U.S. Department of Education
April 2014

Recognizing and conducting opportunistic experiments in education: a guide for policymakers and researchers

Alexandra Resch
Jillian Berk
Lauren Akers

Mathematica Policy Research

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S Department of Education

REL 2014–037

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased, large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

April 2014

This report was prepared for the Institute of Education Sciences (IES) by Decision Information Resources, Inc. under Contract ED-IES-12-C-0057, Analytic Technical Assistance and Development. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Resch, A., Berk, J., & Akers, L. (2014). *Recognizing and conducting opportunistic experiments in education: A guide for policymakers and researchers* (REL 2014–037). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development. Retrieved from http://ies.ed.gov/ncee/edlabs.

Contents

Part I. Making the case for opportunistic experiments 2

What are opportunistic experiments? 2

Opportunities for experimentation in education 3

Part II. Recognizing potential opportunistic experiments 5

Step 1. Look for potential random assignment opportunities 6

Step 2. Consider whether the research opportunity aligns with the research agenda 6

Step 3. Assess the feasibility of the research project 6

Part III. Conducting opportunistic experiments 13

Step 1. Identify sites 13

Step 2. Recruit participants 14

Step 3. Conduct random assignment and monitor compliance 14

Step 4. Collect data 16

Step 5. Analyze data and report findings 16

Implications: the low cost of doing opportunistic experiments—and the potentially high cost of not doing them 17

Appendix. Key questions to consider when recognizing and conducting opportunistic experiments 18

References 21

Figure

Figure 1. Theory of change for the Free Application for Federal Student Aid Completion Project 8

Boxes

Box 1. Close partnerships between researchers and district administrators 3

Box 2. Case study of a literacy intervention pilot in New York City 4

Box 3. Case study of a randomized rollout of the Free Application for Federal Student Aid Completion Project 4

Box 4. Case study of a lottery-based evaluation of Knowledge is Power Program charter schools 5

Box 5. Outcomes from state longitudinal data systems 10

Box 6. Exploring long-term outcomes 10

Box 7. Estimating the size of a research sample 12

Box 8. Small sample sizes can still provide useful information 12
Box 9. Encouraging participants to participate in research using random assignment 14
Box 10. Accommodating exclusions from random assignment 15
Box 11. Monitoring randomization in a lottery study 16

Randomized controlled trials (RCTs) are the gold standard for evaluating an intervention's effectiveness (U.S. Department of Education, 2005). While the number of RCTs conducted in education has increased substantially in recent years (Puma, Olsen, Bell, & Price, 2009), education policymakers and administrators often lack the rigorous evidence of efficacy that they need to make evidence-based decisions for their states and districts. But with the help of a research team, state and district administrators can begin to fill this knowledge gap by taking advantage of opportunities to introduce an experiment into the normal course of action without the high cost or disruption sometimes associated with large-scale RCTs. This cost-effective, minimally disruptive approach—known as "opportunistic experiments"—can generate strong evidence for informing education decisions.

An opportunistic experiment is a type of RCT that studies the effects of a *planned* intervention or policy action; by contrast, other types of RCTs examine an intervention or policy action that is *implemented for the research study*. Opportunistic experiments also typically make use of existing administrative data and willing study participants, whereas other RCTs may collect new data and recruit participants. By embedding research in planned changes and using willing participants and administrative data, opportunistic experiments can reduce the direct financial costs sometimes associated with large-scale RCTs and reduce the disruption experienced by study participants. We define disruption as any activity or effort that uses staff or classroom time to facilitate a study. Other RCTs may include surveys or additional testing that require staff time, disrupt class time, and, importantly, include the implementation of an intervention at the request of the researchers. An opportunistic experiment often uses existing data and always evaluates policies or interventions already planned by the district or state; thus, the policy or intervention is not a disruption created as part of the research but something initiated by entity being studied.

The federal government has recently increased its emphasis on the use of opportunistic experiments. Specifically, in July 2013 the Office of Management and Budget released guidance for 2013 agency budget submissions that encouraged agencies to propose "high-quality, low-cost evaluations" that "should help agencies improve the quality and timeliness of evaluations—for example, by building evaluation into ongoing program changes and by reducing costs by measuring key outcomes in existing administrative datasets" (Burwell, Muñoz, Holdren, & Krueger, 2013, p. 3).

Opportunistic experiments are common in a variety of settings. For example, fundraisers often test messages in their communications with potential donors. A recent presidential campaign embedded experiments in its emails to potential donors and supporters by randomly assigning different email messages to different groups on its mailing list.[1] It then tracked how much money each group donated, determining which messages were most effective and using them for future appeals. Web businesses test website designs to learn which ones work best—to increase business and improve user experience.[2] Government agencies conduct opportunistic experiments to learn how best to use scarce resources to serve citizens. Virginia tested several outreach strategies to learn which was most effective in increasing participation in the Earned Income Tax Credit program (Beecroft, 2012). As a result of the study, Virginia learned that a strategy that combines phone calls and mailings to eligible households increases participation and is a cost-effective use of state resources.

[1] For more details, see Green (2012) and Issenberg (2012).

[2] For more information, see Linden (2006) and Mayer, Hom, and Wiley (2009).

Opportunistic experiments can be just as useful in education. Schools seeking to increase parent enthusiasm for events like parent–teacher association meetings could use a strategy similar to the presidential campaign approach described above. The school district could send parents a letter encouraging their participation by both email and postal mail and see which is more effective. Another opportunity for experimentation in education is when districts pilot a curriculum at a few schools before rolling it out across the district. If the district selects the pilot schools randomly, an opportunistic experiment could be conducted with minimal additional disruption and cost. The results could provide information on whether the new curriculum works better than the old one in the local setting before the district rolls out the curriculum to all schools.

This guide is designed to help researchers recognize and conduct opportunistic experiments. Part I defines opportunistic experiments and provides examples. Part II discusses key steps and issues to consider when identifying potential opportunistic experiments. Part III outlines the critical steps to complete opportunistic experiments. This guide concludes with a discussion of the potentially low cost of conducting opportunistic experiments and the potentially high cost of not conducting them. An appendix consolidates the concepts in this guide into key questions to consider when conducting opportunistic experiments.

Part I. Making the case for opportunistic experiments

What are opportunistic experiments?

Opportunistic experiments are RCTs that take advantage of a planned intervention or policy action—that is, an opportunity—to generate rigorous evidence. These experiments are opportunistic because they typically have one or more of the following characteristics, which may reduce costs or disruption compared to traditional, large-scale evaluations:

- *An opportunity to randomly assign participants to an intervention.* In opportunistic experiments, participants are randomly assigned to an intervention in a way that minimizes the disruption associated with random assignment for the entity being studied (such as a district) and to study participants (such as students in a district). When random assignment is seen as a fair way to allocate limited resources and is planned in a way that does not cause major disruption, it is easier to randomly assign schools, classrooms, or students to treatment and control conditions.
- *Experimentation on a program or policy change initiated by the entity being studied.* Opportunistic experiments typically take advantage of an intervention or policy change that is initiated by the entity being studied. Districts and states often implement new curricula or policies with the belief that these changes will improve student (or other) outcomes but without rigorous causal evidence to support this belief. When administrators and researchers build research into a planned program or policy change, they introduce little disruption for participants and can evaluate whether the change does improve student outcomes. Because the program or policy will be implemented with or without an embedded evaluation, researchers working with state and district staff may have an easier time recruiting sites. For example, if a district is planning to pilot a new program in a subset of schools, researchers can work with district staff to encourage schools to participate in the study and undergo random assignment.

- *Strong partnerships between the research team and state and district administrators.* When research plans align with a state or district's programmatic plans, everyone can benefit. The state or district can help focus the research questions toward actionable results, and because they are invested in the findings, close participation of state and district administrators may also facilitate easier access to administrative data and enhanced cooperation of schools and teachers. If the research is conducted by an internal party, the evaluation might require fewer clearances than RCTs with research implemented by an external party, speeding up the timeline and reducing the disruption introduced by research. Likewise, if external researchers conduct the study in close partnership with district administrators, a district research application might not be necessary. Box 1 discusses the importance of close partnerships between researchers and district administrators in identifying opportunities for research and conducting studies with minimal disruption or added cost.
- *Fairly easy access to data on key outcomes.* The district may already collect data on relevant outcomes (such as through administrative records). Such readily available data make a study less disruptive for participating educators and students than a study for which researchers must collect additional data to measure key outcomes.

Many past studies could be considered opportunistic experiments if they evaluate a program or policy change; an opportunistic experiment need not have all of the preceding characteristics. For example, an opportunistic experiment might exploit an opportunity to randomly assign participants (such as a charter school lottery) but require additional data beyond those collected conventionally (such as data on college attendance of charter-school students). Although this type of experiment may be opportunistic, it may also require additional effort or cost to fully answer questions of interest.

Box 1. Close partnerships between researchers and district administrators

Ongoing, mutually beneficial partnerships can be critical to conducting rigorous opportunistic experiments with little added cost or disruption. Even prior to the initiation of any study, establishing an ongoing "thought-partner" relationship can facilitate the early identification of research opportunities. These thought-partner relationships could be:

- Established through monthly lunch meetings or forums.
- Held with staff from several offices within the district, rather than just one (given that formal interoffice contact may not be frequent enough to identify research opportunities sufficiently early).
- Focused on sharing regular status updates, in which district administrators discuss planned initiatives and seek feedback from a research perspective and in which researchers share results from their latest work and help identify potential opportunities to generate rigorous evidence on the effectiveness of a district's initiative.

Once an opportunity for experimentation has been identified, continuing communication between researchers and district administrators can help ensure that the research opportunity aligns with the district's research agenda and is feasible to conduct. From there, close relationships between researchers and district administrators can help ensure that the experiment minimizes disruption for schools, teachers, and staff by facilitating recruitment, random assignment, and data collection, analysis, and reporting.

Opportunities for experimentation in education

In education, opportunistic experiments can take many forms. Three common situations can facilitate random assignment with minimal disruption to participants—randomized pilot, randomized rollout, and oversubscription lottery. One common concern about random assignment studies is that potential

participants are excluded from the program if they are assigned to the control group. In each of the cases illustrated below, random assignment is a fair way to allocate limited slots. In this section, we briefly describe each of these situations and provide a case study from recent or ongoing research.

Randomized pilot. To test a new program or intervention at a few sites before deciding whether to implement more broadly, a district can randomly assign sites to the pilot (treatment) group or to the business-as-usual (control) group. The case study in box 2 is an example.

Box 2. Case study of a literacy intervention pilot in New York City

In 2011, New York City announced a Middle School Quality Initiative to focus on improving literacy instruction in middle schools (New York City Department of Education, 2013; Anand, 2013). As part of an expansion of the initiative to 40 additional middle schools in 2013/14, New York City is piloting a new literacy intervention designed by The After-School Corporation and Harvard University's Education Innovation Laboratory (EdLabs). The intervention uses an expanded school day, adding 12 school hours per week, to provide students in pilot schools with intensive reading tutoring and other learning activities. This program is modeled on math tutoring programs that EdLabs has found to have positive and substantial impacts on math achievement in Denver and Houston. Because the literacy intervention is untested, New York City seeks to learn about the effectiveness of the program before implementing it more widely.

New York City is taking advantage of the opportunity to embed rigorous research in this pilot program. The 40 volunteering schools were randomly assigned to receive the literacy pilot or to serve as a control. If the new initiative improves student reading achievement, the district will have rigorous evidence to support the expansion of the program to other schools. If it does not, the district can call off the expansion.

This study has several characteristics of an opportunistic experiment. The study is embedded in a planned literacy pilot program initiated by the entity being studied: the New York City Department of Education. Because the key outcome is student achievement for students in grades 6–8, the relevant data are already collected and will be readily available to the district and to the researchers who partner with the district.

Randomized staggered rollout. When a district lacks the resources to rollout an intervention for all sites simultaneously, a randomized staggered rollout design may be used to assign sites to either immediate or delayed receipt of the intervention. For example, the treatment group receives the treatment in year 1 of the experiment, and the control group receives the treatment in year 2. The latter group thus acts as a control to the former group during year 1 (but not after). The case study in box 3 provides an example.

Box 3. Case study of a randomized rollout of the Free Application for Federal Student Aid Completion Project

Research has shown that insufficient knowledge of college costs and financial aid availability may prevent some students from applying to and attending selective colleges (Hoxby & Turner, 2013; U.S. Department of Education, 2011a). In response to concerns that the Free Application for Student Aid (FAFSA) application process is complicated and acts as a barrier to college access for disadvantaged students (Bettinger, Long, Oreopoulos, & Sanbonmatsu, 2009), the U.S. Department of Education simplified the process and sought other ways of improving access to higher education for disadvantaged students. Starting in 2012/13 the department rolled out a program to provide access to lists of specific students who have applied for the FAFSA, allowing schools to focus their outreach and counseling on other grade 12 students.

The department's Institute of Education Sciences used this gradual rollout to generate rigorous evidence on the program's impact. Schools within participating districts were randomly assigned to receive access to information on which of their students had completed the FAFSA. Treatment schools first received this information in 2012/13, while control schools received it after a one-year delay. If the new program is effective in improving FAFSA completion rates, the department will

have rigorous evidence to support the expansion of the program to other schools. If the program is not effective, the department can adjust or modify the program before further expansion.

This study has several characteristics an opportunistic experiment. School districts were eager to receive the intervention and volunteered to participate, so recruitment of districts and schools did not require additional resources. The intervention is simple to implement because it consists of providing treatment schools with access to data on which grade 12 students completed the FAFSA forms. School staff currently provide financial aid guidance to their students; the intervention provides the staff with student-level information for targeting certain students for additional guidance. The intervention does not require any particular action on the part of the school—schools can make their own determination of how to use the data. The relevant outcome data—that is, whether students apply for financial aid, receive Pell Grants, or enroll in college—are already collected by the department's Office of Federal Student Aid and the study does not collect implementation data, so there is no additional cost associated with data collection and minimal intrusion by researchers. (Note, however, that the lack of implementation data precludes the study from exploring how school staff actually use the data to assist students.)

Oversubscription lottery. In cases where interest in an education intervention exceeds capacity (oversubscription), districts may use a lottery to randomly assign interested parties to either an intervention group (offered a spot in the program) or a control group (not offered a spot). Box 4 provides an example.

Box 4. Case study of a lottery-based evaluation of Knowledge is Power Program charter schools

Charter schools—which receive public funding but are not subject to many of the regulations governing traditional public schools—are a big part of current efforts to reform the U.S. public education (Clark, Gleason, Tuttle, & Silverberg, 2011). In 2012/13 more than 2 million students were enrolled in more than 6,200 charter schools across the United States, with nearly 10 percent more new charter schools opening than in the year before (Center for Education Reform, 2013). Most states require oversubscribed charter schools to hold lotteries to determine admissions (McEwan & Olsen, 2010). Despite their increasingly widespread use, many charter school models lack rigorous evidence of their effectiveness on a broad scale (Clark et al., 2011).

One widespread charter school model—the Knowledge is Power Program (KIPP)—typically has oversubscribed schools and so must hold random lotteries to allocate seats to students. KIPP decided to take advantage of this preexisting random assignment mechanism to rigorously test the impact of KIPP middle schools on students. Researchers followed students who were offered seats at KIPP middle schools and those who were not, measuring the impact of being offered the opportunity to attend a KIPP school. The study of KIPP middle school entry lotteries in 11 districts across the country can be thought of as a series of opportunistic experiments analyzed together (Tuttle et al., 2013).

This study has several characteristics of an opportunistic experiment. The students sought to attend KIPP schools, so the "intervention" was occurring with or without the research, and random assignment of individual students to different conditions was occurring with or without the presence of an evaluation. A key outcome of the study—student achievement for students in grades 5–8—is tested in all states and readily available through administrative data. Unlike many opportunistic experiments, however, the study required additional data collection and analysis. In addition to collecting state test scores, the study team also administered its own standardized achievement test as well as student and parent surveys and conducted several years of follow-up data collection to assess longer-term outcomes.

Part II. Recognizing potential opportunistic experiments

This section presents three key steps and other issues to consider when identifying potential opportunistic experiments. The first step is to look for potential random assignment opportunities, such as those in the three case studies described above. Once an opportunity for random assignment is identified, the next step is to consider whether the potential evaluation aligns with the research needs of the state or district. The

final step is to assess the feasibility of the evaluation. Integral to each step is a close relationship between researchers and district administrators (discussed in box 1 above).

Step 1. Look for potential random assignment opportunities

Certain situations provide opportunities to allocate a new program using random assignment through a randomized pilot, staggered rollout, or oversubscription lottery. The following situations for states and districts are common:

- *Pilot test.* A district may be considering a new program or curriculum but want to test the innovation before adopting it more broadly. Often, districts will ask schools to volunteer to become pilot study sites, but a randomized pilot would provide the opportunity to rigorously test the innovation. Districts could randomly assign the pilot and control schools from the population of schools that volunteer.
- *Limited resources.* A district may be planning to introduce a new program or infrastructure enhancement to all schools but not have the resources (financial or professional) to implement the change in all schools at once. This situation lends itself to a randomized rollout. Randomly assigning participants to different phases of the rollout may not be overly disruptive, and it will allow the district to evaluate the change.
- *Excess demand.* If a program or school is overenrolled or is expected to have excess demand, an oversubscription lottery can be used to assign slots. Charter school and voucher evaluations have used this lottery mechanism for rigorous evaluation of overenrolled schools or oversubscribed voucher opportunities. After-school programs with excess demand could likewise be studied in this way.
- *Centralized communication.* A district may be trying to influence student or parent behavior through the centralized distribution of information—for example, by sending families information on college preparation resources. If the district wants to test the effectiveness of different strategies, an opportunistic experiment may be easier to implement if the communication effort is centralized. This can be viewed as a randomized pilot.

Step 2. Consider whether the research opportunity aligns with the research agenda

Potential opportunities for random assignment abound, but before conducting an opportunistic experiment, it is important to consider whether it can answer an important research question. When making this decision, consider whether the research question relates to an established agenda or supports an underlying priority, such as increasing student achievement. While addressing a previously identified priority may not be a prerequisite for the experiment, it is nevertheless important in considering the costs and potential disruption incurred because of the experiment. Even relatively low-cost opportunistic experiments entail costs for districts and their research partners, and not all opportunities to use random assignment for an experiment may be an efficient use of limited resources.

Step 3. Assess the feasibility of the research project

As in all research, the next step is to consider whether the research study is feasible. This section looks at six questions for determining the feasibility of a study:

1. What is the theory of change, and what comparison will be evaluated?
2. What are the specifics of random assignment?
3. What are the critical outcomes and data sources?
4. Is adequate statistical power possible?
5. Is research clearance needed?
6. What is the study timeline?

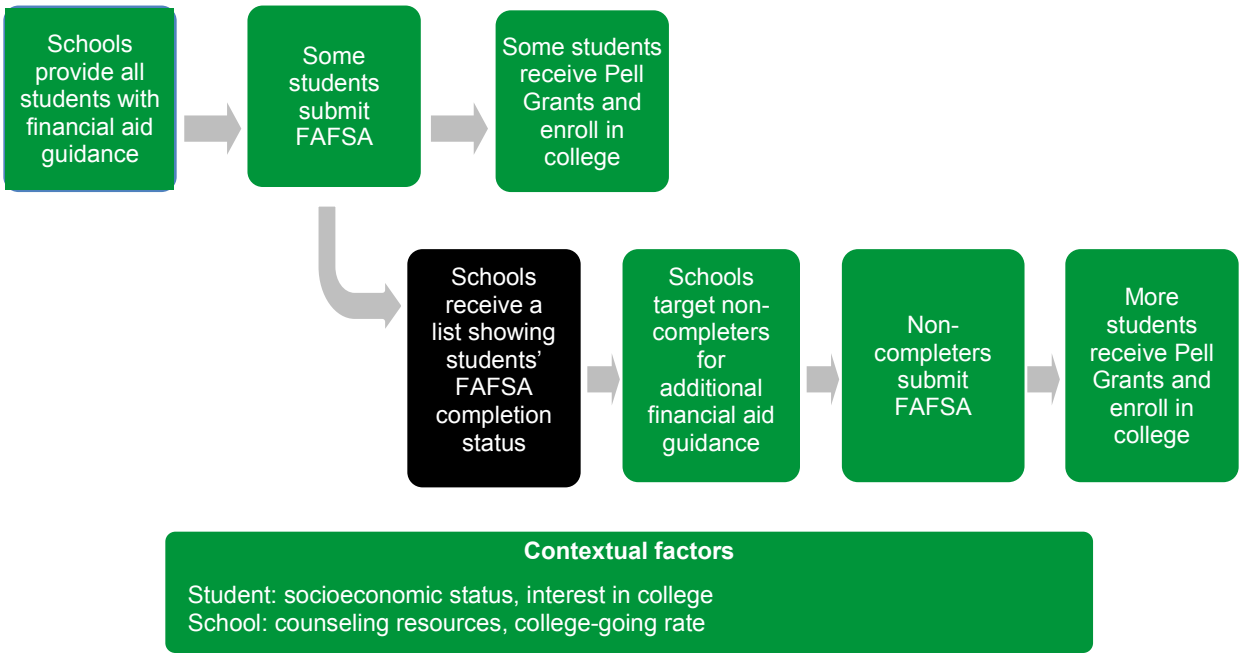
The remainder of this section describes each question in turn.

1. What is the theory of change, and what comparison will be evaluated? Before designing a research study, we need to understand how an intervention works, consider who it will affect, and identify relevant outcomes. In other words, all evaluations of education interventions should begin with a theory of change. A theory of change describes the causal chain of events that leads from the implementation of the intervention to the desired outcome. In some cases the theory of change is simple—targeted literacy tutoring is expected to directly influence literacy outcomes—but in more complex cases it might help to create a diagram that captures the theory of change. This diagram includes the intervention, intermediate outcomes, and final outcomes, and it illustrates contextual factors that can affect this causal chain. Having a diagram provides structure for thinking about *how* the intervention affects final outcomes—or what actions or changes in behavior are expected to occur as a result of the intervention. If a theory of change does not exist, it is important to develop one at this step, although the theory need not be a diagram at this stage. Researchers may need to lead development of the theory of change and then solicit feedback from the district.

Figure 1 presents a simple theory of change for the Free Application for Federal Student Aid (FAFSA) Completion Project (see box 2). The top row of green boxes shows the path of college counseling and student behavior in high schools in the absence of the intervention. Schools provide college counseling to their students, possibly including financial aid guidance. Following the arrows, a subset of grade 12 students submits the FAFSA, attends college, and receives financial aid. The black box in the second row represents the FAFSA Completion Project intervention, which entails providing a set of schools with access to detailed information on which grade 12 students have completed the FAFSA. The remaining boxes in this row show the anticipated results of this intervention: schools are able to target their subsequent counseling to students who have not completed the FAFSA, thereby increasing the chances that those students will complete the FAFSA and enroll in college.

A theory of change should explicitly differentiate between the proposed intervention (to which the treatment group is assigned) and current practice (to which the control group is assigned). In a research study comparing a new program with current practice, the study team refers to current practice as the "counterfactual condition"—that is, what would have happened to members of the treatment group in the absence of the intervention. In figure 1, the top row represents the counterfactual condition. The second row shows the anticipated causal path when the intervention is implemented. A comparison of the two rows illustrates the key research question. In this case, the study examines the impact of the intervention (shown in the black box) on the outcomes in the boxes on the right of each row.

Figure 1. Theory of change for the Free Application for Federal Student Aid Completion Project



FAFSA is Free Application for Federal Student Aid.

Source: Authors' compilation.

2. What are the specifics of random assignment? The second feasibility consideration is to determine the specifics of random assignment. Random assignment can happen in different ways, so researchers should consider the following key questions:

- *What is randomly assigned?* The theory of change will help determine whether the appropriate level for random assignment is schools, teachers, or students. If the intervention is a schoolwide behavior program, random assignment should occur at the school level because it would not be possible to find a control group of students within the school that is unaffected by the intervention. If the intervention is supplemental literacy tutoring, random assignment could occur at the student, classroom, or school level.[3]

  Consider how the level of random assignment is related to "spillover" (where control group members may be affected by the intervention). Spillover can be a particular concern when rolling out certain instructional strategies due to the increasing prevalence of professional learning communities and school-based collaborative planning teams. For example, if teachers develop lesson plans collaboratively, teachers in the control group have access to ideas from teachers in the treatment group. Also consider how the level of random assignment is related to "contamination" (where control group members actually receive the treatment). For example, in a charter school lottery the research team should closely monitor compliance with random assignment to ensure that students assigned to the control group do not enroll in the charter school.

[3] The level of random assignment also has important implications for the data analysis. For details, see Murnane and Willett (2011).

- *Who is eligible to be randomly assigned? Are they representative?* In opportunistic experiments, schools or individuals may volunteer to be randomly assigned to treatment or control groups. They may be students trying to enroll in a charter school or summer program, teachers interested in a new training program, or schools willing to pilot a new curriculum. While randomly assigning volunteers does not affect the internal validity of a study, it can affect the generalizability of study findings. In some studies, such as a study of charter school enrollment, the impact on volunteers may be the key research question. In other words, policymakers want to understand the impact of the charter schools on students who want to enroll. In other studies, such as when a district is piloting a new reading curriculum, policymakers may be more interested in the curriculum's impact in the average school, not its impact in the schools that are willing to participate in the pilot. The researcher must be aware of the implications of who is randomly assigned and interpret and present the results of the study accordingly. Likewise, it may help for researchers to provide district administrators with examples of what a proposed study's results would—and would not—enable them to say about a program's effectiveness.
- *When and where does random assignment occur?* If random assignment occurs at one point in time and in one location, such as by a central administrator after all applications to a program are submitted, it can be fairly straightforward to implement. If the intake for an oversubscribed program occurs on a rolling basis in multiple locations, such as at each charter school during its lottery, random assignment may be more difficult to implement and monitor because more staff will need to be trained on the procedures.
- *How big will the treatment and control groups be?* Random assignment studies do not require that 50 percent be assigned to the treatment group and 50 percent to the control group. In lottery studies the share of students assigned to the treatment group can vary widely on the basis of the amount of oversubscription. While the treatment–control ratio will have implications for statistical power (see the statistical power subsection on page 10), other factors should also be considered. The number of treatment schools or students may depend on budget or capacity. If a district has limited funds to implement the intervention, it may choose to randomly assign only 30 percent of schools to the treatment group, for example.
- *Will random assignment be implemented with integrity?* To ensure the integrity of experimental designs, compliance with random assignments must be monitored (and verified). In some cases the researcher has complete control over compliance with random assignment, as when treatment schools receive tailored information that would not be useful to control schools, thus eliminating the possibility of spillover. In other cases the researcher must monitor compliance with random assignment, as when monitoring a charter school lottery to ensure that lottery results were recorded accurately and that enrollment offers were based on lottery results.

3. What are the critical outcomes and data sources? The third way to assess the feasibility of a research project is to consider ways to reliably measure the key outcomes of interest and assess the difficulty of obtaining the required data. Using outcome measures from administrative records data may be the most cost-effective way to collect outcome measures for opportunistic experiments. Box 5 describes data typically available through one type of administrative record—state longitudinal data systems. Although opportunistic experiments are often thought of as cheap and quick and thus focused on short-term outcomes, the need

for long-term outcomes should not preclude a researcher from starting an opportunistic study. Box 6 further explores long-term outcomes.

Box 5. Outcomes from state longitudinal data systems

Administrative record data are a possible source for outcome data. State longitudinal data systems contain statewide student data that can be linked over time and, potentially, to additional data sources. Since 2006, 47 states have received at least one grant from the Institute of Education Sciences' Statewide Longitudinal Systems Grant Program to support the design, development, implementation, and expansion of the data systems. The key advantage of working with state longitudinal data systems is that they typically include data from all public schools in the state. While the specific data elements vary by state, the data systems can include

- Student, school, and district identifiers.
- Student demographics, accountability subgroups, and participation in federal programs.
- Enrollment, attendance, and completion.
- State assessment data.
- Courses taken, grades, and credits earned.
- College Board and ACT test scores.
- Staff data, including teacher data.

Source: Levesque & Fitzgerald, in press.

When assessing the potential outcomes and data sources, consider the following questions:

- *Are the outcomes cost-effective to collect?* It is important to be realistic about the resources necessary to evaluate impacts on outcomes of interest. With limited resources, evaluations may focus on a specific subset of outcomes that can be collected most easily.
- *Are missing data common? What are the patterns of missing data; that is, are certain individuals more or less likely to have missing data?* If outcome data are frequently missing or missing disproportionately for one group, the outcome may not be appropriate for evaluation.
- *Are the outcomes measured similarly for the treatment and control groups?* If not, the evaluation will not be able to determine whether differences in outcomes are due to differences in effectiveness or differences in how outcomes are measured.

Box 6. Exploring long-term outcomes

Districts may be less eager to conduct an opportunistic experiment if the key outcome of interest is not available for several years. However, the need for long-term follow-up should not necessarily preclude districts and their research partners from starting a study. Intermediate outcomes may provide early evidence of an intervention's success, and implementing random assignment preserves the validity of impacts on intermediate outcomes and options for future follow-up.

In the Free Application for Federal Student Aid (FAFSA) case study (see box 2), for example, the ultimate goal of the intervention is a long-term outcome: increased college attendance. The current study does not include this long-term outcome and instead focuses on the related short-term outcome of submitting a FAFSA. The intervention's theory of change is that some students are not enrolling in college because of financial constraints and that these students are eligible for

financial aid but are not filing FAFSA applications because of lack of information. Documenting the short-term impact of the intervention on FAFSA applications is the first indication that the intervention may be having an impact on the ultimate long-term outcome of college attendance. As long as the district maintains records on the FAFSA experiment sample, it will later be able to obtain records from the National Student Clearinghouse to measure college attendance, the long-term outcome of interest. Researchers could plan for a longitudinal study and craft a data-use agreement that allows for long-term follow-up.

4. Is adequate statistical power possible? The fourth feasibility consideration is whether the experiment will be able to detect an effect of a given size. Conducting a power analysis can help the research team determine how large a sample the opportunistic experiment will need to detect a policy-relevant effect with reasonable precision. Although researchers may not be able to control the sample size in an opportunistic experiment, they should consider statistical power analysis[4] as a tool for determining whether the potential experiment can generate findings with adequate precision. In some circumstances, such as in a school district with a small number of schools, it may not be feasible to evaluate a program implemented at the school level.

Several factors can affect an experiment's statistical power:[5]

- *Size of the expected impact.* If an intervention is expected to have a large impact on outcomes, the impact will be easier to detect, which means that a smaller sample is required to detect that effect. Researchers should assess the size of the expected impact by considering the theory of change and examining prior research. Generally, an intensive intervention is more likely to have a large impact than a "light touch" intervention.
- *Sample size.* Statistical power generally improves with a larger sample size. Researchers should thus ensure that the estimated size of the research sample is sufficient to detect the expected impact. Box 7 discusses how to proceed in cases when the sample size may be difficult to estimate; box 8 discusses how an opportunistic experiment with a small sample size may still provide useful evidence.
- *Level of random assignment.* As mentioned above, statistical power improves with a larger sample size. However, if random assignment is conducted at the school level, the primary determinant of statistical power is the number of schools, not the number of students within the schools. In this case researchers should pay special attention to the number of schools that can be included in the study, not the number of students in each school.
- *Availability of baseline data.* Baseline data on characteristics related to the outcome can improve statistical power by accounting for variation in outcomes attributable to differences in these pre–random assignment characteristics. Prior measures of the outcome—for example, test scores from the previous year—are especially valuable because they tend to explain a large share of the variation in the outcome measure (Bloom, Richburg-Hayes, & Black, 2007; Schochet, 2008).

[4] For more details on statistical power analysis, see Murnane and Willett (2011, chapter 6).

[5] For more details on calculating statistical power for random assignment studies in education, see Schochet (2008).

Box 7. Estimating the size of a research sample

If the random assignment mechanism is a lottery, it may be difficult to estimate the size of the research sample. The primary factor in determining sample size is the number of applicants in the lottery. Examining historical applicant trends may provide some guidance, but researchers should also consider factors that may affect demand for the school, including the school's academic performance in the previous year, the academic performance of neighborhood schools, and changes in supply of slots at other schools. If a researcher learns that a school had to turn students away in the previous year and, as a result, plans to reduce its outreach or recruitment efforts, the school may be a less desirable candidate for a study. Schools sometimes use other modifications to the lottery, including exemptions for certain groups (such as siblings of existing students), that can impact the size of the research group (Tuttle, Gleason, & Clark, 2012). For this reason, schools with many exempted groups are less likely to be good candidates for a study.

Box 8. Small sample sizes can still provide useful information

Although studies with larger samples are better able to detect the impacts of interventions, results from an inexpensive but small study can still be useful to stakeholders. Even if the study does not find statistically significant impacts, the results could offer some indication about the size and direction (positive or negative) of the intervention's impacts and provide important data to guide future research.

5. Is research clearance needed? The fifth feasibility consideration is whether any research clearance is required. Education research may be subject to multiple layers of clearance or review, depending on the details of the study and the policies of the participating states or districts. Many districts and states require researchers to complete a research application before undertaking any research in their jurisdiction. Districts and states may also require that researchers provide evidence of Institutional Review Board (IRB) approval or that researchers complete a data use agreement. The requirements vary widely across states and districts; close partnership with a district or state will ensure that researchers understand what is required and may lead to expedited review. Situations in which multiple layers of clearance and review are required may not be good candidates for opportunistic RCTs, because these reviews can be costly and time consuming.

Research that involves human subjects may require approval from an IRB. An IRB review typically requires the researchers to present the research questions and plans for protecting the participants. There are local variations in requirements, but in some cases, not all research studies with human subjects are required to go through the full IRB process. For example, the federal regulations on protecting human subjects include exceptions for "research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods," so many of the opportunistic experiments that this guide addresses may be exempt.[6]

A data-use agreement defines the terms under which researchers may access certain data that have been collected by another entity. Data-use agreements are usually required before external researchers can gain access to administrative records data. One of the primary reasons for this is the requirements of the Federal Educational Rights and Privacy Act, which prohibits the disclosure of personally identifiable information

[6] For more information, see Protection of Human Subjects (2006).

from education records without written consent, except under the "studies" and "audit or evaluation" exceptions. How the act's requirements are interpreted varies by state. Some states will provide researchers with confidentiality and data security restrictions, some states will provide only deidentified data, and others require researchers to work onsite.

Data-use agreements could also discuss the dissemination of study results and may specify the types of publications that would result from the study, such as public reports, issue briefs, or articles submitted to a peer-reviewed journal. Data-use agreements often specify that the district's identity will be kept anonymous. Agreements also often grant the district the right to review all publications for compliance with the agreement's terms, such as ensuring participants' confidentiality.

6. What is the study timeline? The final feasibility consideration is a matter of timing. As in many other research endeavors, the feasibility of an opportunistic experiment depends critically on the study timeline and whether researchers can implement random assignment at a time that works with the school's or district's decision-making calendar. For example, charter school lotteries occur at a specific point in the school year. The decision to roll out a new curriculum might have to be finalized by the spring in order to have teachers trained to implement the curriculum in the fall. If researchers miss the chance to implement random assignment, they might have to wait for another year, or the opportunity may be lost entirely.

A second timing consideration has to do with results. The nature of the research questions asked when conducting opportunistic experiments means that practitioners and policymakers want to know the results of the study quickly, so they can make decisions based on the results. Researchers should consider these interests while also being realistic about the study timeline. As in any kind of research in schools, factors that can affect the timeline include the need to secure outside funding, time to recruit study participants, and time to negotiate access to data. When results are needed quickly, the opportunistic experiment may still be a feasible option if researchers focus on only short-term outcomes.

Part III. Conducting opportunistic experiments

This section identifies five steps for conducting an opportunistic experiment and uses case studies to illustrate the steps and decisions. The steps are: identify sites; recruit participants; conduct random assignment and monitor compliance; collect data; and analyze data and report findings. All experiments include these features. This guide focuses on aspects that are different in opportunistic experiments than in other RCTs.

Step 1. Identify sites

In many opportunistic studies, the plans for the intervention determine the set of possible sites. This can be seen as an advantage, because the researcher does not need to identify potential sites and can focus immediately on the relevant sample. In the pilot of the literacy intervention in NYC middle schools described in Box 2, for example, the set of potential sites is middle schools in New York City that are not already implementing the program. An opportunistic experiment with a well defined set of possible sites may be more cost-effective because the researchers will require fewer resources for recruiting.

Step 2. Recruit participants

In some cases, experiments do not require recruitment, as in instances of centralized communication (where participants do not need to agree to participate in the intervention) or oversubscription (where participants have already agreed to participate in the intervention). Where recruitment is necessary opportunistic experiments still offer advantages. First, because sites are already planning to implement the intervention or are eager to implement it, researchers do not have to "sell" the study to sites (explaining how the benefits of the intervention outweigh the disruption that comes with it). Framing the experiment as a staggered rollout rather than as a pilot can facilitate recruitment because in a staggered rollout, the control group still receives the intervention, but in a pilot, it may receive the intervention only if the study results are favorable. Second, if the research relies on administrative data, the experiment does not require data collection that disrupts classrooms, which may make it easier to recruit participants. Where random assignment is added for research purposes, however, the research team will need to work with the state, district, or school that implements the intervention to encourage participants to undergo random assignment (box 9).

Box 9. Encouraging participants to participate in research using random assignment

Educators may hesitate to randomly assign schools or students to treatment conditions for many reasons, including ethical or logistical concerns (Cook, 2001). The following strategies can encourage educators to participate:

- Emphasize that because the district is planning to implement the intervention with or without a study, it is in the district's best interest to know whether the intervention works. An experiment can enable districts to discontinue use of an intervention found to be ineffective or expand use of an intervention found to be effective.
- Name the study's potential benefits, such as enabling districts and schools to better serve students and better allocate limited resources by continuing the use of effective interventions only.
- Discuss the ramifications of making decisions without rigorous evidence.
- Convey that in the case of oversubscription or limited resources (or both), randomization is a fair and transparent way to distribute resources and services.
- Assure participants that the study has been designed to minimize disruption for schools, classrooms, and students.
- When necessary, allow for a limited number of exemptions from random assignment.
- If needed, the district can offer incentives, such as financial compensation, to encourage participation.

Step 3. Conduct random assignment and monitor compliance

Depending on what or whom is being randomly assigned, the research team may work with administrators from the state, district, or school to ensure that assignment to treatment and control groups is truly random. Three key issues to consider when conducting and monitoring compliance with random assignment include:

- *Ensuring the system for assignment is truly random.* Where states, districts, and schools have systems for assigning schools, classrooms, teachers, or students to a program or intervention, the research team should ensure that the systems are truly random. For example, some systems use methods that administrators describe as "effectively random" (such as assigning students to a classroom by the

alphabetical order of their names), which the research community does not typically recognize as random because the methods may result in systematic differences among groups.[7] In such cases the research team should work with state, district, or school officials to agree on a truly random mechanism for assignment.

- *Accommodating school or district needs and preferences for random assignment.* The research team may also need to work with state, district, or school officials to accommodate needs and preferences that influence random assignment. For example, in an opportunistic experiment where students are randomly assigned to classrooms, researchers can work with school staff to accommodate school preferences for classroom composition (such as balanced gender or achievement levels across classrooms).[8] Likewise, if a district needs to roll out a program to only a few schools at a time because of resource constraints, researchers can conduct a randomized staggered rollout. Researchers and staff must identify any exceptions in advance of random assignment. Box 10 discusses how researchers can accommodate exclusions from random assignment.

Box 10. Accommodating exclusions from random assignment

Just as in other randomized controlled trials, researchers may face pressures to include favored sites or individuals in the treatment group of an opportunistic experiment. One strategy to alleviate this pressure is to exclude some sites or individuals from randomization (and the study) but let them participate in the intervention. For example, some charter schools admit siblings of current students before the lottery is conducted, exempting these students from the lottery. Similarly, in some education experiments school staff may strongly prefer that a particular child receive the intervention, so random assignment to the control group would be unacceptable. In such instances that child could receive the treatment but not be included in the study.

To preserve both sample size and power, as well as the ability to generalize results to other settings, researchers should minimize such exclusions. When exclusions are necessary, researchers and district staff should discuss and negotiate eligibility for exclusion from random assignment (and thus from the study) well before random assignment occurs. As in any study, research teams should work with district administrators to keep accurate documentation of any exclusions.

- *Monitoring and confirming compliance with random assignment.* After conducting random assignment researchers should work with staff to monitor and confirm compliance with random assignment so that no changes are made after generating assignments. The monitoring task for opportunistic roll-outs and pilots is similar to that for other RCTs: when relevant administrative records are available, researchers should verify that the actual assignments of individual schools or students to conditions match the original random assignment. Even in an oversubscription lottery for which the district or school is already using random assignment to regulate admission to oversubscribed programs, the researcher must still monitor random assignment and compliance. Box 11 outlines how researchers can ensure validity of randomization generated by lottery.

[7] "Effectively random" methods may not meet What Works Clearinghouse guidelines for random assignment if it is unclear that all units in the study were assigned by chance. See U.S. Department of Education (2011b, p. 12) for more detail.

[8] For example, researchers often block or pairwise match the research sample on relevant student or school characteristics when planning and conducting random assignment to ensure balance on key characteristics. In the FAFSA study researchers conducted pairwise matching of schools within districts to ensure balance on these characteristics between treatment and control samples. Once researchers have constructed the pairwise matches, sites or participants are randomly assigned within pairs (often using a random number generator) to treatment conditions.

Box 11. Monitoring randomization in a lottery study

Lotteries for oversubscribed charter schools are conducted with or without a researcher; attention to recordkeeping and compliance with the randomized list may vary across schools. To ensure validity of randomization generated by lottery, the researcher should ideally:

- Participate in the lottery by observing or even generating the randomized lists.
- Compare the randomized lists to actual admissions offers and enrollment records.

Step 4. Collect data

In some opportunistic experiments all relevant data may be readily available (such as through administrative sources), but others will require additional data collection. To preserve the possibility of longer term follow-up, researchers could collect individual identifiers, such as name, birthdate, address, and Social Security number. With this information, researchers may find sample members in other administrative data (such as the National Student Clearinghouse for postsecondary education outcomes) or contact them for a follow-up survey if funding for such research becomes available. However, these data can be difficult for researchers to collect because of privacy protections.

Step 5. Analyze data and report findings

Analysis and reporting activities are similar across opportunistic experiments and other RCTs. Researchers should conduct the same careful analyses that they would in any study, accounting for such issues as spillover, low treatment take-up, control group reactions caused by anticipation of treatment, a lack of baseline data, and attrition.[9] However, these issues should not stand in the way of conducting an opportunistic experiment. Even knowledge gained from an opportunistic experiment with limitations or flaws may be more favorable than the knowledge gap left by the absence of an experiment.

Disseminating the findings from opportunistic experiments, as well as the context in which the experiment took place, is extremely important. When disseminating findings, researchers should describe what is being compared with the intervention. It is much more informative to report that a pilot curriculum improved test scores relative to the existing curriculum than to just describe the pilot curriculum as "effective." While districts may have little incentive to publish results from an opportunistic study that focuses on local decisions, researchers can help the process along. As mentioned above, data-use agreements can help alleviate district concerns about publication by specifying that the district's identity will remain anonymous (thereby alleviating concerns about negative publicity) and granting districts the right to review all publications for compliance with the agreement's terms, such as participant confidentiality. Researchers can also emphasize that dissemination not only contributes important information to the field but also helps the district and other stakeholders build an evidence base relevant to their context and disseminate the results to staff not involved in the study.

[9] Murnane and Willet (2001) and the What Works Clearinghouse guidelines (U.S. Department of Education, 2011b) provide guidance on how to account for these issues in the analysis.

Implications: the low cost of doing opportunistic experiments—and the potentially high cost of not doing them

This guide has shown that an opportunistic experiment need not be large or expensive. Opportunistic experiments can generate rigorous evidence at low cost, with opportunities to randomly assign participants with minimal disruption to participants; interventions initiated by the entity being studied, which will take place with or without experimentation; strong partnerships between the research team and state and district administrators; and fairly easy access to data on key outcomes.

Despite the relative ease of opportunistic experiments, these opportunities are often missed. In such cases a state or district may continue to use resources—or even increase spending—on an ineffective intervention. For example, policymakers in a North Carolina school district missed the opportunity to rigorously test whether providing students with laptops affects student outcomes.[10] The district rolled out a technology program over four school years. By not randomly assigning the students involved in the staggered rollout, the district missed an opportunity to generate rigorous evidence on the program's effectiveness. The district cites increasing proficiency rates on the state assessment as evidence that the program has been successful, but it is impossible to attribute the improvements to the laptop program without a valid counterfactual condition that represents what would have occurred in the absence of the intervention. Without generating data to guide their decision, the district may continue to buy laptops without knowing whether that money is well spent.

Researchers can play a critical role in helping state and district administrators identify and conduct opportunistic experiments and avoid missed opportunities for experimentation. Through close collaboration, researchers can help state and district administrators make evidence-based decisions on education program and policy changes. Ultimately, these partnerships can help improve education quality and student outcomes.

[10] For more details, see Mooresville Graded School District (n.d.).

Appendix. Key questions to consider when recognizing and conducting opportunistic experiments

This appendix lists key questions that summarize the concepts presented in this report. In the best circumstances opportunistic experiments offer a chance to obtain rigorous evidence on policy changes or decisions with minimal disruption.

Part I. Recognizing the potential for opportunistic experiments

A close, ongoing relationship between researchers and district administrators can lay the foundation for recognizing potential opportunities to generate rigorous evidence on a district initiative.

Step 1: Look for potential random assignment opportunities

Are there situations in which some element of randomization can be introduced with minimal disruption for schools and students, such as opportunities created by:

- Oversubscription?
- Limited resources?
- The need to pilot a program?
- Centralized communication efforts?

Step 2: Consider whether the research opportunity aligns with the research agenda

Does the research opportunity answer a valuable question? For example, does it relate to an established research agenda, such as the Regional Educational Laboratories' list of research priorities? Does it inform policy decisions within the district?

Step 3: Assess the feasibility of the research project

- What is the theory of change?
    - Can it explain the causal chain of events that leads from the implementation of the intervention to the desired outcome?
    - Is it explicit about how the proposed intervention is different from current practice?
- What are the specifics of random assignment?
    - What is randomly assigned?
    - Who is randomly assigned?
    - When and where does random assignment occur?
    - How big will the treatment and control groups be?
    - Will random assignment be implemented with integrity?
- What are the critical outcomes and data sources?

- Are the outcomes cost-effective to collect?
- Are missing data common? What are the patterns of missing data; that is, are certain individuals more or less likely to have missing data?
- Are the outcomes measured similarly for the treatment and control groups?

- Given the expected sample size, does the experiment have sufficient power to answer the questions of interest?
  - What is the size of anticipated impact?
  - What is the sample size?
  - What is the level of random assignment?
  - Are baseline data available?
  - If the study is anticipated to have small sample sizes and low power, could it still be useful to stakeholders?
- Does conducting random assignment or accessing the outcome data require approval from an Institutional Review Board or a data-use agreement?
- What is the study timeline?
  - Is the opportunity for random assignment time-sensitive?
  - Does the timeline depend on securing outside funding, recruiting study participants, or negotiating access to data?
  - Are the results of the study needed for immediate decisionmaking?

Part II. Conducting opportunistic experiments

Maintaining close relationships between researchers and district administrators can help ensure that the experiment minimizes disruption for schools, teachers, and staff by facilitating recruitment, random assignment, data collection, and analysis and reporting.

Step 1: Identify sites

What is the set of possible study sites?

Step 2: Recruit participants

- Are participants already planning to implement the intervention (or eager to implement it)?
- Are there any incentives or benefits to the school that can be used in recruiting?
- If needed, will state-, district-, or school-level staff implementing the intervention be able to work with the research team to encourage participants to undergo random assignment?

Step 3: Conduct and monitor random assignment

- Does the research team need to work with state, district, or school administrators to ensure that the system for assignment is truly random?
- Does the research team need to work with state, district, or school administrators to accommodate non-negotiable needs and preferences in the random assignment process (such as balanced gender or achievement levels across classrooms)?
- Will state-, district-, or school-level staff be able to help researchers accurately document any exclusions?
- Will state-, district-, or school-level administrators be able to help researchers monitor and confirm compliance with random assignment to ensure that no changes are made after generating assignments?

Step 4: Collect data

- Are all relevant data readily available (such as through administrative sources), or is there a need to collect additional primary data?
- When possible, will state-, district-, or school-level staff be able to help researchers obtain baseline and outcome data?
- To preserve the possibility of longer term follow-up, will state-, district-, or school-level staff be able to help researchers collect individual identifiers, such as name, birthdate, address, and Social Security number?

Step 5: Analyze data and report findings

- Will state-, district-, or school-level staff be able to help researchers conduct the same careful analyses that they would in any study—accounting for issues such as spillover, low treatment take-up, control group reactions caused by anticipation of treatment, a lack of baseline data, and attrition?
- To generate knowledge for the education community, will state-, district-, or school-level staff be able to help researchers report and disseminate findings from opportunistic experiments?

References

Anand, A. (2013). *City announces the 20 schools that will have longer school days.* New York: Gotham Schools. Retrieved July 15, 2013, from http://gothamschools.org/2013/06/10/city-announces-the-20-schools-that-will-have-longer-school-days/

Beecroft, E. (2012). *EITC take-up by recipients of public assistance in Virginia and results of a low-cost experiment to increase EITC claims.* Richmond, VA: Virginia Department of Social Services. Unpublished manuscript.

Bettinger, E., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2009). *The role of simplification and information in college decisions: Results from the H&R Block FAFSA experiment* (Working Paper No. 15361). Cambridge, MA: National Bureau of Economic Research. Retrieved July 15, 2013, from http://www.nber.org/papers/w15361. http://eric.ed.gov/?id=ED510037

Bloom, H., Richburg-Hayes, L., & Black, A. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29*(1), 30–59. http://eric.ed.gov/?id=EJ782432

Burwell, S., Muñoz, C., Holdren, J., & Krueger, A. (2013). *Next steps in the evidence and innovation agenda* (Memorandum to the Heads of Departments and Agencies). Washington, DC: Office of Management and Budget, Executive Office of the President. Retrieved July 15, 2013, from http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-17.pdf

Center for Education Reform. (2013). "Charter Schools Primer." Retrieved July 15, 2013, from http://www.edreform.com/2013/07/charter-school-primer/

Clark, M. A., Gleason, P., Tuttle, C. C., & Silverberg, M. K. (2011). *Do charter schools improve student achievement? Evidence from a national randomized study* (Working Paper). Princeton, NJ: Mathematica Policy Research. Retrieved July 15, 2013, from http://www.mathematica-mpr.com/publications/PDFs/education/charterschools_WP.pdf. http://eric.ed.gov/?id=ED528381

Cook, T. D. (2001). Sciencephobia: Why education researchers reject randomized experiments. *Education Next, 1*(3), 62–68. http://eric.ed.gov/?id=EJ637042

Green, J. (2012, November 29). The science behind those Obama campaign e-mails. *Bloomberg Businessweek.* Retrieved July 15, 2013, from http://www.businessweek.com/articles/2012-11-29/the-science-behind-those-obama-campaign-e-mails

Hoxby, C., & Turner, S. (2013). *Expanding college opportunities for high-achieving, low- income students* (Discussion Paper No. 12-014). Stanford, CA: Stanford Institute for Economic Policy Research. Retrieved July 15, 2013, from http://siepr.stanford.edu/publicationsprofile/2555

Issenberg, S. (2012, May 22). The death of the hunch. *Slate.* Retrieved July 15, 2013, from http://www.slate.com/articles/news_and_politics/victory_lab/2012/05/obama_campaign_ads_how_the_analyst_institute_is_helping_him_hone_his_message_.html

Levesque, K., & Fitzgerald, R. (in press). *Using state longitudinal data for applied research.* Draft report submitted to the U.S. Department of Education, Institute of Education Sciences.

Linden, G. (2006). *Early Amazon: Shopping cart recommendations.* Retrieved July 15, 2013, from http://glinden.blogspot.com/2006/04/early-amazon-shopping-cart.html

Mayer, M., Hom, K., & Wiley, J. (2009). *Now you see it, now you don't.* Google Official Blog. Retrieved July 15, 2013, from http://googleblog.blogspot.com/2009/12/now-you-see-it-now-you-dont.html

McEwan, P. J., & Olsen, R. B. (2010). Admissions lotteries in charter schools. In J. R. Betts & P. T. Hill (Eds.), *Taking measure of charter schools: Better assessments, better policymaking, better schools* (pp. 83–112). Lanham, MD: Rowman and Littlefield.

Mooresville Graded School District. (n.d.). *Academic success*. Mooresville, NC: Author. Retrieved July 16, 2013, from http://www5.mgsd.k12.nc.us/staffsites/digitalconversion/Digital_Conversion//Academic_Success.html

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research.* New York: Oxford University Press.

New York City Department of Education. (2013). *Chancellor Walcott & Speaker Quinn announce schools selected for extended learning day pilot & expansion of Middle School Quality Initiative.* New York: Author. Retrieved July 15, 2013, from http://schools.nyc.gov/Offices/mediarelations/NewsandSpeeches/2012-2013/061013_qualityinitiative.htm

Protection of Human Subjects, 34 C.F.R., pt. 97. (2006). Retrieved July 15, 2013, from http://www2.ed.gov/policy/fund/reg/humansub/part97-2.html#97.101

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE No. 2009-0049). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://eric.ed.gov/?id=ED511781

Schochet, P. Z. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics, 33*(1), 62–87. Retrieved July 15, 2013, from http://www.mathematica-mpr.com/publications/redirect_PubsDB.asp?strSite=http://jeb.sagepub.com/cgi/reprint/33/1/62. http://eric.ed.gov/?id=EJ788461

Tuttle, C. C., Gill, B., Gleason, P., Knechtel, V., Nichols-Barrer, I., & Resch, A. (2013). *KIPP middle schools: Impacts on achievement and other outcomes.* Washington, DC: Mathematica Policy Research. Retrieved July 15, 2013, from http://www.mathematica-mpr.com/publications/PDFs/education/KIPP_middle.pdf. http://eric.ed.gov/?id=ED540912

Tuttle, C. C., Gleason, P., & Clark, M. (2012). Using lotteries to evaluate schools of choice: Evidence from a national study of charter schools. *Economics of Education Review, 31*(2), 237–253. Retrieved July 15, 2013, from http://www.sciencedirect.com/science/article/pii/S0272775711001051. http://eric.ed.gov/?id=EJ957650

U.S. Department of Education. (2005). Scientifically based evaluation methods. *Federal Register, 70*(15), 3585–3589. Retrieved July 15, 2013, from http://www2.ed.gov/legislation/FedRegister/finrule/2005-1/012505a.html

U.S. Department of Education. (2011a). *The FAFSA Completion Project: An annotated bibliography*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education. Retrieved July 15, 2013, from http://www.schoolcounselor.org/files/FAFSABibliography.pdf

U.S. Department of Education. (2011b). *What Works Clearinghouse procedures and standards handbook (version 2.1)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved July 15, 2013, from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf