Examining Educational Experiments:

A Field Guide for Conducting Scientifically Based Research

CALIFORNIA POSTSECONDARY EDUCATION COMMISSION

VERSION 1.0
FEBRUARY 2008

Examining Educational Experiments:

A Field Guide for Conducting Scientifically Based Research

CALIFORNIA POSTSECONDARY EDUCATION COMMISSION

VERSION 1.0
FEBRUARY 2008

Table of Contents

Table of [Check]Lists

Acknowledgements

This document is one of the ways in which the California Postsecondary Education Commission's Improving Teacher Quality Program seeks to assist the educators with whom it works in the field. It is primarily intended to support the work of project and research directors who implement teacher professional development grants funded under Title II-A of the federal No Child Left Behind Act of 2001. But we hope it may serve a larger audience in providing practical guidance for educational research, especially as it takes place within the context of professional development initiatives serving California's K-12 teachers.

This work is principally the brainchild and product of Howard Levine, who is Director of the program's Assessment and Dissemination Project. That project continuously assesses what we can learn from the Improving Teacher Quality (ITQ) Program activities in California schools and then shares that learning widely in the ITQ community. The need for the *Field Guide* arises from the program's requirement that grant projects include a scientifically based research component. However, our hope is that its utility is not limited to those projects, but will assist others beyond the program who engage in similar work.

In producing it, Mr. Levine received significant editorial and production assistance from Don Hubbard and P.J. Hallam, both members of the Assessment and Dissemination Team, and from Bill Fitzgerald who is completing an SBR project. ITQ Program Assistant Natalie Sidarous is largely responsible for editing and formatting, and was assisted by Ilona Manzyuk and Kim Crittenden. The support of the Commission and Executive Director Murray Haberman for the program and especially for the research component of current grant projects is deeply appreciated.

As acknowledged in the Introduction, this is a "work in progress" and will surely be improved in future editions by the contributions of those above and others. We hope it will be a very useful tool to support high-quality research on teacher professional development projects and will encourage others to engage in such research to improve the quality of teaching well into California's future.

Karen Humphrey, Administrator
Improving Teacher Quality Program
California Postsecondary Education Commission

Introduction

Why this Field Guide

"The unexamined life is not worth living." If Socrates' famous words are a good guide for personal education, then similar words apply equally well to the entire education enterprise (or any other field of knowledge generation): The unexamined experiment is not worth conducting. Among the reasons for thinking this true is more good advice from Socrates – It is not wise to think that I know what I do not know.

Aphorisms aside, there are many sound, evidence-based reasons for believing that research has an important role to play in the conduct of effective modern education. The stakes for determining and implementing the best possible educational practices have never been higher. The United States spends more than half a trillion dollars annually on K-12 education. Globalization puts a premium on national educational attainment (e.g., better trained citizens, patents, other forms of intellectual property). For the individual, the income gap between those with college degrees and those dropping out of high school is ever widening. With so much on the table and at stake, the United States must apply its great knowledge generation engine to the discipline responsible for fueling that engine – education.

Yet, education is far more than a mere instrument of national competitiveness. With over 55 million students and 4 million K-12 teachers, it is the nation's largest profession; and, as Linda Darling-Hammond has persuasively argued, professions must be built upon a shared and growing knowledge base. To be sure, the professionalization of teaching requires more than the mere existence of such a knowledge base, and that base itself will be comprised of many different types of data, information and knowledge. This *Field Guide* is predicated upon a simple belief – the profession of teaching and the process of education must do a better job of generating knowledge about themselves. Other disciplines have made substantial progress by embracing and applying a wide array of exemplary research strategies and then feeding proven results back into the discipline. Education must do the same. The unexamined experiment is not worth conducting.

What is a Field Guide

Taking the long historical view, what the Federal government now calls scientifically based research, "SBR", has its antecedents in the work of Aristotle, the first writer to systematically study causation ("Wisdom deals with the first causes and principles of things"). More proximately, a word that is a direct steal from Aristotle, SBR derives from Campbell and Stanley's 1963 monograph, *Experimental and quasi-experimental designs for research*. To be sure, this seminal work is based on the work of others (e.g., Hume, Mill, Koch), but that intellectual history is of no concern here. Since 1963 that slim, suggestive monograph has grown and morphed into a 600-plus page "bible" – *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* – that is the compendium of our knowledge regarding causal inference. This *Field Guide* is intended as an adjunct to the theoretical knowledge contained in that volume. The field-

generated materials contained in this *Guide* – checklists, timelines, flowcharts, guidelines and tips based on real world experiences – are designed to allow the researcher to apply the deep theory of causal inference to the hectic and messy world of school-based research. Speaking metaphorically, this *Field Guide* serves as a set of variably scaled maps to the SBR process. But like the novice traveler, the researcher, novice or not, is always warned not to confuse the map with the territory. The world of education-based research is far richer, and messier, than any two-dimensional map.

Who this Guide is For

Since its inception, the Improving Teacher Quality (ITQ) Program, and the Eisenhower Program before it, have required a collaboration between Institutions of Higher Education (IHEs) and local education agencies (LEAs). The No Child Left Behind Act's focus on scientifically based research adds a third party – the professional researcher. In order to be successful, ITQ projects must forge a real working collaboration among all of these partners. At a minimum, this requires an understanding of each partner's roles and responsibilities. This *Field Guide* is designed to help all of the ITQ partners meet this requirement by focusing on the critical junctures where the partners must work collaboratively to conduct scientifically based research on the underlying professional development project. In addition to a succinct overview of the SBR process, this *Guide* contains chapters on topics such as: how the professional development partner, typically the IHE, can find and work with research consultants; how the IHE and researcher can work together to meet the requirements of the Institutional Review Board; how the researcher can work with LEAs to secure the most useful data set; and how all can work together to insure that research results are successfully disseminated. Not everyone will need to read every page of the *Guide*, yet there is something of value for every ITQ participant in its pages.

Feedback

The Guide in your hands is a draft based on our early experiences with scientifically based research within the ITQ program. It is our intention to widely circulate this document during 2008, solicit suggestions for improvement, and publish a more polished and complete version in early 2009.

Because this Guide is solidly based in real-world field experience, your comments for improvement are vital. Please send all suggestions, criticisms, and recommendations to cpecfieldguide@gmail.com. While we award no prizes, all those who submit feedback will be acknowledged in the next version of the Field Guide.

PART I
SBR BASICS

1 Defining Scientifically Based Research

The Advent of Scientifically Based Research

Prior to 2002, the phrase "scientifically based research" had hardly entered the vocabulary. After all, wasn't all research scientifically based? What would unscientific research look like? Such questions might seem absurd in the hard sciences, but in education – an applied social science – the research landscape was substantially different. First, most educational practice, and a fair amount of its experiments, simply went unresearched. Second, much of educational research failed to meet the minimal standards of other social sciences (e.g., quantitative data, isolating causes through experimental designs, randomization or carefully selected comparison groups). All this changed with the passage of the No Child Left Behind Act of 2001 (PL 107-110) which set in motion a considerable, and long overdue, challenge to this long accepted state of affairs. In fact, this *Field Guide* might be considered one of the responses to that challenge.

NCLB mentions scientifically based research more than 100 times and ultimately provides a definition in Title IX - General Provisions, Part A-Definitions, Sec. 9101, Definitions, Number 37, subpart A: "As defined here, the term scientifically based research means research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs." Perhaps more important than this simple definition, NCLB required SBR of its awardees across a broad range of educational programs, including Title I and Title II (the program that supports professional development (Improving Teacher Quality) and the math and science partnerships). Washington has its own spin on the golden rule – he who has the gold makes the rules – so one obvious outcome of NCLB was a significant increase in educational research and, in particular, the type of research defined as SBR.

Yet, this impetus toward a more reflective view of education is of greater importance than simply following federal guidelines. For example, the standards of the National Staff Development Council (NSDC) state that the content of staff development should provide educators with "research-based instructional strategies." The NSDC explains,

> The charisma of a speaker or the attachment of an educational leader to an unproven innovation drives staff development in far too many schools. Staff development in these situations is often subject to the fad du jour and does not live up to its promise of improved teacher and higher student achievement. Consequently, it is essential that teachers and administrators become informed consumers of educational research when selecting both the content and professional learning processes of staff development efforts.

The same, of course, is true of every educational decision from curriculum adoptions to school calendars. Scientifically based research is critical for sound educational decision making (see Part I, Chapter 5 for the K-12 Educators' Checklist for Evaluating SBR). Scientifically based research is also critical if education is to continue to grow as a true applied science. The unexamined experiment is not worth conducting.

The Letter of the Law

The phrase "scientifically based research" (SBR) (sometimes referred to as "evidence based research") is defined by the text of the NCLB Act under Title IX - General Provisions, Part A- Definitions, Sec. 9101. Definitions, Number 37, subpart B, as "research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs." The Act then proceeds to focus this definition by establishing six criteria as the hallmarks of SBR (**bold** text is from the Act).

Hallmarks of SBR

1. **Employs systematic, empirical methods that draw on observation or experiment.** SBR requires precise descriptions of the interventions and programs being studied, careful observations of the work itself, and exacting measurements of its outcomes. In the ITQ context, this criterion requires quantitative research employing numerical measurement of teacher and student outcomes.

2. **Involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn.** At a minimum, such data analysis needs to meet the following tests:

   Sample size and representativeness. The sample must represent the targeted population for the intervention and must be large enough to allow reasonable, justified inferences to be made.

   Statistical procedures for data interpretation. Since SBR is comparative (treatments vs. control), it must include statistical tests (e.g., analyses of variance (ANOVAs)) that allow for the justification of conclusions regarding the efficacy of one group's treatment over the other.

***Hallmarks of SBR**, continued*

Significance. Statistical significance is expressed as the probability that the observed differences could have happened by chance. SBR must perform statistical tests to demonstrate that its effects are significant (e.g., typically p-value of .05)

Effect size. This is a measurement of how large a difference the treatment made. It is not enough to be merely statistically significant; learning gains must also be of such a magnitude that they have real world impacts for students.

3. **Relies on measurements or observational methods that provide reliable and valid data across evaluators and observers, across multiple measurements and observations, and across studies by the same or different investigators.** Reliability and validity are key concepts for the conduct of SBR. Essentially, reliability means the consistency with which an instrument or person measures whatever it is designed to assess; validity means that a test or procedure measures what it purports to measure (see Part I, Chapter 2 for much more on reliability and validity).

4. **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.** This criterion recognizes that all SBR is comparative and that this will require a minimum of two research groups – the treatment group and the control group. If assignments are made to the two groups through a randomization process, the research design is called experimental; if the control group is selected to match the treatment group on important criteria, the design is quasi-experimental (see Part I, Chapters 3 and 4 for more on these two research design methods).

***Hallmarks of SBR**, continued*

5. **Ensures that experimental studies are presented in sufficient detail and clarity to allow for replication or, at a minimum, offer the opportunity to build systematically on their findings.** Because education is an applied social science, it is critical that research studies be reported in enough detail to support the growing database of knowledge about successful educational practice. This means that research reports must include all relevant details about the intervention, participants, materials, outcome measures, and statistical procedures that were employed (see Part II, Chapter 5 for details on communicating research results).

6. **Has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review.** As mentioned in (5) above, one goal of SBR is to contribute to education's shared and open knowledge base. This can often be accomplished through publication of results in a peer reviewed journal or by utilizing the What Works Clearinghouse (see below for more about the WWC).

The Field Speaks

Given the sea change in education policy that the focus on SBR represents, it became something of a cottage industry to interpret and operationalize what it all really meant. At a United States Department of Education (USDOE) seminar in early 2002 on scientifically based research, leading experts offered the following guiding principles of scientific inquiry.

- Pose significant questions that can be investigated empirically.
- Link research to theory.
- Use methods that permit direct investigation of the questions posed.
- Provide a coherent chain of rigorous reasoning.
- Be replicable and generalizable.
- Be transparent and serve as a basis for scholarly debate.

While these are all useful principles, they are far from operationalizing guidelines. Based upon the defining characteristics found across a number of government and non-profit sources, the ERIC Digest (167) offered the following definition of SBR: "Persuasive research that empirically examines important questions using appropriate methods that ensure reproducible and applicable findings." The author, Ron Beghetto, then proceeded to amplify his definition.

- **Persuasive** – Research that uses appropriate methodology, is replicable, or is peer reviewed.
- **Empirical –** Based on measurement or observation.
- **Important Questions –** Linked to prior research, theory, and policy and practice.
- **Appropriate Methods –** Using designs, methods, and techniques that match the nature of the research questions being studied.
- **Replicable and Applicable Findings –** Consistent, meaningful, and detailed research results.

Such research goals are useful targets, but how are they to be operationalized in practice?

The What Works Clearinghouse

Perhaps the most concerted attempt to operationalize SBR can be found in the resources offered by the What Works Clearinghouse (WWC). The WWC was established in 2002 by the USDOE's Institute of Education Sciences (IES) as a centralized location to access scientific evidence of what works in education. As part of their mission, the WWC provides

> … reviews of the effectiveness of replicable educational interventions (programs, products, practices, and policies) that intend to improve student outcomes. To do this, the WWC uses standards for reviewing and synthesizing research. The WWC is currently conducting systematic reviews of existing research, and producing intervention and topic reports.

Additionally, the site offers resources for many aspects of the overall SBR process: There are a number of highly relevant technical working papers available to provide some guidance on research and analysis designs, including the WWC's current standards for study design classification with respect to SBR, the substantive interpretation of effect sizes, a tutorial on the mismatch between the unit of assignment and analysis (e.g., assigning interventions to entire classrooms but analyzing specific student changes), and information on reporting research results, especially how such results may become part of the WWC. Also included are research reviews (i.e., meta-reviews of the adequacy of published research studies on a particular topic) in a number of areas including beginning reading, early childhood education, elementary school math, English language learners, and middle school math curricula. There is also an online "Registry of Outcome Evaluators" to help identify potential researchers to conduct the study component of ITQ projects.

As the materials in this chapter suggest, there is general agreement about the broad principles that constitute scientifically based research, but there is no universal agreement on an exact definition or all the specifics that must be included in SBR. Furthermore, science itself is difficult to strictly define because science is, at least in part, a creative and evolving enterprise. No less is true of the applied social science that is education. As indicated on the WWC site, guidance and standards are not etched in stone and will evolve. While there are some generally accepted principles to use as a guide, all SBR projects are site-and context-specific. As such, the role of the researcher is to judiciously apply those principles and tools to conditions on the ground.

Finally, SBR needs to be thought of in at least two, larger contexts. First, it is research about *something*; in the ITQ context that "something" is a professional development intervention that is designed to lead to increased student achievement. This means that the SBR team is part of a larger educational enterprise that involves many stakeholders. Since the SBR cannot succeed without the cooperation of those stakeholders, researchers are cautioned to be sensitive to their needs. SBR requires a fair amount of technical expertise, but it also requires well-honed people skills. Much of the material in Part II of this *Field Guide* is designed as a guide to the nexus of these two, critical SBR requirements.

Second, SBR takes place within the context of the larger enterprise of educational research. It involves much more than carefully developing a research design or performing the proper statistical tests. It is a process that begins with a literature search to ground the SBR in what the field already knows, and it ends with a dissemination effort that enables the common knowledge base to continue to grow. The successful SBR researcher is much more than an educational "technician." He or she is an applied social scientist using a wide array of analytical tools to better understand the educational enterprise and to use that understanding to expand knowledge about what is effective in education.

2 Scientifically Based Research and Causality

The Science of Causation

All of the definitional attempts in the previous chapter might be compressed into one, properly understood sentence: SBR is the process by which one demonstrates convincingly that some educational intervention (A) caused some change in student achievement (B). Simple, A caused B. The problem, of course, is that demonstrating causation is never simple. Almost 300 years ago the Scottish philosopher David Hume elegantly analyzed what happens when one billiard ball strikes another and concluded that "we cannot observe the act of causation." We must, therefore, argue for causation through an inferential process. Furthermore, when we are attempting to demonstrate causation in a field as multi-faceted as education, that chain of inferences can grow long and, correspondingly, complex.

In order to help manage and control that inference chain, scientists have developed a theory of experimental design: Driving every experiment is a causal theory of action. For example, a program believes that teachers with more content knowledge will lead to students with better test scores. Now suppose that testing of teacher content knowledge and student scores both show increases. Does that prove the causal theory of action? Hardly. It might be that the student test scores fluctuated randomly. Or it might be that the student test underwent a revision in scope or scoring. Or it might be that a new curriculum was introduced. In fact, there are dozens of competing explanations for the change. The job of rigorous experimental design is to control for all those competing explanations, to eliminate or to address alternatives, so that the best explanation of the observed phenomenon is the original causal theory of action. It is very much a Sherlock Holmesian enterprise: "Eliminate all other factors, and the one which remains must be the truth."

Validity

The word used to express the strength of inferences is "validity", which is typically defined as the best available approximation of the truth of an inference. In the deductive sciences (mathematics and logic), this is rather straightforward. An inference is deductively valid if, and only if, true premises lead to true conclusions. In the inductive sciences (natural and social), validity is not nearly so cut and dried. Rather than being understood as necessary (if A is true, then A or not A must be true), judgments about inductive inferences must be considered approximate or conditional. In the context of SBR, the claim that an inference is valid does not mean that the conclusion could not be false; instead it is a claim that the conclusion is the best one possible that can be drawn from the premises. Add a new premise (more data) and the conclusion might change. One important implication of this

'contingency of validity' is that no experimental method (including randomization!) guarantees the validity of an inference. Furthermore, given that there are different types of validity (see below), it turns out that any particular experimental method can affect more than one type of validity at the same time. As explained by Shadish, Cook, & Campbell (SCC, see complete citation in Additional Resources), "[t]his is the nature of practical action: our design choices have multiple consequences for validity, not always ones we anticipate." In other words, SBR is very much a multi-level empirical science, and valid inferences are inferences to the best explanation. We are constantly in the position of trying to maximize validity, to improve the quality of that best explanation, but certainty is never obtainable

Of course, assessing the quality of any given SBR project involves much more than ascertaining the validity of a single inference. SBR projects are complex enterprises that involve many parallel inference chains. Because of this complexity, social science researchers have found it useful to create a taxonomy of four types of validity reflected in four types of questions:

Construct Validity

Does the actual intervention accurately reflect the construct (idea) of the intervention; do the instruments accurately measure the underlying constructs? In layman's language, was the intervention implemented as planned? Were the outcome measures accurate representations of the intervention results? Consider the above example, improved teacher knowledge will lead to improved student test scores. How are we to measure teacher content knowledge? Is it simply a matter of a test like the CSET that tests content knowledge, or must the test also measure pedagogical content knowledge? If so, how can this be accomplished?

Statistical Conclusion Validity

Is there a statistical correlation between the treatment and outcome? If so, how large or significant is it? In layman's language, is there a measurable relationship between them? Given any two variables, it is possible that they may have a positive relationship (e.g., the more time one studies, the more likely one is to do well in school), a negative one (e.g., the more time one spends watching TV, the less likely one is to read a book), or none at all (e.g., the type of music students listen to has no relationship to their school performance). Furthermore, researchers need to guard against the two error types that may cause them to make faulty inferences when attributing correlations between two variables: Type I errors occur when we conclude treatment and effect correlate when they do not, Type II errors occur when there is correlation and we fail to note it.

Internal Validity

Do the inferences between the presumed treatment (A) and the presumed outcome (B) reflect a causal relationship from A to B as those variables are manipulated and measured? In layman's language, did A cause B? Staying with our example, how can one be sure of a specific causal relationship? How can one rule out other possible causal explanations such as student maturation or statistical regression?

External Validity

How can we be sure that the validity of the cause-effect inference holds over other persons, settings, treatments, observations, and instruments? In layman's language, can the effects of this experiment be generalized? For example, was there anything special about the teachers who participated in the intervention? Were they volunteers? Were instruments developed especially for this intervention program? If so, are they consistent with other, broadly used, instruments?

A well run SBR experiment must guard against threats to each of these four types of validity.

Specific Threats to Validity

For any given inference or conclusion, there are always specific threats to validity – reasons why the inference or conclusion might be wrong. Inference to the best explanation sets dual requirements on validity: First, an affirmative, plausible argument as to why the chosen explanation is the best. Second, a strong negative argument as to why other possible explanations are not competitive. These other possible explanations are 'threats to validity', and SCC catalogs 37 different kinds across the four types of validity. Of course, it is not the case that every SBR project will be threatened by all thirty seven. In fact, the likelihood that a specific threat to validity will occur varies across contexts, and no list of threats is ever complete. Therefore, projects must determine which threats are implicit in their research designs and take affirmative steps to render them benign (i.e., demonstrate why they are not the best causal explanation of the desired outcome). Rather than simply list all 37 here, we have focused on the internal validity threats that are common in SBR projects. The table below provides a brief introduction to such threats and to some specific fixes.

Common SBR Internal Validity Threats

Threat	Factors to Consider	Potential Fixes
Attrition	Do participants lost over time reflect some form of bias (e.g., motivation, persistence)?	Reduce attrition by adopting retention strategies; analyze attrition to determine if it is a real threat to causation; consider matching cohorts in NEGD.
History	Did any event occur between pre and post measurement that might account for outcome differences?	Reduce plausibility of history threat by selecting groups from the same location/conditions; ensure that testing schedule is the same for both groups.
Instrumentation	Were there any changes in measurement protocols (e.g., instruments, rubrics) that might account for outcome differences?	Avoid changing data collection instruments/protocols; if change is mandated try to develop a calibration protocol between instruments.
Maturation	Were there any changes in the project participants, independent of the treatment, that might account for outcome differences?	Select participants of the same age group and from the same location.
Regression	How can you show that pre-post measurement changes are not statistical artifacts?	Avoid selection based on extreme scores; consider selection measures that are averaged over time; perform diagnostic tests for regression on standardized scores.
Selection	How can you show that systematic differences in participants do not account for outcome differences?	Consider true randomized experiments; use pretests to explore the size and direction of selection bias.
Testing	How can you show that repeated testing does not account for outcome differences?	Use item response theory to create different tests calibrated to yield equivalent ability estimates.
Interactions	How can you demonstrate that more than one threat (e.g., selection and maturation) are not working synergistically to cause outcome differences?	Little is known, quantitatively, about how threats interact. One might expect, however, that selection is likely to work with attrition, history, and maturation and apply multiple fixes.

General Experimental Design Considerations

The best way to avoid specific threats to validity is through a generally robust experimental design. Careful design with the proper experimental controls (as opposed to relying strictly on *ex post facto* statistical techniques) should minimize both the number and plausibility of threats to validity that remain at the end of the study. Careful researchers need to anticipate all reasonably foreseeable threats and account for them in the research design. Yet, even the most thorough, experienced researcher may find that some threats to validity cannot be directly ruled out through design controls because either the logic of design controls simply does not apply or, more likely given the real world constraints of educational research, practical impediments in the field may prevent available design controls from being used. SBR is an empirical science and there can never be certainty. One can never really list all potential threats before the experiment is conducted, and one can never rule out a specific threat with certainty. The goal is always to make the threat a less plausible explanation within the research context. In addition to using design elements to improve the validity of their causal inferences, researchers may also apply one, or all, of the following strategies:

Argument

Although a generally weak approach, it may be possible to marshal a convincing argument that a particular threat is not relevant, or plausible, on specific grounds. For example, suppose you could demonstrate that there is no mechanism for a given threat to apply in your research project. Or suppose you could show that there is no evidence that the given threat is plausible rather than just possible. Or suppose you could argue the given threat causally operates in the opposite direction as the observed effect so that it could not explain the observed findings.

Measurement or Observation

If a specific validity threat can be specified, then it may be possible to measure it and demonstrate that it does not occur at all or that its effect is so small that it really can't be an alternative explanation for the treatment effect.

Analysis

While statistical analysis should not be the first approach, it can be a useful tool to rule out certain validity threats. Trochim and Land (see Additional Resources) provide the following example of using a two-way analysis of variance to study an attrition threat: One factor is the original group designations (treatment vs. control), the other is attrition (dropout vs. retained). The dependent measure could be the pretest or other pre-treatment measure. A main effect on the attrition factor would indicate a threat to external validity or generalizability, while an interaction between group and attrition factors would point to a possible internal validity threat.

Preventive Action

When it is possible to foresee specific threats, it may be possible to adjust the design in ways that will avoid them. For example, making the treatment available to the control at a later time may eliminate the threat caused by the control "adopting" the treatment on their own and confounding pre-post comparisons.

In conclusion, Trochim and Land provide a general strategy for design construction: While no definitive approach for creating experimental designs exists, we suggest a tentative strategy based on the notion of expansion. First, set forth a design which depicts the simple hypothesized causal relationship. Second, deliberately over-expand this basic design by expanding across time, program, observations, and groups. At this step, the emphasis is on accounting for as many likely alternative explanations as possible using the design. Finally, scale back this over-expanded version considering the effect of eliminating each design component. At this point we face the difficult decisions concerning the costs of each design component and the advantages of ruling out specific threats using other approaches. (See Part I, Chapter 5 for a checklist approach to these general ideas).

3 The Basics of Randomized Experiments

Causal Inference and Randomization

A major goal of scientifically based research is to be able to strongly support a causal claim of the type "student scores improved because of the intervention program." The previous chapter's discussion of internal validity makes clear that such a claim cannot be strongly supported simply because the cause preceded the effect. In order to make a strong case for causation, all the various threats to internal validity (i.e., alternative causal explanations) must be shown to be "weaker" explanations of the effect. Such a threat-by-threat analysis is certainly laborious; it may even be practicably impossible. Suppose, however, there was an experimental design that greatly limited the threat from such alternative explanations. This is the major argument for randomized experiments, sometimes referred to in the literature as randomized control trials (RCTs).

In order to establish causation in the Improving Teacher Quality (ITQ) context, we want ideally to demonstrate two cases: (1) students involved in the intervention saw their scores increase and (2) students not involved in the intervention did not see their scores increase. If these cases could be simultaneously shown for the same students, it would be very strong evidence that the intervention caused the increased scores. But how can these two cases occur for the same, identical students? That would be logically impossible, but RCT holds out promise for the next best thing. Imagine that there are two groups of students, comparable across all the relevant variables, and that one group gets the intervention while the other (the control) does not. Now imagine that the scores of the intervention group students improve while those in the control group do not. To the extent that the groups were really comparable, causation would be demonstrated. But where does one find two groups of such comparable students? This is the genius of randomization. Beginning with a large enough single group, researchers may create such "equivalent" groups by randomly assigning members to either the treatment or control group. The remainder of this chapter briefly explains the promises and challenges of such randomized experiments.

Random Assignment Benefits

The central concept of randomized experiments is random assignment. Random assignment procedures assign units of analysis (e.g., students, teachers) to conditions (e.g., intervention, control) only on the basis of chance, where each unit has a non-zero (and usually equal) probability of being assigned to a particular condition. For example, consider the toss of a fair coin. Over the long run, we expect that heads will come up 50% of the time. Thus, if heads come up on the toss for a particular unit, we assign that unit to the treatment condition.

If tails comes up, we assign the unit to the control condition. There are more formal randomization procedures available, including random number tables or the use of computer programs to generate random numbers. Regardless of the sophistication of the randomization technique, the key result is equi-probability.

This key result can be better understood by comparing random assignment to random sampling. Both use a randomization procedure, but do so to obtain different results. Pollsters sometimes randomly *sample* units by chance from the population comprised of all such units to ask their opinions (e.g., registered voters in the United States). The goal here is to create a representative group of the population at large; such results are usually reported within a sampling error. By contrast, random *assignment's* purpose is to make samples similar to each other, thus facilitating causal inference. In both cases, inference is involved, but in the former it is from the sample to the whole, while in the latter it is from the cause to the effect.

A number of complementary statistical and conceptual explanations for why and how random assignment facilitates causal inference can be found in the literature.

- It ensures that alternative causes are not confounded with a unit's treatment condition.
- It reduces the plausibility of threats to validity by distributing them randomly over conditions.
- It equates groups on the expected value of all variables at pretest, measured or not.
- It allows the researcher to know and model the selection process correctly.
- It allows computation of a valid estimate of error variance that is also orthogonal to (i.e., independent of) treatment.

Quite clearly, most of these benefits are related to eliminating threats to internal validity. For example, it helps to eliminate the problem of ambiguous temporal precedence because the temporal structure of the experiment guarantees that cause precedes effect. Another way in which random assignment helps with internal validity is to prevent the threat of selection bias. Selection bias implies that a systematically biased method was used for assigning units to treatment and control conditions. Yet, by definition, random assignment does not have such a systematic bias. In randomized experiments, through the process of random assignment, treatment group units tend to have the same average characteristics as those in the control group. The goal is that this leaves intervention effects as the only plausible explanation for differences between treatment and control groups.

Random Assignment Limitations

It is important to understand that randomization does not mean that systematic differences among units are eliminated. Instead, they are more or less evenly distributed across treatment and control conditions, thus equating the two groups relative to one another before the treatment is administered. For example, suppose we have 50 teachers to randomly assign to conditions, and we intend to administer a Professional Development (PD) intervention to the treatment group teachers. Unknown to us, 10 of these teachers have 25 or more years of teaching experience, far higher than the 10 or less years of experience of the other 40 teachers. In addition, another ten teachers are currently involved in another study where they are receiving PD in the same area as ours. Random assignment of all 50 teachers to the various conditions will not make these differences disappear. What it will do, however, is ensure that these differences are more or less evenly distributed between the treatment and control conditions. Thus, the average years of teaching experience are likely to be more or less the same in each group, as will the number of teachers currently receiving the additional PD.

Another important limitation is that randomization does not prevent the occurrence of other threats to internal validity. It can only reduce the likelihood that these other threats are confounded with the treatment. Continuing the previous example, random assignment of the 50 teachers to the two conditions does not preclude the possibility that some teachers may drop out of the study (i.e., attrition), or that some teachers might opt to sign up for additional PD during the course of the study (i.e., history). But all other things being equal, randomization does make it more likely that such unforeseen events will affect both treatment and control groups to a similar extent. Had the teachers not been randomly assigned to conditions, the retirements of the 10 senior teachers during the study would probably have differentially affected one condition or the other. Randomization helps to ensure that change affects both conditions to a similar extent.

Experimental Designs Employing Random Assignment

Random assignment is simply a statistical technique for creating equivalent condition groups. This technique may be embedded within many different experimental designs to produce different versions of randomized experiments. The list below is a brief introduction to those designs that may be most beneficial in the Improving Teacher Quality context. (A checklist for developing RCTs is found in Part I, Chapter 5). Much more detailed information is available in Shadish, Cook, & Campbell (SCC).

The Basic Randomized Design

In the basic randomized experiment, there are at least two conditions to which units are randomly assigned. A posttest assessment of units is also included. Selection of the type of control group needed is a key issue here, and this depends upon what must be controlled. Types of control groups include the typical no-treatment controls, and also dose-response controls, wait-list controls, expectancy controls, or attention-only controls. In ITQ, one might typically use no-treatment controls since the goal is to assess the efficacy of the PD intervention compared to a lack of treatment.

The Basic Randomized Design Comparing Two Treatments (Alternative-Treatments)

This variation of the basic design compares two treatments instead of a treatment to a control. This design is typically used if the effects of one of the treatments are well-known (i.e., a "gold standard") as compared to no-treatment controls, and one is interested in comparing the effects of a new treatment against a more accepted gold standard.

The Basic Randomized Design Comparing Two Treatments and a Control

This variation extends the prior design in an important way. In this design there is no need to have prior knowledge of the track record of one of the treatments against no-treatment controls. Instead, the addition of a control group can help to assess the efficacy of the treatment at posttest. With the control group as a reference, one gets a better idea of the relative effectiveness of each treatment or lack thereof.

The Pretest-Posttest Control Group Design

This design is probably the most used randomized field experiment. The previous designs did not include a pretest, but a pretest is usually recommended if there is any likelihood of attrition from the study. In most field experiments, including ITQ Scientifically based research (SBR), attrition is a fairly common phenomenon. Adding the pretest to the design increases our ability to cope with attrition as a threat to internal validity. In addition, it can allow for certain statistical analyses that increase statistical power (to reject the null hypothesis).

Alternative-Treatments Design with Pretest

The addition of a pretest to this design is useful if posttests reveal no differences between the two treatment groups. It becomes possible to look at pretest and posttest scores to see whether either group showed improvement. This design is also useful when there are ethical or other considerations that make comparison to a control condition impractical. For example, suppose you are fortunate enough to have the resources to deliver your PD intervention to all teachers in a given school district, and identifying and using controls outside the district is not feasible. Under these circumstances, you could design a study in which you give a more complete intervention to one treatment group, and a reduced intervention to the other.

The Basic Randomized Design Comparing Two Treatments and a Control with Pretest

This design represents an improvement over the basic design with the addition of a pretest. It also represents an improvement over the previous design because it allows a comparison of two treatment conditions to a reference control group. This design may not be practical due to a lack of resources or for logistical concerns. But, in theory it can be extended to include more than two alternative treatment conditions and more than one control group. The more levels of the treatment that can be administered, the finer the assessment of the functional form of the dosage effects of the treatment. More treatment levels also can help to detect effects that could have been missed had only two levels of treatment been used.

Factorial Designs

In its most basic form, this type of design uses two independent variables (called factors), each with two levels. The number of factors and levels can also be extended. Suppose you want to compare two levels of your PD intervention, and you want to compare the delivery of this intervention with respect to whether it is delivered by certified or non-certified district staff. The teachers who are to receive the PD would be randomly assigned to one of the four logical combinations of your treatment (e.g., high-PD and certified, high-PD and uncertified, low-PD and certified, low-PD and uncertified). These types of designs have three major advantages. They often require smaller sample sizes than would otherwise be needed, they allow for the testing of combinations of treatments more easily than other designs, and they allow for the testing of interaction effects. However, they are often difficult to implement in field settings because they require close control over the combination of treatment given to each unit.

Longitudinal Designs

These designs build upon the basic pretest-posttest control group design. They do so by adding multiple observations taken before, during, or after treatment. The feature of multiple observations is well-suited to the ITQ SBR context. Making multiple observations during the delivery of treatment is wholly consistent with the formative assessment that is typically done over the course of an ITQ project, and it aids in judging the fidelity of the treatment implementation. In addition, promising interventions may warrant multiple years of follow-up observations on either teachers or students to assess longer-term impacts of the PD intervention that was delivered. A major problem for these types of designs can be increased attrition with longer follow-up periods after treatment.

Randomized Experimental Design Benefits

According to the text of the No Child Left Behind (NCLB) Act, randomized experimental designs are the preferred design for the ITQ SBR context. In addition to the reasons already presented dealing with causal inference, the reasons below provide additional support for RCTs primacy in the NCLB Act:

When Demand Outstrips Supply

Under these conditions, randomization can be a defensible way to distribute resources fairly. High demand with low supply is practically a general working condition in the context of ITQ SBR since projects work in high-need districts with the relatively limited resources of the grant. Randomization is an unbiased way to decide who gets to receive the treatment during the experiment, but the procedure must be explained clearly and carefully so that everyone sees and understands why this is so, and that no unintended discrimination is taking place.

When an Innovation Cannot Be Delivered to All Units at Once

Sometimes it is logistically or financially impossible to deliver the PD intervention to all the teachers that could be included in the study. Perhaps the PD is complex and intensive, and so can only be delivered during a one-week retreat to a limited number of teachers per year by a limited staff of qualified experts. Under these conditions, randomization is also a potentially useful way to decide who gets the treatment, and still allows a comparison of the relative efficacy of the intervention against a control group.

When Experimental Units Are Spatially Separated or Inter-Unit Communication is Low

This is a condition that can often be found in the ITQ SBR context. For example, it might exist in a large district containing a large number of schools, a rural school district with a number of schools located far apart, or in cases where the PD intervention is delivered across multiple school districts. Because of the separation, randomization of the units will tend to preserve fidelity of the implementation without leakage across units.

When Change Is Mandated and Solutions Are Acknowledged to Be Unknown

This condition could reasonably be regarded as a founding premise for the ITQ Grant Program. Improvements in teacher quality and student achievement are goals of NCLB. We all recognize that change is needed, but none of us are even pretending to have all the answers for what works and what does not work. Therefore, rigorous testing of various solutions through randomized experiments is warranted. In addition, keep in mind that NCLB has a stated preference for these randomized designs.

Randomized Experimental Design Limitations

Although randomized experimental designs are preferred for ITQ SBR, there are some conditions under which using these designs may not make complete sense. In these cases, a quasi-experimental design of some kind (see Part I, Chapter 4) might be preferable.

When Quicker Answers Are Needed

It is possible that, given the nature and content of the PD intervention, research using a randomized design cannot be successfully carried out within the project time frame. In such cases, alternative designs using a quasi-experimental approach and/or qualitative methodologies may be most appropriate.

When the Need for Great Precision is Low

It might also be a waste of resources to conduct an additional randomized experiment when an effect of a treatment has been shown in other studies to be so large and dramatic that there is little doubt it resulted from the treatment. For example, suppose one is studying a PD intervention about which much high-quality information already exists, and the research goal is to try to answer questions about differences resulting from how the same intervention is delivered to teachers. In this case, it may be more useful to focus on

program monitoring procedures designed to insure the fidelity of the treatment implementation than on randomization.

When Randomized Experimentation is Premature

Suppose a project is going to administer a fairly new and innovative PD intervention, or perhaps it is using a method of delivery that is new or innovative (e.g., a wholly online PD intervention). In this case, it may make more sense to use research resources to implement pilot tests that investigate the feasibility and acceptability of the project, rather than to conduct a premature randomized experiment.

There are also a number of more practical problems that can arise in implementing a randomized experiment. They should be carefully considered and kept in mind if you plan to implement a randomized design. We list them here, and refer readers to (Part II, Chapter 6) for potential ways to address them.

- A sufficiently large number of units (e.g., schools, teachers, students) may not exist who are both eligible and willing to receive the treatment if assigned to it at random.
- Randomization procedures are not always properly designed and implemented.
- The treatment assigned is not always the treatment received.
- Variable attrition of research subjects may occur.

4 The Basics of Quasi-Experimental Designs

Quasi-experiments vs. Randomized Experiments

Put simply, quasi-experiments must meet all the standards of randomized experiments except, of course, for random assignment. In other words, causal inferences from quasi-experiments are subject to the same basic requirements as those for randomized (true) experiments: 1) cause must precede effect, 2) cause must covary with effect, and 3) any alternative explanations for causal relationships must be shown to be implausible (or at least more implausible than the proffered explanation). In both types of experiments, the treatment is manipulated to insure that it occurs before the effect. In addition, in both types of experiments statistical analysis can be conducted to assess covariation between cause and effect. However, because quasi-experiments do not use random assignment to determine treatment and control groups, they must rely on other principles to demonstrate the implausibility of alternative explanations for the cause and effect relationship. According to Shadish, Cook, & Campbell (SCC), there are three major principles that may be adduced.

Identification and study of plausible threats to internal validity

The previous two chapters have made clear both the nature of internal threats to validity and the possibility of experimental procedures to rule out given, specific threats. This principle requires that such threats to internal validity be specified and that procedures be developed for their careful study. By so doing, it may become possible to assess how likely it is that any given threat explains treatment-outcome covariation, and it may be possible to demonstrate that, in cases where it is likely, it still turns out not to be the case. For example, recall that for randomized experiments the threat of selection bias was dealt with by the random assignment procedure. This is clearly no longer the case for quasi-experiments. Therefore, selection bias is, prima facie, one of the major threats to internal validity that must be carefully examined as part of the overall study assessment of a quasi-experimental design.

Primacy of control by design

All experiments, even randomized experiments, are vulnerable to the confounding of internal validity threats with the treatment. In general, the best way to avoid this confounding in social science experiments, and especially with quasi-experiments, is to purposively add design elements, e.g., more observations or control groups that directly address such threats. (See the final section in this chapter for a fuller discussion of design elements). Robust experimental designs are a needed antidote to complex experimental conditions. While statistical controls are also a useful tool to try to remove confounds after the study data is collected, the best studies apply both techniques.

Coherent pattern matching

This principle for addressing the plausibility of alternative causal explanations in quasi-experiments utilizes increased complexity in logic and prediction. The general idea is that one may eliminate alternative causal explanations by making a prediction about a given causal hypothesis of sufficient complexity that the prediction cannot be matched by other, simpler, alternative explanations. The more complex the pattern that can be successfully predicted, the less likely it is that the pattern could be matched using alternative explanations, thus increasing the likelihood that the treatment administered had a real effect. An example of this principle is the use of nonequivalent dependent variables in a study – additional dependent variables that are predicted not to change due to the treatment, but that are expected to respond to plausible internal validity threats. Such "internal validity markers" may allow for a much more sophisticated causal analysis.

Understood in this way, randomized experiments (RCTs) and quasi-experiments (QEDs) are quite similar, both needing to meet all the requirements of causal analysis discussed in Part I, Chapter 2. When random assignment is impossible, it becomes more difficult to rule out several internal validity threats, and researchers must be clever when selecting a research design and the way in which it is to be implemented. Generally speaking, there are three classes of QED designs: (1) Non-Equivalent Groups Design (NEGD), (2) Regression Discontinuity Designs (RDDs), and (3) Interrupted Time Series (ITS). The first two are introduced below in brief sections; interrupted time series designs have little application to ITQ work and are not further discussed. Readers interested in more detail, or in variants of the basic designs, are referred to SCC (complete reference in Additional Resources).

Non-Equivalent Groups Design (NEGD)

It is sometimes claimed that Non-Equivalent Groups Design (NEGD) is the most frequently used design in all of social science research. The basic structure is the same as a pretest-posttest randomized design but without the random assignment. There are three reasons why the pre-post is so ubiquitous and important.

- With a pretest, one can directly examine initial differences between treatment and control groups. This helps both immediately (e.g., as a screen against selection bias), and later on as the study progresses (e.g., as a way to understand possible variable attrition).
- With a pretest, one is able to look at the magnitude of initial group differences. This is an important advantage because these observed differences are for a variable that is highly correlated with the intended outcome of the study. This advantage is based upon the assumption that smaller pretest differences between groups implies a reduced threat of initial selection bias. Of course, without random assignment, it cannot be assumed that variables that weren't measured at pretest are unrelated to the intended outcome.
- The availability of pretest data also assists with statistical analysis and more generally with threats to statistical conclusion validity, particularly if the reliability of pretest measures is known. Pretest scores provide a baseline for each group to compare, both across groups and to their posttest performance.

What replaces the random assignment? A process of selective judgment where the researcher tries to create or select a control group that is as similar as possible, along relevant variables, to the treatment group. Because it is impossible to be sure that the two groups are really comparable, this is called the "non-equivalent groups design." One major consequence is that the researcher must be especially vigilant concerning internal validity threat of selection.

What does NEGD look like in operation? In its simplest form, there is a treatment group, and a control group that does not get any exposure to the treatment. Pretest and posttest data are also collected from both groups. For example, a Professional Development (PD) project might deliver a program of mathematics pedagogical content to a group of 7th grade teachers during a summer institute, with school year follow-up. A sister school with similar demographics in the same district could agree to serve as the control, and it is established that no one at the control school received any of the treatment. A pretest is administered in 7th grade mathematics at the start of the school year to the incoming students of teachers in both groups, and this is followed by a posttest of these same students at the end of the school year. Perhaps a pre- and post-content and pedagogical knowledge test is also administered to the teachers. Such a simple design collects enough data to allow for simple causal inferences, though it hardly rules out all other causal explanations.

The combination of a no-treatment control group and pretests allows the assessment of some threats to validity. Without the benefit of randomization, one should assume that selection bias of some kind is present in either the student or teacher samples. Yet, armed with pretest data, one is able to look at the possible size and direction of the potential bias. Later on in the study, this pretest data will also help to assess, and perhaps deal with, issues of attrition since it will be possible to get a sense of group differences with respect to those students or teachers who leave the study (e.g., transiency/mobility, retirement, etc.) and those who remain in the study. But again, one should remember that observed absences in pretest differences is not proof of the absence of selection bias.

Of course, *possible* threats to internal validity do not necessarily mean *probable* threats. One major task when using a QED is to demonstrate the improbability of these possible threats to the study. Given observed pretest and posttest outcomes on both treatment and control groups, it may be possible to use different patterns that emerge in these comparisons to argue that certain threats to internal validity are more or less plausible. Generally speaking, with the NEGD there are five major outcome patterns.

- **Both treatment and control groups grow apart in the same direction:** This is certainly consistent with selection-maturation since initial differences are being reflected over time. On the other hand, it gives little credence to selection-regression because neither group is regressing. (See Part I, Chapter 2 for more on these threats).
- **Treatment group improves, no change in the control group:** This pattern argues against selection-maturation (why no growth for the control?) as a threat but is consistent with a threat from selection history (something other than the treatment caused the treatment group's growth).
- **Initial pretest differences favor the treatment group but they diminish over time:** Selection-regression fits this pattern very well (the treatment group was chosen because of their high scores on the pretest and they have simply regressed) but it is inconsistent with selection-maturation.
- **Initial pretest differences favor the control group but they diminish over time:** Similar to the pattern above, yet this time it is the control group participants who are regressing.
- **Observed outcomes cross over in the direction of relationships:** This pattern provides the strongest evidence for the program effect but it is impossible to set out to create such a pattern. Just count the researcher fortunate whose careful work is so rewarded.

There are numerous design features that can be added to make the "standard" NEGD more robust, and many of them are well covered in SCC. Additionally, it may be possible to combine various experimental designs so that the most efficacious design in a given context turns out to combine RCT, NEGD, and RDD. Yet, regardless how complex or sophisticated

such designs may become, the researcher is cautioned to remember two principles: First, every NEGD must meet all the requirements of randomized experiments except for random assignment. Second, no matter how compelling the comparison group, one must always be alert to threats from selection bias.

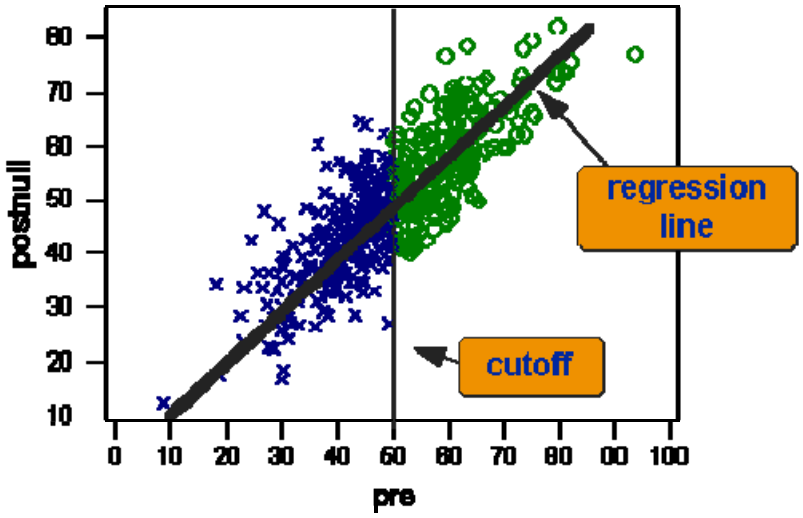
Regression-Discontinuity Design Experiments

Researchers already familiar with the pretest-posttest program comparison group research design (i.e., the basic strategy for both randomized and NEGD experiments) will feel initially comfortable with the basic RDD experiment since it also employs these design features. Yet, this comfort may initially change to unease when they realize that instead of trying to create two equivalent groups (through randomization) or two comparable groups (through an application of judgment), RDD experiments expressly set out to create incomparable groups by using a cutoff score on the pre-program measure as the sole criterion for assignment. Researchers trained in the more familiar experimental designs (RDD experiments are used much less frequently in social science research) generally have two reactions: Why would anybody want to do that? And how could that possibly work?

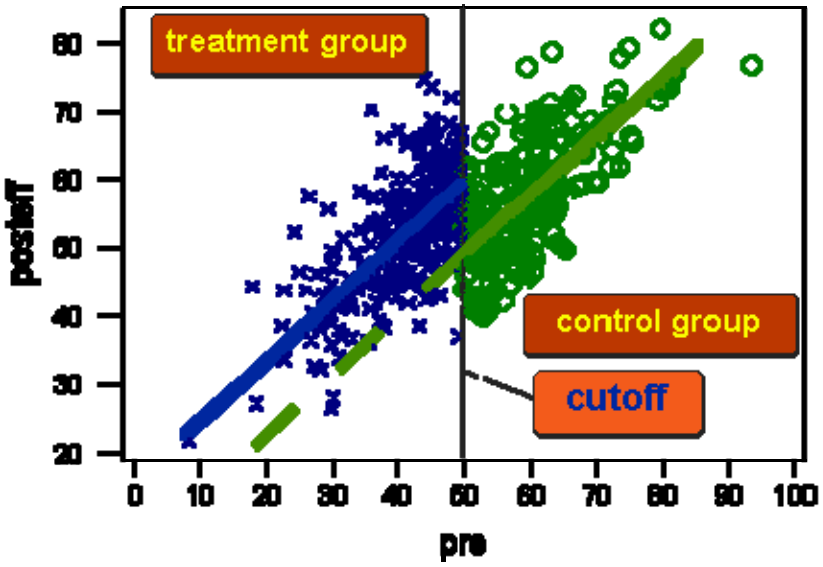
There are four major reasons why a researcher might want to consider an RDD. The first is ethical. A standard argument against randomization is that a scarce resource (e.g., tutoring for students below grade level in reading) should go to those students who are most at need rather than be assigned randomly for the sake of an experiment. RDD is compatible with this argument since the treatment goes to those students who are below a given cutoff score (i.e., presumably those who are below grade level in reading). In fact, one area in which RDD has been widely used is the assessment of compensatory education programs. Second, RDD, because of its requirement for large data sets (e.g., it needs many more participants than a randomized design to be as statistically powerful), can often be used with existing data collection efforts. Seen in this light, it can be a powerful analytical tool. Third, RDD can often be used with, and in support of, other experimental designs. Fourth, and perhaps most surprising, when properly conducted, RDD experiments are comparable in internal validity to conclusions supported by randomized experiments. Proving this requires a fair bit of statistical power (in fact, using RDD requires a solid knowledge of regression techniques); yet, one can begin to appreciate why it might be true by understanding how RDD works.

The basic RDD experiment has two groups (treatment and control), two measures (pre and post treatment) and a cutoff score based on pretest performance. To take a simple, concrete case: Imagine that 100 6^{th} grade students were tested for reading level and imagine that 50 tested below grade level, the rest at or above grade level. The 50 students below grade level receive the treatment (one-on-one tutoring for three months); the others receive no treatment and serve as the control. What will a bivariate graph of this situation look like if there is no treatment effect? There will be a scatter plot of 100 points, all the treatment students below and to the left of the control students, with a vertical line marking the cutoff between the two groups. Most important, there will be a smooth regression line stretching through all the points. This is depicted in the first graph below. Now imagine that the treatment was effective and there was a constant effect raising each treatment student's score by 10 points. The graph would be the same except that all the points to the left of the cutoff score would be

raised ten points on the posttest. Again, most importantly, this time the regression line would jump (be discontinuous) at the cutoff point. This positive jump is the evidence that the treatment was effective. (Pictured in the second graph)



No Treatment Effect



Ten Point Treatment Effect

The key idea is a conceptual shift away from pre-post group differences and similarities which is crucial for randomized and NEGD experiments, to an analysis of pre-post relationships. Specifically, RDD experiments assume that without a treatment the pre-post relationship would be equivalent for the two groups. This means, of course, there is no discontinuity that just happens to coincide with the cutoff point and that the pre-post

relationship can be correctly modeled using statistical analysis. Justifying these assumptions is beyond the scope of this *Field Guide* but such arguments may be found in SCC.

Design Elements

Too often, the task of establishing a research design is akin to ordering off a Chinese menu: The researcher simply chooses one from column A (e.g., the broad category of experiment, NEGD) and one from column B (e.g., a pre-existing design variant within that category, untreated matched controls with multiple pretests and posttests, nonequivalent dependent variables, and removed and repeated treatments). Such taxonomic-driven thinking underestimates the task of research design on two counts. First, it limits design to a finite list of options; second, it imposes the design onto the research context rather than building it up from conditions on the ground. The antidote to such thinking is described by W. Trochim (see Additional Resources) as "a more integrated, synthetic view of quasi-experimentation as part of a general logical and epistemological framework for research." Such a view, "is based on multiple and varied sources of evidence, it should be multiplistic in realization, it must attend to process as well as outcome, it is better off when theory driven, and it leads ultimately to multiple analyses that attempt to bracket the program effect within some reasonable range." Design elements are the tools that allow the researcher to build such an integrated, synthetic experimental design.

According to SCC, there are 25 specific design elements within four broad categories: assignment, measurement, comparison groups, and treatment, and each may be used to make an experimental design more robust by adding design elements that provide additional information which may be useful in ruling out a validity threat. For example, because QEDs rely on nonrandom assignment into treatment and control groups, self-selection is always a threat. One design element that may help to mitigate this threat is matching or stratifying. There is a large literature discussing various matching techniques and it is clear that matching can only take place on observed measures so hidden biases may remain. Still, it represents a useful tool in giving the researcher more control.

The measurement regimen of most educational experiments involves pretests. By incorporating repeated pretests on the same construct it may be possible to uncover maturational trends, detect regression artifacts, and study testing and instrumentation effects. Again, more robustness provides useful information. One of the most difficult aspects of an NEGD experiment is the development of a comparison group. While it may seem counterintuitive, using multiple nonequivalent comparison groups may be both easier to establish (i.e., because one group doesn't have to serve all purposes) and yield better results. The key idea is that multiple groups allow the researcher to explore more, and different, threats to validity and to triangulate toward a more constrained cause and effect relationship. In the ITQ context, cohort controls – groups that move through an institution in cycles – may be particularly useful since they are believed to be more comparable to each other than other nonequivalent comparison groups.

Finally, with the cooperation of those running the professional development, the researcher may be able to vary the timing of the treatment in ways that greatly benefit the research. For

example, the switching replication method applies the treatment at a later time to a group that originally served as a control. Multiple comparison groups (perhaps cohorts) allow the treatment effect to be studied across several groups that receive the treatment at different times, and the repeated treatment method (sometimes called the ABAB design to represent the on-off nature of the treatment) allows for a more precise causal tracking. Whichever design elements are implemented, the goal is always the same: To collect richer, more varied data so as to better understand the complex causal relationship(s) between treatment and effect.

5 Research Checklists

The Checklist Scaffold

The complexity of scientifically based research puts multiple demands on the researchers: They must be expert planners, they must be technically proficient across a broad domain of skills, they must be superior communicators, and they must be efficient administrators. While no single resource can serve as a guide for the total conduct of scientifically based research (SBR), and while there is no substitute for field-based experience, the checklists that comprise a substantial portion of this *Field Guide* represent the best available guide. As a tool, checklists have a number of qualities that make them extremely useful in the conduct of SBR.

- They are an important mnemonic tool that helps to reduce errors of omission.
- They are couched in simple language allowing all stakeholders to understand the process.
- They establish a higher standard of baseline performance by making explicit the minimum requirements of given tasks.
- They reduce the influence of the Rorschach effect (i.e., the tendency to project one's biases onto an information complex) by forcing separate judgments on each, individual step or criteria.
- They incorporate a huge amount of specific knowledge about a specific task domain and are organized in a way that facilitates those tasks.

These checklists are intended to serve as the scaffolding for an SBR project. Individual researchers and their projects must provide the actual "research building materials".

Scientifically Based Research Checklist

While there is no such thing as "the official, government-sanctioned, SBR checklist", the U.S. Department of Education's (USDOE) Institute of Education Sciences (IES) has issued a report, *Identifying and Implementing Educational Practices Supported By Rigorous Evidence: A User Friendly Guide* (see Additional Resources for web address) that contains a "checklist to use in evaluating whether an intervention is backed by rigorous evidence." According to the report,

> The federal *No Child Left Behind Act of 2001*, and many federal K-12 grant programs, call on educational practitioners to use 'scientifically-based research' to guide their decisions about which interventions to implement…This Guide is intended to serve as a user-friendly resource that the education practitioner can use to identify and implement evidence-based interventions, so as to improve educational and life outcomes for the children they serve.

In other words, this checklist is not a procedural instrument intended to guide the research; instead, it is an assessment instrument to be used in determining whether the research meets SBR standards. Of course, it can also be used prospectively to insure that those standards are being met.

SBR Checklist

- ☐ 1.0 Is the intervention supported by "strong" evidence of effectiveness?
  - ☐ 1.1 Is the evidence based on randomized controlled trials?
  - ☐ 1.2 Is the intervention clearly described?
    - ☐ 1.2.1 Who administered it?
    - ☐ 1.2.2 Who received it?
    - ☐ 1.2.3 What did it cost?
    - ☐ 1.2.4 How did the treatment differ from the control?
    - ☐ 1.2.5 How should the intervention affect outcomes?
  - ☐ 1.3 Is the randomization process "fair"?
  - ☐ 1.4 Is there data demonstrating equivalence of the treatment and control groups?
  - ☐ 1.5 Are outcome measures valid?
  - ☐ 1.6 Is there evidence that variable attrition was not a problem?
  - ☐ 1.7 Is data reported on treatment units that did not complete the intervention?
  - ☐ 1.8 Is there data on long term outcomes?
  - ☐ 1.9 Does the study report effect size and analyses showing the result is not due to chance?
  - ☐ 1.10 How are effects on subgroups reported?
  - ☐ 1.11 Are all measured outcomes (not just the positive) reported?

SBR Checklist, continued

- ☐ 1.12 Was the intervention tested in more than one site?
- ☐ 1.13 Were the testing sites "typical"?
- ☐ 1.14 Was the intervention shown to be effective in a site similar to the one you're considering for implementation?

☐ 2.0 Is the intervention supported by "possible" evidence of effectiveness?

- ☐ 2.1 Are the treatment and comparison groups very closely matched?
- ☐ 2.2 Is it clear that the comparison group is not comprised of individuals who opted out of the treatment?
- ☐ 2.3 Were the intervention and comparison groups chosen prospectively?
- ☐ 2.4 Does the study meet the standards of a well run quasi-experimental design?
- ☐ 2.5 Studies that *fail* to meet the threshold for "possible" evidence include:
  - Pre-post studies
  - Comparison-group studies where the groups are not well matched
  - Meta-analyses where the individual studies fail to meet "possible" evidence standards

K-12 Educators' Checklist for Evaluating SBR

The January 2004 issue of District Administration Magazine contained a special supplement, *An Educator's Guide to Scientifically Based Research*, designed to assist K-12 personnel in understanding and using SBR. A portion of that supplement was devoted to a checklist that educators could use to "evaluate research evidence". As such, it shares some of the criteria in the IES checklist above, but it is less interested in technical adequacy and more interested in K-12 utility – How well does a given piece of research meet LEA needs?

K-12 Educators' SBR Checklist

- ☐ 1.0 Relevance
  - ☐ 1.1 Does the evidence provided by the researchers or developers address a question that is important to your needs?
  - ☐ 1.2 Do the developers provide evidence that the research they claim supports their product or program links to and flows from relevant theory and theory-based research?
  - ☐ 1.3 Do the research procedures, analyses, and findings support the researchers'/developers'/ claims?
- ☐ 2.0 Rigor
  - ☐ 2.1 If causal claims are made, was there an adequate comparison group?
  - ☐ 2.2 Was random assignment used?
  - ☐ 2.3 Is sufficient information provided to determine whether the research design, instruments, and procedures, are appropriate for answering the research questions posed by the researchers?
  - ☐ 2.4 Were the research instruments and procedures applied with consistency, accuracy, and for the purpose intended by the developers of the instruments and procedures?

K-12 Educators' SBR Checklist, continued

- ☐ 3.0 Systematic Approach
  - ☐ 3.1 Was the research conducted using carefully planned, logical steps?
- ☐ 4.0 Objectivity
  - ☐ 4.1 Did someone other than the publisher or developer conduct the research attesting to the product's or program's effectiveness?
- ☐ 5.0 Replicability
  - ☐ 5.1 With the information provided, could the same researchers likely repeat the study and obtain the same or highly similar results?
  - ☐ 5.2 With the information provided, could other researchers likely replicate the study's methodology and obtain the same or highly similar results?
- ☐ 6.0 Data Analyses and Interpretation
  - ☐ 6.1 Does the research evidence provided include data or data summaries?
  - ☐ 6.2 Are significance levels and effect sizes reported?
  - ☐ 6.3 Are the conclusions drawn by the researchers/developers clearly supported by the data?

Checklist for Conducting RCTs in Education

If the two previous checklists were designed primarily for users of SBR, the Coalition for Evidence-Based Policy, under contract to the Department of Education, issued a technical paper, "Key Items to Get Right When Conducting a Randomized Controlled Trial in Education" (available on-line, see Additional Resources) intended to guide researchers in the conduct of Randomized Control Trials (RCTs). The checklist below represents the major categories from that technical paper.

Checklist for RCTs in Education

- ☐ 1.0 Planning Phase
  - ☐ 1.1 List specific intervention(s) and measurable outcomes.
  - ☐ 1.2 Is random assignment of individuals or groups preferable in order to determine intervention effects?
  - ☐ 1.3 Conduct a statistical analysis to estimate minimum numbers to randomize in order to determine whether intervention has a meaningful effect.
- ☐ 2.0 Random Assignment Process
  - ☐ 2.1 Protect the integrity of the randomization process.
  - ☐ 2.2 Collect appropriate data on cohort members.
  - ☐ 2.3 Monitor control group and prevent "crossing over" to treatment.
- ☐ 3.0 Measuring Outcomes
  - ☐ 3.1 Use reliable and validated instruments.
  - ☐ 3.2 Corroborate self-reports with independent measures.
  - ☐ 3.3 Implement blind scoring.
  - ☐ 3.4 Seek outcome data for all sample members.
  - ☐ 3.5 Collect outcome data using same procedures for treatments and controls.
  - ☐ 3.6 Collect long term data to determine sustainability of effects.

Checklist for RCTs in Education, continued

- ☐ 4.0 Analysis of Results
  - ☐ 4.1 Retain all individuals in cohorts when studying results.
  - ☐ 4.2 Estimate and report on intervention's effect on all outcomes measured including statistical significance and magnitude.
  - ☐ 4.3 Conduct and report tests to demonstrate equivalence prior to intervention.
  - ☐ 4.4 Conduct and report an analysis to determine if variable attrition created group differences.

Experimental Design Criteria

Good research is more than good experimental design, but the latter is certainly a necessary condition of good scientifically based research. Trochim and Land (see Additional Resources) present the following list as criteria of good design.

- **Theory-Grounded –** Good research strategies reflect the theories which are being investigated.
- **Situational –** Good research design reflects the settings of the investigation.
- **Feasible –** Good designs can be implemented. The sequence and timing of events is carefully thought out. Potential problems are anticipated.
- **Redundant –** Good research designs have some flexibility built into them which may be a consequence of the duplication of essential design elements.
- **Efficient –** Good designs strike a balance between redundancy and the tendency to overdesign. Where reasonable, less costly strategies for ruling out potential validity threats are utilized.

Generalized Research Criteria

Finally, all research, whether it is scientific, or evidence-based, or qualitative, or descriptive, or … must meet the same general, conceptual criteria if it is to be useful research. To be sure, the standards for the application of these criteria will change across different research studies with different goals and methodologies; yet, quality research is research that meets its own standards for the following six criteria.

- **Relevance** – Research must provide information about the questions it was designed to answer.
- **Significance** – Research must provide new and important information.
- **Validity –** Research must provide a balanced picture of the real effects of the intervention being studied.
- **Reliability –** Research must contain evidence that its conclusions are not based on variations in the data due to chance or inconsistency.
- **Objectivity –** Research must not be, or appear to be, biased.
- **Timeliness –** Research must be presented in a usable form when decisions must be made.

PART II
SBR IN THE FIELD

1 The Big Picture: Research Timeline

From Theory to Application

The material presented in Part I of this *Field Guide* serves as an overview to the general theory of causal inference and, in particular, to the role it plays in scientifically based research. Part II – SBR in the Field – contains detailed tools for the application of that general theory to specific research projects. To be sure, every research project is unique and no project can be successful without a thorough understanding of its local context. Still, all successful research projects share a common core of tasks and challenges, and the tools in Part II are designed as heuristics to assist all research stakeholders in accomplishing those tasks and meeting those challenges. The first heuristic, the "map" covering the largest territory and with the greatest scale, is the research timeline.

The Research Timeline

Every task of even modest complexity (and the complexity of social science research in an educational setting is far from modest) is composed of many subtasks. Further, these many tasks need to be sequenced, not necessarily in an absolute sense but at least in a relative sense. Additionally, useful timelines are not ratcheted; that is, tasks may loop back onto one another. For example, it may be necessary to revisit the "Conceptualizing the Research Project" step if there are problems securing Institutional Research Board (IRB) approval in the "Working with Outside Agencies" step. Still, the absence of an absolute, ratcheted research timeline in no way diminishes the utility of the timeline below. In fact, the complexity of the research task only burnishes the utility of such a timeline. While it is in no way intended to be slavishly followed, it does represent an important overview of the research process. A final note about using the timeline: it assumes that a researcher has already been hired. If this is not the case, then the material in Part II, Chapter 2 of this *Field Guide* should be visited before returning to this timeline. Chapter references within this timeline direct the user to additional, more detailed topical material contained in this *Field Guide*.

Research Timeline Checklist

- ☐ 1.0 Conceptualizing the Research Project
  - ☐ 1.1 Develop Questions and Hypotheses
    - ☐ 1.1.1 Clarify the goals and objectives of the research.
    - ☐ 1.1.2 Identify major stakeholders and their informational needs.
    - ☐ 1.1.3 Describe the project that is the focus of the research.
    - ☐ 1.1.4 Generate testable research hypotheses.
  - ☐ 1.2 Identify Research Constraints
    - ☐ 1.2.1 Review previous research about this and other projects.
    - ☐ 1.2.2 Identify and address potential research barriers (see Part II, Chapter 6).
    - ☐ 1.2.3 Determine all research effort resources.
  - ☐ 1.3 Produce Research Plan
    - ☐ 1.3.1 Choose SBR research design (see Part I, Chapters 3 and 4).
    - ☐ 1.3.2 Identify additional qualitative research components (if any).
    - ☐ 1.3.3 Identify data sources and collection strategies (see 2.0 below).
    - ☐ 1.3.4 Identify data analysis techniques (see 3.0 below).
    - ☐ 1.3.5 Clarify the nature and timing of research reports (see 4.0 below).
      - Formative reporting to project
      - Annual reports to CPEC
      - Final reporting to the field and stakeholders
    - ☐ 1.3.6 Review research plan with all stakeholders.

Research Timeline Checklist, continued

- ☐ 2.0 Data Collection and Organization (see Part II, Chapter 4)
  - ☐ 2.1 Data Plan
    - ☐ 2.1.1 Identify all sources of needed information.
    - ☐ 2.1.2 Determine instruments and methods for collecting needed information.
    - ☐ 2.1.3 Specify sampling procedures (if required).
    - ☐ 2.1.4 Ensure, to the extent possible, that research questions are being addressed by multiple methods or data points.
    - ☐ 2.1.5 Develop a schedule for data collection.
    - ☐ 2.1.6 Specify responsible parties for each aspect of the data collection plan.
    - ☐ 2.1.7 Review data plan in relationship to resources and other constraints.
  - ☐ 2.2 Data Organization
    - ☐ 2.2.1 Work with stakeholders to understand the need for anonymity, confidentiality, and security.
    - ☐ 2.2.2 Develop protocols for coding, verifying, filing, retrieving, and securing collected data.
    - ☐ 2.2.3 Determine equipment needs (hardware, software) necessary to process, maintain, and control access to the data.
  - ☐ 2.3 Working with Outside Agencies
    - ☐ 2.3.1 Contact all relevant IRBs (IHEs, LEAs) to determine whether clearance is necessary and what protocols must be followed (see Part II, Chapter 3).
    - ☐ 2.3.2 Specify how controls will be identified, recruited, and data collected.

Research Timeline Checklist, continued

- ☐ 3.0 Data Analysis
  - ☐ 3.1 Determine the necessary procedures, tests, and techniques for analyzing quantitative data.
  - ☐ 3.2 Determine the necessary procedures, tests, and techniques for analyzing qualitative data.
  - ☐ 3.3 Identify standards for interpreting research findings.
  - ☐ 3.4 Synthesize quantitative and qualitative findings to provide the most accurate description of research results.
  - ☐ 3.5 State appropriate caveats for use in understanding research findings.

- ☐ 4.0 Reporting Results (see Part II, Chapter 5)
  - ☐ 4.1 Develop a calendar containing all reporting requirements.
  - ☐ 4.2 Decide on the best communication model to meet each reporting requirement.
  - ☐ 4.3 Make certain to include feedback mechanisms in interim (formative) reports.
  - ☐ 4.4 Insure that reports meet all data restrictions/permissions.
  - ☐ 4.5 As appropriate, provide stakeholders with the opportunity to review all reports to outside audiences.

The Administering Research Outline

Running parallel to the timeline for the actual conduct of the research project is a timeline of the essential tasks for administering a research project. While some of the checkpoints are similar to those for the actual conduct of the research, one should be mindful that scientifically based research (SBR) is a project-within-a-project and, therefore, presents its own administrative tasks and challenges.

Administering Research Checklist

- ☐ 1.0 Planning Phase
  - ☐ 1.1 Identify the research director (see Part II, Chapter 2).
  - ☐ 1.2 Develop the research plan and schedule.
  - ☐ 1.3 Define staff and resource requirements for meeting the plan.
  - ☐ 1.4 Create a detailed research budget.
  - ☐ 1.5 Finalize planning phase with research contract.

- ☐ 2.0 Operational Phase
  - ☐ 2.1 Maintain budgetary control by systematically comparing budget to actual expenditures.
  - ☐ 2.2 Maintain project control by systematically comparing research plan and timeline to events on the ground.
  - ☐ 2.3 Provide opportunities for reviewing and updating the research plan and contract as needed.
  - ☐ 2.4 Keep open lines of communication between all project stakeholders.

2 Working with Research Consultants

Locating Research Consultants

Given the dynamics of the California Postsecondary Education Commission's (CPEC's) Improving Teacher Quality (ITQ) competition process (the need to include a serious research plan within the larger proposal), it seems prudent for projects to secure the services of a research consultant sooner rather than later. Where does one find such individuals/organizations? What is the process through which their services are secured?

In many respects, hiring a research consultant is not much different than hiring other specially trained service providers (e.g., lawyers, accountants). The standard procedure involves generating a list of potential candidates, assessing them against a relevant list of criteria, selecting the best candidate, and negotiating a contract for services. There is, however, one enormous difference between selecting a research consultant and choosing an accountant or lawyer. The research consultant will be a colleague for the next four years; there will be a synergistic relationship between the professional development intervention and the research conducted to measure the impact of that intervention. This means, as emphasized in the section below, the number one assessment criterion must be work style – choosing a candidate with whom there is a great level of trust and comfort.

How is the list of potential researcher candidates generated? Typically, one may begin by seeking recommendations from people whose judgment they trust. At the institution of higher education (IHE), this might include faculty in Schools of Education or departments of psychology. Larger colleges and universities may have organized research units that specialize in educational and/or social science research. Local education agencies (LEAs) may already be using research consultants to help them meet and interpret state and federal testing demands. A different source of names comes from previous ITQ awardees. Whom did they hire as research consultants and how successful was the outcome? Another source of names comes from the research literature itself. What published research seems a good fit for the type of professional development intervention being proposed? Finally, there are nationally-based research firms (e.g., ETS, AIR), state-based firms (e.g., SRI, Inverness Research), and independent contractors who may be located through databases maintained by (e.g., the American Evaluation Association, Registry of Outcome Evaluators) or local researcher networks. Once a short list of potential candidates is generated, it's time for due diligence.

Assessing Research Consultants

At a minimum, due diligence should include reviewing examples of past work, talking with former clients, and a formal interview. Each of these opportunities to gather information about the researcher candidate is informed by the following set of hiring criteria:

Work Style

As much as consultant, the researcher will be a colleague for four years. ITQ projects rely on successful networks and relationships (e.g., at a minimum between IHE departments and the LEA). Therefore, leadership becomes a matter of informal influences based on a fair exchange of needs and resources. It is a collegial system based on knowledge, trust, and reciprocation. Is this the type of system that the researcher seeks? Thrives in? How do they see the role of the researcher in such a collegial atmosphere?

Explicit Experience

There is no substitute for explicit experience in conducting the type of research required by the intervention project. Has the researcher successfully completed the type of study being envisioned? Do they have experience working closely with state data sets? Have they ever worked with the LEA?

Educational/Technical

In addition to formal educational qualifications and publications, does the researcher have the requisite technical skills (e.g., special types of statistical analysis, psychometrics, instrument development) that the study requires?

Communication Skills

The researcher will be required to communicate with many different parties (e.g., the project, the funder, teachers, the field) through many different modalities (e.g., oral reports, written reports, articles). Is the researcher skilled across these modalities? Does the researcher understand the need for feedback loops in an ITQ project? How will the research component be integrated into the communication profile of the whole project?

Past Performance

What is this researcher's reputation for quality of work? Timeliness? Trustworthiness? Based on what data? Would you be happy to have his or her past reports representing your work?

Once the list of candidates has been prioritized, it's time to move to the final negotiation stage.

Research Consultant Agreement Checklist

Just as the underlying professional development intervention cannot succeed without the close cooperation of the partners (IHE and LEA), the research component cannot succeed without the close cooperation of all three partners. To be sure, there will be a formal contract between the individual researcher, or research agency, and the IHE who is the fiscal agent for the ITQ grant. Such a contract should contain all the clauses – work products, due dates, compensation – typical to a work-for-hire situation. Yet, the contract is really the formalization of a complex working agreement between three parties (IHE, LEA, researcher) who must understand each other's needs and appreciate each other's strengths. The following checklist is modified from a list Bob Stake developed for working with evaluation consultants. It is designed as a negotiating tool to help reach a firm understanding of everyone's roles and responsibilities. As such, it is a pre-contract instrument.

Research Consultant Agreement Checklist

☐ 1.0 How well do all the principal parties know each other? What more do they need to find out? How will the interests of other important parties (e.g., funding agency, unions, teachers, students) be represented?

☐ 2.0 What is the best description of the underlying professional development intervention: Whose program is it? What is its setting? Its history? Its goals and objectives? Its scope? Its audience? What changes have been made since the proposal was written? Funded?

☐ 3.0 What is the major research hypothesis to be tested? Other research questions? What research design will be used? What instruments? What data will be collected? From what source? How has the research thinking changed since the proposal was written? Funded?

Research Consultant Agreement Checklist, continued

- ☐ 4.0 What are the primary sources of data? Who is responsible for securing the data? Who will own the data? Who will own the instruments? How will the data be secured? What are the rules for access to the data?
- ☐ 5.0 Who is responsible for securing Institutional Research Board (IRB) approvals? Will the LEAs require separate IRB approvals?
- ☐ 6.0 Who are the audiences for the research findings? Will different audiences (funders, IHE, LEA, teachers, students) have different information needs? What research outputs (reports, presentations, publications) are planned? What are the internal reporting deadlines? Are the researchers free to make presentations and publish study results?
- ☐ 7.0 What resources are available for the conduct of the study? In addition to money, what resources are each of the project partners devoting to research? How will the overall project management plan affect the research? How do the project and research timelines support each other?
- ☐ 8.0 How will further arrangements be negotiated after the research begins? What will be the response to unexpected changes in the program? What misunderstandings may arise between the researchers and the underlying project staff? How will conflicts be resolved?

Four years is a long time and there will be, no doubt, unanticipated changes and challenges to both the intervention and the research project. The hiring process described here is designed to build robust relationships between projects and consultants so that the changes and challenges can be managed. No written agreement can anticipate every contingency. This hiring process focuses less on legal language and more on professional relationships – the foundation of all successful ITQ projects.

3 Working with Institutional Research Boards

Background

Institutional Review Boards (IRBs) are an integral part of biological and behavioral research on all university campuses. Yet, given their ubiquity, they are a quite recent phenomenon. In 1974, recoiling in horror at Nazi atrocities in World War II and the Tuskegee Study's gross abuses of justice, Congress formed the National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research. Five years later, the Belmont Report was issued with its three guiding principles for such research: (1) Respect for persons – individuals should be treated as autonomous human beings including respect for privacy and the granting of informed consent; (2) Beneficence – research must seek to maximize benefits and minimize possible harm; and (3) Justice – the burdens and benefits of the research must be shared equitably. These principles were systematized into law as the Code of Federal Regulations, Title 45 Public Welfare, Part 46, Protection of Human Subjects (45 CFR part 46). The Office for Human Research Protections (OHRP) in the Department of Health and Human Services oversees these regulations and every institution receiving federal research funds must comply with them. As such, it is the institutions which receive Improving Teacher Quality (ITQ) awards from the California Postsecondary Education Commission (CPEC) that are responsible for enforcing these regulations. CPEC's policy is clear: individual grantees must meet all institutional (both institution of higher education (IHE) and local education agency (LEA), where applicable) Institutional Research Board (IRB) requirements.

Campus IRBs

It is the campus IRBs that are responsible for interpreting and enforcing these regulations. As a matter of federal law, all research involving human subjects must be reviewed and approved by the local IRB prior to conducting the research. However, while the work of all IRBs must be responsive to minimum requirements set out by (45 CFR part 46), there is surprising variability in the ways that individual campus IRBs operate. The material below comprises the generic set of information that all IRBs require. The critical rule for successful IRB approval is: **Contact your campus IRB early and follow their procedures and instructions.** Most institutions have websites with online tutorials, procedures, and applications.

The IRB Process

The IRB process will differ on every campus. Some campuses will require a "pre-review" at the departmental level before the IRB will even consider an application. Many campuses now require that all IRB applicants take a short course, frequently on-line, in human subjects research. Some campuses allow for expedited review, rather than full committee review, if threats to privacy seem minimal. IRBs on large campuses may meet weekly, on smaller campuses it might be monthly. The time frame for IRB decision making can vary from weeks to months (especially if changes need to be made to the initial application). Although most IRB forms have an individual campus identity, federal regulations require that a minimum of eight categories of information be addressed in the research application.

These eight criteria aside, there is enough variety across campus IRB processes that the critical rule bears repeating: **Contact your campus IRB early and follow their procedures and instructions.**

Required IRB Application Information

1. A complete description of the proposed research.
2. An analysis comparing potential risks and anticipated benefits.
3. The process and documentation of informed consent.
4. An equitable selection of subjects and a fair distribution of benefits.
5. Safeguards that protect vulnerable populations from pressure to participate.
6. Safeguards for privacy and confidentiality.
7. A plan for handling data in a secure manner.
8. Valid research designs and methods.

Exempt Status

Under federal regulations IRBs have the authority to grant a research project exempt status, and it is also the case that much educational research may fit into this category. But be clear: First, this is not a status that the individual researcher can claim, it is a status granted by the IRB. Second, it does not mean that the research falls outside the purview of either the IRB or federal regulations. At a minimum, exempt status is only granted after IRB review. Third, if a project is granted exempt status, subjects may still need to provide informed consent for their participation in the study.

It may be worthwhile here to note the important differences between dealing with teachers and students as research subjects. Generally speaking, they have different issues with respect to data types, confidentiality of data, and the potential harm from data breaches. IRBs are likely to be more protective of students; research designs requiring extensive, specific student

data collection are not likely to receive exempt status. Therefore, whenever practicable, research designs should avoid having to secure the consent of students (typically from their parents). The best way to accomplish this is to work closely with the partner district(s) and have them supply the needed, de-identified data. Even with these qualifications, for ITQ research exempt status is certainly worth considering. As with the Research Application, campus procedures will differ. The guidelines below are general descriptors (from federal regulations) of research that may be considered exempt.

- Research conducted in established or commonly accepted educational settings, involving typical practices such as research on instructional strategies or research on the effectiveness of instructional techniques or curricula.
- Research involving the use of tests, surveys, interviews, or observations unless the information can be directly linked to subjects and any disclosure of subject responses could place the subjects at risk.
- Research involving the collection or study of existing data if these sources are publicly available or if the subjects cannot be individually identified.

Informed Consent Letter Checklist

Given that even an exempt status IRB approval may still require proper informed consent of research subjects, it is useful to understand the types of information such a letter should contain. Again, campus IRBs may provide their own form. The following checklist is simply offered as a guideline as to what the letter should contain.

Informed Consent Letter Checklist

- ☐ 1.0 Include a brief study statement including purpose, duration, and description of procedures.
- ☐ 2.0 Clearly state any potential risks as well as benefits.
- ☐ 3.0 Explain how confidentiality will be maintained.
- ☐ 4.0 Make certain subjects know participation is voluntary and they may withdraw, without penalty, at any time.
- ☐ 5.0 Include the name, institutional affiliation, address, phone, and e-mail of the Principal Investigator.
- ☐ 6.0 Describe the use of any electronic equipment (audio, video) as needed.
- ☐ 7.0 Include a line for the subject's printed name as well as signature and date line.
- ☐ 8.0 Include the following statement regarding subject's rights: "If you have questions regarding your rights as a participant in this study, contact the _____ IRB at _____.

Research Protocol Checklist

Finally, while it's true that IRB approval represents an additional hurdle in the research process, it also serves as an opportunity to organize research thoughts and procedures. The following checklist may prove to be overkill for an exempt status IRB approval, but it will surely serve to improve the quality of the research, may help in recruiting subjects and lessening attrition problems, and help to meet the high standards of other research checklists included in this volume.

Research Protocol Checklist

- ☐ 1.0 Explain clearly the purpose of the research and describe what it involves.
- ☐ 2.0 Supply details of the data to be collected and procedures for collection.
- ☐ 3.0 Supply the number, description and rationale for the involvement of all participants and highlight any special populations.
- ☐ 4.0 Describe the institutional affiliation of all subjects and any procedures required by the institution for the subjects to participate.
- ☐ 5.0 Describe the procedures to select and recruit participants.
- ☐ 6.0 Describe how subjects will be contacted and informed consent obtained.
- ☐ 7.0 Describe how participants will be informed of their right to withdraw from the study without penalty.
- ☐ 8.0 Explain any potential risks of the research and how they will be mitigated.
- ☐ 9.0 State clearly potential benefits for the subjects and/or public.
- ☐ 10.0 State clearly how confidentiality of subjects will be protected.
- ☐ 11.0 Explain how data will be secured and who will have access to it.
- ☐ 12.0 If subjects are compensated, explain clearly the nature and method of compensation.

4 Working With Local Education Agencies (LEAs)

The LEA – From the Bottom Up

Although it seems paradoxical, educational researchers may find that working with local education agencies (LEAs) is both an exceptionally broad and an extremely narrow experience. It is broad because LEAs, which vary from Los Angeles Unified to single school districts, are composed of many stakeholders (e.g., district staff, principals, teachers, unions, students, parents, community members) and almost all SBR research projects require the cooperation of, at a minimum, several of these groups. It is narrow because the personal dynamics between the researcher and each of these groups tends to be characterized by guarded caution. Putting it bluntly, Institutions of Higher Education (IHEs), and researchers in particular, have a less than stellar reputation within the LEA community. There are several reasons for this: Research protocols often strike LEAs as unreasonable and impractical, requiring too much teacher and district time and disrupting the school schedule. Researchers always seem happy to publish papers and present at conferences, but they fail to make the time to explain the utility of their results to the LEA. The researcher-LEA relationship is often viewed as going strictly one way – the researcher demands and the LEA is supposed to comply. In sum, research rarely meets the needs of the LEA and the LEA is made to feel less a full partner and more a mere conscript. One means of improving this reputation is the formation of a research advisory board (briefly discussed in Part II, Chapter 6). A more systematic approach to the development of a working LEA-research partnership is the collaborative sales model explained in this chapter's final section.

Regardless of how well the LEA-researcher-institutional relationship is working, the actual work of securing and manipulating district data, or the work of teachers filling out questionnaires, will be performed by individuals. One veteran teacher, now a university-based Professional Development (PD) provider, cautions that researchers will find additional resistance at this level: LEA folks are overworked and underappreciated and will view requests for additional work with disdain. In practically every case, the person who approves the work is not the person who will need to carry it out. In practically every case, the person in the LEA who might benefit from the work is not the person who will do it. In almost no case will the LEA staff share your enthusiasm for the research being conducted. While there may be little the researcher can do to encourage enthusiasm, careful communication (see Part II, Chapter 6) and respect can go a long way towards efficiency and effectiveness.

Checklist for Working with LEA Personnel

☐ 1.0 With the approval of the supervisor, arrange for a mutually convenient meeting with the LEA staff who will be working with you.

☐ 2.0 Briefly explain the thrust of the research, what you hope to learn, and how the LEA will benefit.

☐ 3.0 Be sure that the work you're requesting is framed within the context of LEA operations (i.e., avoid acronyms and shorthand phrases) and that it is consistent with LEA data deadlines.

☐ 4.0 Be clear and specific about the requested work (e.g., names of tests, dates of tests, subtests, grade levels, needed demographic categories). If multiple year data sets are being requested, or any data that requires merging, volunteer to perform the merging. District personnel are more likely to send out separate data files quickly and they will appreciate your involvement. Most importantly, leave a printed copy of your agreement with deliverables and due dates.

☐ 5.0 Because many ITQ projects work with data sets generated from statewide instruments, researchers must be knowledgeable about both the instruments and the way that data is reported to the LEA. For example, CST scoring files arrive in late August and LEAs are busy working with them through September. Best to ask for CST data later in October. Because student data has to be anonymized, districts need to create "research IDs" for students and teachers. Researchers need to understand district protocols for access and use of this data.

☐ 6.0 Be certain that a feedback loop is established before leaving the initial meeting. Make sure that all parties understand their responsibilities and obligations, exchange contact information, and establish a call back date to check on progress.

☐ 7.0 If the staff at the meeting inform you that they will require the support of other LEA staff, ascertain the process needed to involve those individuals and follow through yourself.

Checklist for Working with LEA Personnel, continued

- ☐ 8.0 Remember, an ITQ SBR project is a four year process. Things will go wrong, LEA personnel will change. Avoid casting blame and focus instead on problem solving and overcoming barriers.
- ☐ 9.0 Finally, remember that as an outside researcher, you will be viewed as a "visiting fireman" and that the work you request may overload some LEA staff person. Always be cordial, professional, and aware of LEA staff constraints (e.g., time, technical competence).

The LEA – From the Top Down

While the conditions of an ITQ grant insure that cold calls will not be necessary to locate an LEA willing to provide teachers for the treatment cohort, the same can't be said for LEAs that might furnish control groups. Furthermore, there is a wide gap in cooperation between an LEA that simply answers a call and an LEA that acts as a true research partner. What are the criteria that LEAs believe research studies should meet before they agree to sign on? The checklist below is a compilation of information gleaned from the external research applications of several large districts. Even if one is not required to submit such an application, these criteria are good guides for achieving maximum cooperation.

Checklist for Working with LEAs

- ☐ 1.0 Compliance: Has the researcher met all district requirements and submitted all required documentation?
- ☐ 2.0 Significance: Does the research offer potential benefits to our LEA?
- ☐ 3.0 Alignment: Is the research project aligned with LEA initiatives and goals?

Checklist for Working with LEAs, continued

☐ 4. 0 Research Design: Is the design sound? Is there a testable hypothesis with all basic concepts and variables clearly defined? Are the instruments reliable and valid? What data are being requested? Are the controls within/outside this LEA? Will anonymity/confidentiality be preserved? Are all research practices ethical?

☐ 5.0 IRB: Has the research passed IHE IRB approval? If required, has it passed LEA IRB approval?

☐ 6.0 Timeline: Is the proposed schedule feasible? Could the research be potentially disruptive to student learning, teacher classroom responsibilities, or LEA work?

☐ 7.0 Costs: Are there any costs to the LEA?

☐ 8.0 Intended Use: How will the data collected be used by the researcher? Is there a provision for communicating results directly to the LEA?

☐ 9.0 Special Concerns: Some LEAs provide a special warning to researchers that certain types of research procedures present special problems and that such requests are not likely to be approved. Below are some mentioned concerns.

- Research that places an unnecessary burden on participating school(s), programs, students, teachers, or the system
- Sensitive topics such as sex, drugs, gambling
- Research requiring access to archival data
- Research that overlaps the LEA's own research program
- Research requiring physical measurements
- Research requiring contact after school hours
- Assessment schedules that conflict with the LEA's

The LEA-Researcher Partnership

The previous two sections can be summarized by the three Cs: Collegiality, Communication, and Collaboration. In fact, these are important traits for all aspects of a successful ITQ project. But for true effectiveness, they need to be built in from the beginning and not just checked off a list when the opportunity arises. How can a true LEA-researcher partnership be established? One model that has been successful, most notably in the National Science Foundation's Industry/University Cooperative Research Program, is known as collaborative sales (don't be put off by "sales"; the whole idea is to get past the notion that sales means convincing a customer to purchase something whether they need it or not).

Successful SBR researchers must be able to recruit and work closely with LEAs, but that does not mean that they need to face the superintendent with "nothing but a shoe shine and a smile." When Arthur Miller penned that phrase for *Death of a Salesman,* sales people were viewed either as mere order takers or as high pressure hucksters. Today, there is a third model of the salesman's craft – consultative, or collaborative selling – which, by emphasizing client's needs and partnerships, is an excellent fit for an SBR researcher's *modus operandi*. In collaborative selling, solutions are never dictated. Rather, partnerships are formed in which prospects play an active role in the search for best solutions. This joint approach to problem solving helps to ensure that prospects are committed to proposed solutions and reduces the stress of closing by allowing the prospect to commit to the deal when all their buying criteria are met. Collaborative selling puts a decidedly interactive and cooperative face on the five steps in the process:

Contact

The goal of this stage is building a foundation for a cooperative working relationship. Emphasis should be put on exploring needs and opportunities, and explaining advantages, rather than on any preset presentation.

Explore

The key strategy at this stage is two-way communication. Potential LEAs are encouraged to express their needs and how they hope to benefit from the partnership; SBR researchers learn as much as they can about the LEA and begin to suggest ways that the partnership can help meet the expressed needs. In the research context, this stage has two main foci: first, exchanging enough information and answering enough questions so that the LEA is ready to commit to an onsite visit. Second, identifying a "champion" at the LEA who is dedicated to securing the partnership and identifying the LEA decisionmaker who is responsible for authorizing the partnership. Few partnerships are secured without a visit; none are secured without the approval of the decisionmaker.

Collaborate

This is the stage at which "all the cards are on the table". The LEA has a complete picture of the research being proposed as well as its role and benefits, and the SBR researcher fully understands their needs and reservations. What remains is working together, collaborating to produce singular solutions that benefit both partners. A successfully operating partnership is constantly engaged in a dialogue designed to refine the partnership so that it meets both partner's needs. At this point, the SBR researcher and "champion" may need to collaborate in order to determine how the proposed research may help to solve the problems of the decisionmaker (e.g., how the proposed research may assist the LEA's ongoing research efforts).

Confirm

Ideally, the collaborative sales approach is an "agree-as-you-go" process. For example, benefits are not merely explained; rather, they are explained until the potential client agrees that they are valuable for his or her organization. Handling objections is a vital part of the communication process, not a chore held in abeyance until the very end. If all goes well, confirming the partnership agreement is a logical conclusion to the collaborative process. There is no need for a hard close because LEAs will commit to the agreement when all their questions are answered and their needs addressed.

Assure

The level of communication after agreeing to the partnership must intensify, not diminish. LEAs must be assured that need satisfaction and partnering are not just part of the sales pitch but part of the SBR researcher's core operating philosophy. LEAs are encouraged to track their results and communicate any concerns to the site.

In summary, SBR researchers need to understand that successfully working with LEAs is tantamount to creating a partner. One of the best methods to accomplish this goal is to adopt a customer-oriented approach to selling. Such an approach is built on three assumptions: (1) LEAs have latent needs that constitute opportunities for the researcher; (2) LEAs appreciate good suggestions that help them focus and meet their needs; and (3) LEAs will be more responsive to external parties who are interested in long-term relationships. Within this basic framework, LEA-researcher partnerships proceed through the above five steps.

5 Communicating Research Results

Education is a "Social" Science

If the unexamined experiment is not worth conducting, then the unreported experiment is hardly worth conducting. First, because the results of the research are owed to all those who made the research possible: funding agencies, professional development partners, teachers and students who were research subjects, and other stakeholders. Second, because education is an applied social science, its growth and improvement require the free exchange of information. More specifically, its development as an applied social science requires a shared, and open, knowledge base founded on valid and reliable research results. Only through communicating such results to the field can the education profession learn how to do its job more effectively and efficiently.

Communication, in all its varied aspects, is a fundamental aspect of any Improving Teacher Quality (ITQ) project. The Field Lessons chapter (see Part II, Chapter 6) draws particular attention to problems that may occur when communication goes awry, and makes specific recommendations for improving the entire communication process. All of those recommendations apply here as well. Yet, disseminating research results, typically in a formal report but also through oral presentations, published articles, or other media, presents its own challenges and opportunities. The material in this chapter is designed to help minimize the former and maximize the latter.

The Dissemination Plan

Educational dissemination – the transfer of information about an instructional product or process, or research results – is something everyone favors. Unfortunately, it is something that is rarely done well. The major problem is that planning for successful dissemination is a process that must be started at the beginning of the research study, not left until that study has been completed. The following checklist is a useful guide to that dissemination process.

Dissemination Process Checklist

☐ 1.0 Clearly state the goals of the dissemination effort (e.g., useful information will be disseminated to all stakeholders).

☐ 2.0 Link goals to measurable objectives (e.g., the funding agency will receive a formal final report, the local education agency (LEA) will receive an interactive PowerPoint presentation).

☐ 3.0 Describe the scope and characteristics of the potential audiences that your dissemination activities are designed to reach for each of the objectives.

☐ 4.0 Identify the basic elements of the research study you must disseminate to each of the potential audiences.

☐ 5.0 Identify the primary source, or media type, that each audience group relies on for the transfer of information. Consider ways to partner with these sources in your dissemination efforts.

☐ 6.0 Describe the media through which the content of your message can be best delivered to your audience. Describe the capabilities and resources that potential audiences will need to access the content for each media strategy.

☐ 7.0 Describe how you will assess the success of your dissemination strategy. Keep a log of who is accessing, citing, and using your report and results.

☐ 8.0 Describe how you will promote access to your results and how you will archive information that may be requested at a later date. Consider that most people will use your results on a "need to know" basis and not necessarily when you have completed your research project. For example, ITQ projects are strongly encouraged to submit their results to the What Works Clearinghouse.

☐ 9.0 Identify strategies for promoting awareness of the availability of your results and the availability of alternative formats (e.g., written reports as well as Power Point presentations).

☐ 10.0 Identify potential barriers that may interfere with audiences' access or utilization of your results and develop actions to reduce these barriers.

Pre-Report Communication Compass

It has often been said that if you don't know where you're headed, it doesn't matter in which direction you travel. Similarly, communication is a journey of shared meaning, and successful communication requires a clear sense of logical and semantic direction. The set of eight questions below are designed as a communication compass – a tool to help you both set the initial communication course and to keep on it. As such, these questions should be answered, in some detail, before the formal report writing task begins, and they should be consulted as the writing progresses.

Pre-Report Communication Checklist

- ☐ 1.0 What is the major purpose of the research? What major questions and hypotheses does it seek to answer?
- ☐ 2. 0 What are the most important features of the research? Methodology? Subjects?
- ☐ 3.0 What are the most important results? How can they be best represented?
- ☐ 4.0 How does this research fit into the knowledge base? What other work should be cited?
- ☐ 5.0 What are the consequences of the results? For the implementation? For the field?
- ☐ 6.0 Who contributed to the research? Do they have a role to play in communicating the results?
- ☐ 7.0 Who are the audiences (stakeholders) for this report? Will different stakeholders require different report treatments? What background knowledge do different audiences bring to the report?
- ☐ 8.0 How is feedback being assured in the report process? Who will review drafts?

Research Report Template

Many organizations (e.g., APA, MSN) have several research report templates and dozens of academic websites contain guides for writing research reports. The template below is an outline modification of materials developed by the Department of Education's Institute of Education Sciences. It was selected because it is specific to "reporting the results of an evaluation of an educational program or practice 'intervention'." The website for the complete Institution of Education Sciences (IES) guide is found in the Additional Resources section. Yet, regardless of the actual format used, there are steps that the report writer (communicator) may take to improve the quality of the communication. These "report improvement" suggestions follow this template checklist.

Research Report Template Checklist

- ☐ 1.0 Title Page
  - ☐ 1.1 Clear and concise title to facilitate indexing
  - ☐ 1.2 Names and affiliations of author(s)
  - ☐ 1.3 Date of research and report
  - ☐ 1.4 Name of client and/or funder(s)
- ☐ 2. 0 Executive Summary
  - ☐ 2.1 Description of project
  - ☐ 2.2 Research questions and hypotheses
  - ☐ 2.3 Brief description of experimental method and analytical strategy
  - ☐ 2.4 Summary of main findings
  - ☐ 2.5 Recommendations, if appropriate
- ☐ 3. 0 Table of Contents and prefatory material
  - ☐ 3.1 Table of contents containing all first and second level headers in the report
  - ☐ 3.2 Lists of tables, figures and appendices
  - ☐ 3.3 List of acronyms and abbreviations
  - ☐ 3.4 Acknowledgement section mentioning all who contributed to the work (e.g., sponsors, research subjects, research assistants, reviewers, other stakeholders)
- ☐ 4. 0 Introduction and Background
  - ☐ 4.1 Purpose of the research
  - ☐ 4.2 Description of the project undergoing research
  - ☐ 4.3 Identification of the project's target population and relevant audiences and stakeholders for the research
  - ☐ 4.4 Brief review of related research, if appropriate

Research Report Template Checklist, continued

- ☐ 5.0 Methodology
  - ☐ 5.1 Describe the study setting
  - ☐ 5.2 Describe the study sample and recruitment process
  - ☐ 5.3. Describe the intervention and how it differed from what the control/comparison group received
  - ☐ 5.4 Describe how the study sample was allocated to intervention and control/comparison groups
  - ☐ 5.5 Describe how and when outcomes were measured including a description of instrumentation
  - ☐ 5.6 Describe statistical/analytical methods
  - ☐ 5.7 Describe limitations of data and analytical methods

- ☐ 6.0 Results
  - ☐ 6.1 Indicators of successful/unsuccessful study process (e.g., robust data collection/high subject drop out rate)
  - ☐ 6.2 Descriptive data describing index of implementation (i.e., extent to which intervention was implemented)
  - ☐ 6.3 Estimates of the intervention's effect on all outcomes measured
  - ☐ 6.4 Estimates of effects on subgroups within the study sample

- ☐ 7.0 Discussion
  - ☐ 7.1 Interpretation: How best to describe the effectiveness of the intervention
  - ☐ 7.2 Extent to which the results may be generalizable to others who might receive the intervention
  - ☐ 7.3 Significance of the results to educators, policymakers, and others
  - ☐ 7.4 Factors that may account for the intervention's effect(s) or lack thereof
  - ☐ 7.5 Study limitations (e.g., small sample size, variable attrition)

Report Improvement Suggestions

- Since many (most?) readers will not look past the Executive Summary (2.0 above), that is the first, and perhaps last, opportunity to convey the maximum amount of information to the greatest audience. This may be done by including a one-page, bulleted summary of important findings and then referencing supporting material in the body of the report. Similarly, a one-page graphic display of key research data and findings (keyed into the report) may provide more information than several pages of text.
- 3.1 above advises that the table of contents contain all first and second level headers. This advice is both logical and lexicographical: logical, because it suggests that reports be structured hierarchically, from most important to less critical information within each report section; and lexicographical because it suggests that the structure of the report be highlighted through the use of various text devices (e.g., font sizes, font attributes).
- Appendices are mentioned in 3.2. Many scientifically based research (SBR) projects involve a fair amount of sophisticated statistical and analytical methodology (e.g., sampling protocols, validating instruments, hierarchical linear modeling) that most of the report's readership will not be interested in. Best to simply mention that such and such technique was used in the body of the report and then use technical appendices to explain the details and implementation of those techniques for readers who are interested.
- Acronyms and abbreviations are mentioned in 3.3. Education is nothing if not "jargon rich". Rather than spelling them out each and every time they appear (e.g., SBR for scientifically based research), better to include a glossary sheet of acronyms, abbreviations, and special terms used throughout the report.
- Be modest. 5.7 suggests that the report contain a section on limitations of data and analytical techniques. 7.5 suggests a discussion of study limitations. These sections must be consistent with the results reported in section 6.0. Overstating results in the face of technical limitations is one way to lose credibility quickly.
- Section 7.0 is an opportunity to flex one's writing muscles and extend this single research project to a broader discussion of educational research and policy. It is also an opportunity to embed this research to the growing body of SBR through citations to similar work (e.g., in the What Works Clearinghouse). But be cautious: The farther the discussion wanders from the research results, the less persuasive the argument and recommendation.

6 Field Lessons

Avoiding Problems

The conduct of social science research in an educational setting is so complex that even the most careful researcher will not escape problem free. Rather than perfection, the researcher's twin goals must be to 1) minimize problems, especially ones of great consequence, and 2) perceive and correct problems in a timely manner so as to avoid deeper trouble. The best way to accomplish these goals is to be forewarned about the ways that research projects can go awry, and to constantly check the progress of the research against such threats. The following checklist list, a modification of Dan Stufflebeam's Evaluation Plans and Operations Checklist, categorizes research problems into six major areas. It is not a list to be merely checked off at a research project's conclusion; it is a tool to be consulted frequently during the entire project. As such, it is an important adjunct to the Research Timeline in Part II, Chapter 1.

Avoiding Research Problems Checklist

- ☐ 1.0 Conceptual Problems: All stakeholders must establish a shared, collegial understanding of the guiding principles supporting the research.
  - ☐ 1.1 Are all stakeholders agreed on the purposes of the research?
  - ☐ 1.2 Are all stakeholders agreed on the questions/hypotheses the research will answer?
  - ☐ 1.3 Are all stakeholders agreed on the research's overall methodology?
  - ☐ 1.4 Are all stakeholders agreed on the types of data that will be collected?
  - ☐ 1.5 Are all stakeholders agreed on who all the stakeholders are?
  - ☐ 1.6 Are all stakeholders agreed on the research team?

Avoiding Research Problems Checklist, continued

- ☐ 2.0 Socio-political Problems: Project and research partners must identify and meet the concerns of all stakeholders.
  - ☐ 2.1 Is the involvement of all necessary parties assured?
  - ☐ 2.2 Is there an internal communication system to keep all project parties informed?
  - ☐ 2.3 Is there an external communication system to keep all stakeholders informed?
  - ☐ 2.4 Does the research team have the confidence of all stakeholders?
  - ☐ 2.5 Are data and confidentiality of sources secure?
- ☐ 3.0 Contractual/Legal Problems: Researcher and project partners must establish clear working agreements to ensure efficient collaboration and protect involved parties' rights.
  - ☐ 3.1 Are all project roles clearly specified?
  - ☐ 3.2 Are all research products specified?
  - ☐ 3.3 Is there a specific timeline for all research tasks and deliverables?
  - ☐ 3.4 Are there rules for project editing of research reports?
  - ☐ 3.5 Who owns the instruments and data?
  - ☐ 3.6 How will access to data be controlled?
  - ☐ 3.7 Are all financial arrangements clear?
  - ☐ 3.8 How will changes to the original scope of work be made?
- ☐ 4.0 Technical Problems: Researchers must convert the general proposal language into a detailed research plan.
  - ☐ 4.1 Are the research hypotheses and questions clearly specified?
  - ☐ 4.2 Are all data types/sources clearly specified?
  - ☐ 4.3 Are all data gathering strategies specified?
  - ☐ 4.4 Are types/sources of instruments made clear?
  - ☐ 4.5 Are there plans for data storage and retrieval?

Avoiding Research Problems Checklist, continued

- ☐ 4.6 Is the data analysis plan technically and conceptually adequate?
- ☐ 4.7 Are reports designed to meet all audience needs?
- ☐ 4.8 How does the research plan ensure that findings will be reliable and valid?

☐ 5.0 Management Problems: Researcher must direct and control the research effort within the context of the larger, professional development project.

- ☐ 5.1 Are the relationships between all project parties and the research team clearly specified?
- ☐ 5.2 Is the research staffing plan specified?
- ☐ 5.3 Is the research timeframe consistent with the timeframe of the implementation project?
- ☐ 5.4 Is the research budget specific and adequate?
- ☐ 5.5 What information will be communicated to which audiences and on what schedule?

☐ 6.0 Utility Problems: Researcher must plan and execute steps to promote the constructive use of the research findings.

- ☐ 6.1 Does the researcher understand the information needs of all stakeholders?
- ☐ 6.2 Is there agreement among all stakeholders on the general research approach?
- ☐ 6.3 Do dissemination strategies meet the needs of various stakeholders?
- ☐ 6.4 Are there specific plans to disseminate the results to the field?
- ☐ 6.5 Will the research meet the major criteria of utility: relevance, significance, validity, reliability, objectivity, and timeliness?

Listening to the Field

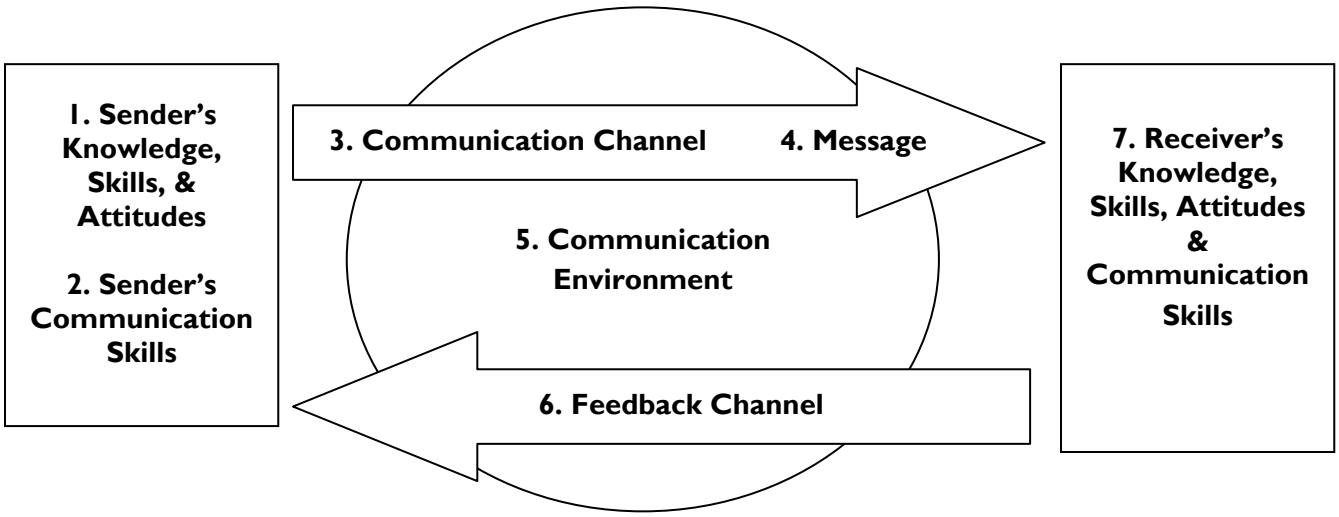
Professor Stufflebeam's checklist is the result of his lifetime of experience as a professional evaluator. While it certainly reflects the input of many colleagues, it is primarily his work. A different strategy in attempting to uncover how research projects may go awry is to ask the researchers themselves what problems have arisen in the conduct of their research. This is exactly what the California Postsecondary Education Commission (CPEC) did. In the fall of 2007, we convened a mini-conference of the ten research projects that had been funded in 2005 and 2006. In addition to discussing their progress, we asked them what lessons they had learned. Here's what they told us: Communication, both internal and external to the project, can be a potential source of serious problems. Approximately one-half of the issues discussed at the conference had communication as a driving force. To cite just two examples: One researcher thought she had a clear understanding with the district as to the student test data required. Unfortunately, she received data completely inappropriate to the project's needs. She concluded, "Obtaining data from district testing departments can be an exercise in miscommunication and the lack of a common vocabulary." Communication between the professional developers and researchers can be equally problematic. A second project substantially revised its Professional Development (PD) plan to meet conditions on the ground but never communicated this to the research team, or the need to develop a new research plan consistent with the altered implementation. Two years from inception, this project was still struggling to get its research component off the ground.

Much of this *Field Guide* is devoted to the proper types of questions to ask in the conduct of a successful research project. Yet, asking these questions (focusing solely on the message) is not enough. Even the right questions must be embedded in a communication context that emphasizes quality information flow. Because of the central role communication plays in all aspects of a successful Improving Teacher Quality (ITQ) project, the section below outlines some of the key findings of communications research that may be applied to improving the quality of ITQ messaging and information.

Communication is Key

Unlike research design, or statistics, or professional development, which we think of as skills requiring special training, few of us ever rigorously study communication. It is simply a given; after all, we've been communicating for as long as we can remember. Yet, if we are honest, we realize that many of our problems (with superiors, co-workers, family) arise from a failure of communication. Sometimes we fail to make ourselves clear, sometimes we fail to properly interpret another's message, sometimes the critical message is drowned out by peripheral noise. However, the complexity of social science research projects, and the premium they place on the flow of many types of information, requires that we attend to our communication skills or risk sub-par results. The following material is intended to help each of us become better communicators.

The Seven Components of Communication



The following seven clusters of suggestions that correspond to the seven numbered steps in the above diagram are designed to help one maximize his or her communication opportunities. To be sure, no one has the time to review each of the steps every time they need to communicate. Still, awareness of their existence, as well as explicit referral during critical communication chores, should help to avoid misunderstandings and maximize the flow of important information.

1 | Sender's Knowledge, Skills, and Attitudes

What is the major purpose of the message? Why is communication taking place? Is it simply to convey information (e.g., the date of an event), persuade someone to take some action (e.g., make data available), begin a dialogue (e.g., how the research project can help the district). Further, what are ramifications of the ideas being communicated? What do they imply concerning both future communication and action?

2 & 3 | The Sender's Communication Skills, & the Communication Channel

Select the best medium to express your message and the one you feel most comfortable communicating through. There are at least sixteen recognizable methods for communication (and technology keeps upping the ante). Each has strengths and weaknesses which need to be matched to the communication context and the communicator's skill set. For example, inter-personal channels (face-to-face, telephone, teleconference) are high on potential for feedback but it is difficult to express complex ideas and they leave no record. Written communication channels (memo, mail, fax) can transmit complex

information but may discourage feedback or seem impersonal. Presentational channels (meetings, video conferencing) can allow a great breadth of information to be communicated and group interaction to take place, but they are expensive and make confidentiality difficult. Each of these media can play a role in a communication strategy, but each must be applied to the appropriate information need within the proper information context.

4 | Message

Produce the simplest, clearest message possible. For many folks, this is all that communication means – "I said what I meant; if you failed to get it, that's not my problem." But communication doesn't take place in a vacuum, it's a social act engaged in for a purpose – to transmit information, to collect information, to persuade, to dialogue – so a failure to communicate is the problem of the communicator. Still, a clear, concise message is a necessary, if not sufficient, condition of successful communication: make sure the message and its intent are consistent. Reduce the number of people through whom a message must travel. Preview and review the material in the message. Report details in a specified order (e.g., chronological, cause-effect). Highlight important information. Consider the consequences of your communication.

5 | Communication Environment

Inventory the communication environment. There is no sense in preparing an excellent message if it can't get through to the intended receiver. Successful communicators need to assess the potential blocks (e.g., time, subordinate screening, delays in feedback) in the environment and devise strategies for avoiding them.

6 | Feedback Channel

Leave a communication channel open for feedback. This may be the most important and overlooked aspect of effective communication. Good communicators are good listeners. At its simplest level, feedback lets you know that your message was received. Feedback shows the receiver that you are open to ideas, and it gives you an opportunity to learn how the ideas in your message might be improved.

7 | Receiver's Knowledge, Skills, Attitudes and Communication Skills

Understand the receiver's ability to receive your message. Effective expression is not equal to effective communication. Successful communicators understand that the receivers of their messages are really interpreters of their messages. The completion of a successful communication act is mutual understanding.

The Field Speaks

While (mis)communication plays some role in almost all research problems, ITQ researchers did encounter 19 other "lessons learned" which can be placed into three categories: experimental design; project management; and testing and instrumentation.

Experimental Design

- Establishing control schools can be problematic, but it is not appropriate to select such a school based solely on internet data since this does not permit intermediate analysis nor correlations to program variables.
- Obtaining data from teachers in the control group may be stymied by the control teachers failing to perceive a personal benefit.
- Establishing the comparability between any two groups is always an imprecise endeavor.
- Teacher attrition is always a possibility and must be guarded against. If it occurs, it must be identified early and addressed.
- Teachers move, get different class assignments, or their plans change. This can disrupt experimental designs.

Because of the real difficulties in selecting proper matching controls, almost all researchers advocate randomized control trials (see Part I, Chapter 3) as the first choice of experimental design. While there seems to be a general argument that Randomized Control Trials (RCTs) are inappropriate, or too difficult, in an educational setting, this conclusion is far from proven. Although this *Field Guide* is not the place to engage is such a full- fledged discussion, it is appropriate to mention two aspects of RCT that may allow it to be more easily applied in educational research. First, random assignment is far more sophisticated than a simple flip of the coin. Shadish, Cook, & Campbell (SCC) discuss seven types of random assignment including restricted random assignment to force unequal sample sizes – a technique that allows for a greater number of control than treatment participants and, when

used properly, can increase the probability of correctly rejecting a false null hypothesis.

Typically, however, it is not the randomization design itself that is being argued against, but the problems inherent in implementing such a design. Boruch and Wothke (see Additional Resources) face this head-on with a detailed discussion of seven lessons learned about implementing random assignment (the list below is from a summary table in SCC).

1. Plan in advance how to explain the nature and purpose of randomization to those who will be affected, how to respond to various arguments about why randomization could or should not be done, and how to provide incentives for doing randomization.
2. Pilot test the randomization procedure to discover problems that can be remedied with further planning.
3. Develop clear procedures for implementing, controlling, and monitoring the randomization process through the entire experiment.
4. Have meetings at which to negotiate the randomization procedures with those who will be affected by them.
5. Develop fallback options that can be used to bolster estimates of program effects in the event that randomization fails.
6. Take advantage of naturally occurring opportunities that facilitate the conduct of randomization.
7. Carefully examine the match between the proposed design and those factors that will make randomization more likely to be successful in the particular context of the experiment.

Such advice might well apply to the implementation of all research designs. Boiled down to its essentials, it advocates that researchers plan, monitor, and communicate.

Project Management

- Union contracts can pose constraints on meeting times and scheduling as well as data gathering on teacher characteristics.
- There can be changes in institutional programs that lead to recruitment difficulties.
- Teachers may need to be incentivized in order for them to follow through on their project responsibilities.

- Changing institutional conditions in personnel and program can impact a project's proposed treatment and timeline.
- Year round schools present a unique set of conditions for the implementation of a project.

Much of the discussion about project management problems can be summarized in a single sentence: District/teacher priorities change over time and move away from those of the ITQ project. While this "threat from change" has always existed, it may be exacerbated by an ITQ research component that does not have complete LEA buy-in. Fortunately, there are several time-tested strategies that mitigate this threat. Most importantly, successful projects must create a genuine collaboration among all project partners. This does not mean that the project workload must be equally shared. It does mean that there be an equally shared respect. All parties must value what they receive from the others, and they must work to insure that each party is rewarded for its unique contribution. For example, the research team will be responsible for conducting most of the research but the district must feel they are a respected partner – that their needs are being met and their input respected (Part II, Chapter 4).

Given the "threat from change", successful projects must adopt strategies to insure strong, continuous leadership. One of the best ways to accomplish this is an active advisory board. Such a board, meeting at least quarterly, provides continuity, guarantees that stakeholder voices are heard, provides for democratic operation, and spreads project ownership and leadership. Many different board models are possible: (1) a broad- based community board representing all stakeholders, (2) an inter-project board with representatives from the many interventions that may be taking place at the school/district, (3) a teacher-participant board composed mainly of project teachers, and (4) a research board with representatives from the district, union, PD project, and/or other research efforts in the district. Whichever model is chosen, advisory boards encourage robust leadership and stakeholder buy-in. As such, they provide a way to "work through change".

Collaborative strategies and advisory boards are manifestations of a collegial leadership style. In such a model, expertise replaces authority, and honesty and trust are key qualities. The formal lines of authority disappear and leadership becomes a matter of informal influences based on a fair exchange of needs and resources. Leadership is manifested by listening, fairness, knowledge, and the ability to help others have their needs fulfilled. Within such a leadership model, there is no sharp division between the ITQ project and the district. "Change" is the result of a shared decision that all parties negotiate and can succeed within.

Testing & Instrumentation

- There may be a tendency for teachers to overstate their competence on pretest measures.
- Self-reporting can yield unrealistically high levels of treatment effects.
- These projects are not designed as instrument development efforts, but in many cases there are no off-the-shelf instruments and the difficult task of developing instruments must be addressed.
- Use of "off the shelf" instruments can produce a misalignment between data gathered and program variables.
- The need to develop instruments can lead to inadequate results.
- Testing students with customized instruments after April is problematic due to year-end testing activities and fatigue.
- Interpreting data from California Standardized Test (CST) scores across several years is not psychometrically sound.
- Student commitment to testing can be lacking.
- It is difficult to get clean data from districts.

These nine issues boil down to two clusters of problems: first, the use of measures to support causal inferences; second, the tension between available measures and the perceived need to develop new measures (i.e., instruments). Regarding the first, prudent practice relies on three principles: (1) Always use the "strongest" measure available for a given measurement task. For example, asking for a self-report about content knowledge competence would be weak because there are numerous valid and reliable tests that would provide a stronger measure. (2) Increase the reliability of the study's measures by increasing the number of measures, or improving the quality of measures, or using advanced techniques like latent variable modeling of several observed measures to determine true scores from error variance. (3) Conceptualize the entire causal model with techniques like path diagrams (see SCC) in order to better understand all the causal factors and measures that may be at work in a given project. Like so much else in the conduct of SBR, being forewarned is being forearmed. There are a vast array of techniques to deal with measurement problems; the prudent researcher needs to flag potential problems as soon as possible in order to employ the best solution(s).

Regarding the second cluster of problems, the ITQ program never intended to support instrument development. First, because of the large, potential expense of such an enterprise; second, because the skill set needed for such development does not necessarily overlap with the skill set needed to run and research a successful ITQ grant. Still, several projects have undertaken this challenge. The verdict is not yet in

on both the quality of the instruments and the quality of the projects they are intended to measure. Even so, it is ITQ policy to encourage grantees to use existing tests, instruments, and measures. This begins with awareness. For example, the California Department of Education (CDE) web site (see Additional Resources) contains a chart of the California Assessment System listing thirteen different instruments in statewide use. It may be possible to develop credible measures using different parts or scales from these instruments. It is also often the case that school districts are engaged in local testing efforts. For example, one of the first SBR projects took place in a large high school district that was part of the Mathematics Assessment Collaborative (MAC). This enabled them to use the well-established MAC instruments to measure student learning. Of course, working closely with the LEA is one goal of ITQ, and Part II, Chapter 4 contains many suggestions for how to do this within the context of SBR.

Additional Resources

The resources annotated here are not intended as a comprehensive bibliography of all the materials consulted in the production of this *Field Guide*. Rather, they represent the author's judgment as to the most useful resources for the application of scientifically based research to the educational, and in particular the Improving Teacher Quality (ITQ), context.

Part I, Chapter 1: Defining Scientifically Based Research

www.ed.gov/policy/elsec/leg/esea02/107-110.pdf Public Law 107-110, also known as the No Child Left Behind Act, all 670 pages of it.

www.ed.gov/admins/lead/account/nclbreference/reference.pdf A Desktop Reference to NCLB put out by the Department of Education in 2002, ~180 pages.

Scientific Research in Education, ed. By Shavelson & Towne, NRC, 2002. A major report from the National Research Council on many aspects of SBR.

www.ericdigests.org/2003-5/based.htm *Scientifically Based Research. ERIC Digest.* Ron Beghetto, April 2003. A clear, and influential, attempt to explain SBR and its potential implications.

www.learningpt.org/pdfs/qkey7.pdf *Scientifically Based Research, Quick Key 7*, Learning Point Associates, 2007. Written primarily for school leaders, an overview of SBR with a very useful bibliography and glossary of common research terms.

www.ies.ed.gov/ncee/wwc Home page for all the What Works Clearinghouse resources.

Part I, Chapters 2-4: Causality and Experimental Design

Experimental and Quasi-Experimental Designs for Generalized Causal Inference, Shadish, Cook, & Campbell, Houghton Mifflin, 2002. The essential work in the field. Referred to in this *Field Guide* as SCC.

www.socialresearchmethods.net/ks/index.php *The Research Methods Knowledge Base*, W. Trochim, 2006. Available on line and in printed form, an especially clear, and oft-times opinionated, presentation of the basics of social science experimentation. The two RDD charts are found in this material.

"Designing Designs for Research," Trochim and Land, *The Researcher*, 1982, v.1 n.1. An important look at what experimental designs can and cannot do in establishing causality. Also available online in a slightly updated version, www.socialresearchmethods.net/kb/desdes.php

www.ies.ed.gov/ncee/wwc/pdf/guide_RCT.pdf *Key Items to Get Right When Conducting a Randomized Controlled Trial in Education*, Coalition for Evidence-Based Policy, 2005. The complete 10 page checklist summarized in the *Field Guide* text.

www.ed.gov/rschstat/eval/resources/randomqa.pdf *Random Assignment in Program Evaluation and Intervention Research: Questions and Answers*, David Myers and Mark Dynarski, 2003. A Q&A concerning the reasons for randomized experiments and some of the procedures. More conceptual than technical.

Advances in Quasi-Experimental Design and Analysis, Ed. By W. Trochim, Jossey-Bass, 1986. An influential volume of essays pushing QED past the taxonomic approach and arguing for a more nuanced, complex, and judgmental approach.

Part I, Chapter 5: Research Checklists

www.wmich.edu/evalctr/checklists Home page of the Evaluation Checklists Project at Western Michigan University. The site's purpose is to improve the quality and consistency of evaluations and enhance evaluation capacity through the promotion and use of high-quality checklists targeted to specific evaluation tasks and approaches. Several of the checklists in this *Field Guide* are based on ones found at this site and one goal of this volume is to do for SBR what they have done for evaluation.

www.ed.gov/rschstat/research/pubs/rigorousevid.pdf *Identifying and Implementing Educational Practices Supported By Rigorous Evidence: A User Friendly Guide*, Coalition for Evidence-Based Policy, 2003. The complete 28 page document summarized in the *Field Guide*.

www.edvantia.org/products/pdf/SBROnlineGuide.pdf *Scientifically Based Research: A Planning Tool for Educators*, Doris Redfiled, 2007. Originally published in *District Administration Magazine* in January 2004, now available online.

Part II, Chapter 1: Research Timeline

This timeline is based primarily on material found elsewhere in the *Field Guide* and referenced within the timeline. Several of the checklists at the Western Michigan site, especially Dan Stufflebeam's work on Plans and Operations and Evaluation Design, were also important influences.

Part II, Chapter 2: Working With Research Consultants

www.wmich.edu/evalctr/checklists/negotiating.pdf *Checklist For Negotiating an Agreement to Evaluate an Educational Programme*, Robert Stake, 1976 The primary source for the Research Consultant Agreement Checklist.

www.ies.ed.gov/ncee/wwc/tech_assistance The Registry of Outcome Evaluators available through WWC.

www.eval.org/find_an_evaluator/evaluator_search.asp The American Evaluator Association's find an evaluator portal.

Part II, Chapter 3: Working With Institutional Review Boards

As mentioned several times in the text, IRBs are highly variable in their operations. It is vital that researchers contact their local Boards as early as possible in the planning process.

www.hhs.gov/ohrp/ Office for Human Research Protections in the Department of Health and Human Services, the agency responsible for regulating IRBs.

Part II, Chapter 4: Working With Local Education Agencies

Concurrent Marketing: Integrating Product Sales and Service, F. Cespedes, Harvard Business School Press, 1995. A clear, but book length, vision of the collaborative sales process by one of its developers.

Part II, Chapter 5: Communicating Research Results

www.researchutilization.org Home page for Research Utilization Support and Help (RUSH) at the Southwest Educational Development Laboratory. A web site containing many tools to assist in dissemination.

www.wmich.edu/evalctr/checklists/makingevalmeaningful.pdf Checklist for *Making Evaluation Meaningful to All Stakeholders*, P. Gangopadhyay.

www.ies.ed.gov/ncee/wwc/pdf/guide_SRF.pdf Report guide developed by the Department of Education's Institute of Education Sciences.

Part II, Chapter 6: Field Lessons

www.wmich.edu/evalctr/checklists/plans_operations.pdf Checklist for *Evaluation Plans and Operations*, Dan Stufflebeam.

Communicating for Managerial Effectiveness, P. Clampitt, Sage Publications, 1991. A well written, practical discussion about strategies to improve communication effectiveness.

Randomization and Field Experimentation, ed. By Boruch and Wothke, Jossey-Bass, 1985.

www.cde.ca.gov/ta/tg/sa/caassessment.asp Chart of the California Assessment System.

Glossary of Acronyms

AIR	American Institute for Research
APA	American Psychological Association
CDE	California Department of Education
CPEC	California Postsecondary Education Commission
CSET	California Subject Matter Examinations for Teachers
CST	California Standardized Test
ETS	Educational Testing Service
IES	Institute of Education Sciences
IHE	Institution of Higher Education
IRB	Institutional Research Board
ITQ	Improving Teacher Quality
LEA	Local Education Agency
MAC	Mathematics Assessment Collaborative
NCLB	No Child Left Behind Act of 2001
NEGD	Non-Equivalent Group Design
NSDC	National Staff Development Council
OHRP	Office for Human Research Protections
PD	Professional Development
QED	Quasi-Experimental Design
RCT	Randomized Control Trial
RDD	Regression Discontinuity Design
SBR	Scientifically Based Research
SCC	Shadish, Cook, & Campbell (see Additional Resources)
USDOE	U.S. Department of Education
WWC	What Works Clearinghouse