counterparts are among today's most important policy issues, yet they cannot be addressed through research based on the method of random assignment (Furman and Rubin 2007).

Nor is random assignment useful in understanding health crises that occur as a result of malfunction in several institutions at the same time, as is often the case with natural disasters. For example, in his analysis of the hundreds of heat-related deaths in Chicago during one week in 1995, Eric Klinenberg (2002) found that multiple social institutional factors contributed to the high rates of death of the elderly: the para-military organization of the emergency health services, the social isolation of the elderly, lack of cooling facilities in the apartments and buildings housing the elderly, and the slow identification of the health-care crisis by political organizations and the media. Many of these patterns resurfaced in the Hurricane Katrina disaster. These disasters' rare nature as well as the multiple institutions involved in shaping the (inefficient) response do not permit study via the method of random assignment (Erickson 1978; Klinenberg 2002). Similarly, schooling is a complicated process that is affected by multiple institutions, of which only one is schools themselves.

Moreover, the randomized control trial gives short shrift to the meaning of events in the lives of individuals. Take, for example, the vast amount of research that has been done on parent involvement in schooling—research that identifies parent involvement as an important influence on children's achievement. There is little doubt that teachers value parent involvement highly (Epstein 2002). They call for "partnerships" in education with parents. Most surveys also reveal that parents want to take an active, helpful role in their children's schooling.

There have been interventions to increase parent involvement in schooling; parent involvement was promoted in the What Works Clearinghouse (U.S. National Library of Medicine).
Department of Education 2006). But these studies do not sufficiently recognize
that social class appears to influence what parents mean by the term "parent
involvement." For example, in selecting schools, middle-class parents appear to
collect very detailed information about educators, and these parents oversee
their children's educational lives closely (Diamond and Gomez 2004; Lareau
2000). Working-class parents, on the other hand, also see themselves as heavily
involved in schooling, but they often interpret involvement to mean preparing
children to go to school and deferring to the educational professional expertise
of educators (Lareau 2000, 2003).

It is important to remember that the study of what people do is only one
piece of the puzzle. Complex social processes undergird educational reform.
Small, intensive, non-random case studies are crucial for identifying these
processes and shedding light on the conditions associated with successful
reform initiatives.

Barriers to Randomized Controlled Trials

In practice, the emphasis on "rigor" incorporated into No Child Left Behind
has come to be equated with a commitment to and reliance on randomized
controlled trials (see the appendices to this volume). Indeed, Institute of
Education Sciences Director Grover "Russ" Whitehurst has termed randomized
controlled trials the "gold standard." The approach has many advocates.
Yet despite the desire of researchers to implement a medical model, there are
problems. One is difficulty in selecting and sustaining control groups.

When Is a Control a Control?

Sustaining the quality and quantity of the control group is challenging, as the
authors of a review of diabetes research note (Montori et al. 2006):

In conducting systematic reviews of RCTs [randomized controlled
trials] in diabetes we have noticed that researchers seem to pay little
attention . . . to methodological safeguards that limit the introduction
of bias into RCTs. As a result, these potentially biased RCTs could
mislead clinicians. . . . When reports leave out critical information
about methodological safeguards against bias, readers cannot ascertain
if these safeguards were present.

(p. 1833)

The authors reviewed RCTs published in general medical journals (e.g., the New
England Journal of Medicine), in those that specialize in diabetes research (e.g.,
Diabetes), and in those that address metabolics and nutrition (e.g., American
Journal of Clinical Nutrition). They described their evaluation criteria this way:

We considered trials to be of low methodological quality when they
had three or more of these criteria: inadequate (or not reported)
allocation concealment, inadequate (or not reported)
patients and caregivers, failure to blind (or not report)
failure to blind (or not report)
participants, or a reporting bias (or not report)
participants, or a reporting bias (or not report)
outcome measures, or failure to describe (or not describe)
outcome measures, or failure to describe (or not describe)
outcome measures, or failure to describe (or not describe)
outcome measures, or failure to describe (or not describe)
outcome measures, or failure to describe (or not describe)
(p. 1833)

The authors judged 53% of the included studies as low-quality trials.
That figure is itself a cause for concern. They focused on a broad array
of studies across a broad array and found a higher rate of low-quality
studies. Similar quality-related problems are found among the most basic assumption
that the external validity of studies in the experimental or control
groups. While the experimental group will not report the treatment "dosage" will be
changed in the field of education, these assumptions are difficult for
researchers who have attempted to use randomized
trials.

When educational reforms are implemented elsewhere, for example, the
promoting education at-risk children is sought to improve education for
these children. Implementing a study based on historical data and
tially difficult in a tumultuous environment. It was experiencing
disturbances in the reform effort, the Chicago school environment. It was experiencing

Thomas Cook was among those who
reporting the results of these research
carried out the project at the Chicago
sponsored a similar study to evaluate
development program.

Unfortunately 5 of the 24 school
students F participate in the program, at
appointed who did not want to
continued. Of the Phase II school
allocation concealment, inadequate (or not reported) blinding of
patients and caregivers, failure to adhere to the intention-to-treat
principle, or a reporting >10% (or did not report information to
calculate) of loss to follow-up.

(p. 1834)

The authors judged 53% of the 199 trials they evaluated to be of "low quality." That figure is itself a cause for concern, but the authors suggest that it may be an underestimate. They focused on work published in "top journals"; a review of studies across a broader sample of scientific journals would likely show a higher rate of low-quality studies (Montori et al. 2006).

Similar quality-related problems have surfaced in education studies. Indeed, among the most basic assumptions of any randomized trial are that the control group and the experimental/treatment group are highly similar before the start of the study, that the experimental group will receive a full course of treatment, while the control group will not receive any exposure to the treatment, and that the treatment "dosage" will be delivered consistently over time. In school settings these assumptions are difficult or impossible to meet, as researchers who have attempted to use randomized controlled trials have discovered.

When educational reforms prove successful in specific settings, they often are implemented elsewhere. For example, James Comer's complex reform for promoting education of at-risk youth was adopted in Chicago. The program sought to improve education through community involvement, intensive parent involvement programs, and community programs. The introduction of this highly regarded reform into the Chicago public schools was subject to an unusually elaborate array of randomized controls and program evaluation. Implementing a study based on randomized control methodology is exceptionally difficult in a tumultuous organizational environment, however. At the start of the reform effort, the Chicago public school system was just such an environment. It was experiencing high rates of teacher turnover, changing district politics, changes in administrator leadership, and turnover in the student population.

Thomas Cook was among those studying the reform. In published papers reporting the results of this research, he and his colleagues (to their credit) explicitly acknowledge some of the barriers they faced trying to use a randomized control approach to evaluating the impact of the Comer school development program:

Unfortunately 5 of the 24 schools dropped out of the study at different times for various reasons. . . . In two cases, a new principal was appointed who did not want the program; in the third, the principal did not want the quantitative research component (although his school stayed in the program and the ethnographic component continued). Of the Phase II schools, only one dropped out, [but it was]
almost immediately after learning it had been assigned to control status... (Cook, Murphy, and Hunt 2000: 544)

Such selective attrition vitiates the randomized experiment because more Comer than control schools dropped out and because the treatment is clearly confounded with principal turnover. (pp. 544-545)

That "a new principal... did not want the program" and that one-fifth of the schools dropped out of the study for periods of time are critical problems. Other difficulties Cook and his co-authors do not mention include researchers' lack of control over subject turnover (in addition to student and teacher turnover, top leadership at participating schools can change); and researchers' limited power either to shore up principals' flagging interest in participating in a study or to overcome their initial resistance to such participation. As a result, randomized controlled trials that focus on longitudinal change in schools face daunting, and arguably insurmountable, challenges.

Of even greater concern is the fact that the double-blind character of the medical model, made possible by a clinical environment in which a treatment pill and a "sugar pill" look identical, has no counterpart in education. In the social and political world of schools, there is no way to create a double-blind reform. School reforms are discussed by teachers, principals, district leaders, and, in some schools, by parents and students as well. These parties do not live in a vacuum, and educators in the "control" and the "experimental" schools have social and professional relations. As a result, studies are easily contaminated: In a real-world educational setting, it is not possible for researchers to prevent the control group from being exposed to the treatment. As Cook et al. (1999) acknowledge in the Chicago Comer reform study:

Three comparison schools borrowed some program elements: in one case, a husband and wife worked in different schools, one with and one without the program, in another case the daughter of a senior New Haven program principal was worked as an administrator at a control school; and, in the last case, a comparison group principal likewise the program, studied it for himself, and discussed it with principals from program schools. And further diffusion between program and control schools probably occurred during district-wide in-service training sessions conducted by the county program coordinator. (p. 584)

The fact that a principal in a "control" school liked the program and the program was diffused in district-wide training sessions is a sign of the social character of education. Still, Cook and his co-authors discount the impact of this source of contamination on their results:

But we judge the degree of directness of the major program elements, the three schools, namely the in-school training sessions, and training on trial reclassifying these three controls of mid-student effects.

But it is clear that the most basic element of a new principal's Есть не adequately met. Noto capability would be consistently defined.

Midway through the study the policy was putand his appointees... reduced.

The new emphasis made test school effectivity (rather than management). Because of the schools were put on probation.

Hence, the vigor with which basic reform across time. This pattern has been a little.

In sum, medical research shows what was found for subjects in "treatment" and "control" comparison. Large numbers of subjects are vastly more complex and "natural evaluation had an ethnographic character." (Before the end of the study) the data of the programs as faithfully performed, although some were very close. (Concern being faithfully followed, then it is developed.

Translating Legislation into Practice Implementation

Finally, it is worrisome that the programs of educational reform have paid so little intervention. There is ample evidence of policy and its implementation. (Mehan 2006; McLaughlin 1987).
But we judge the degree of diffusion from these sources to have been relatively minor. The three schools in question did not have access to the major program elements that build and sustain the program in a school, namely the in-school facilitator, local program in-service training sessions, and training sessions in New Haven. Moreover, reclassifying these three control schools made no difference in analysis of student effects. (p. 584)

But it is clear that the most basic element of the design, notably that there would be a distinctly different experience for “control” and “experimental” groups, was not adequately met. Nor could researchers be confident that the “treatment” would be consistently delivered to the treatment group.

Midway through the study the political context changed. The mayor and his appointees . . . reduced the power of local school councils . . . . The new emphasis made test scores the dominant criterion for judging school effectiveness (rather than success in creating decentralized management). Because of their low test scores, one sixth of all city schools were put on probation. (Cook et al. 2000: 541)

Hence, the vigor with which basic elements of the reform were enacted differed across time. This pattern has been found in many other studies.

In sum, medical research shows that it is difficult to sustain trials where subjects in “treatment” and “control” groups have radically different experiences. Large numbers of subjects are frequently non-compliant. School settings are vastly more complex and “noisy” than medical studies. Indeed, the Cook evaluation had an ethnographic component which offered a powerful critique: “[Before the end of the study] the ethnographers were not willing to classify any of the programs as faithfully following all of the program guidelines although some were very close” (Cook et al. 2000: 564). If the programs are not being faithfully followed, then it is difficult to accurately evaluate their effectiveness.

Translating Legislation into Practice: The Problem of Implementation

Finally, it is worrisome that the proponents of randomized controlled studies of educational reform have paid so little attention to the nature of the policy intervention. There is ample evidence that the road between the passage of policy and its implementation is long and rocky (Hubbard, Stein, and Mehan 2006; McLaughlin 1987). As Milbrey McLaughlin (1987: 172) has pointed out:
The overarching, obvious conclusion running through empirical research on policy implementation is that it is incredibly hard to make something happen... It's hard to make something happen primarily because policymakers can't mandate what matters. We have learned that policy success depends critically on two broad factors: local capacity and will... But will, or the attitudes, motivation, and beliefs that underlie an implementer's response to policy's goals or strategies, is less amenable to policy intervention.

“Will” cannot be mandated but it plays a crucial role in implementation. Yet advocates of randomized controlled trials have failed to come to terms with the fact that the factors that can easily be manipulated experimentally, and thus whose effects can be studied in a randomized controlled trial, may be precisely those elements that matter the least in the grand scheme of schooling and learning.

Literature on implementation of public policy also suggests that local context significantly shapes implementation. For example, in the randomized trial evaluation of the Chicago School Development Program researchers were unable to achieve a standardized implementation. Instead, the reform varied to “maximize the fit to local circumstances” in Chicago. Cook explains:

Elements of the implementation of [Come's] SDP [School Development Program] vary by both district and school in order to maximize the fit to local circumstances. The Chicago program is unique in many ways. It was introduced in the middle of a citywide educational reform initiated by local politicians, businesses and philanthropies... At first, the reform emphasized local school councils; ... councils were encouraged to set their own school goals and establish their own instructional designs... [although this pattern]... did not arise in practice.

(Cook et al. 2000: 541)

If policy implementation is not standardized, then the validity of an assessment of its impact using a randomized controlled trial is compromised.

Several studies suggest that principals play a critical role in policy implementation. One study of “teacher accountability” in California reported:

Our research revealed that principals often played a pivotal role in how teachers experienced accountability policies. Principals either acted as a “buffer,” shielding teachers from test-score pressures, or as an added source of pressure for teachers, emphasizing the need to raise test scores and the school's ranking.

(PACE 2006: 8)
The role of principals appeared to be key in understanding how teachers experienced a reform:

Principal's attitudes towards accountability and their leadership styles also influenced teachers' experiences with district and state mandates. Principal's emphasis on or avoidance of certain district and state instructional reforms, for example, were reflected in teachers' acceptance or critiques.

(PAGE 2006: 8)

The authors of this study also note that rather than simply looking at schools, we need to look more directly at the mediating role of districts:

Currently, accountability policy assumes that individual schools are the primary locus of reform. While school-based reform is essential, our research shows that districts, through various programs and policies, have the power to facilitate or hinder school improvement. The state must recognize district influence in the implementation of accountability policy.

(PAGE 2006: 9)

The message is clear: There are many mediating factors that shape the implementation of educational policy. This is why it is crucial that reform initiatives incorporate procedures and "feedback loops" so that during the implementation, mandated changes can be modified to accommodate institutional constraints.

The more fundamental issue, however, is the degree to which institutions will respond to results produced by randomized controlled trials if the implementation process is not closely monitored. Here too, the evidence from medicine is not encouraging. Consider, for example, the surprisingly large numbers of hospital patients who are exposed to an infection from germs on the hands of doctors and/or nurses. Research by Sean Beensholtz and colleagues (2004) and reporting by Atul Gawande (2007) indicate that contamination of "lines" (such as a line or a catheter) from the germs on caregivers' hands can lead to infection. In intensive care units, patients who have a line in for more than ten days have an infection rate of 4%. Fatalities from line infections range from 5% to 28%, depending on how sick patients are when they become infected. Those who survive line infections spend more days in intensive care. Thus such infections are not a serious problem. They result in more deaths, more days in the hospital, more dollars spent on medical costs, and more days of missed work.

Johns Hopkins researcher and physician Peter Pronovost and his colleagues designed a policy intervention focused on the steps that health practitioners
take when putting in a line: they wash their hands with soap, clean the patient’s skin with an antiseptic, put sterile drapes over the patient’s entire body, wear a sterile mask, hat, gown, and gloves, and put a sterile dressing over the catheter site once the line is in. The researchers discovered that about one-third of the time, the clinicians skipped one of these essential steps (Berenholz et al. 2004; Gawande 2007). Eventually, Dr. Pronovost’s team developed a checklist for a new protocol that empowered nurses to stop doctors to question them if they missed a step. Infection rates plummeted. Compared to an earlier period, the ten-day line infection rate went from 11% to 0%. In a 15-month period, there were only two line infections. Thus, adhering to the new protocol probably prevented 43 infections and eight deaths, and saved $2 million in costs (Berenholz et al. 2004).* 

Yet when hospital administrators attempted to "scale up" this new model and implement it widely, the policy intervention was resisted rather than embraced (Berenholz and Pronovost 2003; Gawande 2007). This happened, in part, because, as in schools, the intervention required a change in organizational processes. The hierarchy of authority in medicine has been weakened due to changes in insurance and managed care. Nevertheless, doctors have considerable power; nurses have lower professional status and autonomy compared to doctors. The checklist required that nurses or technicians stop the procedure (in a hectic, often understaffed intensive care units) to ask, "Doctor, have you washed your hands?" This reversal of the usual hierarchy of authority was difficult to implement. In addition, many hospitals and clinics have been subject to severe budget cuts and staffing reductions, both of which make routine procedures more rushed and thus undermine attempts to add even a simple new safeguard to the standard repertoire.

How long it will be before the average doctor or nurse is apt to routinely use a checklist? Peter Pronovost is not optimistic: "We could get I.C.U. checklists in use throughout the United States in two years, if the country wanted it. [but] ... [a]t the current rate, it will never happen" (quoted in Gawande 2007: 8). The resistance is rooted not in science or technology but in social and institutional factors:

If someone found a new drug that could wipe out infections with anything remotely like the effectiveness of Pronovost’s lists, there would be television ads with Robert Jarvik extolling its virtues, detail men offering free lunches to get doctors to make it part of their practice, government programs to research it, and competitors jumping in to make a newer, better version. That’s what happened when manufacturers marketed central-line catheters coated with silver or other antimicrobials—they cost a third more, and reduced infections slightly—and hospitals have spent tens of millions of dollars on them. But with a checklist, what we have is Peter Pronovost trying to see if maybe, in the next year or so, New Jersey will give his idea a try.

It is rare that a policy intervention takes the resistance to its implementation for educational researchers. It is hard to see how the findings are not enough; teachers, parents, and the media are "brought along," to be helped by another approach is na"}

**Concluding Thoughts**

There are variations in how stringent the criteria for randomized controlled trials. But in Chapter 1, the use of this methodology was not the focus. Increasingly, funding priorities are shifting towards training resources privilege randomized controlled trials and academic career assessments often focus on "quality" of education research. There is a trend toward increased accountability in education, and that different questions with a randomized controlled approach among many in the repertoire.

It also would be a less troubling manner to trolled trials offered a more clear-cut result with carrying out randomized controlled trials of the combination of a narrow vision in reference to RCTs as a "gold standard" ways, inferior. That such sweeping structurally limited opportunities schizophrenia controlled trials of great concern. Editors to do research. The conditions of the sometimes dramatically.

It is also troubling that advocates of that, even if conditions could be met, methodology can accommodate is not education research, as in medicine, an all approach. Key factors in education research does not make them any less important processes of educational delivery system of a random assignment approach does.
see if maybe, in the next year or two, hospitals in Rhode Island and New Jersey will give his idea a try.

(Gawande 2007: 7)

It is rare that a policy intervention is as clearly effective as the checklist. Thus the resistance to its implementation is a particularly important cautionary tale for educational researchers. It is hard to implement research results. Valid findings are not enough; teachers, parents, students, and administrators need to be “brought along,” to be helped to “buy into” the proposed change. Any other approach is naive.

Concluding Thoughts

There are variations in how strident researchers are in insisting on the need for randomized controlled trials. But, as Pamela Barnhouse Walters has noted in Chapter 1, the use of this methodology in education research has made headway. Increasingly, funding priorities for research grants and for professional training resources privilege random assignment; journal reviews, grant reviews, and academic career assessments consider this criterion in evaluating the “quality” of education research. The term “gold standard” is common.

This trend would be less alarming if federal department of education decisionmakers acknowledged that there are many varied research questions in education, and that different questions call for different methods. From this perspective, randomized controlled trials would be one, and only one, approach among many in the repertoire available to social science researchers. It would also be a less troubling movement if advocates of randomized controlled trials offered a more clear-eyed assessment of the problems associated with carrying out randomized controlled trials. Thus what is problematic is the combination of a narrow vision and the belief, signaled by the frequent reference to RCTs as a “gold standard,” that all other methods are, in crucial ways, inferior. That such sweeping claims are made despite the flawed and structurally limited opportunities schools offer for carrying out randomized controlled trials is of great concern. Educational institutions are not ideal places to do research. The conditions of the “laboratory” change frequently, and sometimes dramatically.

It is also troubling that advocates of RCTs do not sufficiently acknowledge that, even if conditions could be made ideal, the range of questions this methodology can accommodate is narrow. Many important questions in education research, as in medicine, are not amenable to a random assignment approach. Key factors in education research are difficult to manipulate, but this does not make them any less important. Studying the meaning and social processes of educational delivery systems is crucial, and it is an area advocates of a random assignment approach do not sufficiently acknowledge.
The current movement toward the use of randomized controlled trials in education research fails to offer an umbrella sufficiently broad to both welcome and respect diverse methodological approaches. This narrowness of spirit is not only a tactical mistake. Its short-sightedness will, in the end, impede random assignment advocates from realizing the educational reform goals they hold most dear.

Notes


2. Given the evidence that exists about the effectiveness of policy interventions of the last 40 years, it is foolishly for search for a single policy intervention or "silver bullet" to improve the educational achievement for all American children in all schools all of the time. There are also likely to be interactions. Doris Entwisle and Karl Alexander at the Johns Hopkins University, for example, showed that teachers who themselves had low socioeconomic origins did particularly well in producing achievement from students with low socioeconomic origins; these teachers from low socioeconomic origins outperformed other teachers who had been raised in more prosperous families (Entwisle, Alexander, and Olson 1997). Thus the impact of policies that are universally implemented is not universal in the student population. Some students benefit more than others.

3. In the development of clinical trials for drug, neither the physician nor the patient know if the patient was in the subgroup that was receiving a "placebo" or "sugar pill" or if he/she was actually receiving the treatment; hence the term "double-blind." In this research tradition, every effort is made to select individuals who are so similar to one another that any difference in their outcomes could be attributed to the treatment. It is this approach that has been adopted for schools.

4. Compliance is higher for acute illnesses for chronic conditions. In clinical trials, patients tend to be more dutiful due to the follow-up, but even under these conditions medication adherence rates are reported to be only in the range of 40% to 78% (see Osterberg and Blaschke 2005).

5. Some teachers are more effective in guiding student involvement than others; see Becker and Epstein's work (1982) or teacher leaders.

6. Proponents and colleagues stress the importance of retaining professional expertise rather than dictating a complex set of procedures. They explain: "We reduced our rate of infection using relative simple and inexpensive interventions, as opposed to implementing more expensive interventions, such as antibiotic/antiseptic catheters. For interventions to work in the busy world of clinical practice, they should be simple to implement. By changing systems rather than enrolling providers to comply with guidelines, we can help ensure that patients receive effective therapies. For example, it was difficult to write a detailed guideline regarding the need for central venous catheter; there are too many decisions in account for. It is unlikely that detailed guidelines would be practical for complex decisions, such as ICU admission and discharge, examination, and use of catheters. Rather, we simply asked physicians to consider daily whether central catheters could be removed, highlighting the risk of catheters yet allowing physicians to use their clinical judgment" (Berenholz et al. 2004, 2020).

7. A perspective that most first-year graduate students in the social sciences and in education encounter in their beginning research methods courses, by the way that the trick is to match the research method to the question the researcher wants to investigate and not to find one that fits an "all" method.

References and Further Reading


Rumberger, Russell W. and Scott L. Thern among Urban and Suburban High 3


