Eisenhart, M. (2006). Qualitative science in experimental time. *International Journal of Qualitative Studies in Education, 19*(6),697-707.

Qualitative Science in Experimental Time

Margaret Eisenhart*

Pre-publication Copy

**** School of Education, University of Colorado, USA;***
margaret.eisenhart@colorado.edu

Abstract

I address the "state of qualitative inquiry" in the sense of how that inquiry is being positioned in the current construction of a U.S. national policy agenda for "scientifically-based" education research. In my view, qualitative inquiry is being drowned out in the national agenda despite its ability to provide the kinds of answers about education that policymakers and others want. The drowning out is accomplished in part by discursive conflating of "experimental research" and "scientifically-based research" and by use of the phrase "gold standard" to position experiments as *the* exemplar of good and rigorous research in education. I critique these political-linguistic moves and suggest two predominantly qualitative forms of inquiry--interpretive and practical science--as alternatives to experimental science. I end with some thoughts, based in cultural anthropology, for improving the position of qualitative inquiry in the current political environment surrounding education research.

Author's Biographical Note

Margaret Eisenhart is University Distinguished Professor and Charles Professor of Education, School of Education, University of Colorado, Boulder, CO 80309-0249, USA; office: 303-492-8583; fax: 303-492-7090; email: margaret.eisenhart@colorado.edu.

Qualitative Science in Experimental Time

Introduction

"Scientifically-based research" has become the watchword in educational research in the United States today. Everyone from teachers and faculty to curriculum developers, policy makers, Congressional leaders, and the President of the United States is talking about what it means to do scientifically-based research in education and how it can be accomplished. I have had a part in this conversation because I, a cultural anthropologist and qualitative researcher, served on the National Research Council Committee that produced the report, *Scientific Research in Education* (NRC, 2002); my perspective here is colored by that involvement. In this article, I want to develop further two points about the meaning of "scientifically-based research" that I first made shortly after the publication of the NRC report.

The first point is that scientifically-based research has erroneously come to be described and acted on as if it were synonymous with experimental research (research that answers questions about "causal effects") by President George W. Bush and his leadership appointees in the U.S. Dept. of Education and its research arm, the Institute of Education Sciences (IES) (Angrist, 2004; Eisenhart & Towne, 2003). This is a very unfortunate development, not only for qualitative researchers but also for education.

The current government's preference for experimental research is abundantly evident—in the language of federal legislation, e.g., the No Child Left Behind Act of 2001, in the establishment of the What Works Clearinghouse (www.w-w-c.org), and in the funding priorities of IES (www.ed.gov/programs/). This preference is being further promoted by academic references to experimental research as the gold standard for education research (e.g., Raudenbush, 2005; Slavin, 2002). A lead standard might be a more fitting moniker. While I think that a few well-chosen, federally-sponsored experimental studies could be worthwhile, giving priority to experimental research in education is seriously misguided.

The second point I want to make is that there are other research designs that are much more promising, and qualitative research is central to them. To some people, arguing about the value of qualitative research in education is old news from the 1980s. I know this: I was there, and I did that. But I think that the present political and social climate in the U.S. requires qualitative researchers to try again, even more forcefully if possible, to articulate why experimental research in education will always be relatively limited and why qualitative research promises more satisfying answers to many important research questions (see also Smyth, 2006).

The Limits of Experimental Research

The idea that there is only one way to do scientifically-based research—in education research or in any other field—is ridiculous (see also Berliner, 2002). No thoughtful person can possibly believe that experimental designs or randomized field trials are, or should be, the “gold standard” for addressing all the questions in need of answers in education. That experimental designs are now required by law in No Child Left Behind, privileged in other federal legislation governing education, and singled out for funding by the U.S. Department of Education and the Institute for Education Sciences, is absurd. Particularly with regard to public-interest questions about educational policy and practice, experimental research will almost certainly miss the mark. Other forms of empirical study, including interpretive, practical, critical, and historical, are just as likely if not more likely to produce useful results in education, to do so more quickly, and to do so for less money.

Before saying more about these other empirical designs, I must set the stage by saying a little more about the NRC report, *Scientific Research in Education (SRE)*. In *SRE*, the committee argued that "scientifically-based research in education" attempts to answer *three kinds* of research questions with empirical evidence (NRC, 2002, pp. 99-123; see also Eisenhart, 2005):

1. Descriptive questions. Descriptive questions ask about what is going on. Some examples are: What are students learning about reading or math or identity or democracy in this class? What is actually occurring in the classroom when an innovative mathematics program is being implemented? What are the students doing while the teacher is presenting the curriculum? What does it mean to students to be labeled "lacking proficiency" in math or reading?
2. Causal questions. Causal questions ask about cause-effect relationships. For example, what is the relationship between a source (a program, a person, an event, etc.) and an outcome (an achievement, a result, a consequence, etc.)? What is the relationship between class size and student achievement in reading? What is the relationship between identifying oneself as a "weak student" and dropping out of school?
3. Process (or explanatory) questions. Process questions ask about how or why something happened. For example, why or how did a source (see #2) produce

> a particular outcome (i.e., what is the process by which a source produced an outcome)? Why or how does class size affect student achievement (i.e., what is the explanation for the affect of class size on achievement)? How or why does identifying oneself as a weak student lead to dropping out of school?

To me, these question types are crucial to understanding the meaning of scientifically-based research as described in *SRE*, yet they seem to have been overshadowed by the debate about whether various forms of qualitative research can or should be considered "scientifically-based."

I now realize that the section of *SRE* in which these question types appear can be read in such a way as to suggest that the second type is the most important (Erickson & Gutierrez, 2002). In the report, the second question type is glossed as the one that directly answers the "what works" question, the one that "establishes causation" or the "underlying causes," and the one for which randomized field trials are described as the "ideal method" (pp. 108-112). In this framing of the questions, answers to questions 1 and 3 become either preliminary (Question 1 becomes a precursor to formulating good hypotheses for experimental testing) or secondary (Question 3 becomes a useful addition if you can figure out a plausible process, mechanism, or explanation). I do not agree with this ranking of the three questions. In fact, I believe that answers to descriptive and process questions are far more useful to educational policymakers, practitioners, and the public than answers to questions about causal effects.

Before I proceed to develop this point, I should say that I do think that answering questions about causal effects and conducting experimental research to identify causal effects can be valuable in some cases. For example, if you want to know whether reducing class size is likely to be associated with increased achievement in a particular context, experimental research can provide an answer, assuming the study is feasible and all else is equal or constant. Experimental research about causal effects isolates variables, such as class size, and investigates their relationship to other variables, such as achievement. Repeated experiments with the same variables lead to reliable estimates of their effect sizes. This procedure will work if the variables being measured do not change over time, are not variably influenced by circumstances, and are not affected by human intention or desire.

The problem is that in education and other social practices all else is not equal, and little if anything remains constant. As Erickson & Gutierrez have written:

> Educational treatments…are locally constructed social ways of life involving continual monitoring and mutual adjustment among persons, not relatively replicable entities like chemical compounds or surgical procedures or hybrid seed corn or manufactured airplane wings. (2002, p. 21)

The way causal effects must be operationalized in experimental research designs cannot accommodate these complex contingencies.

> The variety and changeability of the hierarchically embedded contexts of social life are such that simple, consistent associations between generic cause and generic effect of the sort tested in formal social experiments are not likely to occur. From this point of view the level of abstraction in the operational definition of aspects of social process as unitary variables—which characterizes large-scale social experiments—results in knowledge

> that at best can be characterized as rough approximation or guesswork.
>
> (Erickson & Gutierrez, 2002, p. 23)

The limitations of experimental research were clearly demonstrated in the aftermath of the Tennessee class size experiment (Finn & Achilles, 1990). In this state-wide Tennessee experiment, the researchers found that when class size in schools was manipulated, achievement scores went up as class size went down. But when the experiment was conducted again in California, achievement levels did not increase with class size reductions (Angrist, 2004; Biddle & Berliner, 2002). There are numerous reasons for this difference: The most important is that the conditions of manipulating class size in Tennessee were not replicated in CA, e.g., “small” classes in CA were bigger than “small” classes in TN. In consequence, what “worked” in Tennessee did not work in California—through no fault of the experimental design but because the requirements of the design (replication of the implementation) could not be met in the practical context of educational intervention in California.

There is no reason to think this is an isolated case. As many others have shown (e.g., Berliner, 2002; Erickson & Gutierrez, 2002; Howe, 2004), educational contexts are inevitably varied, dynamic, over-determined, and interrelated; thus, experiments done in one context using necessarily finite numbers of variables are quite unlikely to be reproducible in others, and their results are quite unlikely to be realized elsewhere. The claim that experimental research will allow policymakers and practitioners to distinguish educational interventions that work “at scale” and over time is spurious. And since the promise of replication and generalizability are prime bases for elevating experimental research to the gold standard in education research, those bases are eroded.

This situation also raises questions about the amount of money and effort spent to set up educational experiments and how long it takes to obtain experimental results. Experimental studies are difficult and expensive to conduct (Boruch, 2002)—in the case of the Tennessee class size study, the study required state legislative authorization, millions of dollars of state funding, more than a million dollars of research funding, and more than a decade of work. Now, years later, reports of the study have been peer-reviewed and published, yet disputes continue about the best way to analyze the data, the significance of the findings, and their implications (Angrist, 2004). Further, in Tennessee where class size reduction "worked," the state could not afford to reduce class size to the level that worked in the experiment. Even if there were cases where it is possible to specify all the relevant variables and measure and model them properly, huge efforts must be mounted to set up experimental studies, large amounts of money must be spent to conduct them, long periods of time are likely to pass before anything definitive is known from them (if it ever is), and the implementation of their results may be infeasible. Meanwhile, the entire context of application will change. So, in addition to the fact that experiments do not produce the widely applicable results claimed for them, the large amounts of money, effort, and time devoted to them may well be wasted.

I have argued elsewhere that "scientifically-based research" applies to much more than experimental research (Eisenhart, 2005). So, what about other ways of doing science? Are they any more likely than experimental science to yield useful information for educational policy makers or practitioners? The answer is yes.

Interpretive Science

Interpretive science focuses on the webs and layers of meaning (the complex of interrelated symbols) that people who interact in a group draw on to make sense of their world and act purposefully in it (Geertz, 1973). The work of the interpretive scientist is (primarily) to do ethnography: i.e., to learn how to translate symbol systems that are meaningful to one group into terms that are understandable to another group. Interpretive research questions about education might include: What does school achievement mean to students who are low achievers? How does doing science or engineering come to have such negative meanings for U.S. women? How does "…the meaning of schooling get contested, negotiated, or re-invented across multiple streams…of practice?" (Nespor, 1997, p. xii). In interpretive research, questions of this type provide answers about *what* is going on (question type 1 from *SRE*) in a locally inflected and complex sense, thereby overcoming the narrow focus of isolated variables. These questions also can produce answers to *how or why* things happen as they do (question type 3 from *SRE*) in a particular situation. In effect, they erase the need for a separate investigation of causal effects. If you know in detail *what happened* and you know *how or why* it happened, it seems to me that you are usually informed enough to take action—as a teacher, researcher, a policy maker or a president. For example, if we in the US know *what happened* in Iraq: we went to war; and we know *why* we got into this war: our leaders believed that Saddam Hussein's regime was a threat to our economy and our way of life, do we really need to know what particular cause or set of causes (fear of weapons of mass destruction, worries about losing oil supplies, or concern about human rights) was the main or underlying one *in order to decide what we should do now?* Why? What difference does it make for our actions and plans *now* to know the main *cause* or the *effects of that cause now*? If it does make a difference, is that difference large enough to justify large amounts of money and long periods of time devoted to studies that might determine the causes and their effects? (Note that I am not disputing the value of having good information in order to make a decision about launching the war in Iraq. My point here is that determining causal effects (always done after the fact) often is not useful for making decisions about policy or practice.)

Similarly, if we know that a child is struggling to read, and we know why (he believes that it is uncool to be seen as someone who reads), is it more important for us to know what caused him to develop this belief or how to act now that he has this belief? It seems to me that finding answers to *SRE*'s question types 1 and 3 provides perfectly adequate information for making decisions about what to do next in most educational situations.

Interpretive researchers have already provided helpful answers to educational question types 1 and 3, although they seem largely invisible to policy makers. Anthropologist Signithia Fordham's book, *Blacked Out* (1996), for example, describes in detail how and why black students at a Washington, DC, high school struggle with the meaning of school work and achievement. The particular actions and beliefs of the students in her study may not appear elsewhere, but the forms of resistance to schooling that she identified are widely applicable to minority students in US schools. Anthropologist Norma Gonzalez, in her book, *I Am My Language* (2001), describes how Mexican-American students in Tucson make sense of using both English and Spanish in their lives. Here again, the particular characteristics of the students are not likely to generalize, but the emotional loading of language use that she identified is clearly relevant to bilingual education wherever it takes place. Sociologist Mary Metz (1990) describes an ideology of schooling that she discovered in a study of several Midwestern high schools in the 1980s. At that time she found a common script for a legitimate U.S. school (which she referred to as "The Real School"); today that same script is evident in NCLB and other current policy initiatives (Metz, 2005).

These patterns—of resistance, language loading, and taken-for-granted understandings of what makes a "real school"—are the kinds of generalizations that result from interpretive science. They represent (in some detail) descriptions that capture and recurring processes that explain (in part) major educational problems, including low achievement, weak second language skills, and sometimes excessive discipline, in the U.S. context. These processes do not lend themselves to easy solutions or quick fixes, but they could be widely exposed, publicly debated, and addressed with respect to particular schools or communities. Some local initiatives to address the processes might be costly and long term, but many would not be. Presumably, they would all be intentional and relevant in context (at least for awhile). It seems that time, effort, and money are less likely to be wasted than with experimental studies and attempts to implement them.

Practical Science

Practical science is another promising possibility for education research. Taking his cue from Aristotle's intellectual virtues, Bent Flyvbjerg (2001), Danish Professor of Planning, describes practical science as empirical inquiry that aims to understand how knowledge comes to be important or consequential in practice. The value of practical science lies in its potential to illuminate how rationality (knowledge) is constructed and acted upon in public deliberations and decision making (pp. 142–143). Flyvbjerg advocates historical case study methods and analyses of power dynamics as the methods of inquiry for practical science.

Flyvbjerg uses an example from his work on urban planning in Aalborg, Denmark to illustrate his argument for the value of practical science. I have summarized his example here so it can serve as a resource for educational researchers interested in similar kinds of research. The context described by Flyvbjerg was a decision on the part of the Aalborg City Council, backed by the public, to improve the quality of the downtown urban core. To this end, the City Council approved a proposal prepared by urban planners to prohibit cars and to increase public transportation, bicycle paths, and walking corridors. However, as time passed and changes were actually implemented, cars were not prohibited, the number of cars increased, and the quality of the downtown core (in terms of air pollution, traffic, and health concerns) deteriorated even further. In this case, Flyvbjerg knew what had happened; his research was intended to find out how and why it happened.

His research design was to trace historically the interests and actions of key persons and groups, their deliberations and decisions, the policies that were implemented, the compromises that ensued, representations of events in the media, and the outcomes that were produced. Using data from documents, interviews, surveys, and analyses of business trends and traffic patterns, he found that the power dynamics between two key groups — the City Council and the Chamber of Commerce — privileged the Chamber's position that the vitality of the downtown core depended upon shoppers who drove there. Although survey data later showed otherwise, the Chamber held firmly to its position and used its power to influence decisions and to ensure that media accounts supported its view. Over time, this led to changes to the original plan and to outcomes (lower quality of life) that neither the City Council nor the public desired. He writes:

> The fate of the Aalborg Project would be decided by these two rationalities [the knowledge/ position of the City Council and that of the Chamber of Commerce] fighting it out, and the group who could place the most power behind their interpretation of what was rational and what was not would win…. Distorted relations of power produced a distorted project. Power thus defined a reality in which the real Aalborg Project, that which has become a reality, deviates from and on principle objectives directly counteracts the formal [original] Aalborg Project, which was ratified by the City Council with a vote of 25–1, but which exists only on paper (pp.147–148, 154).

The kind of research questions Flyvbjerg addressed and his approach to conducting research are germane to many research questions in education: How and why does a decision to require more standardized testing lead to desirable outcomes in some cases but undesirable ones in others? How and why does the adoption of a constructivist math program lead to desirable (or undesirable) outcomes? How and why does a school choice policy lead to desirable (or undesirable) outcomes? Practical science studies of this kind can reveal how consequences are produced over time by decisions made and actions taken in the minutiae of everyday practice, by who is involved and who is left out, and by whether the outcomes are desirable to those involved.

Similar to the case for interpretive science, the chronologies and processes revealed by practical science bring to light information that can be acted on by local leaders, parents, teachers, and others. Compared to what good experiments cost, how long they take, and how inconclusive or infeasible they are (as described by Finn & Achilles, 1990; Angrist, 2004; Biddle & Berliner, 2002), good ethnographies, historiographies, and narratives of interpretive and practical science seem to be a wiser investment. Although I do not know of any national figures on this topic, personal experience and an informal survey of qualitative researchers I know personally suggest that most interpretive and practical studies (even multi-sited ones) are done for less than $500,000 and are completed enough to be useful within ten years.

What's Next?

The absence of powerful discourses that can undermine the conflation of "scientifically-based research" and "experimental research" in the dominant political discourse about education research in the U.S. is a serious problem for qualitative researchers today. As I hope I have demonstrated here, interpretive and practical science offer alternatives that challenge experimental science, making it a costly, irrelevant, and ineffective approach to learning what needs to be known to take action on pressing educational problems. Yet, the dominant discourse continues to prevail, drowning out the contributions and benefits of these valuable alternatives.

To improve the state of qualitative research, those who favor it must find more potent ways to inform and educate policy makers and the public about interpretive and practical research. If we (qualitative educational researchers) do not agree with decisions to limit scientific research to experimental designs, or with the value of spending limited federal research dollars on expensive, non-generalizable and non-replicable experimental studies, then we need to find persuasive ways to speak out about what is wrong with this approach, what alternatives exist, and why they are valuable. I do not think qualitative and other "non-experimental" researchers should find this terribly hard to do. As I have tried to suggest with a few examples above, it should be relatively easy to provide evidence and good arguments for the relevance, applicability, generalizability, and cost-effectiveness of qualitative inquiry. Yet, such successful persuasion has not occurred; we still lack discursive forms that "work" to communicate our priorities.

Unfortunately, our efforts are hindered by the reputation of education research for weak scholarship, professional organization, and political presence, especially on the national level (Kaestle, 1993; Lagemann, 2001). In developing the Reading Excellence Act (1998) and No Child Left Behind (2001), for example, Congressional staffers developed standards for scientifically-based research in education without any widespread contact with education researchers and no contact (as far as I know) with qualitative researchers. This process continued for months with few education researchers, including qualitative researchers, taking any notice (Eisenhart & Towne, 2003). As a community, qualitative researchers in education are not a political player to be reckoned with. As with other dominated groups, we will likely have to work twice as hard to get half the respect we deserve. This is too bad, but I believe it is our present reality.

During the past several years while I have been focused on the issue of scientifically-based research in education, I have asked myself: What would an anthropologist studying this situation, with an eye toward changing the direction of federal policy toward a more favorable positioning for qualitative inquiry, make of it, and do? Of course, I cannot speak for everyone in a field as large and diverse as cultural anthropology. And of course, it is very hard to analyze what is going on as it unfolds. But in the traditions of cognitive anthropology (e.g., Frake, 1981; Tyler, 1969), applied anthropology inspired by cognitive anthropology (Agar, 1996), and diffusion research (Rogers, 1995), it is commonplace to find the advice to use key actors' framing of an issue (in language, representations, context, and interests) as a point of departure for initiating the changes you want them to make. Translating desired changes into an indigenous (or hegemonic) framing that makes the change worth considering or necessary, especially to those in a position to encourage or impede it, is often a first step in a successful project. This general idea is exemplified in the activist work of some academics, e.g., David Berliner and Bruce Biddle (1996) and George Lakoff (2004), who rely on common discursive forms to provoke or shame powerful Americans (including their fellow academics) into social action. (This does not imply that those involved will stay, or even can stay, within the confines of the initial framing—alternative framings are crucial if the project is to move forward. There can be no momentum for change unless desirable alternatives—new pieces of equipment, new ways of doing business, new goals, new ways of ordering priorities, new ways of talking about what is desirable or good, etc.--are available for consideration and use.) Research in the traditions of cognitive anthropology and diffusion suggests that identifying discursive forms that resonate with powerful actors in the context at issue is key to successful change.

Other anthropologists stress that cultural forms, including new discursive forms, become salient and consequential when they can be used productively in direct social interactions with other people (e.g., Carrithers, 1992; Erickson, 2004; Wortham, 2006). Serving on committees composed of policy makers with limited or different views, organizing professional meetings with methodological adversaries, or inviting your local legislator to lunch are some ways of creating the conditions for face-to-face exchanges between people who otherwise might never talk to each other. Once direct channels of communication are opened up, new ideas and new understandings also are likely to emerge.

Finally, some anthropologists have found that in hierarchical societies (like the U.S.), there are patterned ways in which knowledge and expertise that were once low status become prestigious (e.g., Carrithers, 1992). A common sequence is the following:

1. Define a body of knowledge that is the group's special province.
2. Establish the nature of rigor and standards that will be applied to experts in that knowledge.
3. Make it hard, not easy, to get access to that knowledge and to become an expert.
4. Build social capital with powerful people and groups.
5. Create tight bonds among the experts.
6. Add something recognized as valuable to society with the group's expert knowledge.
7. Create an organization of experts whose voice is clear and who cannot be ignored in public debates about their area of expertise.

Taking these steps may sound calculating or manipulative, but qualitative education researchers might usefully think of them as ways to work against the threat being posed by the current hegemonic discourse favoring experimental research in education. There is no doubt in my mind that qualitative inquiry is under attack (again) in the current political climate. It is in danger of being lost, ignored, and unfunded as more federal attention and limited monies go to experimental studies. It has not found a prominent place in public discourse or imagination about educational information (data), research, and improvement. Debates among qualitative researchers, although valuable for advancing research as a whole, are not likely to be the discursive form that sparks public interest in qualitative inquiry. If we wish to alter the current state of qualitative inquiry, we must find ways to position qualitative research (including those who do it) as indispensable to informed action on behalf of educational improvement.

Acknowledgements

Portions of this article have appeared previously in Eisenhart, 2005, and were presented at AERA, 2006. Special thanks to Beth Graue, Joe Harding, and the *QSE* reviewers for helpful comments on earlier versions.

References

Agar, M. (1996). *The professional stranger*. San Diego: Academic Press.

Angrist, J.D. (2004). American education research changes tack. *Oxford Review of Economic Policy, 20* (2), 198-212.

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

Berliner, D. & Biddle, B. (1996). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. New York: Perseus Books.

Biddle, B. & Berliner, D. (2002) Small class size and its effects, *Educational Leadership* (February), 12-23.

Boruch, R. (2002). The virtues of randomness. *Education Next*. Retrieved from http://www.educationext.org/20023/36.html, September 20, 2002.

Carrithers, M. (1992) *Why humans have cultures: Explaining anthropology and social diversity*. Oxford, UK: Oxford University Press.

Eisenhart, M. (2005) Hammers and saws for the improvement of education research, *Educational Theory, 55* (1), 245-261.

Eisenhart, M. & Towne, L. (2003) Contestation and change in national policy on "scientifically based" educational research, *Educational Researcher, 32* (7), 31-38.

Erickson, F. (2004). *Talk and social theory: Ecologies of speaking and listening in everyday life*. Cambridge, UK: Polity Press.

Erickson, F. & Gutierrez, K. (2002) Culture, rigor, and science in educational research, *Educational Researcher, 31* (8), 21-24.

Finn, J.D. & Achilles, C.M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27* (3), 557-577.

Flyvbjerg, B. (2001). *Making social science matter. Why social inquiry fails and how it can succeed again.* Cambridge, UK: Cambridge University Press.

Fordham, S. (1996). *Blacked out: Dilemmas of race, identity, and success at Capital High.* Chicago: University of Chicago.

Frake, C. (1981). *Language and cultural description.* Stanford: Stanford University Press.

Geertz, C. (1973). *The interpretation of cultures*. New York: Basic Books.

Gonzalez, N. (2001). *I am my language: Discourses of women and children in the borderlands.* Tucson: University of Arizona Press.

Howe, K. (2004). A critique of experimentalism, *Qualitative Inquiry, 10* (1), 42-61.

Kaestle, C. (1993). The awful reputation of educational research. *Educational Researcher, 22,* 23-31.

Lagemann, E. (2002). *An elusive science: The troubling history of education research.* Chicago: University of Chicago Press.

Lakoff, G. (2004). *Don't think of an elephant: Know your values and frame the debate.* White River Junction, VT: Chelsea Green.

Metz, M. (1990). Real school: A universal drama amid disparate experience, in: D. Mitchell & M. Goertz (Eds) *Educational politics for the new century*. Philadelphia: Falmer Press.

Metz, M. (2005). NCLB as a tool to build the myth of a single "American" school system, Paper delivered to the American Sociological Assocation, August 12, 2005.

National Research Council (2002) *Scientific research in education.* Washington, DC: National Academies Press.

Nespor, J. (1997) *Tangled up in school: Politics, space, bodies, and signs in the educational process.* Mahwah, NJ: Erlbaum Associates.

Raudenbush, S. (2005). Learning from attempts to improve schooling: The contribution of methodological diversity, *Educational Researcher*, *34* (5), 25-31.

Rogers, E. (1995). *Diffusion of innovations.* 4th edition. New York: The Free Press.

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

Smyth, J. (2006) Researching teachers working with young adolescents: Implications for ethnographic research, *Ethnography and Education*, *1* (1), 31-51.

Tyler, S. (1969). *Cognitive anthropology*. New York: Holt, Rinehart & Winston.

Wortham, S. (2006). *Learning identity: The joint emergence of social identification and academic learning*. Cambridge, UK: Cambridge University Press.