

EVIDENCE MATTERS

Randomized Trials in Education Research

Frederick Mosteller

Robert Boruch

Editors

BROOKINGS INSTITUTION PRESS

Washington, D.C.

Overview and New Directions

ROBERT BORUCH

FREDERICK MOSTELLER

THIS BOOK PRESENTS some histories and current status of research on practices in education from several points of view. Our main goal is to help improve education in schools in the United States by encouraging the gathering of better evidence on the impact of educational practices.

That governments have become seriously interested in the quality of research in the education sector is plain. The U.S. Congress's Committee on Education and the Work Force, for example, has been concerned about the wide dissemination of flawed, untested educational initiatives that can be detrimental to children. In 2000 this concern led to the development of a bill designed to address problems in this arena: the Scientifically Based Education Research, Evaluation, and Statistics and Information Act of 2000. Among other things, the bill called for increased scientific rigor in education research and tried to specify what this meant. It called for controlled experiments and the use of properly constituted comparison groups in quantitative research and for scientifically based qualitative research standards.

That particular bill did not pass. But it is a clear signal of congressional interest. Partly as a consequence, the National Research Council has convened a committee titled Scientific Principles in Educational Research to help make explicit the standards of evidence that are appropriate. The National Academy of Education and the Social Science

Research Council have initiated a different related effort, focusing on the intellectual organization and state of educational research.

Examples from other countries are not difficult to identify. In the United Kingdom, for example, the Department for Education and Employment, in cooperation with the Economic and Social Research Council, has set up a new Centre for Evidence-Informed Policy and Practice at London's Institute of Education to improve the knowledge base. This move was driven partly by the country's concerns about ideology parading as intellectual inquiry and about the relevance and timeliness of research and the intelligibility of its results. The biennial Conferences on Evidence-Based Policy at Durham University (United Kingdom) anticipated much of this more recent governmental concern.

An appropriate approach to studying the value of an intervention depends on the question that is asked. This book focuses on evidence about what interventions work better. Some of the best evidence to address this question can be generated in randomized field trials (RFTs). By this, we mean situations in which individuals or entire organizations are randomly assigned to one of two or more interventions. The groups that are so constructed do not differ systematically. That is, there are no hidden factors that would lead the groups to differ in unknown ways. When the groups are statistically equivalent at the outset, we can then be assured that comparison of the relative effectiveness of the interventions will be fair. That is, the properly executed RFT will produce estimates of relative effects that are statistically unbiased. Furthermore, one can make legitimate statistical statements about one's confidence in the results.

This book also focuses on empirical studies of the relative effectiveness of programs in education. Other kinds of research are important in building up to controlled studies of program effectiveness, in augmenting such studies, and in adding to what we understand about education and its effects. These other kinds of research include sample surveys, designed to understand which children need assistance and to what extent. They include narrative studies of how children, their parents, their teachers, and policymakers make decisions about their needs. The other kinds of studies include exploratory research on new ideas about what might work. Even throat-clearing essays at times contribute to understanding.

Overview of the Chapters

Chapter 2, by Judith Gueron, discusses the difficulties in practice and their resolutions when randomization is used. The chapter includes counsel about carrying out randomized field trials, using the experiences of the Manpower Development Research Corporation as a guide. Judith Gueron's long and successful experience with such randomized trials brings reality to her advice about carrying out studies. Among the problems is the task of persuading the sponsors and participants of the value of this approach. Her organization has executed many such investigations by persuading sponsors that there is no easier way to get the answers to the right questions, by meeting legal and ethical standards, and by ensuring that each study is large enough in numbers and duration not to miss the effects being studied. Impacts often emerge over time. Gueron emphasizes the importance of giving people the services to which they are entitled, of addressing previously unanswered questions of sponsors and participants, and of having adequate procedures to assure participants of data confidentiality, and to ensure that the innovations offer more than the usual treatment. Although Gueron's advice is especially oriented toward RFTs, most of it has value for any investigator trying to study a sensitive social problem in the turmoil of the real world.

Chapter 3, by Robert Boruch, Dorothy de Moya, and Brooke Snyder, presents many examples of evaluation in education and elsewhere using RFTs. The authors emphasize the systematic study of field trials themselves in the world literature and offer a variety of viewpoints for considering these investigations. Internationally, the number of RFTs is increasing, presumably because the quality of evidence from these studies seems stronger than that from other kinds. Chapter 3 introduces several ways of considering the value of such trials: by looking at the severity of the problem, the realism of the investigation for policy guidance, and the ability to estimate costs and benefits. Attention is also given to the international Campbell Collaboration, which prepares systematic reviews of comparative studies in education, social science, criminal justice, and other areas. Cooperation between the older Cochrane Collaboration in health care and Campbell may be of help to both groups.

In chapter 4, David Cohen, Stephen Raudenbush, and Deborah Ball offer a historical view of progress in education since about 1960. For researchers in education, they argue, the question "Do resources matter?" is far less important than "What resources matter, how, and under

what circumstances?” They push on to ask, “What instructional approaches, embedded in what instructional goals, suffice to ensure that students achieve the goals?” Adding resources to schools, they point out, does not automatically bring improved performance to the children. We need to attend to the environment that accompanies resources and to the interactions among such variables as money spent on instruction, years in school, depth of teachers’ subject matter preparation, time on task, class size, and how instruction is actually carried out. The authors think that we need to move away from thinking about resources as money to considering environment, students, teachers, and their interactions. They want to see more research not on the components of instructional approaches, but on coherent instructional regimes. These regimes need to be compared for their effect on learning.

The remaining chapters turn to diverse topics, beginning with the pros and cons of using RFTs. The authors recognize the advantages of RFTs in relation to alternatives and ask why they are infrequently used in education. Thomas Cook and Monique Payne suggest that the rarity of RFTs is due to educational researchers’ objections to randomized trials. They present arguments to refute these objections. Maris Vinovskis focuses on the institutional history of federally sponsored evaluations, arguing that the rarity of good studies is a function of weak political and administrative support for rigorous research. Gary Burtless focuses on chance, social convention, and major political impediments to randomized trials. And Carol Weiss discusses the practical difficulty of community-wide trials and identifies ways to study education in anticipation of RFTs or in parallel with them. A second topic of discussion is alternatives to experiments. Burtless and Cook and Payne cite evidence on defects in the design and results of nonrandomized trials, especially the difficulty of ensuring that estimates of relative effects are interpretable and statistically unbiased. All of the authors recognize that randomized trials are impossible at times, and they identify alternatives that may suffice, depending on the context of the evaluation. Weiss introduces some approaches that help to build understanding “before the randomizer arrives.”

The essays also introduce refreshed ways of thinking about these topics and others. In chapter 5, Maris Vinovskis comments on the checkered history and role of the U.S. Department of Education (USDE) in rigorous evaluation of federal programs. He outlines major contributions such as the Coleman report and those of various political administrations, such as President Lyndon Johnson’s enhanced investments in education. He also

notes that other administrations, such as President Richard Nixon's, reduced investments in research. He describes how various institutional arrangements have emerged and then disappeared, institutions such as the National Institute of Education. To judge from Vinovskis's description, the National Center for Education Statistics has made substantial progress since the 1980s in the collection and dissemination of descriptive statistics, including student achievement data. According to Vinovskis, some senior executives at the U.S. Department of Education have supported rigorous evaluation, although the record in producing such evaluations has been spotty. The Even Start Family Program is used to illustrate a remarkable effort to sponsor and to execute a randomized field trial.

One reason for the scarcity of randomized trials, says Vinovskis, is that the USDE's Planning and Evaluation Service has focused on many short-term studies that assist the Department of Education. Hence the unit's resources are spread thinly. Similarly, he is concerned that regional laboratories and R&D centers spread their limited resources among many different small-scale projects rather than concentrating them on a few larger and longer-term initiatives. Vinovskis therefore urges policy-makers to consider reconfiguration of the institutions responsible for program development and evaluations, beginning perhaps with a new National Center for Evaluation and Development in the USDE to handle large-scale evaluations and development projects. He would structure this new entity so as to reduce political influences on evaluations, for example, by providing for a six-year term of office for the center's commissioner. The proposed center would help consolidate evaluations across education programs and reduce what Vinovskis views as a fragmentation of federal monetary investments in evaluation and development.

In chapter 6, Thomas Cook and Monique Payne recognize that education research is strong in that it has produced high-quality description, survey methods and data, and achievement testing. They believe, however, that education research has been weak in establishing causal relationships. This weakness is based, they say, on some contemporary researchers' objections to using RFTs to understand such relationships. They present counterarguments and empirical evidence to meet these objections.

In response to criticism that RFTs stem from an oversimplified theory of causation, Cook and Payne point out that the purpose of such trials is "more narrow and practical—to identify whether one or more possible causal agents cause change in a given outcome." Opponents of RFTs,

they add, have stressed qualitative methods that depend on different assumptions and involve a different objective: hypothesis generation rather than hypothesis testing. Cook and Payne also disagree that RFTs are inappropriate in the complex U.S. education system, noting that experiments can be designed and executed to take into account both the similarities among schools and their heterogeneity. Furthermore, they do not believe that randomized assignment is premature until a good program theory and mediational processes are developed, as some claim. They maintain that the value of randomized field trials depends not on having good program theory, but on making a fair comparison and protecting against statistical bias, regardless of the particular theory. Pointing to trials that have been mounted successfully, they disagree that randomization has been used and has failed. Such criticisms, they note, are often based on *non*randomized trials and on the imperfections of randomized trials in fields other than education.

Claims that RFTs involve trade-offs that are not worth making sometimes take the form of rejecting internal validity in favor of external validity. That is, the opponents of trials attach higher value to the generalizability of a study's results (external validity) and lower value to unbiased estimates of the relative effectiveness of programs (internal validity). Cook and Payne recognize that a given trial may or may not be designed to ensure that the results are generalizable from a particular sample to larger populations. They encourage the use of designs that "minimize the external validity loss."

In considering the complaint that RFTs are unethical, Cook and Payne recognize the legitimate tensions between a lottery and measures of merit as a device for allocating scarce resources. They maintain that randomization can be justified when we do not know which intervention is more effective, especially when resources for interventions are limited. They also examine the view that there are good alternatives to RFTs. They encourage the reader to recognize qualitative case studies as adjuncts to RFTs, rather than as substitutes for a randomized trial. They criticize another purported alternative, theory-based evaluation, which Weiss encourages, pointing out that theory-based approaches are rarely sufficiently explicit. Weiss offers some counterexamples. Furthermore, when there are multiple ways of making theory explicit, the theories are unspecific about timelines, reciprocal causal influences, and counterfactuals. Their concern about others' claims that quasi experiments are an alternative to randomized field trials lies in the fact that the phrase "quasi

experiment” is often used promiscuously and that the quasi-experimental designs are often poorly conceived and executed.

In chapter 7, Gary Burtless notes that sizable RFTs have often been mounted to evaluate certain economic programs started on a pilot basis. Among the basic reasons for trials in these areas, he says, is that randomization ensures that we know the direction of causality and helps remove any systematic correlation between program status and unobserved characteristics of participants and nonparticipants. Burtless argues that RFTs can be designed to meet ethical standards while permitting policymakers to test new programs and new variations on old programs. In the area of welfare, says Burtless, some RFTs have been done where economists have found important defects in some of the statistical alternatives. These defects have led to wrong conclusions about the effects of programs.

Burtless also discusses political and professional influences on our willingness to do RFTs in the area of welfare and employment and training, as opposed to education programs. First, Butler notes that economists knowledgeable about RFTs have held important cabinet positions in the U.S. government in agencies such as the U.S. Department of Labor, but not in the U.S. Department of Education. Second, the influence depends on the extent of the federal role in financial, political, and regulatory environments. For instance, the federal government has a substantial role in drugs that are put on the commercial market partly because of regulations that require good evidence on the effects of new drugs. It does not have the same arrangement in place to determine whether education interventions are safe or effective. Burtless also recognizes that some target populations are politically weak, suggesting that their members then are not well positioned to oppose an RFT.

RFTs appear infrequently in education, Burtless suggests, partly because the federal financial share in education is small in relation to state and local investments. Consequently, the federal government has limited influence on mounting RFTs to test innovations in education. Another reason is that economists and other social scientists with a strong interest in rigorous evaluation have not held important positions in federal education departments or the legislative forums that shape policy on education programs. In addition, argues Burtless, teachers, parents, and others in the education community view RFTs as potentially denying benefits to children who need assistance. Furthermore, school administrators and teachers believe that they will lose control in some respects when a trial

is undertaken. Finally, because the education community is politically influential, states Burtless, its members have the capacity to impede or stop a trial that they consider unethical or apt to reduce their authority.

In Carol Weiss's view, outlined in chapter 8, in some situations the sampling of individuals is not an appropriate basis for measuring program success. This would be the case in the study of community programs where the purpose is not to change the behavior of the individuals in the community but to change the community itself. One could sample communities, but this approach leads to large studies and the further difficulty that initial conditions in the chosen communities may differ substantially. Weiss offers a brief history of community studies in the United States. She also suggests a variety of alternatives to randomization, including the use of qualitative information to get at the rich structure of the activities and their outcomes.

Carol Weiss's main recommendations as alternatives to randomization are theory-based evaluation (TBE) and Ruling Out Alternative Explanations. By way of example, she describes a program expected to create jobs in a depressed community and a theory of change for an initiative to create a healthier atmosphere for adolescents (considering both positive and negative aspects). By spelling out the theories, Weiss argues, we are able to focus on events and consequences that demonstrate its weaknesses. We know what to pay attention to. We can gather data appropriate to the theory from the beginning and see if it is supported or whether some alternative theory is needed.

Weiss's second approach is to rule out alternative explanations, which she does by employing the usual requirements for indications of causality. For example, we may know the order in time that things should occur, and when they do not, we can eliminate an explanation. Or when it appears unlikely that a sequence of requirements can all be true, we can again eliminate an explanation.

New Directions

The central question addressed in this book is what areas of education might be suitable for mounting randomized trials or other studies of effectiveness? How do we go about selecting these areas? At least three kinds of answers are sensible. The first depends on the efforts of government and its advisers, including people who can put research into practice. The second depends on our own studies. The third kind of answer is

a promissory note that international collaborative efforts such as the Campbell Collaboration can help in identifying research targets.

Government Efforts

The U.S. Department of Education's Strategic Plan for 2001–05 identifies three major goals that are pertinent here. It also identifies the specific objectives within each goal, and the programs that are being mounted to meet the objectives. More to the point, the strategic plan describes some of the evaluations that the department is planning or supporting, as described in a section of the plan called "Evaluation Highlights."

Goal 1 is to build a solid foundation of learning for all children. To this end, the department is supporting RFTs of the Even Start Family Program, for example. This program, intended to raise the literacy skills of families and young children, is being evaluated in twenty sites using RFTs. The department is also undertaking feasibility studies to determine whether RFTs of new programs can be mounted under the Reading Excellence Act. The department's trials on the Upward Bound Programs are well advanced. Some RFTs are being mounted in collaboration with the Department of Health and Human Services, notably on Head Start. Other programs targeted for evaluation under goal 1 pertain to Comprehensive School Reform, Talent Search, Career Academies, School to Work, and Title 1.

Goal 2 of the department's strategic plan is to reform the U.S. education system. A variety of evaluative activities are under way to assess whether objectives within this goal are being met. One program, the 21st Century Community Learning Centers, has been targeted for randomized trials supported by the department and at least one private foundation. This program funds after-school activity designed to raise academic achievement in schools in high-poverty areas. The department has also decided to study teacher development programs in mathematics and reading, teacher recruitment and retention programs, and programs that are sometimes controversial, such as charter schools and magnet schools.

Goal 3 of the Strategic Plan is to "ensure access for all to a high-quality post-secondary education and life-long learning." None of these evaluations involve a randomized trial. Instead, other kinds of studies are being deployed to generate evidence about the effectiveness of TRIO's Student Support Services for college students at risk, Vocational Rehabilitation Services, the Adult English as a Second Language Program, and the Adult Basic Education Program.

Evidence from Past Trials

Many topics in education merit study, in addition to the ones identified in the Strategic Plan. Some are already receiving attention. The effectiveness of summer school, for example, has been a subject of growing interest ever since it was established that children in lower-income families lose more of their educational attainment during the long summer vacation than children from higher-income families. Consequently, the effectiveness of summer school is important in our attempts to maintain equal opportunity. A recent meta-analysis of this subject by Harris Cooper and his colleagues collects the results of many studies and suggests that summer school can be effective in reducing the gap in achievement created by the long summer vacation.¹

This summer loss in achievement also raises the question of whether some rearrangement of the school calendar might be more productive than depending on summer school.² For example, the year could be broken up into several periods of equal length with short vacations after each. Such a plan may be very difficult to implement and evaluate because it requires more agreement among the states and school districts than we ordinarily have in this country. On the other hand, similarly difficult challenges have been met in trials that randomly assigned entire schools to different schoolwide interventions.

The field of special education for handicapped or disabled students offers many opportunities for investigation. Recently, the trend has been to include the handicapped in regular classes as much as possible while also offering some special time for them alone. Various questions need to be addressed here: for example, is it appropriate to have multiple teachers in a class, are specially trained teachers required, and how many handicapped can be included in a class of given size?

We have had one large long-term investigation of the effect of class size on children's achievement. The STAR trial carried out in Tennessee suggests that reducing class size from about twenty-four to sixteen for the earliest years of schooling, kindergarten through grade 3, improved the scores of children.³ This improvement occurred not only during the years that class size was reduced, but also in later years when class size was

1. Cooper and others (2000).

2. See Cooper and others (2000); Entwisle, Alexander, and Olson (2000).

3. Mosteller, Olson, and Youtz (1996).

restored to its previous level. Including a teacher's aide with class size did not seem to improve scholastic performance very much. What was it about smaller class size that improved scores in the beginning and later grades? To answer that question, it would be helpful to study how resources are reallocated in an environment of reduced class size and what the effects of reallocation are. For example, it might just be that less confusion in the classroom leads to more effective education, but we await more information on the cause of improvement.

Yet another topic of widespread interest is cooperative learning, which also appears to produce improvements in performance.⁴ The basic idea here is to group students in a classroom into teams, and to give each team responsibility for making sure that all its members master the material in each lesson. In other words, members of a team work together to improve their performance. Extensive data and many investigations strongly suggest that this approach does produce improvements in performance.

Place-Based Trials

Many people do not realize that entire classrooms, schools, and districts have at times been randomly assigned to different interventions. Typically, the study's purpose is to estimate the relative effectiveness of a program that is deployed in many classrooms, schools, or districts. Tests of the Drug Abuse Resistance Education (DARE) program in Illinois, for instance, involved randomly allocating half the schools in twelve pairs of schools to DARE and half to a control condition. This study and similar randomized trials helped to show that, contrary to people's expectations of schoolwide DARE programs, the programs had no appreciable effect on substance abuse by the young.⁵ The findings from this trial and earlier ones led to controversy and stiff opposition from DARE advocates. In 2001 DARE leaders decided to revise the program in ways that might enhance its effectiveness.

Other studies further illustrate the feasibility of such trials. Fifteen school districts in Minnesota and Wisconsin agreed to be randomly assigned to a specialized theory-driven schoolwide program to reduce under-age alcohol abuse and to a control condition in which ordinary education on the topic might or might not be provided.⁶ Trialists have

4. See Slavin (1995).

5. Rosenbaum and Hansen (1998).

6. Wagenaar and others (1994).

randomly assigned classrooms to new sexual risk reduction programs to compare them to conventional handling of the topic in middle schools in California and elsewhere, to determine whether the new intervention did indeed reduce the risky youth behavior below ambient levels.

Still other trials have been mounted to test community-wide interventions (on smoking, for example), hospital-wide changes in handling certain illnesses, and AIDS risk reduction involving work sites as the units of allocation.⁷ All of these examples can be counted as relevant here because they include education and information as program components, although they may not deal directly with youth in schools.

Such trials are institutionally complex because they demand the cooperation of different organizations and groups of people. The analysis of resultant data can also be complex, although statistical and methodological advances have made this exercise more tractable and orderly.⁸ Nonetheless, they are extremely important where interventions with potential social benefits must be deployed at the institutional or community level. Theory, empirical evidence, and ordinary experience indicate that social norms influence human behavior. If these norms are changed at the institutional level through an intervention, then we may discern the effect of this on people's behavior through randomized trials that involve institutions as the units of allocation. The beneficial or harmful effects of schoolwide programs on children's education achievement can be discerned before the programs are adopted at the regional or state level.

These precedents are relevant to current debates about how to evaluate comprehensive school reforms, how to generate evidence about the effectiveness of various models of reform and effective practice, and how to enhance the well-being of children and youth. Evidence from randomized trials permits us to estimate the relative effect of the effort in the least ambiguous way possible and thus to engage in thoughtful debate about what ought to be done next.

Lists of the sort given here that direct attention to the topics in which more evaluation is needed will depend heavily on the people and the institutions that draw up the list. The choices for the list presented in this chapter, for example, were influenced by the people who advise the U.S. Department of Education, our own judgment, and the people who contributed to this volume. We can map the terrain of evidence on the effects

7. Boruch and Foley (2000).

8. See Murray (1998); and Donner and Klar (2000).

of interventions in other ways. One way is to generate good evidence on where the dry ground is: where educational interventions have been evaluated well. This involves searching for empirical evidence on where the sea is deep, where few or no interventions have been evaluated. And, of course, to map the evidential terrain we need to know where the swamps are, that is, where the evidence on the effectiveness of education interventions is ambiguous.

Collaborative Efforts

This kind of empirical mapping can be facilitated by international collaboration, such as the evolving Campbell Collaboration. This collaboration's main mission is to generate high-quality systematic reviews of studies of the effects of interventions, in other words, to find out what works, what does not work, and what is promising, according to good standards of evidence. The older sibling of the Campbell Collaboration is the Cochrane Collaboration in health care.⁹

The Campbell Collaboration's systematic reviews can be configured to provide a map of where research on the effects of interventions has and has not been done, and where such research may be sorely needed. This map will provide policymakers, practitioners, and researchers with an informed basis for delineating these areas. Since this is public information, people can add to it or offer constructive criticisms. To be useful for research policy, such mapping must be both prospective and retrospective in its orientation. Therefore it is essential to develop registries of evaluations that are just *beginning*, as well as collaborative systems for tracking their progress. These registries must also be updated continuously to keep track of studies that have been completed.

At present, the Campbell Collaboration is developing a Social, Psychological, Educational, and Criminological Trials Registry (C2-SPECTR) whose contents will be used to map studies in each of these arenas. C2-SPECTR is designed to be prospective and retrospective. In the educational arena, entries are coded along various dimensions, including the types of outcomes examined in each study (such as math achievement, dropout, or socioemotional development) and the kinds of interventions under study (such as tutoring or class size). This coding will make it possible to tailor the maps to the special interests of people who will be using them.

9. See www.campbell.gse.upenn.edu and www.cochrane.org.

Generating better evidence for better education is not an easy task. Unless researchers and practitioners work at it, they will, in Walter Lippmann's words "imperil the future by leaving great questions to be fought out between ignorant change on the one hand and ignorant opposition to change on the other."¹⁰ The contributors to this volume give good advice on how to reduce our ignorance and generate better evidence. We trust that readers will take seriously what they have to say so as to enhance the education of children and youth.

References

- Boruch, R., and E. Foley 2000. "The Honestly Experimental Society: Sites and Other Entities as the Units of Allocation and Analysis in Randomized Trials." In *Validity and Experimentation: Donald Campbell's Legacy*, edited by L. Bickman, 1:193–238. Thousand Oaks, Calif.: Sage Publications.
- Cooper, H., K. Charleton, J. C. Valentine, and J. Muhlenburck. 2000. "Making the Most of Summer School: A Meta-Analytic and Narrative Review." *Monographs of the Society for Research in Child Development*, Serial 260, 65(1). Malden, Mass.: Blackwell.
- Donner, A., and N. Klar. 2000. *Design and Analysis of Cluster Randomization Trials in Health Research*. Oxford University Press.
- Entwisle, D. R., K. L. Alexander, and L. S. Olson. 2000. "Summer Learning and Home Environment." In *A Notion at Risk*, edited by Richard D. Kalenberg, 9–30. New York: Century Foundation Press.
- Lippman, W. 1963. "The Savannah Speech." In *The Essential Lippmann*, edited by C. Rossiter and J. Lare. Random House. (Originally published in 1936.)
- Mosteller, F., M. Olson, and C. Youtz. 1996. "Notes on Cooperative Learning, Center for Evaluation." *Initiatives for Children*. American Academy of Arts and Sciences (April 10): 1–7.
- Murray, D. M. 1998. *Design and Analysis of Group Randomized Trials*. New York: Oxford University Press.
- Rosenbaum, D. P., and G. S. Hanson. 1998. "Assessing the Effects of School-Based Education: A Six Year Multilevel Analysis of Project DARE." *Journal of Research in Crime and Delinquency* 35(4): 381–412.
- Slavin, R. E. 1995. *Cooperative Learning*. 2d ed. Allyn & Bacon.
- U.S. Department of Education. 2000. *Strategic Plan, 2001–2005*. Government Printing Office.
- Waganeer, A. C., D. M. Murray, M. Wolfson, J. L. Forster, and J. R. Finnegan. 1994. "Communities Mobilizing for Change on Alcohol: Design of a Randomized Community Trial." *Journal of Community Psychology*. Monograph Series/CSAP (Special Issue): 79–101.

10. Lippmann (1963/1936, pp. 495–97).