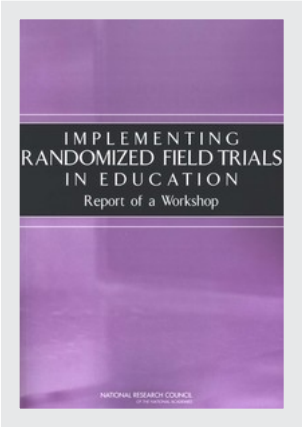

Implementing Randomized Field Trials in Education: Report of a Workshop

DETAILS

62 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-09192-3 | DOI 10.17226/10943

AUTHORS

Lisa Towne and Margaret Hilton, Editors, Committee on Research in Education, National Research Council

IMPLEMENTING RANDOMIZED FIELD TRIALS IN EDUCATION

Report of a Workshop

Committee on Research in Education

Lisa Towne and Margaret Hilton, Editors

Center for Education

Division of Behavioral and Social Sciences and Education

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract No. ED-00-CO-0088 between the National Academy of Sciences and the U.S. Department of Education, Grant No. 2002-7860 from the William and Flora Hewlett Foundation, and Grant No. 200200225 from the Spencer Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Education, the William and Flora Hewlett Foundation or the Spencer Foundation.

International Standard Book Number 0-309-09192-6 (Book)
International Standard Book Number 0-309-53155-1 (PDF)

Additional copies of this report are available from National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu

Printed in the United States of America

Suggested citation: National Research Council. (2004). *Implementing Randomized Field Trials in Education: Report of a Workshop.* Committee on Research in Education. L. Towne and M. Hilton, Eds. Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON RESEARCH IN EDUCATION
2004

Lauress L. Wise (*Chair*), Human Resources Research Organization (HumRRO), Arlington, VA
Linda Chinnia, Baltimore City Public School System
Kay Dickersin, Department of Community Health, Brown University
Margaret Eisenhart, School of Education, University of Colorado
Karen Falkenberg, Division of Educational Studies, Emory University
Jack McFarlin Fletcher, University of Texas-Houston Health Science Center and Center for Academic and Reading Skills
Robert E. Floden, College of Education, Michigan State University
Ernest M. Henley (emeritus), Department of Physics, University of Washington
Vinetta C. Jones, School of Education, Howard University
Brian W. Junker, Department of Statistics, Carnegie Mellon University
David Klahr, Department of Psychology, Carnegie Mellon University
Ellen Condliffe Lagemann, Harvard Graduate School of Education
Barbara Schneider, Department of Sociology, University of Chicago
Joseph Tobin, College of Education, Arizona State University

Lisa Towne, *Study Director*
Margaret Hilton, *Senior Program Officer*
Tina Winters, *Research Associate*

Preface

The central idea of evidence-based education—that education policy and practice ought to be fashioned based on what is known from rigorous research—offers a compelling way to approach reform efforts. Recent federal trends reflect a growing enthusiasm for such change. Most visibly, the 2002 No Child Left Behind Act requires that "scientifically based [education] research" drive the use of federal education funds at the state and local levels. This emphasis is also reflected in a number of government and nongovernment initiatives across the country. As consensus builds around the goals of evidence-based education, consideration of what it will take to make it a reality becomes the crucial next step.

In this context, the Center for Education of the National Research Council (NRC) has undertaken a series of activities to address issues related to the quality of scientific education research.[1] In 2002, the NRC released *Scientific Research in Education* (National Research Council, 2002), a report designed to articulate the nature of scientific education research and to guide efforts aimed at improving its quality. Building on this work, the Committee on Research in Education was convened to advance an improved understanding of a scientific approach to addressing education prob-

[1]Other NRC efforts—especially the line of work that culminated in the recent report *Strategic Education Research Partnership* (National Research Council, 2003)—offer insights and advice about ways to advance research utilization more broadly.

lems; to engage the field of education research in action-oriented dialogue about how to further the accumulation of scientific knowledge; and to coordinate, support, and promote cross-fertilization among NRC efforts in education research.

The main locus of activity undertaken to meet these objectives was a year-long series of workshops designed to engage a range of education stakeholders in discussions about five key topics:

- *Peer Review in Federal Education Research Programs*. This workshop focused on the purposes and practices of peer review in federal agencies that fund education research. Federal officials and researchers considered a range of models used across the federal government to involve peers in the review of proposals for funding and discussed ways to foster high-quality scientific research through peer review.
- *Understanding and Promoting Knowledge Accumulation in Education: Tools and Strategies for Education Research*. With a focus on how to build a coherent knowledge base in education research, researchers and federal officials analyzed several elements of the research infrastructure, including tools, practices, models, and standards. Fundamental questions about what such a knowledge base might look like were also considered in this context.
- *Random Assignment Experimentation in Education: Implementation and Implications*. The evidence-based education trend has brought to the fore decades of debate about the appropriateness of randomized field trials in education. Far less consideration has been devoted to the practical aspects of conducting such studies in educational settings; this workshop featured detailed descriptions of studies using randomized field trials in education and reflections on how the current trend to fund more of these studies is influencing states, districts, and students.
- *Journal Practices in Publishing Education Research*. Following the more general discussion of how to build a coherent knowledge base in education in a previous workshop, this event took up the specific case of journals that publish education research. Editors, publication committee members, and others involved in the production and use of journal articles considered ways to promote high-quality education research and to contribute to the larger body of knowledge about important areas of policy and practice.
- *Education Doctoral Programs for Future Leaders in Education Research*. A final workshop focused on the professional development of edu-

cation researchers, with a specific emphasis on doctoral programs in schools of education. Deans, graduate study coordinators, foundation officials, and policy makers came together to share observations and chart potential paths for progress.

Additional information on each of these events, including transcripts of presentations and discussions, can be found at http://www7.nationalacademies.org/core/.

This report is a summary of the third workshop in the series, on the implementation and implications of randomized field trials in education. Educators and researchers have debated the usefulness of these methods for conducting research in education for decades. As many more of them are being funded in education than ever before, our objective in convening this workshop was to provide a venue for researchers and practitioners who have been involved in this kind of study in educational settings to share their experiences. The event took place on September 24, 2003, at the National Academies' Keck Center in Washington, DC.

This report summarizes common issues and ideas that emerged from the presentations and discussion during the workshop (see Appendix A for the workshop agenda and Appendix B for biographical sketches of the committee members and speakers). These issues included *why* researchers use randomized field trials, *when* such a design is appropriate for answering questions about education, and *how* to implement this kind of research in an educational setting. In discussing these issues, workshop speakers identified challenges to successfully carrying out randomized field trials in schools and described strategies for addressing those challenges. Although investigators conducting any type of research in schools would encounter many of these challenges, some are unique to this research design.

While this report represents our synopsis of the key issues aired at the workshop, it does *not* contain conclusions or recommendations. We will issue a final report with recommendations for improving scientific research in education based on the series of five workshops. In addition, because the one-day workshop that is the subject of this report necessarily included only a small number of practitioners and researchers, this summary cannot be construed as representative of all experiences and views of those who have been involved in randomized field trials in educational settings. We did take care to invite individuals who were experienced and knowledgeable about implementing this kind of research in social settings and believe

that the insights they shared are useful. Our aim is to help investigators, funders, and educators involved in the next generation of randomized field trials in education to avoid common pitfalls and to carry out best practices.

This workshop report would not have been possible without the stellar group of speakers who shared their expertise with the committee. We would like to thank each of them for their contributions: Robert F. Boruch, professor, Graduate School of Education, University of Pennsylvania; Wesley Bruce, assistant superintendent, Indiana Department of Education; Linda Chinnia, Area Academic Officer (Area 1), Baltimore City Public School System; Donna Durno, executive director, Allegheny Intermediate Unit; Olatokunbo S. Fashola, research scientist, Johns Hopkins University Center for Research on the Education of Students Placed at Risk; Judith Gueron, president, MDRC; Vinetta C. Jones, dean, Howard University School of Education; Sheppard Kellam, public health psychiatrist, American Institutes for Research; Anthony (Eamonn) Kelly, professor of instructional technology, Graduate School of Education, George Mason University; Sharon Lewis, director of research, Council of the Great City Schools; Loretta McClairn, family, schools, and communities coordinator, Dr. Bernard Harris Elementary School, Baltimore City Public School System; David Myers, vice president, Mathematica Policy Research; and Richard J. Shavelson, professor, School of Education, Stanford University.

Of course, without the generous support of our sponsors, neither the workshop nor this report would be possible. We extend our gratitude to the former National Educational Research Policy and Priorities Board and the Institute of Education Sciences, the William and Flora Hewlett Foundation, and the Spencer Foundation.

Finally, we thank each of the members of the Committee on Research in Education. We especially appreciate the efforts of the workshop planning group, led by Kay Dickersin, who designed an outstanding event that has made a unique contribution to an important debate. Finally, we wish to acknowledge the contributions of Richard Nelson of Columbia University, who participated in early planning for the event but later resigned from the committee.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objec-

tivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report: Mark Dynarski, Education Research Department, Mathematica Policy Research, Inc., Princeton, New Jersey; Susan Fuhrman, Graduate School of Education, University of Pennsylvania; Julia Lara, Division of State Services and Technical Assistance, Council of Chief State School Officers, Washington, DC; Patricia Lauer, Principal Researcher, Mid-continent Research for Education and Learning, Aurora, Colorado; and Jean Williams, Center for Research in Education, RTI International, Research Triangle Park, North Carolina.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by Milton Hakel, Department of Psychology, Bowling Green State University. Appointed by the National Research Council, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Lauress L. Wise, *Chair*
Lisa Towne, *Study Director*
Committee on Research in Education

Contents

1

What Is a Randomized Field Trial?

People behave in widely varying ways, due to many different causes, including their own individual volition (conscious choices). Social scientists often seek to understand whether or not a specific intervention may have an influence on human behavior or performance. For example, a researcher might want to examine the effect of a driver safety course on teenage automobile accidents or the effect of a new reading program on student achievement. But there are many forces that might cause a change in driving or reading skills, so how can the investigator be confident that it was the intervention that made the difference? An effective way to isolate the effect of a specific factor on human behavior and performance is to conduct a *randomized field trial*, which is a research method used to estimate the effect of an intervention on a particular outcome of interest.

As a first step, investigators hypothesize that a particular intervention or "treatment" will cause a change in behavior. Then they seek to test the hypothesis by comparing the average outcome for individuals in the group who were randomly assigned to receive this intervention with the average outcome for individuals in the group who do not. This method helps social scientists to attribute changes in the outcome of interest (e.g., reading achievement) to the specific intervention (e.g., the reading program), rather than to the many other possible causes of human behavior and performance.

MAJOR FEATURES

In this section, we sketch the defining features of randomized field trials. In particular, we focus on the two key concepts of randomization and control and then briefly situate randomized field trials within the broader context of establishing cause-and-effect relationships.

A research design is randomized when individuals (or schools or other units of study) are put into an "experimental" group (which receives the intervention) or a "control"[1] group (which does not) on the basis of a random process like the toss of a coin.[2] The power of this random assignment is that, on average, the two groups that result are initially the same, differing only in terms of the intervention.[3] This allows researchers to more confidently attribute differences they observe between the two groups to the intervention, rather than to the myriad other factors that influence human behavior and performance. As in any comparative study, researchers must be careful to observe and account for any other confounding variables that could differentially affect the groups after randomization has taken place. That is, even though randomization creates (statistically) equivalent groups at the outset, once the intervention is under way, other events or programs could take place in one group and not the other, undermining any attempt to isolate the effect of the intervention.

Randomized field trials are also controlled; that is, the investigator controls the process by which individuals (or other entities of study) are assigned to receive the intervention of interest. If the assignment of individuals or entities is outside the investigator's control, then it is generally

[1]A control group is a comparison group in a randomized field trial that acts as a contrast to the group receiving the intervention of interest. In randomized field trials involving humans, research participants in the control group typically either continue to receive existing services or receive a different intervention.

[2]Tossing a coin is a useful way of explaining the situation in which the participants have a 50-50 chance of being assigned to either of two groups: the experimental or the control group. Randomized field trials can have more than two groups; as long as the assignment process is conducted on the basis of a statistical process that has known probabilities (0.5 or otherwise), the groups will be balanced on observable and unobservable characteristics.

[3]It is logically possible that differences between the groups may still be due to idiosyncratic differences between individuals assigned to receive the intervention or to be part of the control group. However, with randomization, the chances of this occurring (a) can be explicitly calculated and (b) can be made very small, typically by a straightforward manipulation like increasing the number of individuals assigned to each group.

much more difficult to attribute observed outcomes to the intervention being studied. For example, if teachers assigned some students to experience a novel teaching method and some to a comparison group that did not experience it based on their judgment of which students should experience the method, then other factors (such as student aptitude) may confound or obscure the specific effect of the novel teaching method on student learning outcomes.[4]

Thus, randomization and control are the foundation of a systematic and rigorous process that enables researchers estimating the effect of an intervention to be more confident in the internal validity of their results—that is, that differences in outcomes can be attributed to the presence or absence of the intervention, rather than to some other factor. External validity—the extent to which findings of effectiveness (or lack of effectiveness) hold in other times, places, and populations—can be established only when the intervention has been subjected to rigorous study across a variety of settings.

The ultimate aim of randomized field trials is to help establish cause-and-effect relationships. They cannot, however, uncover all of the multiple causes that may affect human behavior. Instead, randomized field trials are designed to isolate the effect of one or more possible treatments that may or may not be the cause(s) of an observed behavioral outcome (such as an increase in student test scores) (Campbell, 1957). Furthermore, a single study—no matter how strong the design—is rarely sufficient to establish causation. Indeed, establishing a causal relationship is a matter of some complexity. In short, it requires that a coherent theory predict the specific relationship among the program, outcome, and context and that the results from several studies in varying circumstances are consistent with that prediction.

A few final clarifications about terminology are in order. Some observers consider the term "randomized field trial" to be limited only to very

[4]In some cases, an investigator may conduct a randomized field trial when an intervention is allocated to individuals based on a random lottery. As discussed in Chapter 3, some school districts have used randomized lotteries to allocate school vouchers, in order to equitably distribute scarce resources when demand exceeds available funding for vouchers. In these cases, the investigator typically does not directly control the random assignment process, but as long as the process is truly random, the statistically equivalent groups that result isolate the relationship between group membership (treatment or control) and outcome from confounding influences and the essential features of a randomized field trial are retained.

large medical studies or studies conducted by pharmaceutical companies when testing the safety and efficacy of new drugs. Randomized designs, however, can be part of any research in any field aimed at estimating the effect of an intervention, regardless of the size of the study. In this report, we use the term "randomized field trial" to refer to studies that test the effectiveness of social interventions comparing experimental and control groups that have been created through random assignment. Although most of the workshop discussions focused on large-scale randomized field trials, the key elements for education research do not involve the size of the study, but the focus on questions of causation, use of randomization, and the construction of control groups that do not receive the intervention of interest. Indeed, even small "pilot" studies can use randomization and control groups to determine the feasibility of scaling an intervention.

CURRENT DEBATES AND TRENDS

At the workshop, University of Pennsylvania professor Robert Boruch described how randomized field trials have been used in a range of fields over time. Since World War II, he explained, randomized field trials have been used to test the effectiveness of the Salk polio vaccine and the antibiotic streptomycin, and these designs are now considered the "gold standard" for testing the effects of different interventions in many fields. Boruch went on to describe the growing use of randomized field trials to evaluate social programs since the 1970s (Boruch, de Moya, and Snyder, 2002) and noted that the World Bank, the government of the United Kingdom, the Campbell Collaboration, and the Rockefeller Foundation, all held conferences promoting the use of randomized field trials during 2002 and 2003.

Trends in other fields notwithstanding, scholars of education have long debated the utility of this design in education research. Those who question its usefulness frequently argue that the model of causation that underlies these designs is too simplistic to capture the complexity of teaching and learning in diverse educational settings (e.g., Cronbach et al., 1980; Bruner, 1996; Willinsky, 2001; Berliner, 2002). Others, in contrast, are enthusiastic about using randomized field trials for addressing causal questions in education, emphasizing the unique ability of the design to isolate the impact of interventions on a specified outcome in an unbiased fashion (e.g., Cook and Payne, 2002; Mosteller and Boruch, 2002; Slavin, 2002).

In the past five years, as calls for evidence-based education have become common, these debates have intensified and expanded beyond aca-

demic circles to include policy makers and practitioners. Most visibly, the No Child Left Behind Act, passed by Congress in 2001 and signed by the President in 2002, includes many references to "scientifically based" educational programs and services. The law defines scientifically based research as including research that "is evaluated using experimental or quasi-experimental designs in which individuals, entities, programs or activities are assigned to different conditions and with appropriate controls to evaluate the effects of the condition of interest, with a preference for random-assignment experiments, or other designs to the extent that those designs contain within-condition or across-condition controls."

Furthermore, in its strategic plan for 2002-2007, the U.S. Department of Education has established as its chief goal to "create a culture of achievement" by, among other steps, encouraging the use of "scientifically based methods in federal education programs." The strategic plan also aims to "transform education into an evidence-based field" (U.S. Department of Education, 2002, pp. 14-15). The Institute of Education Sciences, the Department of Education's primary research arm, has established the What Works Clearinghouse to help reach these goals by identifying "interventions or approaches in education that have a demonstrated beneficial causal relationship to important student outcomes" (What Works Clearinghouse [2003], can be found at http://www.w-w-c.org/july2003.html). The expert technical advisory group guiding the clearinghouse has established quality standards to review available research on such critical education problems as improving early reading and reducing high school dropout rates. These standards place high priority on randomized field trials, which are seen as "among the most appropriate research designs for identifying the impact or effect of an educational program or practice" (What Works Clearinghouse [2003], can be found at http://www.w-w-c.org/july2003.html). They also acknowledge that there are circumstances in which they are not feasible, suggesting that quasi experiments (which are comparative studies that attempt to isolate the effect of an intervention by means other than randomization) may be useful under such circumstances.[5]

[5]In a quasi-experimental study, researchers may compare naturally existing groups that appear similar except for the intervention being studied. In this research design, investigators often use statistical techniques to attempt to adjust for known confounding variables that are associated with both the intervention and the outcome of interest, thus invoking additional assumptions about the causal effects of the intervention. While these statistical techniques can address known differences between study groups, they may inadequately adjust unknown

The National Research Council report *Scientific Research in Education* (National Research Council, 2002) was designed to help clarify the nature of scientific inquiry in education in this rapidly changing policy context. That report links design to the research question and, for addressing causal questions (i.e., "what works") about specified outcomes, highlights randomized field trials as the most appropriate research designs when they are feasible and ethical. This report summarizes a workshop in which participants addressed the question: When randomized field trials are conducted in social settings like schools and school districts, how can they be implemented and what procedures should be used in implementation?

confounding variables. The major drawback of quasi-experimental designs is the possibility that the groups are systematically different (a problem known as "selection bias"), and thus investigators may be less confident about conclusions reached using these methods (National Research Council, 2002, p. 113). In contrast, randomization theoretically creates groups that are not systematically influenced by both known and unknown confounding variables.

2

Why Are Randomized Field Trials Used?

Workshop speakers suggested that investigators choose a randomized field trial research design to answer important questions about the effects of social programs or services and to obtain credible results that are likely to be used by policy makers.

ANSWERING IMPORTANT QUESTIONS

In his opening remarks, Stanford University professor of education Richard J. Shavelson, who chaired the committee that produced *Scientific Research in Education* (National Research Council, 2002), emphasized that researchers use study designs appropriate to answer "the important questions." Summarizing the main findings from *Scientific Research in Education*, Shavelson argued that important questions can be divided into three classes: "What is happening [in education]?" "Is there a systematic effect [e.g., of an educational program]?" and "Why or how is it happening?" The first set of questions can be answered best using descriptive research methods. Descriptive research methods can help researchers and policy makers to identify and describe particular education problems (e.g., dropout rates) and may also aid in designing interventions, he said. Once an intervention has been proposed, a randomized field trial is often the best method to help researchers understand whether the intervention has the intended (causal)

effect on an educational outcome of interest, in Shavelson's view. Through this approach, investigators may be able not only to establish a systematic effect of the intervention, but also to reliably estimate the magnitude of the effect.

To answer the third set of important research questions in education—"How or why is it [the effect] happening?"—Shavelson suggested that researchers combine several methods, including randomized field trials, quasi-experimental designs, and descriptive techniques. For example, classroom observations, interviews, and ethnographic studies in California (Borhnstedt and Stecher, 1999) complemented randomized field trials in Tennessee (Achilles, 1999) to identify teacher behavior as the key factor in explaining why smaller class sizes improved student achievement.

You ought to design a study to answer the question that you think is the important question, not create the question to fit the design.

–Richard J. Shavelson, Stanford University

Later in the day, three studies that feature randomized field trials in educational settings were described in detail. In each case, investigators followed the approach outlined by Shavelson, using previous descriptive and quasi-experimental research to identify and articulate important research questions, and choosing methods appropriate to answer their questions. Described in detail in Chapter 3, all three studies build on previous research. The designers of all three studies chose a randomized field trial to assess whether various interventions targeted to elementary school students had a systematic effect on specific academic and behavioral outcomes.

YIELDING USEFUL RESULTS

Workshop speakers observed that researchers choose a randomized field trial design not only because it can answer one class of important questions, but also because of its capacity for generating valid and reliable results that are trusted and used by policy makers. Judith Gueron, president of MDRC, a large nonprofit research corporation specializing in randomized field trials, made this case convincingly. Over the past 30 years, she said, her com-

pany has conducted 35 to 40 large-scale randomized field trials involving about 400,000 people in 250 locations around the United States, and "we've never had a serious challenge to their credibility." With alternative research methods, she said, "this is much less true."

> *With randomized field trials, you can more confidently separate fact from advocacy, and with alternative [designs] this is much less true.*
>
> ***–Judith Gueron, MDRC***

Acknowledging that "people don't wake up in the morning wanting to be in a randomized field trial," Gueron explained that MDRC has built a constituency over time by persuading participants that these methods are essential to improve policy. She said that strong statements about the validity of findings from randomized field trials by prestigious groups, including National Research Council committees (e.g., National Research Council, 1985, 1994), have helped to convince policy makers of the credibility of their findings. This credibility, in turn, has helped to ensure that these findings are translated into laws and programs. For example, when Congress amended the Social Security Act in 1988 (the Family Support Act of 1988, P.L. 100-485), it continued to allow states to waive provisions of the Aid to Families with Dependent Children law in order to test new approaches to welfare reform, but the waivers were available only if states assessed these new approaches. From the early 1980s through 1996, under both Republican and Democratic administrations, the U.S. Department of Health and Human Services interpreted this law as requiring states to conduct randomized field trials of the new approaches (Gueron, 1997). Outside observers agree that the studies conducted by MDRC have had a strong impact on welfare policy and practice, particularly on the Family Support Act of 1988 (e.g., Baum, 1991; Haskins, 1991).

Although most workshop speakers agreed that randomized field trials can potentially yield valid and reliable results, George Mason University professor Anthony (Eamonn) Kelly raised the most pointed questions about their viability in educational settings and thus their ultimate utility for improving policy and practice. Kelly argued that researchers face significant barriers translating the "ideal" design of a randomized field trial into a real-

life education study.[1] He illustrated these real-world "threats to internal and external validity" by describing problems he identified in a randomized field trial used to evaluate the Comer School Development Program in Chicago schools (Cook, Hunt, and Murphy, 2000). The Comer program aims to improve student achievement by improving the social climate in the school. In this study, researchers randomly assigned schools to the Comer program or to a control group. Kelly reported that due to high turnover of school principals, all elements of the Comer program were not carried out faithfully in the experimental group, which reduced the validity of the results. More generally, he argued that results from randomized field trials must be implemented and "diffused," and the field knows little about the factors that guide successful implementation. Kelly also suggested that it is often difficult to generalize the findings from randomized field trials because local factors may differ significantly from those of the schools in the study.

> *There are randomized field trials as intended, and there are randomized field trials as carried out. You're working with a system that's in flux.*
>
> ***–Anthony (Eamonn) Kelly, George Mason University***

Gueron, however, argued that designing randomized field trials involves trade-offs between internal and external validity that must be made in light of the goals of the study as well as other considerations, such as the state of knowledge in a field. For education, Gueron argued that since the use of randomized field trials is in its early stages, the first challenge is to show that such studies can be successfully conducted. In that context, she urged researchers to give priority to generating internally valid results, ar-

[1]Kelly sketched some supplementary methods that could inform the design of randomized field trials as well as implementation and diffusion studies. Describing what is typically referred to as "design research," he argued that studying, understanding, and improving educational practice must be framed "in terms of exploration and prototyping" until the operative factors and variables in the complex reality of schooling are better understood. He cited a recent special issue of *Educational Researcher* (Kelly, 2003) as providing further elaboration and criticism of these ideas.

guing that "it's better to learn something with confidence versus push for external validity and not understand at the end what you've got."

In response to a question at the workshop, Gueron described the disadvantages of *not* using this design. During the 1970s, she said, Congress passed a law guaranteeing jobs for young people in selected poor communities, if the young people would stay in—and make progress in—school. To carry out the law, the U.S. Department of Labor saturated the selected communities with funds to create jobs and MDRC tracked the young people's participation in school and work. When a National Research Council committee (1985) later reviewed the MDRC study, it agreed with the study's conclusion that the program produced a dramatic increase in work while the job guarantee was operating. However, committee members questioned the conclusion that it also produced a longer term increase in employment rates, asking whether more young people were working because of outside variables, such as changes in local labor markets in Baltimore, Cleveland, and other cities. In retrospect, Gueron said, she wished that the study team had used random assignment in some communities to yield stronger conclusions about the long-term effectiveness of the program.

3

When Are Randomized Field Trials Feasible?

As summarized in the previous chapter, workshop speakers suggested that investigators choose research designs using randomized field trials because they can help answer important questions about the systematic nature of effects and yield credible, useful findings. Since there have been very few randomized field trials conducted in educational settings to date, however, the field as a whole is in the early stages of learning how to conduct them.

In each of the three studies described in detail at the workshop, a researcher-practitioner team described how they implemented randomized field trials in educational settings. Boxes 3-1, 3-2, and 3-3 briefly summarize their experiences, illustrating the different approaches to implementing the same underlying logic of randomized field trials in urban school environments. The lessons they learned about what it takes to successfully conduct these studies are highlighted thematically in this and the final chapter.

Many of the challenges are not particularly unique [to implementing randomized field trials].

–Judith Gueron, MDRC

As the discussion that follows makes clear, there are a number of pragmatic issues that must be addressed. Challenges include meeting ethical and legal standards, recruiting and randomly assigning students or schools

or both, understanding the local educational context, and having adequate resources. Although Gueron pointed out that most of these challenges are not unique to randomized field trials, they all affect the extent to which investigators and educators can successfully implement randomized field trials in schools.

MEETING ETHICAL AND LEGAL STANDARDS

As described in Chapter 1, in a randomized field trial researchers attempt to create groups that are (statistically) equivalent except for the intervention or interventions assigned. For an investigator focusing on the reliability and validity of the results, this enables comparisons between the two groups that can support inferences about the effect of the controlled factors on any observed differences in specified outcomes. From an educator's perspective, however, any research (including randomized field trials) that involves controlled assignment to an intervention requires them to relinquish power in determining what type of education their students (or classrooms or schools) experience.

Some workshop speakers expressed ethical concerns about controlled assignment, while others suggested ways to address it. For example, Howard University dean of education Vinetta C. Jones said that educators in underserved schools with large numbers of minority students often believe (whether correctly or not) that randomized field trials "involve denying beneficial services or interventions to some students." In her remarks, Jones acknowledged that such perceptions may or may not be true, but nonetheless urged researchers that these concerns were real and must be taken seriously.

To avoid "poisoning the environment for future work" by failing to meet ethical standards, Gueron warned researchers: "Do not deny entitlements. Do not reduce service levels." Sharon Lewis, research director for the Council of the Great City Schools (a coalition of 60 of the largest urban school systems in the United States), agreed that researchers should not reduce service levels. She said that large urban school districts support randomized field trials when they are used to test program variations but oppose them when they involve excluding students from promising interventions in order to create control groups. However, committee member Jack Fletcher, in moderating the presentations and discussion of the three studies earlier in the day, argued that if interventions have not been subject to rigorous scrutiny, it is impossible to know whether the services are beneficial, have no effect, or may even be harmful.

In his presentation of the Power4Kids study (see Box 3-2), David Myers, vice president of Mathematica Policy Research, noted that this study was designed to partially address the concern that students may not receive a promising intervention. Each of the participating schools has been assigned to test one of four tutorials, so the study design does not exclude any school from these interventions. Nevertheless, in each participating school, some students will receive the tutorials, while a control group will continue to receive conventional instruction.

Furthermore, Gueron argued that a randomized field trial can provide a fair and objective way to simultaneously allocate the new service (since the process of random assignment ensures that everyone has an equal chance of being selected) and investigate its impact. For example, some cities (e.g., New York) have instituted the use of random lotteries to allocate limited school voucher funding to interested parents when demand outstrips funding levels. The use of these lotteries enables investigators to conduct a randomized field trial by comparing the outcomes of students who received vouchers to attend private schools with those students who applied but did not win the lottery and continued to attend public schools.

Workshop speakers emphasized that it is imperative for researchers to meet high ethical and legal standards today in order to overcome negative perceptions that have their roots in the history of social research. For example, Gueron said that some people have referred to MDRC researchers conducting randomized field trials of welfare programs as "perpetrators of Tuskegee," and Myers said that he and others at Mathematica have also been accused of "doing something like [the] Tuskegee" study. Their comments refer to the well-known Tuskegee Syphilis Study. For four decades, beginning in 1932, the U.S. Public Health Service carried out a longitudinal study of the natural history of syphilis among black men in the Tuskegee, Alabama, area. In that study, the researchers withheld treatment from 399 men with late-stage syphilis as well as 201 men free of the disease in order to study its progression (Reverby, 1998). The men in both groups were told they were participating in an experiment and receiving treatment.

Although the Tuskegee study was a natural history study and not a randomized field trial, the fact that investigators withheld treatment that was known to be effective from the sick men has influenced the public's response to many forms of research, including randomized field trials in school systems. As discussed further below (see Chapter 4), most workshop speakers agreed that it is possible to overcome these negative views by forg-

ing the respectful partnerships necessary to carry out successful and ethical research in a school setting.

Gueron argued that researchers conduct a randomized field trial only when ethical and legal standards can be met in both the design and implementation of the study. For example, when randomly assigning individuals or families in a school-based study, she said, ethical and legal standards require that investigators inform parents of the research and obtain their consent. She also emphasized that investigators must take steps to ensure that individual data and identifying information are kept confidential. Indeed, partly in response to public concern about the Tuskegee study, the federal government now regulates research involving human participants under the "common rule" (45 Code of Federal Regulations, Part 36.102i) (National Research Council, 2003). Under these regulations, universities and other research organizations have established institutional review boards to ensure that researchers provide human participants with full information about the potential risks and benefits of participating in certain kinds of research and obtain their informed consent. Because education research often involves human participants, the need to meet ethical and legal standards (including the standards imposed by institutional review boards) applies not only to randomized field trials but also to other types of education research.

ESTABLISHING ADEQUATE SAMPLE SIZE AND RECRUITING PARTICIPANTS

Several workshop participants highlighted the importance of the plan for randomization and analysis in randomized field trials. Two related concepts were discussed: the sample size (that is, the number of participants) and the unit of randomization (that is, what entity or entities are assigned to the experimental and control groups—student, classroom, school, or some combination). Gueron made the general point that for studies designed to address policy questions, the sample size must be large enough to detect policy-relevant effects of the intervention. In the presentations of each of the specific studies described at the workshop, sample size was an important topic of discussion.[1] Later in the day, Jones suggested that ensuring adequate sample sizes in urban settings may be difficult due to the

[1]A technique called power analysis can help determine the sample size necessary to detect effects of varying sizes in a particular study.

high mobility rates in central cities. Indeed, in at least one of the studies featured at the workshop, this was a problem: Olatokunbo (Toks) Fashola, research scientist at the Johns Hopkins University, described the specific problems she encountered recruiting enough participants into the Baltimore After-School Program Study and the small sample size that resulted (see Box 3-1).

A related issue is what unit is randomized. As Shavelson argued in his talk, all aspects of design depend heavily on the particular question that is posed. However, workshop discussions made clear that there are important nuances and constraints that influence the choice of unit of randomization in conducting randomized field trials in education. In a line of questioning related to the Baltimore After-School Program Study (in which 70 students were initially randomized) and the Power4Kids study (in which 52 schools and 772 students in those schools were initially randomized), a member of the audience observed that focusing on students as the unit of randomization may well be preferable from a cost perspective (for example, it is easier and cheaper to collect adequate data on 40 students than it is to collect adequate data on 40 schools). He raised concerns, however, about potential drawbacks. As he described it, the basic problem is that such a design typically requires the researcher to make the (often unrealistic) assumption that the effect of teachers on students is the same across different classrooms, leading to questionable conclusions.[2]

Once the plan for randomization and analysis has been established, the next step of the process is recruiting the participants. When required, the process of obtaining informed consent of participants in randomized field trials (and other research) in educational settings involves both technical and political challenges.[3] As described in Box 3-1, the first case study pre-

[2]This issue is a common methodological challenge associated with how to model outcomes in schools that are by their very nature "nested" (students within classrooms, classrooms within schools, and so on), and the role of "fixed" and "random" effects in multilevel modeling in particular. Regardless of whether the unit of randomization is the school, the student, or other entity, the effect of interventions in such nested environments can be estimated if sampling is conceptualized as multilevel. See Bryk and Raudenbush (1992) for a detailed treatment of this issue.

[3]Some education research is exempt from human subjects regulations because it does not present risk to the participants. As discussed in *Scientific Research in Education* (National Research Council, 2002): "education research, including evaluation studies, rarely presents any true risk to the participant so long as care is taken to protect identities and that researchers understand and are responsive to the needs of individual participants" (p. 152). Other stud-

sented at the workshop was a randomized field trial designed to estimate the effect of a one-on-one reading program for first graders who needed remediation and were enrolled in the Child First Authority (CFA) after-school program in Baltimore. First, the research plan was reviewed by the Johns Hopkins Institutional Review Board. Fashola, the principal investigator of the study, explained that, because Johns Hopkins "has been in and out of the news" due to concerns about the protection of human research participants, dealing with the institutional review board was hard, taking away time that she had planned to use to implement and study the program. Next, during the fall of 2002, she obtained approval from the Baltimore City Public School System, which she described as another tedious process. Fashola said that she could not provide informed consent forms to the teachers until she had obtained these approvals from the institutional review board at Johns Hopkins and the school system.

Finally, after the 2002-2003 school year had begun, teachers recruited students into the study, distributing consent forms to them to obtain their parents' written consent. Fashola explained that the schools had difficulty obtaining parent signatures, particularly from the parents of first graders most likely in need of the one-on-one reading tutorial, even though she extended the period to sign up for the program. Outside factors (see Box 3-1) slowed communication with parents. Ultimately, only 50 students (including experimental and control groups) remained in the study, limiting its ability to detect (statistically significant) effects.

As a technical issue, other workshop speakers suggested that such challenges associated with consent might be addressed by allowing ample time to communicate with students and their parents. Donna Durno, executive director of the Allegheny Intermediate Unit, who was in the early stages of the large Power4Kids study at the time of the workshop (see Box 3-2), said that even though their team started meeting with groups of parents and other stakeholders between six and eight months before the study began, the process was rushed. Indeed, Shep Kellam, public health psychiatrist of the American Institutes for Research, and Linda Chinnia, of the Baltimore City Public School System described a partnership that took two to three years to build in advance of the study of the whole-day program in Baltimore (see Box 3-3).

---

ies do not require written informed consent because no risk is involved or bias will be present in terms of who can provide consent. In all these cases, an institutional review board may approve the study under an exempt or expedited category.

BOX 3-1
Baltimore After-School Program Study

Olatokunbo (Toks) Fashola, a Johns Hopkins University research scientist, and Loretta McClairn, Child First Authority (CFA) program coordinator at Dr. Bernard Harris Elementary School, outlined a study of an after-school tutorial program. The program was based on reading curricula used in the Success For All school reform model as implemented in the Baltimore City Public School System. The study built on previous research indicating that the reading programs used by Success For All were effective in helping disadvantaged children learn to read as well as research on after-school programs (Fashola and Slavin, 1997; Ross, Smith, and Casey, 1997; Fashola, 2001).

Fashola chose a randomized field trial design to answer the question, "What is the effect of a one-to-one tutorial reading program on the standardized test scores of first grade students in need of remediation attending an after-school program in Baltimore city?" She said she and her colleagues at the Johns Hopkins University Center for Research on the Education of Students Placed at Risk have advocated for randomized field trials as a way to provide scientifically based evidence of program effectiveness. Fashola noted that it was timely to study an after-school program, because the Baltimore City Public School System master plan had established the goal of increasing academic achievement by implementing academically oriented after-school programs.

Noting that it is very difficult for researchers to enter a school "cold turkey," Fashola said she focused on first graders who were enrolled in the CFA after-school program with which she had a professional relationship. Fashola told the audience that, when describing the study to the CFA director, she explained that the funding was not adequate to provide the one-on-one tutorials to all CFA students, but that the study would pay to hire and train teachers to deliver the tutorials to an experimental group and to provide the schools (at no cost) with tutorial materials they could keep beyond the time of the study. The executive director, principals, and program coordinators welcomed the study as a way to help achieve the city school system's goal of providing academically oriented after-school programs.

McClairn explained that CFA offers academic enrichment, cultural enrichment, and homework help to about 170 students in eight Baltimore schools, from 2:30 pm to 5:00 pm, four days a week. Although CFA has many first-grade teachers in place (to keep a low

student-teacher ratio), McClairn said, these teachers didn't feel that they were in competition with the additional teachers hired with study funds to deliver the tutorials "because we're teammates."

The study was conducted during the 2002-2003 school year in four schools. Due to the unexpectedly long process of obtaining required approvals from Johns Hopkins University and the Baltimore City Public School System, the CFA teachers did not begin recruiting students to participate until after the school year began. Other problems, including a fatal fire in a CFA family's home that killed four CFA students and one parent volunteer, a change in the Maryland governorship that temporarily closed the CFA after-school program, and snow days in the severe winter of 2002-2003, also hurt recruitment. Ultimately, sample sizes in the four participating schools ranged from 8 to 16, and the total number of students enrolled in the study from beginning to end was 70, although due to attrition the final sample size was 50.

In the fall of 2002, participating students were pretested using three subtests of the Woodcock Reading Mastery Test (letter identification, word identification, and word attack). They were then randomly assigned according to school into either an experimental or a control group. From November 2002 until May 2003, students in the experimental group were provided with individual tutoring sessions lasting 30 minutes three times per week. Students in the control group had opportunities for academic activities that included homework help, group tutoring, and enrichment programs, but they did not receive the Success For All individual tutoring intervention. As one way to ensure that students in the control group did not receive the tutoring intervention, only teachers in the experimental group were trained. In addition, in order to minimize "transfer" of elements of the program to participants in the control group, the regular school-day first-grade teachers were not allowed to deliver tutorial sessions after school. The specially hired teachers did not interact with students in the control group.

At the end of the tutorial period, students in both groups were administered post-tests. Results of the study showed that although all students performed better on the post-tests, and although the experimental group outperformed the control group students on all measures, the differences between the two groups were statistically significant only on the word attack subtest.

Funding for the study was provided by the U.S. Department of Education's Office of Educational Research and Improvement, now the Institute of Education Sciences.

BOX 3-2
Power4Kids Study

David Myers, vice president of Mathematica Policy Research, and Donna Durno, executive director of the Allegheny Intermediate Unit, described the background and implementation of this ongoing study. Descriptive research indicates that 40 percent of fourth graders in the United States do not read at their grade level. Further research by Florida State University professor Joseph Torgesen indicates that, by the time students reach grades three through five, there is a large gap in reading ability between students who read at their grade level and those who do not. On the basis of that research, Torgesen has called for intensive interventions that could bring students up to grade level, possibly within a single school year.

The Power4Kids study builds on this research, Myers said. As a first step, the Haan Foundation for Children convened 15 to 20 publishers of reading tutorials to "show their wares." Following a review, Torgesen and other members of the research team (which includes Mathematica Policy Research and the American Institutes for Research) selected four tutorials for inclusion in the Power4Kids study. In making its selections, the team considered previous research on the effectiveness of the programs, including small randomized field trials and quasi-experimental studies. Myers said that the study was designed to address the following questions:

- Can children who have reading difficulties in middle to late elementary school acquire adequate reading skills in a short period of time if they are taught with intensity and skill?
- Can intensive interventions affect all critical reading skills, such as accuracy, comprehension, and fluency?
- Do some children benefit more or less from these intensive and well-implemented reading interventions?

Power4Kids includes evaluation of the four selected tutorials in a pullout program, an impact study, a fidelity study, and a cost study. The impact study is currently under way in several suburban school districts near Pittsburgh, all affiliated with the Allegheny Intermediate Unit. In order to assess the effect of the four reading tutorial interventions, Mathematica researchers chose a "scientifically rig-

orous" randomized field trial design that will include collection of longitudinal data for three years.

Durno described the challenges of recruiting over 40 schools to be randomly assigned to one of four different reading interventions. She explained that, as executive director of the Allegheny Intermediate Unit, which provides resources, instruction, and education services to the affected school districts, she had a "credible" relationship with the schools, so "it works out well if we do the recruiting." Nevertheless, there have been challenges in implementing this large-scale study. Each school district has its own climate and culture, and there are philosophical differences that had to be overcome. In addition, each school district has its own informal power structure and different decision-making processes regarding instructional programming. Ultimate authority could rest with the school board, the superintendent, the school principal, or the curriculum coordinator. Through frequent and consistent communication, these challenges have been addressed.

Myers explained his strategy for ensuring that each group receives only one of the four interventions. He said he had asked each participating school to nominate one teacher to provide the remedial reading interventions, replacing these teachers with long-term substitutes for the entire school year. These teachers were given specific training and materials, to help ensure that they carry out the alternative reading programs as designed. They are unlikely to "accidentally" provide the remedial reading tutorials to students not in their experimental groups, because they will not be acting as regular classroom teachers during the 2003-2004 school year.

Researchers tested third and fifth grade students to identify readers with reading proficiencies below the 20th percentile. Among those eligible to participate, researchers randomly assigned some to receive the form of tutorial assigned to the school, and others to a control group. Those receiving the tutorial will work in small groups of no more than three children with one teacher for one hour each day, receiving a total of about 100 hours of instruction. Since this study is just under way, the results are not yet known.

Power4Kids is funded by the U.S. Department of Education. The Haan Foundation for Children helped formulate the idea for the study and brought the partners together.

BOX 3-3
Baltimore Whole-Day First-Grade Program Study

Public health psychiatrist Sheppard Kellam of the American Institutes for Research and Linda Chinnia of the Baltimore City Public School System described this study in the context of the larger, long-term Baltimore Prevention Program. Kellam explained that the program is based on research and theory in child and adolescent development (Kellam and Van Horn, 1997).

The goal of the prevention program is to get children off to a good start in school in order to prevent later school failure, substance abuse, and mental and behavioral disorders among teenagers (Johns Hopkins University School of Public Health, 2003). The program is based on earlier studies that show how children behave, learn, and feel about themselves in first grade are good indicators of whether they will have problems as teenagers (e.g., Boyd et al., 2003). Previous randomized field trials have assessed the impact of different first-grade interventions designed to reduce these behaviors and learning problems (e.g., Kellam et al., 1998).

Kellam said that the current randomized field trial is designed to assess the effects of an integrated set of preventive first-grade interventions. The interventions are directed at improving (1) teachers' classroom behavior management, (2) family-classroom partnerships regarding homework and discipline, and (3) teachers' instructional practices regarding academic subjects, particularly reading. These interventions have been combined in a single whole-day first-grade classroom program.

Other speakers suggested that the challenge of obtaining informed consent from large numbers of students requires addressing not only technical but also deeper political issues. For example, Chinnia noted that the study of the whole-day program in Baltimore schools requires "highly intrusive activities" that disrupt normal school and family patterns. These include the investigation of teaching practices and curricula, teachers contacting families about student progress, and classroom behavior management strategies that affect peer group relationships. In addition, random assignment of teachers and students is a change in normal school practices. However, she said that "strong institutional and community partnerships" and "shared values and mutual respect" helped sustain support for random assignment and other aspects of the study. Building on Chinnia's remarks, Kellam said,

In the current study, researchers will assess the effect of the program by randomly assigning children and teachers to intervention classrooms or standard program classrooms (control condition) in elementary schools. They will also measure variation in the impact of the program that may be due to variation both in the ways teachers implement the specific elements of the program and in the children themselves.

Chinnia explained that the Baltimore City Public School System supports this study because it lays the foundation for translating its findings into policy and practice. In addition to assessing the impact, as outlined above, the researchers will follow the first-grade children as far as the end of third grade, and they will also follow their first-grade teachers over two subsequent cohorts of first graders. This long-term observation will allow researchers to test whether the multiple levels of support and training for teachers sustain high levels of program practice. The study will also test in the fourth year whether the support and training structure is successful in training nonprogram teachers. Another element of the study—which has not yet been funded—could potentially be very useful to the schools. If funding is obtained, the researchers will conduct a cost-effectiveness study of the program, comparing program costs with the potential long-term cost savings that would result from reductions in drug use, behavior problems, and dropping out during the teenage years.

The study is supported by the National Institute on Drug Abuse and other funders.

"It's a mistake to think of the IRB (institutional review board) process and the . . . informed consent process separately from the partnership." The importance of such partnerships was a recurring theme in the workshop, and we return to it in more detail in Chapter 4.

GROUNDING THE STUDY IN THE RELEVANT EDUCATIONAL CONTEXT

Many workshop speakers agreed that a randomized field trial is most appropriate when it is responsive to the current political and economic context of schools. For example, Chinnia suggested that researchers take time to analyze the social and political structure of the school district, learn

the school's vision, and understand its challenges and goals. Because Kellam and his colleagues took this approach, they have formed a partnership with the Baltimore schools that has successfully supported three generations of randomized field trials. The studies are designed to evaluate approaches to improved management and interaction in first-grade classrooms and to assess the effect of these approaches in reducing later drug abuse, crime, and academic failure.

Some workshop speakers suggested that the No Child Left Behind Act may discourage schools from participating in randomized field trials, even though the law explicitly encourages "scientifically based" education research. Wesley Bruce of the Indiana Department of Education said, in trying to comply with the goals of the act, teachers and administrators in his state are focusing less on students and more on test scores as measures of schools' performance. To comply with the law, the Indiana Education Department has identified the 25 percent of all schools that have failed to make "adequate yearly progress" and has assigned departmental employees to work with these schools. Bruce said that the education department staff will be able to help only the poorest performing of the bottom 25 percent of schools, leaving the top schools in the group with only a list of best practices. Although the What Works Clearinghouse (see Chapter 1) may be helpful in this regard, Bruce questioned whether there was enough time to conduct randomized field trials to identify, learn more about, and implement best practices within the time periods set by the legislation. Referring to the adequate yearly progress requirements, Bruce said, "you don't have three years, four years to conduct research and get results back to schools about good practice," because every year "the bar has been raised higher for the level of performance."

> *When we look at how you conduct this research [randomized field trials] in schools, researchers need to understand that for every single school the stakes are high and have gotten higher.*
>
> ***–Wesley Bruce, Indiana Department of Education***

Bruce went on to note several challenges to implementing randomized field trials that paralleled earlier comments: (1) the political reality that schools, school districts, and teachers like local control and may not welcome federally funded research or researchers (echoing a similar statement

by Durno); (2) teachers who are satisfied with their own teaching methods may not faithfully implement the educational intervention being studied; (3) "good teachers will share ideas that seem to be working," which may mean that they share ideas from the intervention, which could lead to the control group students receiving intervention strategies; and (4) it may be difficult to provide adequate review to ensure that ethical and legal standards are met. Finally, he said, researchers would need buy-in throughout a school system when conducting a randomized field trial, because superintendents change positions frequently.

Lewis agreed with Bruce that national requirements for improvement in test scores may discourage school districts from participating in randomized field trials. She said that such participation can take time away from instruction at a time when No Child Left Behind requires improvement in test scores.

SECURING ADEQUATE RESOURCES

Gueron suggested that a randomized field trial is more likely to succeed when resources are adequate to address the feasibility criteria and challenges outlined above. Reflecting on Shavelson's presentation, Gueron agreed that researchers should select research methods appropriate to the questions being asked. She cautioned participants, however, that answering policy-relevant questions about the effectiveness of interventions depended in part on having adequate resources. Although questions about whether a widely used educational intervention has a systematic effect may best be answered with a large-scale randomized field trial, she said that such studies "can't be done on the cheap in terms of resources or time." Noting that a successful randomized field trial requires creativity, flexibility, and "operational and political savvy," Gueron said that funding should be adequate to support the salaries of "very senior people" who possess these abilities. Financial resources are needed to successfully carry out the random assignment process and to gather data on both control and experimental groups over "an adequate length of time," Gueron said. When funding is available, she said, it is also useful to replicate a randomized field trial of a promising intervention in several different areas, to test effectiveness in diverse settings.

Gueron explained that although large-scale randomized field trials require considerable resources, they may be more cost-effective overall than studies using alternative research designs (e.g., quasi experiments). She

> *High-quality large-scale field studies with primary data collection are very expensive, and actually they're getting more expensive. And this is true for randomized field trials but it is also true for the alternative [designs].*
>
> ***—Judith Gueron, MDRC***

noted that any large study involving the collection of primary data in the field is expensive, but a randomized field trial requires higher costs in the start-up and operational stages than designs that do not feature randomization. This cost differential stems from the start-up phase of designing, selling, and initiating the study, which Gueron called the "make it or break it" point for success. In this start-up phase, researchers and educators develop what Kellam described as "the essential partnership." In the operational phase, randomized field trials have some added costs (for policing the implementation of random assignment, which Gueron called an "all or nothing process") but otherwise are comparable to quasi-experimental studies in the resources required to obtain equivalent, high-quality data on both experimental and control/comparison groups, as well as to gather information on the process and context for the program. However, she maintained that the later stages, including the analysis of data and diffusion of study results, were less expensive than in other types of large-scale studies of social programs. When looked at from the broader perspective of policy impact per dollar spent, Gueron concluded randomized field trials may be less expensive than quasi-experimental research designs that have high "political and financial costs" when they "end in methodological disputes."

4

How Can Randomized Field Trials Be Conducted in an Educational Setting?

The researchers and practitioners who spoke at the workshop offered several practical lessons for planning and conducting randomized field trials in education. At the conclusion of the workshop, committee member Kay Dickersin summarized these lessons, suggesting that success in conducting a randomized field trial in an educational setting involves four interdependent steps: (1) develop a true partnership with the school community, (2) ensure the internal validity of the study design and implementation, (3) focus on recruiting sufficient numbers of study participants, and (4) plan for future implementation of an intervention if the study shows it is effective.

DEVELOPING A PARTNERSHIP

By far, the single strongest message that emerged from the dialogue during the workshop is that developing and nurturing a true partnership between the research team and the relevant education communities is critically important to success in carrying out any research in schools, including randomized field trials. In each of the three studies featured at the workshop, researchers were able to gain entry to the schools, to ensure cooperation in faithfully carrying out the interventions, and to make progress toward mutual goals only by establishing trust and encouraging open communication. Their experiences suggest that it is nearly impossible for

researchers to conduct randomized field trials in schools unless both researchers and education service providers take time to understand each others' goals and develop a study design that will help both parties to reach them.

> *It is very difficult for researchers to walk into any type of school setting cold turkey and say we would like to engage in a randomized field trial study.*
>
> ***–Olatokunbo (Toks) Fashola, Johns Hopkins University***

All three studies showed the value of partnering (see Boxes 3-1, 3-2, and 3-3). For example, the Baltimore After-School Program Study was facilitated by the existing relationship between Fashola (the chief researcher) and key staff members of the Child First Authority (CFA) after-school program where the study was conducted. The Power4Kids study is based on a formal partnership between Myers and the research team and the Allegheny Intermediate Unit, a group that works with a consortium of school districts outside Pittsburgh that is participating in the study.

The Baltimore Whole-Day First-Grade Program Study demonstrates most clearly the value of taking the time to identify the education community's goals and interests. In their presentation, Kellam and Chinnia described how their partnership helped both the education community and the research team meet their goals. Kellam asserted that when a partnership is in place based on "mutual self-interests at multiple levels," then "consent sounds like a silly word"—illustrating how key implementation tasks such as recruitment are facilitated by the relationship. Chinnia described some of the "self-interests" that led to the long-term partnership between the Baltimore City Schools and the American Institutes for Research. She explained that randomized field trials helped to meet several of the school system's goals, including intervening early in elementary school to enhance and maintain student achievement, identifying best practices for instruction and classroom management, and promoting parent involvement in students' progress. She noted that the current study could help to sustain best practices in a whole-day first-grade program, and that the goal of creating and sustaining whole-day first-grade programs is included in the Baltimore City Public School System's master plan.

> *In order to do this type of research, especially randomized field trials, it's important that we have very strong partnerships. Not only do we have partnerships within the school system but also with the community at large. Some of this [partnership building] has taken two to three years of planning.*
>
> ***–Linda Chinnia, Baltimore City Public School System***

Workshop speakers also suggested that in large, heavily minority urban school districts, researchers must be sensitive to issues of race and power when seeking to develop such partnerships. For example, Jones said that, partly because of the lack of diversity in the research community, some minorities do not trust researchers because they feel that "poor and minority groups are the most evaluated and researched populations while they have no input into the process." Lewis expressed similar views, saying, "We are difficult to work with, yes we are, but many of us have been burned, so we have reason to be difficult." As Dickersin put it in her closing remarks, creating "culturally competent" research teams who have experience working with urban schools is critically important to the success of the research.

Some workshop speakers argued that creating racially and ethnically diverse research teams can be an important step toward enhancing cultural competence, building trust, and developing partnerships. Lewis suggested that, to accomplish this goal, research organizations could, for example, find competent black researchers through the American Educational Research Association's special interest groups focusing on urban and minority education (American Educational Research Association, 2003). In addition, she proposed that these more diverse research teams collaborate closely with researchers employed by urban school districts. She said that several large urban school districts (including Houston and Atlanta) have outstanding research staff who would welcome the opportunity to collaborate in randomized field trials. Jones echoed these sentiments, suggesting that such efforts can help to break down historical power dynamics between researchers (often white) and students and education professionals in urban school districts (often nonwhite), which might otherwise pose barriers to establishing mutual trust. Questioning the capacity of the current cadre of researchers to develop partnerships with inner-city schools in particular, Lewis

argued that researchers are frequently unfamiliar with inner-city schools and unsure of how to work with them.

> *Many of the researchers are not skilled in working with people in urban centers.*
>
> ***–Sharon Lewis, Council of the Great City Schools***

Finally, establishing close partnerships can help schools not only to identify but also to implement scientifically based best practices. These best practices may in turn help schools make the adequate yearly progress required by the No Child Left Behind Act. Developing and nurturing this partnership between the researchers and professionals working in the study sites facilitates *all* other implementation tasks—including the three described next.

ENSURING INTERNAL VALIDITY

Dickersin also pointed out that several workshop speakers emphasized the need to spend time in the early design and planning phases to ensure internal validity of the study. Indeed, in debating the merits of randomized field trials in education, many scholars have focused on the trade-off that is necessary between internal validity—that is, the extent to which it can be concluded that the treatment led to the effect, or difference, between one group and another on a particular outcome—and external validity—the extent to which the findings of a particular study hold in other times, places, and populations. For example, critics argue that the strict protocols of the studies, which are required to maximize internal validity (e.g., program options are consistently and comprehensively implemented in both the experimental and control groups), do not reflect typical school operations, and thus the usefulness of the results for real life is questionable (Cronbach et al., 1980; National Research Council, 2002).

Several workshop speakers and audience members raised questions about a particularly important aspect of comparative studies, including randomized field trials, in ensuring internally valid results: how to measure and account for the implementation of the experimental and control treatments. Gueron argued that it is important for investigators to monitor and to understand how the interventions are being applied during the study to

assist with the interpretation of the comparison of outcomes between the groups. She cited the collection of "equivalent and high-quality data" on process and contextual factors in both groups as essential to the success of a randomized field trial. Shavelson, too, urged researchers to spend the time and money required to observe and characterize implementation, since how a program plays out in schools can often be quite different from what was planned or expected at the outset of the study. Kellam, too, argued that it is extremely important to measure implementation in the control group "with the same intensity" as in the experimental group.

Two speakers offered specific strategies, based on early planning, that helped to ensure that the intervention is carried out faithfully and that the randomized field trial would yield internally valid results. In the Baltimore After-School Program Study, Fashola arranged for the CFA to hire new non-first-grade teachers who were not a part of the regular CFA program to provide one-on-one tutoring to the experimental group of first graders. Using new teachers helped to ensure that the existing CFA did not provide similar tutoring to students in the control group. In the Power4Kids study in the Pittsburgh area, Myers described a similar approach. He asked each participating school to nominate one teacher to provide the remedial reading interventions, replacing these teachers in their usual positions with long-term substitutes for the entire school year. These teachers were given specific training and materials to help ensure that they carry out the alternative reading programs as designed. In this way, Myers argued that they are unlikely to "accidentally" provide the remedial reading tutorials to students not in their experimental groups, because they will not be acting as regular classroom teachers. Researchers employing randomized field trials can also use observational methods to detect factors that might contaminate the results of the study. For example, in the Comer study described by Kelly, many ethnographers were hired to study and characterize program implementation, a major factor in the researchers' ability to identify threats to internal validity.

RECRUITING SUFFICIENT NUMBERS OF STUDY PARTICIPANTS

As summarized in Chapter 3, Dickersin reiterated that recruiting study participants is critical to ensure sufficient sample sizes and to be able to draw valid and reliable conclusions from a randomized field trial in turn. The implementation of this step, too, depends in large part on the broader

partnership between researchers and educators. According to several workshop speakers, success in recruiting study participants can be aided by key intermediaries. At the highest level, a board including school, community, and research officials—such as the Baltimore City Community and Institutional Board, which oversees the long-term research program there—can help to both oversee and develop support for the study. When many school districts are involved in a large-scale study, an intermediary such as the Allegheny Intermediate Unit led by Durno, who has strong working relationships with the schools, can help to forge partnerships. At the level of the individual school, a designated site coordinator can help to communicate with teachers and parents and ensure that random selection and experimental and control processes flow smoothly, and that differences between the experimental and control groups are preserved over the course of the study. In describing the use of such a strategy, Durno put it this way, "the on-site coordinator must be an excellent communicator who can meet the school on that school's terms." Chinnia trained a community organizer, who was already trusted by parents, to help explain the study and win the support of students and parents. Now in a full-time position as a "family classroom partnership aide," the coordinator has won consent among 100 percent of parents when recruiting students to participate in the most recent generation of randomized field trials.

> *One of the challenges that we experienced in working with the schools is the difference in the climate and the culture of each school.*
>
> ***–Donna Durno, Allegheny Intermediate Unit***

PLANNING FOR IMPLEMENTATION

When educators consider the possibility of participating in a randomized field trial, workshop speakers suggested that the pivotal factor is often whether the study is likely to lead to an actual improvement in their own schools. Researchers can help educators reach this objective by designing the study to support future implementation of an intervention if it has been shown as effective and, after the study is completed, by working with the school and funders to provide the effective intervention to all students. The Baltimore Whole-Day First-Grade Program Study plans for and sup-

ports future implementation of the intervention by including analysis of the cost-effectiveness of the whole-day program, in light of its potential to reduce the future costs associated with drug abuse and school failure.

In all three of the studies described at the workshop, research teams provided materials and training to administrators and teachers, at no cost to the schools. Workshop speakers describing these studies suggested that investigators wishing to conduct a successful randomized field trial should consider whether they have adequate funding to provide such services. With adequate funding, researchers can form a research partnership that will help schools reach their goals, which include implementing (not just studying) educational interventions if they are found to be successful. Providing a proven program to all students following the study ensures that the study results are used and useful, not only by policy makers, but also by classroom teachers.

LOOKING AHEAD

Taken as a whole, the workshop presentations and discussions suggested that the outlook for continuing and expanding randomized field trials in schools, based on strong community-research partnerships, is good. In her concluding remarks, Dickersin said that one of the most important things she had learned from the workshop was that some educators were "very positive about participating" in randomized field trials "if the barriers . . . can be overcome." She referred to an earlier session, in which she had asked education representatives how they would respond if she proposed to conduct a randomized field trial on an issue important to their schools, with adequate resources and a culturally sensitive research team. In response, Lewis said, "I would jump at it," while Jones said that her students (studying to become teachers and education researchers) would also want to be partners in the research.

The three studies featured at the workshop also indicate that when designed to help meet their goals, educators and school officials would welcome the opportunity to participate in, and help carry out, randomized field trials. In presenting each case, representatives of the education communities described why they viewed participating in such studies as beneficial for their schools. Their enthusiasm at the workshop appeared similar to what Gueron said she had encountered among state and local welfare officials who have willingly participated in repeated randomized field trials over the years, because "they actually believe it makes a difference and . . . can bring visibility and resources to their community."

References

Achilles, C. (1999). *Let's put kids first, finally: Getting class size right.* Thousand Oaks, CA: Corwin Press.

American Educational Research Association. (2003). *Special interest groups of AERA, 2003-2004 directory.* Available: http://www.aera.net/sigs/siglst.htm#v [December 23, 2004]. (Special Interest Groups focusing on urban and minority children include the SIG "Research Focus on Black Education," "Urban Learning, Teaching, and Research," and "Education and Student Development in Cities.")

Baum, E.B. (1991). When the witch doctors agree: The Family Support Act and social science research. *Journal of Policy Analysis and Management, 10*(4), 603-615.

Berliner, D. (2002). Educational research: The hardest science of all. *Educational Researcher, 31*(8), 18-20.

Bohrnstedt, G.W., and Stecher, B.M. (1999). *Class size reduction in California 1996-1998: Early findings signal promise and concerns.* Palo Alto, CA: CSR Research Consortium, American Institutes for Research.

Boruch, R., de Moya, D., and Snyder, B. (2002). The importance of randomized field trials in education and related areas. In R. Boruch and F. Mosteller (Eds.), *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

Boyd, R., Lambert, S., Cooley-Quille, M., and Ialongo, N. (2003). First grade risk behaviors for community violence exposure in middle school. *Journal of Community Psychology, 31*(3), 1-18.

Bruner, J. (1996). *The culture of education.* Cambridge, MA: Harvard University Press.

Bryk, A.S., and Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Campbell, D.H. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54,* 297-312.

Cook, T.D., and Payne, M.R. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller and R. Boruch (Eds.), *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution Press.

Cook, T.D., Hunt, H.D., and Murphy, R.F. (2000). Comer's School Development Program in Chicago: A theory-based evaluation. *American Education Research Journal*, *37*(2), 535-597.

Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F., and Weiner, S.S. (1980). *Toward reform of program evaluation.* San Francisco: Jossey-Bass.

Fashola, O.S. (2001). *Building effective after school programs.* Thousand Oaks, CA: Corwin Press.

Fashola, O.S., and Slavin, R.E. (1997). *Effective and replicable programs for students placed at risk in elementary and middle schools.* (Paper written under funding from the Office of Educational Research and Improvement, U.S. Department of Education.) Available: http://www.successforall.net/resource/researchpub.htm [December 8, 2003].

Gueron, J. (1997). Learning about welfare reform: Lessons from state-based evaluations. *New Directions for Evaluation, 76*(Winter*),* 79-94.

Haskins, R. (1991). Congress writes a law: Research and welfare reform. *Journal of Policy Analysis and Management, 10*(4), 616-632.

Johns Hopkins University School of Public Health. (2003). *The prevention program.* Available: http://www.bpp.jhu.edu/index.htm [December 23, 2003].

Kellam, S.G., and Van Horn, Y.V. (1997). Life course development, community epidemiology, and preventive trials: A scientific structure for prevention research. *American Journal of Community Psychology, 25*(2), 177-187.

Kellam, S.G., Ling, X., Merisca, R., Brown, C.H., and Ialongo, N. (1998). The effect and level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology, 10*(2), 165-185.

Kelly, A. (Ed.). (2003). The role of design in educational research [Special issue]. *Educational Researcher, 32*(1), January/February.

Mosteller, F., and Boruch, R. (2002). *Evidence matters: Randomized trials in education research.* Washington, DC: Brookings Institutions Press.

National Research Council. (1985). *Youth employment and training programs: The YEDPA years.* Committee on Youth Employment Programs. C.L. Betsey, G. Hollister, and M. Papgeorgiou, eds. Washington, DC: National Academy Press.

National Research Council. (1994). *Preparing for the workplace: Charting a course for federal postsecondary training policy.* Committee on Postsecondary Education and Training for the Workplace. J.S. Hansen, ed. Washington, DC: National Academy Press.

National Research Council. (2002). *Scientific research in education.* Committee on Scientific Principles for Education Research. R.J. Shavelson and L. Towne, eds. Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2003). *Protecting participants and facilitating social and behavioral sciences research.* Panel on Institutional Review Boards, Surveys, and Social Science Research. C.F. Citro, D.R. Ilgen, and C.B. Marrett, eds. Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Reverby, S.M. (1998). History of an apology: From Tuskegee to the White House. *Research Nurse, 3*(4). Available: http://www.researchpractice.com/archive/apology.shtml [October 21, 2003].

Ross, S.M., Smith, L.J., and Casey, J.P. (1997). Preventing early school failure: Impact of Success for All on standardized test outcomes, minority group performance, and school effectiveness. *Journal of Education for Students Placed at Risk, 2*(1). Available: http://www.csos.jhu.edu/jespar/vol_2_no_1.htm [March 4, 2004].

Slavin, R. (2002). Evidence-based educational policies: Transforming educational practice and research. *Educational Researcher, 31*(7), 15-21.

U.S. Department of Education. (2002). *Strategic plan 2002-2007*. Washington, DC: Author. Available: http://www.ed.gov/about/reports/strat/plan2002-07/plan.pdf [November 24, 2003].

What Works Clearinghouse. (2003). *Introduction to the What Works Clearinghouse evidence report process and the role of scientific standards.* Rockville, MD: Author. Available: http://www.w-w-c.org/july2003.html [November 25, 2003].

Willinsky, J. (2001). The strategic education research program and the public value of research. *Educational Researcher, 30*(1), 5-14.

Appendix A

Workshop Agenda

Randomized Field Trials (RFTs) in Education:
Implementation & Implications—September 24, 2003
The National Academies—Keck 100

8:00 am **Continental Breakfast**

8:30 am **Workshop Objectives & Overview**
Lauress (Laurie) Wise, HumRRo, Committee Chair
Lisa Towne, National Research Council, Study Director

Session 1. RFTs in Context

What is the role of RFTs in research and research methods? And how are they implemented in social settings, including educational sites? Two lead presentations will place the role of RFTs in context for the day's discussion.

Committee Moderator: Brian W. Junker

8:45 am **Nature of Education Research & Methodology**
Richard J. Shavelson, Stanford University

9:15 am **Implementing RFTs in Social Settings**
Judith Gueron, MDRC

9:45 am **Q&A**

10:30 am **Break**

Session 2. RFTs in Educational Settings: Lessons Learned

This session will explore implementation issues associated with RFTs in educational settings, with a focus on how implementation influences the provision of education (e.g., student access to interventions, teacher/administrator workloads) as well as research process and products (e.g., design features, data collection, nature of inferences). Discussions of three studies—led by researcher/policymaker/practitioner teams—will address these issues by describing relevant political, policy, legal, and ethical contexts, outlining research questions and methods, and participant recruitment, costs, and attrition.

Committee Moderator: Jack Fletcher

10:45 am **Case 1: Baltimore After-School Program Study**
Olatokunbo (Toks) Fashola, Johns Hopkins University
Loretta McClairn, Baltimore City Public School System

11:15 am **Case 2: Power4Kids Study**
David Myers, Mathematica Policy Research
Donna Durno, Allegheny Intermediate Unit

11:45 pm **Case 3: Baltimore Whole-Day First-Grade Program Study**
Sheppard Kellam, American Institutes for Research
Linda Chinnia, Baltimore City Public School System

12:15 pm **Lunch and Q&A**

Session 3. Implications for Research & Practice

Given the current push for more RFTs in federal education law, what do these implementation issues mean for education and education research? Experts will address this question with respect to a handful of key stakeholder groups: education researchers, states, urban districts, and student populations who have been traditionally underserved.

Committee Moderator: Robert E. Floden

1:45 pm	**Implications for Education Research & Researchers** Robert Boruch, University of Pennsylvania Anthony (Eamonn) Kelly, George Mason University
2:15 pm	**Q&A**
3:00 pm	**Break**
3:15 pm	**Implications for States** Wesley Bruce, Indiana Department of Education
3:30 pm	**Implications for Urban Districts** Sharon Lewis, Council of the Great City Schools
3:45 pm	**Implications for Traditionally Underserved Populations** Vinetta C. Jones, Howard University
4:00 pm	**Q&A**
4:30 pm	**Wrap-Up Discussion of Themes & Implications** Kay Dickersin, Committee Member
5:00 pm	**Adjourn**

Appendix B

Biographical Sketches of Committee Members and Workshop Speakers

COMMITTEE MEMBERS AND STAFF

Lauress L. Wise (*Chair*) is president of the Human Resources Research Organization (HumRRO). His research interests focus on issues related to testing and test use policy. He has served on the National Academy of Education's Panel for the Evaluation of the National Assessment of Educational Progress (NAEP) Trial State Assessment, as coprincipal investigator on the National Research Council's (NRC) study to evaluate voluntary national tests, and as a member of the Committee on the Evaluation of NAEP. He has been active on the NRC's Board on Testing and Assessment, the Committee on Reporting Results for Accommodated Test Takers: Policy and Technical Considerations, and the Committee on the Evaluation of the Voluntary National Tests, Year 2. At HumRRO, he is currently directing an evaluation of California's high school graduation test and a project to provide quality assurance for NAEP. Prior to joining HumRRO, he directed research and development on the Armed Services Vocational Aptitude Battery for the U.S. Department of Defense. He has a Ph.D. in mathematical psychology from the University of California, Berkeley.

Linda Chinnia is an educator with the Baltimore City Public School System. During a 32-year career, she has served as an early childhood teacher, a senior teacher, a curriculum specialist, an assistant principal, a principal, and the director of elementary school improvement. Currently she serves as

an area academic officer, supervising 35 elementary and K-8 schools. She has been an adjunct instructor at the Baltimore City Community College, Coppin State College, Towson University, and Johns Hopkins University. She has taught courses in early childhood education, elementary education, and educational supervision and leadership. She has B.A. and M.A. degrees from Towson University.

Kay Dickersin is a professor at the Brown University School of Medicine. She is also director of the U.S. Cochrane Center, one of 14 centers worldwide participating in The Cochrane Collaboration, which aims to help people make well-informed decisions about health by preparing, maintaining, and promoting the accessibility of systematic reviews of available evidence on the benefits and risks of health care. Her areas of interest include publication bias, women's health, and the development and utilization of methods for the evaluation of medical care and its effectiveness. She was a member of the Institute of Medicine's Committee on Reimbursement of Routine Patient Care Costs for Medicare Patients Enrolled in Clinical Trials, the Committee on Defense Women's Health Research, and the Committee to Review the Department of Defense's Breast Cancer Research Program. She has an M.S. in zoology, specializing in cell biology, from the University of California, Berkeley, and a Ph.D. in epidemiology from Johns Hopkins University's School of Hygiene and Public Health.

Margaret Eisenhart is professor of educational anthropology and research methodology and director of graduate studies in the School of Education, University of Colorado, Boulder. Previously she was a member of the College of Education at Virginia Tech. Her research and publications have focused on two topics: what young people learn about race, gender, and academic content in and around schools; and applications of ethnographic research methods in educational research. She is coauthor of three books as well as numerous articles and chapters. She was a member of the NRC's Committee on Scientific Principles in Education Research. She has a Ph.D. in anthropology from the University of North Carolina at Chapel Hill.

Karen Falkenberg is a lecturer in the Division of Educational Studies at Emory University. She is also the president of the Education Division of Concept Catalysts, a consulting company that has a specialization in science, mathematics and engineering education reform. She works both nationally and internationally. She was the program manager for the National

Science Foundation funded local systemic change initiative in Atlanta called the Elementary Science Education Partners Program, and has been a mentor for SERC@SERVE's Technical Assistance Academy for Mathematics and Science and for the WestEd National Academy for Science and Mathematics Education Leadership. She also served on the National Academy of Engineering's Committee for Technological Literacy. Earlier, she was a high school teacher of science, mathematics, and engineering and was featured as a classroom teacher in case studies of prominent U.S. innovations in science, math, and technology education. Before she became an educator, she worked as a research engineer. She has a Ph.D. from Emory University.

Jack McFarlin Fletcher is a professor in the Department of Pediatrics at the University of Texas-Houston Health Science Center and associate director of the Center for Academic and Reading Skills. For the past 20 years, as a child neuropsychologist, he has conducted research on many aspects of the development of reading, language, and other cognitive skills in children. He has worked extensively on issues related to learning and attention problems, including definition and classification, neurobiological correlates, intervention, and most recently on the development of literacy skills in Spanish-speaking and bilingual children. He chaired the National Institute for Child Health and Human Development (NICHD) Mental Retardation/Developmental Disabilities study section and is a former member of the NICHD Maternal and Child Health study section. He recently served on the President's Commission on Excellence in Special Education and is a member of the NICHD National Advisory Council. He was a member of the NRC's Committee on Scientific Principles in Education Research. He has a Ph.D. in clinical psychology from the University of Florida.

Robert E. Floden is a professor of teacher education, measurement and quantitative methods, and educational policy and is the director of the Institute for Research on Teaching and Learning at Michigan State University. He has written on a range of topics in philosophy, statistics, psychology, program evaluation, research on teaching, and research on teacher education. His current research examines the preparation of mathematics teachers and the development of leaders in mathematics and science education. He has a Ph.D. from Stanford University.

Ernest M. Henley is a professor emeritus of physics at the University of Washington. He has served as the dean of the College of Arts and Sciences

at the University of Washington and as director and associate director of its Institute for Nuclear Theory. The focus of his work has been with space-time symmetries, the connection of quark-gluons to nucleons-mesons, and the changes that occur to hadrons when placed in a nuclear medium; at present he is working in the area of cosmology. He was elected to membership in the National Academy of Sciences in 1979 and served as chair of its Physics Section from 1998-2001. He is a Fellow of the American Academy of Arts and Sciences, and served as president of the American Physical Society and as a member of the U.S Liaison Committee for the International Union of Pure and Applied Physics. He has a Ph.D. in physics from the University of California, Berkeley.

Margaret Hilton (*Senior Program Officer*) has contributed to consensus reports at the National Academies on monitoring compliance with international labor standards and on the national supply of Information Technology workers. Prior to joining the National Academies in 1999, Hilton was employed by the National Skill Standards Board. Earlier, she was a project director at the Congressional Office of Technology Assessment. She has a B.A. in geography, with high honors, from the University of Michigan (1975), and a master of regional planning degree from the University of North Carolina at Chapel Hill (1980).

Vinetta C. Jones is an educational psychologist and the dean of the School of Education at Howard University. During a 30-year career in public education, she has maintained a singular focus: developing and supporting professionals and creating institutional environments that develop the potential of all students to achieve high levels of academic excellence, especially those who have been traditionally underserved by the public education system. She has written and lectured widely on issues related to the education of diverse populations, especially in the areas of academic tracking, the power of teacher expectations, and the role of mathematics as a critical factor in opening pathways to success for minority and poor students. She served for eight years as executive director of EQUITY 2000 at the College Board, where she led one of the largest and most successful education reform programs in the country. She has served on numerous boards and national committees and was inducted into the Education Hall of Fame by the National Alliance of Black School Educators in 2000. She has a B.A. from the University of Michigan and a Ph.D. in educational psychology from the University of California, Berkeley.

Brian W. Junker is professor of statistics, Carnegie Mellon University. His research interests include the statistical foundations of latent variable models for measurement, as well as applications of latent variable modeling in the design and analysis of standardized tests, small-scale experiments in psychology and psychiatry, and large-scale educational surveys such as the NAEP. He is a fellow of the Institute of Mathematical Statistics, a member of the board of trustees and the editorial council of the Psychometric Society, an associate editor and editor-elect of *Psychometrika*. He also served on the NRC's Committee on Embedding Common Test Items in State and District Assessments. He is currently a member of the Design and Analysis Committee for the NAEP. He has a Ph.D. in statistics from the University of Illinois (1988).

David Klahr is a professor and former head of the Department of Psychology at Carnegie Mellon University. His current research focuses on cognitive development, scientific reasoning, and cognitively based instructional interventions in early science education. His earlier work addressed cognitive processes in such diverse areas as voting behavior, college admissions, consumer choice, peer review, and problem solving. He pioneered the application of information-processing analysis to questions of cognitive development and formulated the first computational models to account for children's thinking processes. He was a member of the NRC's Committee on the Foundations of Assessment. He has a Ph.D. in organizations and social behavior from Carnegie Mellon University.

Ellen Condliffe Lagemann is an education historian and dean of the Harvard Graduate School of Education. Dr. Lagemann has been a professor of history and education at New York University, taught for 16 years at Teachers College at Columbia University, and served as the president of the Spencer Foundation and the National Academy of Education. She was a member of the NRC's Committee on Scientific Principles in Education Research. She has an undergraduate degree from Smith College, an M.A. in social studies from Teachers College, and a Ph.D. in history and education from Columbia University.

Barbara Schneider is a professor of sociology at the University of Chicago. She is a codirector of the Alfred P. Sloan Center on Parents, Children and Work and the director of the Data Research and Development Center, a new $6 million initiative of the Interagency Education Research Initiative.

Her current interests include how social contexts, primarily schools and families, influence individuals' interests and actions. She has a Ph.D. from Northwestern University.

Joseph Tobin is a professor in the College of Education at Arizona State University. Previously he served as a professor in the College of Education at the University of Hawaii. His research interests include educational ethnography, Japanese culture and education, visual anthropology, early childhood education, and children and the media. He was a member of the NRC's Board on International Comparative Studies in Education. He has a Ph.D. in human development from the University of Chicago.

Lisa Towne (*Study Director*) is a senior program officer in the NRC's Center for Education and adjunct instructor of quantitative methods at Georgetown University's Public Policy Institute. She has also worked for the White House Office of Science and Technology Policy and the U.S. Department of Education Planning and Evaluation Service. She has an M.P.P. from Georgetown University.

Tina Winters is a research associate in the NRC's Center for Education. Over the past 10 years, she has worked on a wide variety of education studies at the NRC and has provided assistance for several reports, including *Scientific Research in Education*, *Knowing What Students Know*, and the *National Science Education Standards*.

WORKSHOP SPEAKERS

Robert Boruch is university trustee chair professor in the Graduate School of Education and the Statistics Department (Wharton School) at the University of Pennsylvania. He has received awards for his work on randomized trials and on privacy of individuals and confidentiality in social research from the American Evaluation Association (Myrdal Award), American Educational Research Association (Research Review Award), and the Policy Studies Association (Donald T. Campbell Award). He has a Ph.D. in psychology from Iowa State University.

Wesley Bruce is the assistant superintendent for the Center for Assessment, Research, and Information Technology in the Indiana Department of Education. Previously he served in several administration positions over

the 9 years he was with South Bend Community School Corporation, and also served 11 years in the Kanawha County schools of Charleston, West Virginia. He has a B.A. in psychology from Rice University and a Ph.D. in computer science from the University of Charleston, West Virginia.

Linda Chinnia was appointed to the committee after the workshop was held. Her biographical sketch appears earlier.

Donna Durno is the executive director of the Allegheny Intermediate Unit, a service agency that provides resources, instruction and education services for schools, families, and communities through collaborative partnerships with local school districts, institutions of higher education, government agencies, and foundations. She has over 30 years of educational experience and expertise, culminating in 1987 when she was named commissioner for basic education for the Commonwealth of Pennsylvania. She has a B.S. from Seton Hill College, an M.Ed. in guidance and counseling from Indiana University of Pennsylvania, and a Ph.D. in educational administration from the University of Pittsburgh.

Olatokunbo S. Fashola is a research scientist at the Johns Hopkins University Center for Research on the Education of Students Placed at Risk. Her research interests include reading, after-school programs, language development, emergent literacy, program evaluation, educational policy issues, problem solving, school-wide reform, and bilingual education. She has a Ph.D. from the University of California, Santa Barbara.

Judith M. Gueron is president of the nonprofit, nonpartisan MDRC, where she has directed many large-scale demonstrations and evaluations of social policy innovations and developed methods for rigorously studying real-world programs. The author of *From Welfare to Work* and numerous other publications, she has served on many advisory panels in the areas of employment and training, poverty, and family assistance. She has a Ph.D. in economics from Harvard University.

Vinetta C. Jones was appointed to the committee after the workshop was held. Her biographical sketch appears earlier.

Sheppard G. Kellam is a public health psychiatrist at the American Institutes for Research, where he developed the Center for Integrating Educa-

tion and Prevention Research in Schools. Since 1983, in partnership with the Baltimore City Public School System and Morgan State University, he has led three generations of epidemiologically based randomized field trials. He has an M.D. from Johns Hopkins University.

Anthony (Eamonn) Kelly is professor of instructional technology in the Graduate School of Education at George Mason University. He coedited the *Handbook of Research Methods in Mathematics and Science Education*, and edited the special issue on research methods in education in the *Educational Researcher* in 2003. He has a Ph.D. in psychological studies in education from Stanford University.

Sharon Lewis is the director of research for the Council of the Great City Schools, a research program that articulates the status, needs, attributes, operation, and challenges of urban public schools and the children whom they serve. She has worked for 30 years in the Detroit public schools and served as the assistant superintendent for research and school reform. She has an M.A. in educational research from Wayne State University.

Loretta McClairn is the family, schools, and communities coordinator at Dr. Bernard Harris elementary school (#250) in Baltimore. She is also the program coordinator for the Child First Authority at the school. She has a B.A. from Bowie State University in elementary education and has been teaching for more than 30 years.

David Myers is a vice president and the director of human services research in Mathematica Policy Research's Washington, DC office. He has directed three large random assignment studies in education: the National Evaluation of Upward Bound, the Evaluation of the New York City School Choice Scholarship Program, and an evaluation of remedial reading programs for elementary school students—the Power4Kids Initiative. He has a Ph.D. in sociology from Washington State University.

Richard J. Shavelson is a professor in the School of Education and the Department of Psychology (by courtesy) at Stanford University and past dean of the School of Education. For more than 20 years, he has sought new techniques for measuring performance in ways that contribute to educational and workplace goals, exploring, for example, alternatives to multiple-choice tests in schools, on the job, and in the military. His recent

research has focused on new assessment tools for science and mathematics achievement; measuring individual and group performance in science and mathematics; statistically modeling performance assessment; and addressing policy and practice issues in measurement reform. He has chaired the NRC's Board on Testing and Assessment. He has a Ph.D. in educational psychology from Stanford University (1971).