

Guiding Principles for Mathematics and Science Education Research Methods: Report of a Workshop

November 19-20, 1998
Arlington, Virginia

Larry E. Suter
Division of Research,
Evaluation and Communication
National Science Foundation
and
Joy Frechtling
Education Studies
Westat



June 2000

**Guiding Principles for
Mathematics and Science Education Research Methods:
Report of a Workshop**

**November 19-20, 1998
Arlington, Virginia**

**Larry E. Suter
Division of Research, Evaluation and Communication
National Science Foundation
and
Joy Frechtling
Education Studies
Westat**

June 2000

Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

Acknowledgments

This report was drafted from comments written or submitted by the chairs of working groups. Larry Suter organized the conference and edited the final version of the report. Joy Frechtling of Westat was instrumental in arranging for the workshop and preparing a first draft from the submitted materials. Brian Kleiner of Westat drafted the description of existing research guidelines. Special efforts were made by Thomas Romberg, Marcia Linn, Leona Schauble, Judith Sowder, Joe Krajcik, and Kathy Borman to prepare materials from the workshop. Eric Kirkland of Cosmos Corporation provided materials about the analysis of grants awarded by the Division of Research, Evaluation and Communication (REC). Materials from specific research projects were provided by Marcia Linn, Paul Cobb, Barbara Schneider, and Rosalind Mickelson. The workshop on research methods was recommended by William Sibley, the Acting Director of REC, and was held in November 1998. Program Directors Eamonn Kelly and Elizabeth VanderPutten contributed to the organization of the workshop, and Nora Sabelli provided comments on the text. A list of workshop participants is included in Appendix A.

Guiding Principles for Mathematics and Science Education Research Methods: Report of a Workshop

**Division of Research, Evaluation and Communication
National Science Foundation**

Intent

The purpose of this report is to present a brief review of research methods employed in recent studies and to propose, for discussion purposes, a number of guiding principles for designing research studies and evaluating research proposals in the area of mathematics and science education. Research on science and mathematics education is supported by the Directorate for Education and Human Resources (EHR) of the National Science Foundation (NSF). That directorate is responsible for “the health and continued vitality of the Nation’s science, mathematics, engineering, and technology education and for providing leadership in the effort to improve education in these areas” (www.ehr.nsf.gov). Thus, research projects supported by the directorate are intended ultimately to help ensure that a high-quality education in science and mathematics is available to every child in the United States and that the educational level is sufficient to enable those who are interested to pursue technical careers of any kind.

The members of the REC research staff decided to seek the advice of leading researchers in the field regarding the message that should be conveyed to submitters and reviewers to improve the quality and utility of both research proposals and funded projects. They invited about 30 investigators to discuss the variety of appropriate methods for high-quality research proposals on mathematics and science education (see the list of participants in Appendix A). The workshop participants were either investigators in NSF-supported educational research projects or researchers who had served on review panels for the Division’s programs.

Review panels do not always agree on research designs or on the quality standards by which proposals will be judged. The members differ in their special expertise and in their use of different methodologies because they have conducted research in many different disciplines (e.g., education research, education technology, the natural sciences, mathematics, and the social sciences). The guiding principles presented here are intended to help provide a common basis for reviewing many research proposals.

Much of education research is criticized for not having achieved high standards of scientific merit (Labaree, 1998). Without established standards for high quality, reviewers struggle with their own personal experiences and often judge new systems on an inappropriate basis. Reviewers of NSF proposals especially struggle with reaching agreement on proposed research topics that use emerging methodologies. For example, research projects that use new technologies for data capture and analysis,

such as video or computer-assisted data collection, present new problems to the research community. Reviewers debate the absolute merits of quantitative and qualitative approaches.

This report is meant to open further discussion into what is meant by, and desired in, high-quality research. No single report can provide absolute standards for judging creative investigations. The principles identified here are selected to be broadly applicable to the wide variety of approaches that could be supported by the Directorate for Education and Human Resources. The intent is to promote high-quality research, relevant to teaching mathematics and science, that is innovative in design, or uses cutting-edge techniques, or addresses difficult-to-study topics.

The report begins by describing the kinds of research that have been supported by EHR; second, it reviews existing guidelines from some research experiences; third, it presents a set of guiding principles that build on both the existing guidelines and a vision of what is meant by high-quality research in mathematics and science.

Education Research

In a recent effort to examine the variety of education research topics and research methods, Eamonn Kelly and Richard Lesh (Kelly and Lesh, 2000) concluded:

We are now at a point where the growing maturity of mathematics and science education research has shifted attention from strict adherence to traditional experimental methods as the best path to scientific insight to renewed interest in the development of alternative methods for research. In the past few number of decades, educational researchers have moved into school systems, classrooms and workplaces and have found a complex and multifaceted world that they feel is not well described by traditional research techniques. In the past, educational phenomena derived their status by surviving a variety of statistical tests. Today, nascent educational phenomena are accorded primacy, and the onus is on research methods to describe them in rich and systematic ways.

Moreover, they say that the research products are increasingly the result of design studies that involve contributions from teachers, curriculum designers, and students. A summary of their observations on changes in educational research is presented in Table 1. Kelly and Lesh point out that agreement on basic issues, such as the outcomes of education, is not easily achieved. Educational researchers have an important role to play in the continued development of theory and general models of schooling.

Table 1. Some Shifts in Emphasis in Educational Research in Mathematics and Science
(from Kelly and Lesh)

Less emphasis on:	More emphasis on:
Researcher remoteness or stances of “objectivity”	Researcher engagement, participant-observer roles
Researcher as expert; the judge of the effectiveness of knowledge transmission using prescribed measures	Researcher as co-constructor of knowledge; a learner-listener who values the perspective of the research subject, who practices self-reflexivity
Viewing the learner as a lone, passive learner in a classroom seen as a closed unit	Viewing the learner both as an individual and social learner within a classroom conceived of as a complex, self-organizing, self-regulating system that is one level in a larger human-constructed system
Simple cause-and-effect or correlational models	Complexity theory; systems thinking; organic and evolutionary models of learning and system change
Looking to statistical tests to determine if factors “exist”	Thick, ethnographic descriptions; recognition of the theory-ladenness of observation and method
The general applicability of method	The implications of subjects’ constructions of content and subject matter for determining meaning
One-time measures of achievement (often summative or pre-post)	Iterative cycles of observations of complex behaviors involving feedback; design experiments; engineering approaches
Multiple-choice or other standardized measures of learning	Multisensory/multimedia data sources; simulations; performance assessments
Average scores on standardized tests as learning outcomes	Sophistication of content models; the process of models; conceptual development
Singular dependence on numbers; apparent precision of numbers	Awareness of the assumptions of measurement; understanding the limitations of measures; extracting maximum information from measures; involving interactive, multi-dimensional, dynamic and graphic displays
Accepting curricula as given	Scientific and systematic reassessment of curricula reconceptualization of curricula given technology and research

Source: Kelly, A. E., and Lesh, R. (2000). *Handbook of Research Design in Mathematics and Science Education*. Mahwah, NJ: Erlbaum.

The Research Program in EHR: 1992-98

A wide variety of subjects and methodological approaches were supported by the research programs of EHR between 1992 and 1998. While all projects were intended to help understand how to improve the quality of existing practice in mathematics and science education in the United States,

the investigators and reviewers represented diverse fields such as educational psychology, sociology, school administration, statistics, education technology, and science fields.

Prior Funding Patterns

The Division of Research, Evaluation and Communication supported about 350 grants in five different programs between 1992 and 1998. These funds were awarded to grantees who submitted proposals to the programs of Research on Teaching and Learning, Applications of Advanced Technology, Studies and Indicators, and Networking Infrastructure for Education. Three programs were merged into one, the Research on Education Policy and Practice Program (REPP), in 1997. Additionally, about 25 research awards were granted between 1994 and 1998 through Learning and Intelligent Systems (LIS), which was part of a cross-directorate program.

The funding levels for the research program remained at about the same level—\$22 to \$28 million—between 1994 and 1997. Additional research awards made in the LIS program raised the total level of funding to \$38 million each year. With growing interest in finding practical answers about how to improve student achievement, funding levels for education research are expected to remain at these levels or to grow in order to support new initiatives.

Content Areas of Investigations

Abstracts of the research projects supported by REC between 1992 and 1998 were used to identify trends in the division's support patterns, and analysis revealed that all projects funded by the program, as expected, had an emphasis in either mathematics or science education. Before 1998, projects in science fields outnumbered those in mathematics, but since then an equal number of mathematics and science projects have been awarded. Two other trends in funding patterns suggest changes that have been underway in these programs. First, since 1995, the research program has supported a declining number of projects involving studies of teaching strategies. Second, a growing number of projects used multidisciplinary teams that involve principal investigators or research team members representing different disciplines or areas of expertise, such as physical sciences and education. This trend toward multidisciplinary teams is reflected in the review panels that are selected to permit in-depth discussion of the content of their proposals.

Methods Used in Education Research Awards

A summary of methods used in 100 NSF education research awards that ended between 1990 and 1998 is shown in Table 2. This analysis shows that the “traditional” educational psychology methods of experimental design or quasi-experiment were not very common. The most common method was a descriptive case study (41 grants out of 100) and survey (24 grants). Quasi-experiments were reported in only 12 grants.

Table 2. Research Method Used in NSF supported Education Research Grants that Ended between 1992 and 1997

Method	Number of grants
Total grants	100
Descriptive case study	41
Survey	24
Quasi-experiment	12
Meta-analysis	8
Action research	6
Causal case study	5
History	5
Ethnographic description	5
Research synthesis	3
Experimental design	0
Other methods	13

Many projects used more than one method of research. A high proportion of projects used both qualitative and quantitative methods, reflecting the fact that many research teams are multidisciplinary. Clearly, the education research community served by NSF does not rely on a single method of investigation to address research issues.

In 1997, nearly all of the 42 active awards in the REPP program were classified as “applied” research, and only 7 awards were classified as “basic” research. This is consistent with the program announcement that encouraged research projects intended to lead toward the improvement of instructional practice or school management. The distinction between applied and basic research is only useful here in that it captures the intention of the researcher to address immediate or long-range educational issues. In fact, education research projects sponsored by the EHR seek to accomplish both. A recent analysis of basic and applied research by Donald Stokes helps clarify the goals of basic and applied research supported by scientific funding agencies. He points out that the researcher is most often driven by curiosity, while funding agencies are more often driven by effective use (that is how they ultimately justify their budgets). Thus, the distinction between applied and basic is used here as a rough indicator of the different goals of research projects (Stokes, 1997, p. 102).

Another review of the repertoire and accepted range of research approaches in mathematics was conducted by Romberg (1992). Romberg briefly describes about 20 research approaches and points out that the choice in method has become “increasingly diverse” over the last two decades. The prevailing notions of acceptable research in education research originally grew out of the logical positivist philosophy that characterized behavioral psychology. The strategy held in highest esteem during the 1960s was the pre-post design with randomly assigned experimental and control subjects. This thinking began shifting in the 1970s, Romberg notes, because the field of educational research had grown such that many research projects included a wider variety of disciplines on the project teams. The number of perspectives maintained by those involved in educational research was also growing, and researchers began to acknowledge that students, teachers, and education institutions are not as amenable to “empirical-analytic” research traditions as are the fields of psychology or agriculture, which were frequently used as models for education research (Romberg, 1992).

In summary, the REC research programs have supported research that often is oriented toward informing practice or resulting in applications. The projects used a mixture of research methods.

Research projects that rely entirely on educational experimental designs were rarely found in the 1990 to 1998 portfolio.

Existing Statements on Standards for Education Research

Several reports intended to provide guidance for education research were identified and shared with the participants of the workshop on methodology. Some reports address the range of research approaches appropriate for education studies without providing guidance on standards. For example, Romberg (1992) provides some excellent advice to graduate students or beginning researchers on factors to consider in developing research studies in the area of mathematics that are generalizable to other subject areas. Other reports suggest standards for educational research on initial design, stages of research implementation, and report generation, but, unfortunately, do not provide a specific set of standards that has been widely endorsed. The October 28, 1998, issue of *Education Week* reported that the search for such a set of standards by a group of outstanding researchers of the National Academy of Education had not been successful after an initial 3 years of work. The National Academy of Education established a Commission on the Improvement of Education Research, chaired by Ellen Lagemann and Lee Schulman, which produced a report that provides an “overview of the tensions, dilemmas, issues, and possibilities that characterize education research” (Lagemann and Schulman, 1999).

To become acquainted with the approaches that have been taken to develop standards, the workshop participants reviewed a number of documents that were attempts at this task. Existing standards for education research frequently separate quantitative and qualitative approaches. In some standards documents, only one approach is addressed. In others, a single document puts forward dual sets of standards, one for each of these main types of social science research. This dichotomy of standards probably reflects the traditional bifurcation within the community of education researchers, given varying aims, methodological backgrounds, and assumptions about how knowledge is best acquired.

Less common are attempts to provide a single set of general standards that are meant to serve as guidelines for all kinds of education studies. Proponents of the single set of standards stress that a common core of issues needs to be considered regardless of the methods espoused. Although not a central feature of most discussions, an underlying message seems to be that mixed method approaches are not only possible, but may be preferable in many instances.

This section will briefly describe four representative examples of standards along these lines in order to illustrate the range of past collaborative efforts to develop guidelines and procedures for education research. As an example of proposed standards for quantitative research, the SEDCAR (Standards for Education Data Collection and Reporting) (U.S. Department of Education, 1991) will be discussed, although other similar documents could equally have been presented. The work of Spindler and Spindler (1992) will then be presented as an instance of standards put forward for qualitative, ethnographic education research. Next, the standards proposed by the FINE (First in the Nation in Education) Foundation (Ducharme et al., 1995) will be described as a representative example of efforts to treat both qualitative and quantitative research designs, though separately, within a single document. Finally, the work of Eisenhart and Howe (1992) will be discussed as an example of how a single set of general standards has been proposed to cover all types of education research. It will become apparent that the guiding principles proposed by members of the NSF Workshop on Education Research Methods are most akin to the more general ones of Eisenhart and Howe, but reflect the special concerns and interests of researchers in the field of mathematics and science.

Standards for Quantitative Research

With the aim of improving the quality of data collected on the condition of education in the nation, the National Center for Education Statistics (NCES) initiated the Cooperative Education Data Collection and Reporting (CEDCAR) Standards Project, which ultimately led to the SEDCAR document. Since these standards are most relevant to data collection activities within the National Cooperative Education Statistics System, the document is predictably geared toward large-scale quantitative studies.

According to the authors, the standards set forth in the SEDCAR document are intended to serve as guidelines for different phases of a research project. Also, they “identify the qualities that characterize good measures and describe the process of selecting and evaluating appropriate measures that will result in data of the highest quality—data that provide useful, timely, accurate, and comparable information” (U.S. Department of Education, 1991, p. xi). SEDCAR proposes six interrelated phases of a large-scale study, which serve as a conceptual framework for the development and organization of the standards. They are management of data collection and reporting, design, data collection, data preparation and processing, data analysis, and reporting and dissemination of data.

Standards proposed within each major phase of data collection and reporting contain a statement of purpose followed by associated guidelines that suggest the “best practice” for satisfying the purpose of the standard. For example, the design phase includes the “Standard for formulating and refining study questions.” The stated purpose of this standard is “to ensure that the study questions are well chosen, well stated, and empirically answerable.” The associated guidelines are presented here to give some indication of their relation to the standard and their degree of specificity:

- Study questions should be formulated to address the identified information needs.
- Study questions should be clearly defined, articulated, and reviewed to ensure that they address all aspects of the issues under investigation.
- The study questions should:
 - Reflect a knowledge of relevant literature,
 - Anticipate and respond to unintended outcomes,

-
-
- Be capable of further refinement as research planning proceeds,
 - Be clear in their meaning, implications, and assumptions,
 - Eliminate bias as fully as possible to avoid any tendency to predispose the findings,
 - Attempt to break down problems into their constituent parts,
 - Be capable of being answered through practical data collection activities,
 - Focus on the information needs,
 - Be prioritized in order of importance, and
 - Be broad enough in scope to cover the needs of the data requestor and, when possible, the needs of secondary data users.

Most of the standards and guidelines within SEDCAR are relevant to quantitative research designs support by REC.

Standards for Qualitative Research

With the emergence and mainstream acceptance of qualitative and ethnographic education research, some have argued that the varied approaches within this domain should be held accountable to a set of standards particular to this type of research. In addition, there has been an assumption that qualitative research requires a distinct set of standards. Spindler and Spindler (1992) propose standards for qualitative, ethnographic education research that are very different from SEDCAR both in content and form. They were not geared toward the broad collection and analysis of nationally representative data, but rather toward a narrowly focused, in-depth study of interaction in a particular environment with a particular set of participants.

Spindler and Spindler provide criteria (standards) for what they call a good ethnography of education. The first three criteria (out of 11) are as follows:

- Observations are placed in context, both in the immediate setting in which behavior is observed and in further contexts beyond that setting.
- Hypotheses emerge *in situ* as the study continues in the setting selected for observation. Judgment on what may be significant to study in depth is deferred until the orienting phase of the field study has been completed.
- Observation is prolonged and repetitive. Chains of events are observed more than once to establish the reliability of observations.

These criteria, quite different from the standards proposed for most quantitative research, reflect the aims, issues, and methods of ethnographic research. For example, the second criterion recommends that hypotheses emerge only after the researcher has embarked on the study and made detailed observations and notes on the setting and participants. In quantitative studies, research questions (and hypotheses) usually drive the design of the work since instruments must be prepared in advance of data collection.

Standards for Both Quantitative and Qualitative Research

The FINE Foundation, established by the Iowa Legislature in 1985, has proposed distinct sets of standards for quantitative and qualitative education research. Their standards are lists of criteria that are useful to remind new researchers what kind of questions are raised by reviewers of proposals. However, the use of lists may lead some researchers into believing that merely satisfying these aspects of research design is sufficient to preparing a good proposal. Furthermore, the list of standards seems to assume that a particular study would choose either one method or another, rather than use a variety of methods to answer a complex question.

The FINE criteria for quantitative studies include those pertaining to four aspects of this kind of research. First, there are criteria (in question form) having to do with the research “problem” (or question): Is the stated problem clear and researchable? Has a thorough review of literature informed the procedures and discussion? Are hypotheses/research questions explicitly and clearly stated? Second, there are criteria relating to research procedures that involve sampling issues, data gathering techniques, and appropriateness of research design (given specific research questions). Third, there are criteria involving discussion of results: Are results appropriate and clear? Do the results of the data analysis support conclusions of the study? Are recommendations for future action asserted? Fourth, there are method-specific criteria for quantitative studies, including criteria for survey/questionnaire studies, correlation studies, causal-comparative studies, and so on (Ducharme et al., 1995).

Criteria for qualitative studies recommended by the FINE group include the same four general categories, but with slightly different subparts. Interestingly, components of the first category, “introduction to problem,” are almost identical to those for quantitative studies, which is an indication of the features taken to be common to all good education research. The criteria begin to diverge, however, with respect to the categories “research procedures” and “discussion,” and naturally the “method-specific criteria for qualitative studies” (including interview/focus group studies, observation studies, historical studies, etc.) are completely unlike those proposed for quantitative research.

One Set of Standards for All Methods

Given the differences in methods and assumptions of quantitative and qualitative research designs, providing a single set of standards to cover both may not seem appropriate. However, the work of Eisenhart and Howe (1992) (continued in Eisenhart and Borko, 1993) suggests that a single set of standards is not only possible, but is also preferable. The standards they propose are united under the notion of “validity,” which generally has to do with the “trustworthiness” of inferences drawn from data. Eisenhart and Howe propose that both qualitative and quantitative research be subject to the same general standards of validity, though all research studies will have to satisfy design-specific standards as well.

Eisenhart and Howe (1992) assert that general standards for the conduct of education research should, with respect to validity, transcend specific disciplines and research designs. They propose five general interrelated standards for validity in education research.

-
- Standard 1 asserts that the research methods should ideally fit and be driven by the research questions.
 - Standard 2 states that data collection and analysis techniques should be competently applied. Connected to this is the requirement that researchers locate their methods within the historical, disciplinary, or traditional contexts in which they were developed.
 - Standard 3 requires that studies demonstrate their link to a background of existing theoretical, substantive, or explicit practical knowledge.
 - Standard 4 addresses what the authors call “value constraints.” “External” value constraints have to do with whether the research is demonstrably worthwhile in addressing concerns and issues in educational practice. That is, researchers must show that their work is important and useful. “Internal” value constraints have to do with the ethical conduct of the research.
 - Finally, standard 5 involves the balancing of the first four standards and the achievement of overall clarity, coherence, and competence.

Eisenhart and Howe (1992, p.657) assert that far from being ephemeral and vague, articulated standards provide for three significant benefits:

- They allow economy of thought in designing and evaluating educational studies.
- They provide the starting point for reflection on and improvement of the educational research enterprise.
- They serve as the vehicle both for communicating within and across research traditions and for orienting newcomers.

Their standards, which were not written specifically to fit the types of science and mathematics education topics that are addressed by REC, can be very useful to the beginning researcher.

Relation of Existing Standards to Guiding Principles

The guiding principles for NSF proposals generated at the Workshop on Education Research Methods and introduced in the next section share much in common with Eisenhart and Howe’s criteria in terms of substance and level of generality. For instance, they address issues having to do with situating the study within the context of prior knowledge; showing the import, value, and usefulness of the work; demonstrating a link between research questions and methods; and carrying out the work in an ethical manner. Also, both are general enough to transcend particular disciplines and (qualitative and quantitative) research designs, yet are concrete enough to be relevant and applicable in practice.

This similarity may reflect a consensus in the education research community that good research on education issues frequently is a judicious blend of qualitative and quantitative approaches and that high-quality studies must include, but transcend, technical accuracy. The guiding principles discussed in the next section are explicitly designed for the development and evaluation of proposals for mathematics and science education research; they reflect the composition of the workshop members and the perceived

pressing research needs. The principles for proposals in mathematics and science education research will be more limited in scope than standards reviewed in this section and will not address those aspects of a study that are unforeseeable. Another significant difference between most existing standards and the guiding principles presented in the next section is the strong focus on the potential applicability and relevance of (proposed) research projects to educational practices.

Research Approaches

As the brief review of research funded by REC in the past few years showed, education research studies follow a wide variety of philosophical and research paradigms. The workshop participants strongly believed that making new discoveries about the practice of teaching and learning requires many different approaches that extend far beyond the confines of a single model. Some of the alternative research approaches that are being explored by serious researchers will be described to illustrate the range of models that are respected today. The list is not meant to be exhaustive, but rather to provide some idea of the range of possibilities that the research community might expect to find. Any single project may include any one or all of these research approaches in the same project.

Design Experiments

Allan Collins and Ann Brown used the term “design experiments” to describe education research studies that attempt to engineer educational environments and simultaneously conduct experimental studies of those innovations. The idea was borrowed from the design sciences such as aeronautics (Brown, 1992). A design experiment features cyclical interaction between two complementary aspects of design and research. Working from a base of previous research and theory, researchers craft and implement the design of a learning environment (which may vary in scope from a computer-based tutor to a teacher, classroom, entire school, or a district). The design experiment entails conducting a systematic program of research on the learning that results from the classroom (or school, or teacher) experiment. The design experiments are created to emphasize deep understanding of how student or school outcomes are related to the production of learning, in contrast to evaluation studies or clinical trials, which examine a relationship without deep explanation. An assumption of the design experiment approach is that many forms of learning that are important targets of inquiry cannot, in fact, be studied unless the conditions for their generation are supported first.

Proponents of design experiments feel that they have several distinguishing features: they are firmly grounded in disciplinary subject matter; they focus on emergent ideas, rather than well-articulated visions; they recognize the unique patterns and structures that characterize different layers of the educational system; and they employ multiple and converging methods (Brown, 1992). Many proposals submitted to the education research program are likely to involve problems that do not have a well-articulated vision of the “big ideas” that should drive instruction.

Allan Collins (1999) distinguished design experiments from psychological methodology in these ways:

1. Laboratory setting versus messy settings. Experiments usually use presentations that are one-directional, rather than relying on interactions between teachers and learners. Design experiments are set in real-life learning situations to avoid the distortions of a laboratory.

-
2. Single dependent variable versus multiple dependent variables. Most psychological experiments have one dependent variable. Design experiments have dependent variables that matter: climate variables, outcome variables, and system variables.
 3. Controlling variables versus characterizing the situation. Psychological experiments use a methodology of controlling variables borrowed from early physics. Design experiments seek to identify all the variables and seek to identify the nature and extent of effect of the variables.
 4. Fixed versus flexible design. Psychological experiments have fixed procedures that are documented to permit replication. Design experiments start with plans that are not completely defined and are revised depending on their success in practice. The goal is to progressively refine a teaching method and to modify the refinements when appropriate.
 5. Social isolation versus social interaction. Experiments present material in a standardized manner. Design experiments are conducted in complex social situations such as classrooms.
 6. Testing hypotheses versus developing a profile. An experiment tests one or more hypotheses, systematically varying the conditions of learning. The design experiment's goal is to see what conditions lead to different effects. It might look at many different aspects of the design and develop a qualitative and quantitative profile of the practice. Evaluation is best when done with respect to a number of dimensions in a comparative fashion.
 7. Experimenter versus co-participant design and analysis. Control of design is maintained by the experimenter. In design experiments, different participants are involved in developing the design in order to bring their different expertise together such as technology experts, cognitive psychologists, teachers, curriculum designers, and anthropologists.

Examples:

About 20 percent of projects awarded by REC between 1996 and 1998 may be called design experiments, although the term is not widely used as a descriptor. Design experiments may involve the application of multiple techniques such as case study, interview, video taping, and standardized student assessment. They are often means of developing an improved hypothesis. A project by Marcia Linn at the University of California at Berkeley and another by Paul Cobb, Kay McClain, and Koeno Gravemeijer at Vanderbilt University provide examples of recent use of this design (Linn 1995, Cobb 1999).

The purpose of the study conducted by Linn was to understand how to guide students in the process of “knowledge integration,” which she defined as the process of “making diverse ideas explicit, negotiating among them, and building new understanding” (Linn, 1999). She explains that “knowledge integration involves seeking alternative perspectives, distinguishing among these ideas, gathering empirical, experimental, or observational data, discussing alternatives, and designing new approaches.” This approach was used to understand the process of science partnerships so that individuals brought their own ideas to the mix to “create a design, gather evidence, restructure, reorganize, or reconceptualize the task, and repeat some or all of the steps again.” The “partners” in her project were science teachers who contributed classroom activities and targeted goals for students, and natural scientists who contributed an understanding of science content and knowledge of current controversies. The investigation used software tools, such as a SenseMaker to “make visible the process of organizing warrants to support an argument.” This software helped students to see all of the thinking processes and arguments of scientists

as they solved a problem. A science problem for which no accepted scientific explanation was available was introduced to the students as the project began. The problem she used for this study was explaining the existence of frogs with deformed limbs.

During the design phase, the team developed a vision for presenting scientific knowledge at the level that fit understanding and vocabulary of students in middle school, which proved to be a difficult and time-consuming process. The project found that methods for helping science and school participants communicate about deformed frogs were successful when they made thinking visible with software. The partnerships succeeded when they were able to define their failures as well as successes. The study developed a series of design principles about how students and scientists approach a study of a scientific phenomenon. One principle that came out of the study was that recognition that students approach a problem with a wide array of loosely connected ideas and language that require support to be useful for enhancing their understanding.

The study conducted by Cobb, McClain, and Gravemeijer focused on statistical data analysis at the middle-school level. The design experiments were oriented about statistical distributions. The research team wanted students to view data sets as entities that are distributed in a space of possible values. In order to support how students develop the idea that mean, median, mode, and skewness are characteristics of univariate distributions and that directionality and strength are characteristics of bivariate distributions, the team developed a series of three data analysis tools. In addition, they designed sequences of instructional activities that supported the emergence of significant statistical ideas while students investigated “in the spirit of” genuine data analyses.

Analyses of the design experiments indicate that distribution is a feasible instructional goal at the middle school level for both univariate and bivariate data. The analyses also indicate that students at this level can begin to investigate both the characteristics of data sets that are relatively stable across samples and the relations between sample statistics and population parameters. As it transpired, this approach enabled students to come to appreciate how the legitimacy of conclusions drawn from data depends on the soundness of the data generation process. A retrospective analysis revealed, for example, that they developed an understanding of both the need for procedures such as stratified random sampling and for means of controlling extraneous variables.

Controlled Experiments

“True experiments” follow the classic design that characterized logical positivist philosophy. Such experiments typically include treatment and control or comparison groups, ideally with randomized assignment of subjects to treatment groups. One group is given the treatment of interest, such as a particular curriculum, teaching strategy, professional development experience, etc., and another group is not provided the treatment. Some outcome measure of interest, such as student test scores, instructional practices, or understanding of diversity, is compared to that of the control group.

Studies of this type attempt to make strong causal arguments for the effects of a particular treatment by isolating treatment effects from other possible determiners of outcomes through this use of control comparison groups (Romberg, 1992). There is a long-held belief in educational research (especially by those trained in educational psychology) that this method best provides evidence for making causal statements about education practices. Donald Campbell and Julian Stanley, writing in 1963 about these methods in their influential *Experimental and Quasi-Experimental Designs for Research*, said that their “chapter was committed to the experiment: as the only means for settling disputes regarding educational

practice as the only way of establishing a cumulative tradition in which improvements can be introduced without the danger of a faddish discard of old wisdom in favor of inferior novelties.”

However, very few research projects that involve students or teachers are able to randomly assign students or teachers to particular schools, or even classrooms. Furthermore, choosing a group to be a “comparison” that has all the qualities of the “treatment” group with the lone exception of those factors that are being tested is almost impossible in a live school situation since factors that have not been controlled might easily intervene. Researchers working in school settings cannot “cleanly” manipulate variables as they can in a chemistry laboratory. For example, teachers selected as a control group may instead choose to adapt the lessons in a new text to their prior teaching practices. Few researchers can establish sufficiently strong control over a school administration to maintain control over all aspects of teaching and presentation of materials in a classroom setting.

“Alternate treatment quasi-experiments” are more typical of research projects carried out in schools. They are characterized by using intact natural treatment groups (classrooms or schools) without random assignment and alternate treatments rather than experimental and placebo treatments. Such experiments are done because it is usually difficult to arrange student and school settings to locate causal paths between schooling practices. Many modifications of this strategy have been carried out. For example, the quasi-experimental designs described by Campbell and Stanley provide descriptions of designs that can be carried out in live school settings to test whether rival interpretations of events have credibility. If the potential sources of invalidity are considered and attended to, these designs approach the rigor of randomized experiments.

Examples:

An analysis of 122 National Science Foundation awards that ended between 1996 and 1998 found that no awards were given for classical controlled experiments but that a number of awards, perhaps 10 percent of the total, were for quasi-controlled experiments. A Ph.D. dissertation project provides an example of how strictly controlled experiments might rarely be carried out in school settings.

A Ph.D. candidate from Stanford who had been a teacher in a school system carried out a controlled experiment to determine whether group experiences for children resulted in increases in performance levels (Schultz, 1999). The design involved random assignment of 140 students into four classrooms. The teachers selected were also randomly assigned to the four classes. Two of the classes were instructed to teach a section of biology with procedures that involved group work, and two classes were instructed to present the same material in the same time period, but the students did not work in group settings. To test the changes in student performance on the material, three different types of tests were given before the period of instruction began and after it ended. Thus, the outcome measures for the experiment included a multiple-choice test and performance tests of specific aspects of the unit being taught. This tight design thus permitted a specific test of a specific hypothesis: that group learning experiences using a method of Complex Instruction (Cohen and Lotan, 1997) was likely to lead to increased learning for more students than learning experiences that did not permit students to interact in a group setting.

The results clearly showed higher levels of student performance for the randomly assigned classrooms using group procedures. However, the study could not provide immediate explanations for the large differences that were observed from evidence of teacher practices or other activities.

Another project by Romberg demonstrates the problems and possibilities of experiments when carried out in live school settings. Researchers frequently do not have sufficient power to convince the administrators to maintain consistency in experimental and control groups. In order to test whether a

specific mathematics curriculum would influence growth in student performance, a group of classrooms were selected to receive a new mathematics curriculum and they would be compared with classrooms in the same schools that did not use that textbook and accompanying procedures. The analysis was intended to monitor the exposure to specific content areas, teacher knowledge, classroom events, and pedagogical decisions of the teachers and the students in both settings over a 5-year period. The study was intended to establish whether student growth in achievement, when measured for the same students over a period of time, could be detected more for the new materials. It involved a number of data collection instruments completed by teachers and students to inform the researchers of the level of learning that had occurred during the period.

The investigators were not able to carry out the design of the study as they had intended because of administrative decisions by the schools. First, the school districts did not permit randomly assigned teachers or students to the treatment groups. The principal insisted on choosing the level of students for each setting. Secondly, the schools selected as controls in this longitudinal study did not maintain their original assignment after the first year. The principal said that he did not want to wait for another year to make some changes in instruction. “So, hopefully we won’t be a good control group if what happens is what I intend to happen.” Third, the incentives for those who volunteered, professional development opportunities, were not considered sufficient compensation for continuing involvement with the study. Finally, teachers in the control samples regarded themselves as “lab rats” and would not agree to participate after the first year unless they were provided with the new school materials that were being tested in the experiment classes. Thus, the original intent to carry out a controlled experiment failed in a live school setting for reasons that affect the daily lives of administrators and teachers (Romberg and Shafer, in press).

Representative Sample Surveys

Sample surveys provide descriptive information on the status of a process, value, or perception. They provide data used for descriptive and policy purposes since they can provide information on the changes in adoption of strategies or student achievement levels. Such surveys rely primarily on quantitative techniques and are designed using carefully constructed rules for sampling, data collection, and analysis. While populations are usually relatively large, they may be defined in a variety of ways. That is, they may be broadly defined or segmented into specific subgroups of interest (geographic regions, demographic subgroups, public and private schools, etc.).

Surveys are used in the research program to monitor changes underway in large school systems, such as entire states or the nation as a whole. Recent uses of survey methods include the studies of teaching practices and of student achievement in the Third International Mathematics and Science Study (TIMSS). However important survey techniques are for measuring change in large systems, they cannot provide sufficient information on a broad number of factors that may be the underlying causal influence of change in a system. Thus, major studies are now typically constructed with combinations of survey research and qualitative methods that provide richer descriptions of the underlying events in a school system.

Example:

An analysis of education research projects funded by NSF for the years 1996 to 1998 found that about 20 percent of the research projects included some form of survey. The Third International Mathematics and Science Study was the largest study conducted during this period and provides an example of the types of questions and methods that are used in such studies.

TIMSS was carried out in 1994-95 to attempt to answer the question of why previous studies had shown large differences in student performance between countries. A serious hypothesis for the comparatively low achievement of U.S. middle school students in contrast with those of other countries, developed from prior studies, was that the mathematics and science curriculum in the United States was not demanding. The study was designed to test the causal connection between curriculum policy and student performance. TIMSS data collection instruments provided measurement of the *intended*, the *implemented*, and the *achieved* curriculum with the intent of linking performance on specific topics to policies in the country that could be responsible for coverage of the curriculum topics. This design required a common classification scheme for the topics in mathematics and science that would be used to classify textbook content, national standards documents, teacher classroom coverage, and a new student achievement test. It was impossible to obtain agreement from all participating countries for a longitudinal followup of students through the school system to better examine causal relationships between curriculum and achievement. Thus, the study was conducted at grades 3 and 4, 7 and 8, and 12 (using U.S. definitions of grades for this purpose) in spring 1995 and 1999. This design permits estimates of change between 1995 and 1999 for a cohort (grade 4 in 1995 to grade 8 in 1999), and it includes estimates of prior student performance at grades 4 and 8. For example, student performance at grade 3 can act as an estimate of prior performance for 4th grade students, so that country differences in growth patterns can be related to the material introduced in each country during grade 8.

A sample of schools and students was selected in each country for participation in the testing. The sample was representative of the school system and of sufficient size to establish reliable national estimates of curriculum coverage and student performance. Analysis of the possibility of a causal relationship between curriculum and achievement would be conducted by developing reliable estimates of coverage of mathematics and science topics for each country and grade from the content of textbooks, reports from classroom teachers, performance on the TIMSS tests, and country characteristics.

Causal Modeling

The exploration of relationships between school policies, teacher practices, and student outcomes is occasionally conducted through the development of statistical models of large-scale surveys. Recently developed techniques permit the simultaneous estimation of relationships of events within schools, classrooms, and students, even though these events occur at different points of aggregation within the school system. The development of such models depends on the appropriate classification and nesting of survey data for school systems. Statistical models are especially useful in studies of the causal paths toward increasing student test scores. Correlational analysis of individual differences has been a common method for exploring performance on psychological factors of personality, aptitude, and ability. Such studies of student achievement may require further development of new techniques to better capture the interaction of learning behavior with classroom practices.

Examples:

Education studies that use large data sets to produce models of educational processes represent a small proportion of all projects funded by NSF. Model building for mathematics and science education is limited by the availability of survey data about mathematics or science activities in schools.

Two studies by Schneider conducted of the NELS:88 longitudinal survey provide insight into the power and limitations of correlation analysis for the study of causal relationships between school characteristics and student performance (Schneider, Swanson, and Riegle-Crumb, 1998; Swanson and Riegle-Crumb, 1999). One study examined the causal connection between secondary school curriculum and postsecondary school performance by relating the courses taken in high school in this longitudinal study to later college performance. This found that one of the strongest predictors of continuation in 4-year college attendance is rigorous high school mathematics, science, and foreign language courses but not advanced history.

Another study used large-scale databases to investigate how school and family context variables influence student outcomes including academic performance, college entrance, and psychological well-being. It related courses taken by students in high school to their later performance in college by using a nationally representative sample of students who were followed from 8th grade through college. By analyzing differences in courses reported on high school transcripts, the study investigators found that taking rigorous courses in science, mathematics, and foreign language during high school was related to the likelihood that a high school graduate will attend a 4-year college. Statistical models, such as hierarchical linear models and logistic regressions, were used to fit differences across students in the national survey. Quantitative methods have the value of providing estimates for the general population, but the scope and depth of analysis are constrained by the quality of items used on the instruments for collecting the data.

Case Study and Other Qualitative Methods

An example of a qualitative method that has been used for education research is the case study. Case studies are intensive studies of specific instances. Yin (1994) defines a case study as an “empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident” (p. 13). He observes that the type of research question that might be addressed by a case study are “how” and “why” questions but not “what” questions, which would be answered by analysis of surveys or archives. Case studies are used when the “researcher has little control over events and when the focus on a contemporary phenomenon within some real-life context” (Yin, p. 1). He points out that case studies may be used to “*explain* the causal links in real-life interventions that are too complex for the survey or experimental strategies” (p. 15).

Other qualitative methods might explore aspects of educational activities with yet other means. Investigators who work in this tradition may analyze life histories or use ethnographic methods to describe the features of a home, classroom, school, or school organization. Some methods of inquiry that are often used in such research include thinking aloud, stimulated recall, journal keeping, policy capturing, and the “repertory grid technique” for describing how constructs are created and related to each other (Gall, Borg, and Gall, 1996). Detailed descriptions of investigation techniques are described in the *Handbook of Qualitative Research*, edited by Denzin and Lincoln (1994).

Commenting on the process of making valid inferences from events outside the laboratory with qualitative methods such as the case study method, Donald Campbell made these important observations:

More and more I have come to the conclusion that the core of the scientific method is not experimentation per se but the strategy connoted by the phrase plausible rival hypotheses. This strategy may start its puzzle-solving with “evidence” or it may start with “hypothesis.” Rather than presenting this hypothesis or evidence in the context-independent manner of positivistic “confirmation” (or even postpositivistic “corroboration”), it is presented instead in extended networks of implications that (while never complete) are nonetheless crucial to its scientific evaluation (foreword to Yin, 1994).

He correctly points out that the nature of science is not in the choice of the method, but in the endless tasks carried out by the scientific community to make explicit how available data fit with existing hypotheses.

Example:

A significantly large number of awards to researchers in mathematics and science education use case study methods. Approximately a third or more of all projects involve either a use of descriptive case studies or causal analysis from case studies.

The Business Leaders and School Reform project is a case study of school reform in Charlotte, North Carolina (Mickelson and Smith, forthcoming). The project addressed the relationship between local schools and economic development; the nature and consequences of school reforms influenced by the corporate agenda; and the ways in which adolescents from different races, classes, and genders respond to the opportunity structure they perceive as awaiting them. The project conducted interviews with educational, civic, and business leaders; made observations at conferences and forums; and collected documents that describe the context in which Charlotte launched a school reform initiative in the early 1990s.

Data collected included a survey about employer satisfaction from a sample of business leaders. The study also included the collection of documents from the school system related to curricular and instructional reforms; plans for enrollment growth and pupil assignment; district-wide indicators of student achievement, attendance, retention, and graduation rates; and school- and individual-level indicators of opportunities to learn and outcomes. Focused interviews with key school system personnel further illuminate the patterns the quantitative data suggest. Also, the study conducted a survey of 8th and 12th grade students in the school district. The survey instrument assessed students’ attitudes toward education, work, and the future (educational and occupational aspirations), as well as individual and family background indicators. School system electronic data regarding achievement, test scores, attendance, and the prior schools students attended were merged with survey data. The case study involved integrating all of these sources of information to produce a holistic account of school reform in one city.

Guiding Principles for Research Proposals

This section presents a draft set of guiding principles for proposals submitted to the research program in EHR. These guiding principles were developed from the discussion of the investigators, all of whom had been previously funded by NSF or had participated in reviews of proposals and had attended the workshop on methodology. The principles should apply to research projects regardless of the scale of the project, the methodological approach taken, or whether the project might be classified as either basic or applied. Meeting the standard applied by a given guiding principle (validity, for example) may be carried out differently in projects that use different strategies (e.g., design experiments versus nationally representative surveys).

Research projects are judged on the basis of the match between the approach taken and the ideas, outcomes, or models that the research is trying to explore. No single research model can be selected as more successful than another without clarity on the problem and the theoretical approach that will be involved. Williams James is reputed to have said, “You can’t pick up rocks in a field without a theory” (Agar, 1980). The notion of theory may extend from a guess or conjecture to grand theories. Every research proposal must have a statement of how and why things are put together somewhere in its text. Thus, we begin with a statement about defining a research problem.

The Problem

Every proposal must be clear about the issues, understandings, or practices that are to be addressed. While, in the broadest sense, all projects are expected to deepen our understanding of how students learn in mathematics and science and what can be done to improve this learning, there is great variety in the specifics of what any single work addresses. Descriptions of the problem are expected to indicate what the project is intended to do, as well as why the set of activities is worth doing. The researcher should be able to answer why the proposed study is worth funding and why it is significant in relation to other work and to current issues of importance to education researchers. The merit or value of a research problem should not be assumed or asserted; it should be justified, explained, supported, and in other ways explicitly rationalized.

Researchers should examine the program announcements of the research programs before submitting their proposals. The announcements list criteria for selecting awards and suggest areas for researchable questions. For example, the Research on Education Policy and Practice program suggested areas for investigation such as, How do people learn? How does technology change how people think, learn, approach, and solve problems? What does a constructivist class look like? How can schools be reorganized to encourage this kind of instruction? Proposals should address the following topics in clarifying the nature of the research problem.

Relevance to important sociopolitical research issues of the day. As highlighted in the NSF program announcement, projects funded by EHR are intended to lead toward the improvement of instructional practice or school management. For example, researchers and practitioners have been concerned with developing a more complete understanding of reasons for performance differences among students from different racial and ethnic or gender groups. This interest was reinforced by the civil rights movement and the desire to eradicate the unacceptable gaps in performance between groups that had been observed. More recently, research studies have been influenced by the movement to develop education

standards and the programs and products that they have spawned. Studies that shed light on the extent to which the standards have been implemented and whether they had an impact on mathematics and science learning have been encouraged.

Projects should be clear about how their research problem relates to salient educational policy issues. How does what is being done have the potential for affecting teaching and learning in both the short and longer term? Who will benefit? What use is the study expected to be to the field? Good projects have part of their focus on an issue that can be eventually applied to improved performance in school systems. Another part of the justification for the study should set the research problem in a broader context that demonstrates how it would add cumulative knowledge to understanding education practice in the long term. The justification for any project should provide clear connections between what is being proposed and how the exploration of the topic might be expected to inform the current educational debate.

Importance of deepening our understanding of the content of education; how students learn in mathematics and science. Central to NSF’s mission is helping students to learn mathematics and science with understanding so that all students have the skills and abilities to solve nonroutine problems in varying contexts and situations. This has two important implications for research studies submitted to NSF: they should be strongly anchored in the specific and unique features of the mathematics and science domains; and they should explore how students learn for understanding.

Special attention should be given to developing and justifying studies that can be called “pivotal studies” (Linn, comment at workshop). Pivotal studies challenge our traditional concepts of learning in a discipline and help us to interpret facts and behaviors in a new light. They may challenge strongly held beliefs about how subject matter should be presented, for example, a focus on depth over breadth or how skills should be clustered or sequenced as in how integrated math is taught as compared to the more traditional mathematics sequences.

The value system underlying the research proposed. All research is embedded in a value system or set of beliefs of how the world is structured. For example, the positivist tradition had strong impacts on what was seen as an acceptable research model, data collection techniques, and ways of interpreting what was found. Today it is clear that while positivism still has a strong influence on the thinking of many researchers, multiple paradigms and value systems, sometimes conflicting, co-exist in our educational practices and strongly influence an individual’s approach to designing and exploring a research question. For example, adoption of constructivist models of learning may affect the very questions and vocabulary used to frame the study questions and, almost certainly, the measurement techniques to identify student learning. Thus, research proposals should be as explicit as possible in identifying the framing assumptions being made about the exploration to be undertaken.

Place of investigation in a developing line of research; evidence of linkage to future studies. New research should be related to relevant prior work, and prospective investigations should provide documentation that places underlying philosophies, specific research questions, methodologies, and outcomes in the context of that existing research. This connection to the field is done to incorporate ideas from others and to make a case for the value of the work being proposed. Many new research projects are cross-disciplinary, and methodologies from one field are being applied to another. In such cases, it is important to introduce the new methodologies so that all reviewers will be acquainted with the approach. The proposal could present a short history of previous uses of the approach, explain why application to a new area might be expected to be successful, and discuss differences or adaptations that are being considered.

Just as research projects emerge from previous research efforts, they also have implications for investigations that have yet to be fully formed. Few projects are stand-alone events; more frequently, projects are best understood as that part of a research program or series of efforts designed to more fully explore a broad idea or approach. Consequently, in presenting a proposal for a research study, it is important to show how the study fits with other ongoing or planned events. Such explanations provide a stronger rationale for the particular strategies being proposed. The argument for a case study or investigation focused on a small sample may be strengthened by providing a context for the study as part of a series of in-depth explorations that, taken together, might be considered a more robust examination of the problem. In a similar way, the value of a nationally representative survey may be enhanced if it is understood that other efforts may provide a more detailed examination of critical exemplars explored in a more limited way through survey techniques.

The Research Procedures

The procedures section of a proposal presents the overall approach to carrying out the study, taking into account theoretical, technical, practical, and ethical concerns. The emphasis should be on describing not only what is proposed, but also on why the procedures advance understandings of both substantive and methodological issues.

Overall approach and coherence. Proposals should have a strong internal coherence in terms of questions, design, and data analysis; procedures should be explained and justified as compared to other procedures; and proposal writers should show awareness/understanding of new and emerging ideas/techniques, which may be statistical, methodological, or conceptual.

For example, Marcia Linn spoke at the workshops about studying the inquiry method of investigation as implemented in science classrooms. She explained that researchers need to develop methods appropriate for understanding how students learn to think broadly and not just memorize topics that can be investigated with simple tests. The methods of developing tests and conducting surveys of student opinions are not adequate for describing the classroom experiences of students and teachers under these models of student learning. Rather, what is needed are more in-depth descriptions of the learning process that explore how teachers develop and support inquiry-based learning, how they balance constructivist and didactic approaches, and how they can take multiple paths to support student learning for understanding rather than recall. Similarly, in such a study, a sampling frame, data collection schedule, and analytic techniques must be proposed that provide for rich narrative descriptions of how learning occurs. These techniques should be related back to the conceptual model of teaching and learning under investigation. In such intensive studies of learning behavior, large samples of students and teachers or one-time sampling of behavior are likely to be of marginal relevance because they do not capture the dynamic nature of inquiry.

Research design. A proposal should describe the design and explain why or how it is appropriate for the questions to be addressed. It should also discuss the developmental status of research (how the project will change as it proceeds) and the constraints of the situation in which it will be carried out. How will the research design support the goal of describing more accurately and fully what it takes to increase our knowledge of how students learn with understanding?

Specific attention should be given to describing the treatment, the samples, the time frame, and the analytic techniques that will be used. Proposals should explicitly address whether control or comparison groups will be employed and discuss the rationale for inclusion or exclusion. Some designs can apply

control groups. In other cases, it may be more appropriate to appeal to “standards” rather than control groups to judge the efficacy of the education activity. (A helpful discussion of types of research designs can be found in Romberg, 1992.)

Instrumentation. Proposals should clearly specify the types of quantitative and qualitative techniques that will be used to collect data, along with a rationale for why the technique was selected. Researchers are encouraged to try new ways of collecting data, drawing on approaches from other fields and making advances in the use of new technologies.

When discussing new types of instruments, it is important to provide evidence that the instruments meet the quality standard for the field. Existing instruments to be used should be those with established soundness for the research questions and populations to which they will be applied. Where new instruments are proposed, procedures for establishing their soundness should be described.

Procedures will vary with the instrumentation used. Classical reliability and validity measures, as defined by quantitative researchers, are useful only for quantitative measures. Fairness or lack of population bias is also a high priority. Qualitative researchers have long taken exception to the way quality has been assessed, positing the importance of the soundness of data-gathering procedures over stability of outcomes. They argue that validity is the only meaningful criterion and reliability should be de-emphasized (Denzin and Lincoln, 1994). Indeed, the value of stressing validity over reliability is a theme that today cuts across the traditional quantitative and qualitative methods distinction.

- As new types of assessment techniques have replaced standardized, multiple-choice tests, even quantitative researchers have begun to question traditional approaches to establishing quality and have argued for validity over reliability.
- Researchers whose approach focuses on providing in-depth descriptions of changes in learning processes in relationship to changes in learning conditions also find the notion of reliability to be inappropriate. Instead, evidence of authenticity in the situations being presented and the measurement techniques used is seen to be of paramount importance.
- The instruments are based on new and emerging technologies, the challenge becomes more complicated, and some new areas may need to be considered. For example, if video techniques are proposed, the researcher should consult existing handbooks by those who have conducted extensive studies (see Fernandez, Rankin, and Stigler, 1997).

While the absence of a single set of criteria for quality makes the task of judging the soundness of instrumentation more difficult, and even to some extent subjective, it does not make the need for assuring quality any less important. Researchers should show that they have a deep understanding of the criteria commonly accepted for soundness for the instruments proposed and present evidence that these criteria have or will be met.

Feasibility. Research proposals should document an awareness of what needs to be done to carry out studies in a situation the researcher does not control. Special emphasis will be placed on practical concerns that need to be addressed in researching school settings.

Researchers familiar with the constraints of school settings recognize that it is frequently necessary to make some tradeoffs in the requirements of research designs in order to be allowed to conduct studies in school and other real-life settings. Preplanned activities in schools, such as excursions, may conflict

with scheduled exams, and teachers participating in a study may choose to not participate after weeks of involvement. Procedures that may be possible in a laboratory or some other setting may be impossible to implement in the school setting, where the business of educating students has the highest priority.

- While, for example, from a research point of view, multiple lab sessions of 1 to 2 hours might be desirable for studying the acquisition of a particular concept, it may be impossible to remove students from their regular instruction for such lengths of time. Shorter or fewer sessions may be the only choice possible.
- Those wishing to employ techniques such as videotaping of teachers' instructional practices may find that the videotaping procedure is not possible or would not be effective because of the layout of the room or learning spaces. What is videotaped may have to be altered to suit the physical shape of the setting.
- A research proposal that includes alteration of a curriculum in mathematics or science in a school may not be permitted by the school administration unless all the objectives included on the state's high-stakes testing program are covered. What is taught may have to be determined by accountability concerns before the research concerns.

In developing proposals for research in such settings, researchers should provide convincing evidence that they are aware of and have strategies for dealing with constraints that may be placed on their activities. A proposal for school-based research that appears to assume the control available in a laboratory is likely to be questioned. Some factors to address include (1) the timing and duration of the research activities; (2) plans for obtaining permissions that need to be granted to work with schools and students; (3) provisions for review of instruments, procedures, or reports; (4) constraints on the kinds of questions that can be addressed; and (5) acknowledgment of the requirements that may be imposed regarding interaction with special needs students.

Generalizability. Proposals should discuss how they will address issues related to defining or establishing the generalizability of their research and findings to other settings, that is, how the study will address concern for the potential impact of research on other sites, in other situations, moving from the research setting to real-life applications, etc.

Many forms of replicability should be considered. Some are “local,” so that conjectures originally developed in one classroom or with one student are then further explored for robustness and replicability with another. Other forms of replicability are more distal, for example, a reform at one school that is implemented in another. In all cases, the researcher must show sensitivity to the importance of identifying the right description of what it is that one would expect to replicate. Listing the observable materials and activities does not constitute that kind of description, although most discussions of replication assume it does. Successful sustainability and scaling up require the capability to capture the germ of the reform, idea, or product.

Ethics. Research proposals should show an awareness of ethical issues that may be involved with the project. The researcher should show how decisions may be made throughout the research project, from planning and implementation to analysis. The proposal, or related human subjects certification, should discuss how such issues related to privacy and confidentiality will be addressed, that is, what safeguards will be put into place to make sure that the work and reports that come out do not damage the individuals who have participated. The integrity and anonymity of subject—teachers, administrators,

children, and parents—must be respected. Clear statements need to be made regarding who will own the data and who will report on them.

Researchers are well aware of the need to safeguard the privacy of human subjects and to make sure that their participation in a research project does not place them in any personal jeopardy because of what they do or say. Indeed, many projects would be impossible to conduct if participants felt that their opinions, perceptions, or behaviors were to be attributed to them in any specific way. In large-scale studies, it has been fairly easy to provide confidentiality by reporting data only at the group level and by placing limitations on the size of the group for which data will be presented in any disaggregated form. (Usually, the requirement is at least 10 subjects in a group.)

Where small samples are used, assurance of confidentiality may pose a significant challenge. Proposals should address the issue of confidentiality and explicitly address how the rights of subjects will be protected, even if that protection may limit some aspects of the research itself. If only a small number of people will be able to recognize the identity of a respondent, that recognition may be sufficient to cause personal embarrassment and harm. Sowder (1998) points out that some research has led to, and perhaps even rested on, a relationship of trust between the researcher and the subject. Thus, the researcher is duty bound to address the manner in which the data will be presented since presentation can have serious personal consequences.

Researchers who collect large data sets that might be used by others should explain in the proposed statement of work that they have plans for making the data available to others for secondary analysis. It is recommended that all data sets be released to other researchers, with complete documentation, by 1 year following the publication of results.

Final Comments

This workshop was an effort by the Division of Research, Evaluation and Communication to systematically engage principal investigators of funded projects in a discussion of qualities that define the best research studies. The conversation was lively and productive, although it was not completed at the end of the scheduled 2 days. The result was an understanding that educational research must not be limited by a single set of methods and that research results should reflect the rich nature of education experienced by students. This report has attempted to report the essential findings of that workshop to a broader audience to stimulate further efforts to improve the quality of education research. The work to improve educational research is an ongoing effort.

REFERENCES

- Agar, M. (1980). *The professional stranger. An informal introduction to ethnography*. San Diego, CA: Academic Press.
- Bell, P., and Linn, M.C. (1999). Scientific arguments as learning artifacts: Designing for learning from the Web with KIE. Submitted to the *International Journal of Science Education*.
- Brown, A.L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions. *The Journal of the Learning Sciences*, 2, 137-178.
- Campbell, D.T., and Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin Company.
- Cobb, P. (1999). Individual and collective mathematical learning: The case of statistical data analysis. *Mathematical Thinking and Learning*, 1, 5-44.
- Cobb, P. (in press). Supporting the improvement of learning and teaching in social and institutional context. In *Cognition and Instruction: 25 Years of Progress*, edited by S. Carver and D. Klahr. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cohen, E.G., and Lotan, R.A. (1997). *Working for equity in heterogeneous classrooms: Sociological theory in practice*. New York: Teachers College Press.
- Collins, A. (1999). The changing infrastructure of education research. *Issues in education research*, chapter 13, edited by E.C. Langemann and L.S. Schulman, San Francisco, CA: Jossey-Bass.
- Cronbach, L., and Suppes, P. (Eds.). (1969). *Research for tomorrow's schools: Disciplined inquiry for education. A report of the Committee on Educational Research of the National Academy of Education*. New York: Macmillan.
- Denzin, N., and Lincoln, Y. (Eds.). (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- Ducharme, M.K., Licklider, B.L., Matthes, W.A., and Vannatta, R.A. (1995). *Conceptual and analysis criteria: A process for identifying quality educational research*. Des Moines, IA: FINE Foundation.
- Eisenhart, M.A., and Borko, H. (1993). *Designing classroom research: Themes, issues, and struggles*. Boston: Allyn and Bacon.
- Eisenhart, M.A., and Howe, K.R. (1992). Validity in educational research. In *The handbook of qualitative research in education*, edited by M.D. LeCompte, W.L. Millroy, and J. Preissle. New York: Academic Press.
- Fernandez, C., Rankin, S., and Stigler, J. (1997). *Videographics handbook: Video tape procedures for TIMSS*. International Association for the Evaluation of Educational Achievement (IEA). Duplicated.

-
-
- Flinders, D.J., and Mills, G.E. (Eds.). (1993). *Theory and concepts in qualitative research: Perspectives from the field*. New York: Teachers College Press.
- Gall, M.D., Borg, W.R., and Gall, J.P. (1996). *Educational research: An introduction*, Sixth Ed. New York: Longman.
- Gephart, W., and Ingle, R., (1969). *Educational research: Selected readings*. Columbus, OH: Charles E. Merrill.
- Kelly, A. E., and Lesh, R. (2000). *Handbook of Research Design in Mathematics and Science Education*. Mahwah, NJ: Erlbaum.
- Labaree, D.F. (1998). Educational researchers: Living with a lesser form of knowledge. *Educational Researcher*, 27, 8, 4-12.
- Lagemann, E.C., and Schulman, L.S (1999). *Issues in education research*. San Francisco, CA: Jossey-Bass.
- Linn, M.C. (1995). Designing computer learning environments for engineering and computer science: The scaffolded knowledge integration framework. *Journal of Science Education and Technology*, 4(2), 103-126.
- Mickelson, R.D. (Forthcoming). The effects of segregation and tracking on African American high school achievement. *Journal of Negro Education*.
- Mickelson, R.D., and Smith, S.S. (Forthcoming). *All that glitters is not gold: The outcomes of educational restructuring in Carolina, North Carolina*. Education Evaluation and Policy Analysis.
- Romberg, T. (1992). Perspectives on scholarship and research methods. In *Handbook of research on mathematics teaching and learning*, edited by D.A. Grows, Ch. 3. New York: Macmillan.
- Romberg, T., and Shafer, M. (In press). Mathematics in context: *Evidence about student outcomes*, in NSF curriculum projects, edited by Senk and Thompson.
- Schneider, B., Swanson, C.B., and Riegler-Crumb, C. (1998). Opportunities for learning: Course sequences and positional advantages. *Social Psychology of Education*, 2, 25-53.
- Schultz, S.E. (1999). To group or not to group: Effects of group interaction on students' declarative and procedural knowledge in science. Unpublished dissertation, Stanford University.
- Sowder, J.T. (1998). Ethics in mathematics education research. In *Mathematics education as a research domain: A search for identity*, Book 2, edited by A. Sierpiska and J. Kilpatrick, pp. 427-442. Kluwer Academic.
- Spindler, G., and Spindler, L. (1992). Cultural process and ethnography: An anthropological perspective. In *The handbook of qualitative research in education*, edited by M.D. LeCompte, W. L. Millroy, and J. Preissle. New York: Academic Press.
- Stokes, D. (1997). *Pasteur's Quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institute.
-
-

Swanson, C.B., and Riegle-Crumb, C. (1999). Early steps to college success: High school course sequences and postsecondary matriculation. Presented at the American Educational Research Association Annual Meeting, Montreal.

U.S. Department of Education, National Center for Education Statistics. (1991). *SEDCAR (Standards for education data collection and reporting)*. Washington, DC: U.S. Department of Education.

U.S. Department of Education, National Center for Education Statistics. (1992). *NCES statistical standards*. NCES 92-021r. Washington, DC: U.S. Department of Education.

Yin, R.K. (1994). *Case study research. Design and methods*, Second Ed. Thousand Oaks, CA: Sage Publications.

Appendix A:

**Participants in Workshop on Research Methods
National Science Foundation
November 19-20, 1998**

Organizer: Larry E. Suter

Classroom Teachers Group

Kathy Borman	KBorman813@aol.com
Leona Schauble	leona@vms2.macc.wisc.edu
Rich Lehrer	rlehrer@facstaff.wisc.edu
Paul Cobb	cobbp@ctrvax.Vanderbilt.Edu
Ron Marks	ronmarx@umich.edu
Valerie Williams	Valerie_williams@rand.org

Curriculum Group

Joe Krajcik	krajcik@umich.edu,
Judith Sowder	jsowder@sciences.sdsu.edu
Rodney McNair	rmcnair@udel.edu
Kenneth Forbus	forbus@aristotle.ils.nwu.edu
Yasmin Kafai	kafai@gseis.ucla.edu
Robert Donmoyer	rdonmoye@magnus.acs.ohio-state.edu
Jan Hawkins	jhawkins@edc.org
William Sowers	sowers@cosmoscorp.com

Multilevel Group

Ron Anderson	rea@iea.soc.umn.edu
Barry Sloane	FCSLOANE@hewitt.com
Thomas Hoffer	thoffer@norcmail.uchicago.edu
Jim McLean	jmclean@uab.edu
Curtis Tatsouka	tatsuoka@pstat.ucsb.edu
Hugh Cline	hc401@columbia.edu
Uri Wilensky	uriw@media.mit.edu

Student Learning Group

Tom Romberg	Tromberg@facstaff.wisc.edu
Marcia Linn	mclinn@socrates.berkeley.edu
Dick Lesh	Rlesh@purdue.edu
Dick Venezky	venezky@mail.eecis.udel.edu
Angela O'Donnel	angelao@rci.rutgers.edu
Ricki Goldman-Segall	ricki.goldman-segall@ubc.ca
Joseph Conaty	joseph_conaty@ed.gov

NSF Participants

Eamonn Kelly, Elizabeth VanderPutten, Bernice Anderson, Eric Hamilton, John Cherniavsky, William Sibley, Diane Scott-Jones , Eugenia Toma, John Hunt

Westat: Joy Frechtling frechtj1@westat.com

About the National Science Foundation

The National Science Foundation (NSF) funds research and education in most fields of science and engineering. Grantees are wholly responsible for conducting their project activities and preparing the results for publication. Thus, the Foundation does not assume responsibility for such findings or their interpretation.

NSF welcomes proposals from all qualified scientists, engineers and educators. The Foundation strongly encourages women, minorities, and persons with disabilities to compete fully in its programs. In accordance with federal statutes, regulations, and NSF policies, no person on grounds of race, color, age, sex, national origin, or disability shall be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving financial assistance from NSF (unless otherwise specified in the eligibility requirements for a particular program).

Facilitation Awards for Scientists and Engineers with Disabilities (FASED) provide funding for special assistance or equipment to enable persons with disabilities (investigators and other staff, including student research assistants) to work on NSF-supported projects. See the program announcement or contact the program coordinator at (703) 306-1636.

The National Science Foundation has Telephonic Device for the Deaf (TDD) and Federal Relay Service (FRS) capabilities that enable individuals with hearing impairments to communicate with the Foundation regarding NSF programs, employment, or general information. TDD may be accessed at (703) 306-0090 or through FRS on 1-800-877-8339.

The National Science Foundation is committed to making all of the information we publish easy to understand. If you have a suggestion about how to improve the clarity of this document or other NSF-published materials, please contact us at plainlanguage@nsf.gov.

NATIONAL SCIENCE FOUNDATION
ARLINGTON, VA 22230

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE \$300

RETURN THIS COVER SHEET TO ROOM P35 IF YOU DO NOT WISH TO RECEIVE THIS MATERIAL , OR IF CHANGE OF ADDRESS IS NEEDED , INDICATE CHANGE INCLUDING ZIP CODE ON THE LABEL (DO NOT REMOVE LABEL).

**PRESORTED STANDARD
POSTAGE & FEES PAID
National Science Foundation
Permit No. G-69**

NSF 00-113