# SCIENTIFIC RESEARCH IN EDUCATION: AN ANALYSIS OF FEDERAL POLICY

By

Chad Robert Lykins

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Leadership and Policy Studies

August, 2009

Nashville, Tennessee

Approved:

Professor James Guthrie

Professor Robert Crowson

Professor Christopher Loss

Professor Denis Phillips

DEDICATION

*To everyone who has ever wanted to help someone*

*teach or learn something, but didn't know quite how.*

ACKNOWLEDGEMENTS

I could not have written this dissertation without the support and encouragement of a great number of individuals and institutions. My committee gave sage counsel which helped insure this project did not go off the rails. James Guthrie and Robert Crowson read early drafts of each chapter and helped correct some of my clumsier attempts to approach the topic. Christopher Loss guided me through the process of piecing together the historical narrative that eventually became the second and third chapters. Denis Phillips helped me see more clearly through the often confusing (and often confused!) criticisms and defenses of federal policy. I would also like to thank Stephen Heyneman, who took me on as his graduate assistant for two years and taught me a great deal about the craft of education research.

In addition to the support of these and other faculty members, I have been surrounded by smart, hard-working, funny, and generous classmates. Collectively they have provided thoughtful and honest feedback on every significant idea in the dissertation—and a great many insignificant ones. Moreover, their friendship made my years in Payne 207 highly enjoyable.

My doctoral studies were made possible by generous fellowships from Vanderbilt University, the Peabody College of Education and Human Development, and the Department of Leadership, Policy, and Organizations. These groups also provided travel grants to present drafts of various portions of the dissertation at conferences for the American Education Research Association, the American Education Studies Association, the University Council of Education Administration David L. Clark Graduate Student Seminar, the Philosophy of Education Society, and the Southeast Philosophy of Education Society. In addition, a Dissertation Writers Fellowship from the Vanderbilt Center for Ethics helped support me during my final year of writing.

Above all, I would like to acknowledge my wife, Ashley Lykins. In addition to being a loving partner and my best friend, she has gone with me into the intricacies of each argument as well as the tedious process of editing and revising each chapter. My appreciation of everything she has given goes beyond words.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

| | |
|---|---|
| CRA | Cooperative Research Act of 1954 |
| CRP | Cooperative Research Program |
| DOE | Department of Education |
| DWW | Doing What Works |
| ERIC | Education Resources Information Center |
| ESEA | Elementary and Secondary Education Act of 1965 |
| ESRA | Education Sciences Reform Act of 2002 |
| IES | Institute of Education Sciences |
| NAEP | National Assessment of Educational Progress |
| NCES | National Center for Educational Statistics |
| NCLB | No Child Left Behind Act of 2001 |
| NDEA | National Defense Education Act of 1958 |
| NERPPB | National Educational Research Policy and Priorities Board |
| NIE | National Institute of Education |
| NRC | National Research Council |
| NSF | National Science Foundation |
| OE | Office of Education |
| OERI | Office of Educational Research and Improvement |
| ORAD | Office of Reform Assistance and Dissemination |
| RAC | Research Advisory Council |
| REA | Reading Excellence Act of 1999 |
| SRE | *Scientific Research in Education* (2002) |

| | |
|---|---|
| WWC | What Works Clearinghouse |

# CHAPTER I

# THE POLITICAL ECONOMY OF EDUCATION RESEARCH

## Introduction

It is difficult to find an invective that has not been directed at those who produce education research. The field is said to have a "troubled history" (Lagemann, 2000), "awful reputation" (Kaestle, 1993), and "parochial concerns" with "narrow problems" (Goldhaber & Brewer, 2008). The researchers themselves are thought to be poorly trained. Some lament that they lack, "solid grounding in statistical methods" (Goldhaber & Brewer, 2008). Others state that they are naively "scientisitc" and incapable of critical thought (Hyslop-Margison & Naseem, 2007). The former director of the Institute of Education Sciences suggests that education researchers are more inclined to "postmodern musings" than serious reflection upon genuinely important educational problems (Whitehurst, 2003). A policy-maker once remarked that education research simply repeated what his fourth-grade teacher already knew (Kaestle, 1992).

The torrent of criticism extends to those who consume and disseminate education research. There seems to be a fair amount of suspicion that the consumers of education research use it the way a drunkard uses a lamppost—for support rather than illumination. They are said to lack the sophistication or inclination to appreciate differences in research quality (Goldhaber & Brewer, 2008). Those who have participated in recent federal efforts to improve the field have been labeled as "procedurally irrational" (Feuer, 2006). Their policy prescriptions have been equated to "quixotic quests" (Phillips, 2008), a fruitless "quest for certainty" (Baez & Boyles, 2009), or worse, as "endorsing ''terrorist' ideals of consensus" (St. Pierre, 2002) and "racialized, masculinist" assumptions (Denzin, Lincoln, & Giardina,

2006). It has been categorically stated that the most recent federal emphasis on what works simply "won't work" (Atkinson, 2000; Biesta, 2007).

This rhetoric often comes down to the accusation that education research lacks either relevance or rigor. The federal government has a long history of stepping in to try to correct the perceived deficiencies in education research. These interventions have touched every aspect of the research enterprise, including the topics, questions, designs, and methods, as well as the population of those engaged in education research. These interventions have aimed at making education a science which is both *relevant* to federal problems and *rigorous* enough to reliably solve them. However, notions of rigor and relevance are subject to "contestation and change" (Eisenhart & Towne, 2003).[1] Hence, the justifications behind these federal interventions are controversial.

## Purpose

The purpose of this dissertation is to evaluate the ethical and epistemological justifications for various proposed federal interventions in education research. The federal government supplies a significant amount of the total funds available for education research. Researchers have become increasingly dependent both on federal grant money and on the large databases maintained by federal agencies. Thus, federal policy on education research has the potential to affect the topics researchers choose, the questions they ask, the methods they use, and perhaps even the conclusions at which they arrive. Federal policy may be among the strongest levers by which significant changes can be effected (for better or worse) in the rigor and relevance of education research.

[1] Despite the burgeoning reverence for the terms "relevance" and "rigor," it is rare for an author to clarify their meaning and usage. For the purposes of this dissertation, I define relevant research as inquiry which pertains to some individual or institutional interest. I define rigorous research as inquiry which is deliberate, considered, thorough, focused, and withstands scrutiny. It is contrasted with inquiry which is casual, haphazard, lackadaisical, and unable to withstand scrutiny.

Those involved in education have a great deal to benefit from relevant and rigorous education research. It may help them locate the strengths and weaknesses in their current efforts. It may highlight unintended and surprising consequences. It may provide information useful for reevaluating both whether they are meeting their goals and whether their goals are the right ones. Thus, those interested in education have a significant stake in the extent to which federal policies influence the quality of education research.

## Research Questions

The research questions guiding this dissertation are as follows:

1. How has the federal government intervened in the production, consumption, and dissemination of education research? What alternative interventions have been proposed?
2. What are the ethical and epistemological justifications for these interventions? What are the ethical and epistemological objections?
3. How well do the various actual and proposed interventions stand up to these ethical and epistemological objections?

## Methods

The purpose of this dissertation is both descriptive and evaluative. I employ two interconnected strategies. The first is keyword searches of terms related to education research and education research agencies. I first run these through a number of search engines, including Web of Science, ERIC, Google, the search engine for the catalog of the Vanderbilt University Library, as well as through the search engines located within the websites of the Department of Education, Institute of Education Sciences,

National Academy of Sciences, National Resource Council, and the archived website of the Office of Educational Research and Improvement.

My second strategy may be referred to as bibliographic mining. This consists of using the references of each entry as a way of searching for other potentially useful sources. I then track down these sources and use their bibliographies as a way of finding other sources. This occasionally turns up new search terms which are then run back through the various search engines. I continue this process until a point of saturation, understood as the point at which new articles seem to reference articles I already have, or seem to take the same or similar positions to those already accounted for in my outline. These efforts are more of an attempt to be inclusive rather than to be comprehensive. There are many rival conceptions of the proper federal role in education research. While it is unlikely that any single dissertation could adequately represent every view, this approach to assembling a bibliography makes it more likely that my dissertation can address many of the views which have been influential enough to generate reactions from other writers.

Given the selection of research questions and methods, it should be clear that this dissertation gives little in the way of evidence of whether or not federal interventions actually have any impact on the political economy of education research. For instance, some have claimed that federal policy has caused a shift in doctoral training toward quantitative rather than qualitative methods. This is an empirical claim and thus beyond the scope of this dissertation. Instead, I look at whether federal policy *should* favor one kind of research over another. This sort of question draws more on epistemological and ethical theory than on empirical data.

To put it differently, this dissertation is intended to examine *justifications* rather than *explanations* of federal actions. The first chapter articulates a theoretical framework by which one can evaluate justifications for federal interventions. It then argues that justifications for federal interventions must meet certain ethical and epistemological requirements. The second and third

chapters illustrate some of the ethical and epistemological justifications that have been given to justify federal interventions. The fourth chapter examines whether these justifications stand up to relevant objections.

The dissertation does not attempt to prove any controversial empirical claims (the controversial claims it examines are primarily normative in character). It consciously seeks to avoid the genetic fallacy of confusing a claim's origin with its warrant. It treats the social, political, and economic origins of the various arguments concerning federal policy as largely irrelevant to whether the arguments are sound. Unlike the many micro- and macro-level explanations of the origins of federal policy regarding education research, I examine justifications rather than causes. For instance, Kaestle (1992), Sproull, Weiner, and Wolf (1978), Vinovskis (1998, 2001, 2009), and Warren (1974) attempt to show how federal policy is influenced by interactions among specific political parties, interest groups, and personalities. Others, such as Baez and Boyles (2009), Hyslop-Margison and Naseem (2007), and Ross (1991) attempt to show how federal policy is influenced by broader trends in the political economy, such as the growth of scientific management, consumerism, and technology. These accounts offer useful insights into how federal policy is produced. They can also help explain trends in education research. However, they go only so far in explaining whether the ethical and epistemological justifications given for these interventions are *warranted*. My analysis of the political economy of education research is premised on the notion that even if one's claims have deep social, political, or economic causes, it is still possible for one to produce warrants for these claims which transcend such idiosyncrasies.

## Chapter Outline

The dissertation is divided into four chapters. In addition to introducing the topic, methods, and research questions, the first chapter builds on previous attempts to theorize the incentives which drive the production, consumption, and dissemination of education research. It then draws from this framework general ethical and epistemological criteria by which justifications for proposed federal interventions can be evaluated.

The second chapter reconstructs the history of federal interventions in the political economy of education research through the 1990s. I begin by tracing these interventions back through (1) federal involvement in other fields of research and (2) the rise of education and education research in the nineteenth century. This history shows that federal intervention in education research—and their attendant justifications—are not a recent novelty. The novelty of recent efforts is in the *mechanisms* by which the federal government attempts to execute this role.

The third chapter looks at the transformation of the federal role which has taken place from the 1990s to present. I focus on the ways in which two recent interventions, No Child Left Behind (2001) and the Education Sciences Reform Act (2002), can be seen as attempts to narrow the range of activity undertaken with the education research economy.

In the fourth chapter, I evaluate the ethical and epistemological challenges faced in current debates over federal intervention. Specifically, I examine four central claims surrounding current efforts to focus education research on questions of "what works." My analysis shows that arguments on both sides of the debate concerning the current focus on randomized control field trials often rest on rather shaky ethical and epistemological claims. I argue that this may have the unintentionally effect of actually undermining the rigor and relevance of education research. I then attempt to sketch out a role

for federal policy which has a better chance of creating the "unrelenting criticism" necessary for a healthy education research economy.

## Conceptual and Theoretical Framework

### *Education Research*

There is no authoritative, neutral, and widely accepted definition of education research (Johanningmeier & Richardson, 2008, pp. 59-79). Thus, for the purposes of this dissertation, the term is used rather broadly. However, in taking a diverse set of activities under one single heading, there is considerable risk that one could say little of interest that would apply to each case. As one of the key points of contention is exactly what qualifies as legitimate research, I accept this ambiguity rather than risk a having a too narrow definition constrain the investigation. Thus, throughout the dissertation, the term research applies equally to work done in the physical sciences, social sciences, humanities, and arts.

The boundaries between education research and other research are similarly vague. This is in part because education researchers are resourceful, borrowing (with or without permission) topics, questions, methods, and techniques from a wide range of disciplines, from economics, sociology, anthropology, philosophy, history, medicine, epidemiology, and studio art. Given this wide range of resources, it seems unlikely that any definition that focuses on a specific methodology or tradition is likely to succeed. Let it suffice to say that education research is inquiry into the various factors which influence and are influenced by the creation of beliefs, attitudes, dispositions, habits, values, abilities, and skills.

*The Political Economy of Education Research*

This dissertation examines the role of the federal government in the *political economy of education research,* a term I use to refer to the people and institutions engaged in the production, consumption, and dissemination of education research. It is important to first explain the sense in which education research is both "political" and "economic." My explanation draws upon the work of others who have sought to elucidate the market-like dynamics of research, most notably Bourdieu's (1991) analysis of the sciences and Goldhaber and Brewer's (2008) description of the incentives which drive education researchers.

I begin with a description of the individuals and institutions (collectively referred to as "agents") that make up the political economy of education research.[2] Education research is produced and disseminated by universities, think-tanks, private firms, popular media, as well as well as local, state, and federal agencies. These various research products are consumed by students, parents, teachers, principles, administrators, politicians, bureaucrats, program officers, and the producers themselves. Consumers use research both to make decisions and to justify their decisions.

In the case of both production and consumption, the decision to engage with education research can be described as an *investment* of capital made with the hope of receiving some form of *profit*. Profit can be seen as the amount of capital returned to the agent in excess of his or her investment. It is the perceived potential for profit which provides the incentive for agents to invest in education research. These profits are not solely or even primarily financial. Rather, they are often best described of a *social*, *cultural*, or *symbolic capital* (Bourdieu, 1986, 1991). Social capital can be roughly

[2] My use of term "agent" is meant simply to describe a thing that has its own interests and some capacity for pursuing them. This is somewhat different than how the term has come to be used by some political scientist and economists as a contrast to "principals."

conceived of as the durable relationships an agent has with other agents. These relationships can be used to acquire various goods and to establish one's status in a group. Cultural capital consists of the various attitudes, beliefs, and manners which signify his or her social status. Symbolic capital consists of the ways that others classify an agent (for instance, as "lay" or "expert").

On the classical view of economic activity, agents seek to receive the maximum return for the minimum investment. The market of education research is *efficient* to the extent that it facilitates such transactions. Conversely, market failures can be thought of as instances of either *overinvestment* or *underinvestment*. Overinvestment occurs when an agent could invest less in education research while not suffering any loss in his or her profit. Underinvestment occurs when an additional investment in education research would yield additional profit.

In addition to failures in efficiency, the market can also be said to fail in instances in which there are sizable positive or negative *externalities*, understood as consequences which extend beyond those directly involved in the transaction. For instance, it may be the case that the private returns to investments in some forms of longitudinal research are quite low while the public returns might be quite high. For a variety of reasons (some of which are discussed in Chapter Four), these failures may not be self-correcting. In the case of such endemic failures, one can begin to make a place for some form of government intervention. This government intervention can manipulate the rates of return for various investments in education research, thus changing the way that agents perceive the alignment between certain decisions regarding research and their own interests.

As in other spheres of economic activity, agents in the market of education research compete against one another for scarce resources. Success depends to some extent on the ability to dominate others through the appropriation of maximum returns for minimum investments. However, these investments cannot be so low as to jeopardize the perceived quality of the research products. Rival

producers continuously threaten to invent more efficient means of production and better goods. Consumers, sensitive to both price and quality, refuse to purchase goods which are irrelevant to their needs or which are of too poor quality or too expensive to meet them.

Each agent is subject to a species of market discipline. This discipline occurs through the system of "crisscrossing checks and balances" from other producers and consumer of education research (Bourdieu, 1991). In a well-functioning research economy, the profits sought by education researchers—research grants, publications, tenure, and so forth—can only be acquired by submitting their work to the criticism of other similarly self-interested parties. These self-interested parties have a vested interest in framing their rivals' efforts as inadequate or incomplete. If they can demonstrate that the research is importantly flawed, they can improve their own position in a competitive market. If they cannot put forth such a public demonstration, then the producer of the research product is able to appropriate whatever capital is at stake.

This description of education research may lack some of the charm of the "hagiographical" portrayal of scientific inquiry as disinterested investigation into a value-independent world (Bourdieu, 1991, p. 3). On Bourdieu's view, it is not meant to undermine belief in the objectivity of scientific inquiry. Rather, it is meant to explain how exactly scientific inquiry—and to a great extent humanities and arts-based inquiry—can achieve objectivity even if researchers themselves are influenced by values. This objectivity is possible only when the acquisition of capital is dependent on surviving public, free, generalized, and diverse criticism from self-interested rivals (Bourdieu, 1991, p. 22). That said, I do not wish to make too much of the claim to scientific objectivity. Rather than arguing that these economic processes produce objective inquiry, I am content to state that they produce inquiry which is more rigorous and relevant than that which is produced through some other economic processes.

The problem then is that researchers need not limit their profit-seeking to this sort of public competition. Like firms in other economic spheres, agents can affect the perceived value of research products through practices such as marketing and lobbying. These practices can work to undermine the role of scientific criticism as the sole conveyer of capital. If this happens, the various struggles to gain market position among researchers lose the potential to transcend the narrow interests of each party. The rigor and relevance of education research depends on the extent to which the research economy distributes capital according to critical versus non-critical competitions (Bourdieu, 1991, p. 20).

If we follow Bourdieu in assuming that researchers are self-interested, it seems likely that they have an interest in controlling who enters the market. If the "wrong" kinds of producers are granted entry, they can flood the market and negatively affect the perceived value of research products. Each producer has powerful incentives to erect barriers to entry into the market. In a well-functioning market, these barriers take the form of unrelenting criticism. That is, each agent works to expose the ethical or epistemological flaws of his or her opponents. This benefits consumers by improving the overall quality of research products. However, in a dysfunctional market, producers are able to erect extraneous barriers to entry, ones that serve no purpose other than protecting their own market position.

The scathing comments with which I began this dissertation can be described as evidence of market failure. The exact nature of the failure is unclear. Researchers inevitably claim that there is relative overinvestment in some forms of research and underinvestment in others. These claims are typically justified by appeals to rigor and relevance. Most times, the diagnoses can be fairly well predicted by the agent's own research interests.

This dissertation argues in support of two theses. The first is that the primary aim of interventions in the political economy of education research must be to get agents to pursue capital

through the process of mutual criticism rather than through political maneuvering. Only through the former can these interventions be assured to increase the rigor and relevance or research. The second is that the proper intervention strategy is to try to shape the "durable dispositions" of those involved in education research, rather than to force adherence to technical rules or ethical norms (Bourdieu, 1991). The chief drawback to the latter approach is that the incentive to comply with these rules is purely external. They simply add to the existing pile of interests, rather than fundamentally change the ways individuals perceive their interests. That is, agents comply with the rules for reasons completely external to their own understanding of the integrity of their research. The advantage to the focus on dispositions rather than technical rules or ethical norms is that the former addresses the need for congruence between an agent's understanding of his or her own interests and the incentives provided within the social context.

As the following chapters demonstrate, many federal interventions may inadvertently undermine the incentives for education researchers to undergo the "unrelenting censorships of well-armed criticism" in the pursuit of capital (Bourdieu, 1991, p. 22). Rather, they seem to effectively disarm groups of researchers and diminish their ability to affect the perceived value of each other's research products.

Other authors have hinted at such a role for the federal government, arguing either for a change in the incentives to produce and consume research (Goldhaber & Brewer, 2008) or to the creation of a "scientific culture" (Feuer, Towne, & Shavelson, 2002). These sorts of external sanctions, if they can change objective social conditions, might have the effect of changing subjective dispositions. But it is not clear that these changes would be for the better. They might create conditions in which researchers no longer have to face the critique of rivals whose own self-interest depends on subjecting their research to intense scholarly scrutiny. Federal policy which restricts participation in the political

economy of education of research might "disarm" one's opponents, and thereby unintentionally weaken the incentives to produce rigorous and relevant research.

*The Production, Consumption, and Dissemination of Education Research*

Real-world constraints make it such that no one can fully compute what would be the optimal investment in education research. The rigor and relevance of any given research project are often unclear upon its completion and even cloudier at its conception. It is extremely difficult even for researchers to determine the value of various pieces of research. These cognitive limitations do not affect only education researchers but all agents involved with education research—students, parents, teachers, administrators, superintendents, program officers, politicians, as well as the institutions through which they function. This has important consequences for any attempt to craft legislation concerning education research. We can divide these consequences into three categories: production, consumption, and dissemination.

The production of research can be seen as an activity in which agents seek to gain the maximum return for a minimal investment (or, following the behavioral economists, a satisfactory return for a satisfactory investment). However, this rather simple formula belies the complexities in the irreducible variety of kinds of capital invested and profits expected. Producers can invest their time, money, influence, reputation and so forth. It does not particularly matter whether one sees the expectation of profit in the form of "doing good" or "doing well," that is, one's effect on others or one's own well-being. In actuality, the ability to do research which contributes to the public good is dependent on the same economic maneuvering as the ability to do research which contributes to one's private wealth. In both cases, producers must learn which research projects are considered "marketable," that is, which

projects are sufficiently in demand that the producer can attract investors. This requires that the agent gains entry into a competitive market and accumulates financial, social, cultural, and symbolic capital.

As described above, a well-functioning research economy places "unrelenting criticism" as the primary mechanism for the distribution of capital. Entry into the market then depends upon one's preparation to produce and withstand such criticism. This incentivizes research which is both relevant and rigorous. Goldhaber and Brewer (2008) argue that there is evidence that these incentives are currently lacking:

> [A] cursory examination of the latest issues of education research journals or the program of the AERA [American Education Research Association] annual conference suggests that….the proportion of the entire education research enterprise that would pass muster for scientific rigor in other fields is shockingly small (p. 198).

This can be seen as a sort of *modus tollens*: if the research economy functioned according to the proper incentives, then the bulk of research would be rigorous and relevant. The bulk of research is not rigorous and relevant. Thus, the research economy must be distorted by some sort of perverse incentives.

Goldhaber and Brewer go on to describe the ways in which incentives can become distorted. The first is the absence of an informed consumer. Consumers, "have little knowledge of what constitutes sound research design," but they often have a clear idea as to what they would prefer for a study to conclude (p. 199). This means that nuanced findings have no "natural constituency" (p. 213). Researchers then have an incentive to pick research approaches which will give results, "that fit a popular ideological perspective" (p. 199). The ability to reach these consumers through electronic media decreases the incentive to focus on publications which are meant for popular consumption, rather than, "the establishment of a scholarly track record," (p. 199). These perverse incentives have caused the market to become flooded with research, allowing consumers to pick products that serve their preconceived tastes (p. 201).

Even if consumers are open-minded and well-intentioned—that is, if they seek simply *some* answer rather than a specific answer—it is still difficult to coordinate their demands with what producers supply. This is in part because there are, "few mechanisms to ensure that the message about what is being demanded is clearly conveyed to many potential providers of that research" (p. 203). Unlike most markets, there is no clear price mechanism to bring equilibrium between supply and demand. Those who pay for education research are often not the main consumers.

A great amount of research is conducted within universities. Traditions of academic freedom allow university-based researchers great latitude in the work. Some researchers are willing to trade this freedom in exchange for contract-based grant work. Johanningmeier and Richardson (2008) finds evidence that at least some education researchers directly respond to shifts in federal priorities. However, many are willing to sacrifice the external funding necessary for expensive projects in exchange for the preservation of their autonomy. According to Goldhaber and Brewer (2008), the autonomy of academics means that much research may be "fragmented, disparate, and parochial," or focused on "narrow problems" (p. 209). Furthermore, they claim that the dominance of education schools by qualitative researchers leads to overproduction of "small, fragmented studies that focus mainly on process, rather than on large, definitive studies that emphasize outcomes" (pp. 209-210). In their view, producers should not have such a large influence on what gets produced, as they are neither sufficiently "attuned to the needs of the field" nor in possession of "high-quality standards." The implicit assumption then is that the federal government can return some of this power to consumers and that together they can better demand rigorous and relevant research.

According to Goldhaber and Brewer (2008), the final source of incentive perversion arises from the "public monopoly" of K-12 education (p. 214). Due to a lack of competition, there is no form of market discipline to encourage consumers to, "use research-tested education strategies and to demand

better research" (p. 215). If all education providers were held accountable to national standards, researchers could develop common measures which would then facilitate comparable and replicable studies (p. 216).

These features of the production of education research suggest significant limitations in the effectiveness of federal efforts to impose technical rules or ethical norms. Such interventions do not directly influence the bulk of research which is conducted without external funding. Yet they can exert a powerful indirect influence by affecting the durable dispositions of producers and consumers. That is, federal interventions need not be limited to the current interests of producers and consumers, but can also affect how producers and consumers perceive their interests.

This can be done in part by concentrating capital on agents whose durable dispositions align with the federal government's understanding of relevance and rigor. This capital can be leveraged into research which produces highly visible publications. As Barbara Schneider (2009) argues, publications are the "currency" of academia, being the means of acquiring "prestige, recognition, and access to research funding" (p. 83) Drawing on empirical work by Robert Merton (1968) and Arthur Diamond (1986), she states that having one's articles cited by others is vital to achieving "promotion and influence." Part of the problem is that research products—chief of which is the journal article—often fail to appreciate the cognitive limitations of their audiences.

Producers and policy-makers often argue that some group is not consuming enough education research or enough of the right kinds of education research. By converting the products of agents who are aligned with federal interests into a form which is more marketable to influential consumers, the federal government can increase the rates of return to the agent producing the research. This both increases the profitably of such research and lowers the costs of consuming it. From the early Annual Reports of the Commissioner of Education to ERIC and the What Works Clearinghouse, the federal

government has used dissemination as a tool for controlling production and consumption. The conferring of publicity onto certain research projects also confers capital to those who produce them. This capital can then be reinvested into additional research which can further strengthen one's position relative to his or her competitors.

An alternate strategy is to attempt to partially shut down the production of research which does not serve federal priorities. This is effectively the way that lawmakers encouraged the purchase of seat belts—it is now impossible to purchase a new automobile without also purchasing a seat belt. Similarly, the federal government can make it impossible for states to purchase curricula without also purchasing (for instance) a randomized control trial to demonstrate their effectiveness.

*Ethical and Epistemological Considerations*

If it is true that the rigor and relevance of education research depends on researchers being subject to each others' unrelenting criticism, then the basic standard by which proposed federal interventions can be evaluated is the degree to which they encourage researchers to undergo such criticism.

A full justification for any given federal intervention depends on establishing a number of claims. First, one must demonstrate evidence of some educational problem. Second, one must show that the solution of this problem depends at least in part on the production of a certain kind of research. Third, one must argue that in the absence of federal intervention, it is unlikely that such research will be produced. Fourth, one must demonstrate that the proposed intervention is adequate to cause the production of such research.

As the aims of this dissertation are primarily philosophical rather than empirical, though I do take some time to elucidate the various efforts to establish claims one through three, my focus is on efforts to establish the fourth claim. I assume for the sake of argument that there are many legitimate educational problems and that relevant and rigorous education research can contribute to their solutions. I also assume that some education research—for instance, the assembly of large-scale longitudinal datasets—is unlikely to take place absent federal intervention. However, there is always the danger that federal intervention will overcorrect, for instance, focusing on program evaluation. However, this increased emphasis on program evaluation may come at the cost of the development of new ideas or the development of longitudinal studies aimed at informing long-range policy.

My concern is that many proposed interventions designed to increase the relevance and rigor of education research may actually undermine it by opening up channels for the investment of capital other than the unrelenting criticism of other self-interested producers and consumers. This concern is at once ethical and epistemological. It highlights the importance of the effect of proposed policies on participation in the education research marketplace. Limits to participation can raise ethical concerns about fairness. It puts some individuals at the center while pushing others to the periphery. Others have written at length about how federal policy has altered the population involved in the production of education research (Lagemann, 2000; Travers, 1983; Vinovskis, 2009; Warren, 1974). However, an equally important effect of federal policy is the way in which it positions certain people as consumers of education research. For instance, a great deal of current research is geared toward managers and administrators rather than teachers or parents. Teachers and parents are less likely to be treated as the natural consumers of education research.

The issue of access and participation also has an important epistemological dimension. Each act of exclusion runs the risk of weakening the "unrelenting criticism" through which science transcends idiosyncratic interests and gains a degree of objectivity. Attempts to increase rigor and relevance which

are predicated on concentrating capital on a select breed of producers or consumers must provide some evidence that this risk has been mitigated. As discussed in the fourth chapter, this problem remains significant.

Federal interventions can affect the ways that researchers choose topics, problems, designs, and methods. It can affect the way they conceptualize and operationalize key variables. It can also affect the way they interpret the relevance of certain findings. One danger is that this might cause researchers to narrow the range of phenomena they consider in their investigations. This loss in information may have negative epistemological consequences. At its extreme, the inability to study those phenomena that do not fit into predetermined molds may lead to "scientific arrest" (Dewey, 1984, p. 23). Without the unrelenting criticism of the effects of each research decision on the warrants generated, there is less incentive for empirical sophistication. In addition to this epistemic loss, there is the ethical problem that empirically informed debate over various education aims can become impossible. Federal funding can effectively open or close the supply and demand of information regarding the connection between educational processes and outcomes of importance.

*Degrees of Intervention*

There are degrees of federal intervention, from disengagement to absolute control. If the federal government completely disengages from education research, it is likely that much longitudinal, international and comparative, and large-scale research would cease to be possible. By removing the conditions for the possibility of such research, the government can effectively limit the kinds of questions researchers are capable of investigating, such as the differences in pedagogy across districts, states, and countries, long-term trends in student achievement, the relationship between education and health, and so forth.

There are limits to the efficacy of federal intervention. The federal government cannot possibly control all of the incentives which stimulate the production and consumption of research. Even if such control were possible, it is likely to be undesirable in most cases. The evaluation depends to some extent on which aspect of research the intervention is intended to affect, such as access and participation or the selection of research methods and questions. Stronger "nudges" may be less likely to produce negative consequences in some areas than others, a point I return to in the fourth chapter.[3]

[3] For an extended discussion of using "nudges" as interventions, see Richard Thaler and Cass Sunstein (2009), *Nudge: Improving Decisions about Health, Wealth, and Happiness,* New York: Penguin Books.

# CHAPTER II

## FEDERAL INVOLVEMENT IN EDUCATION RESEARCH, 1840-1990S

The basic institutions of the education research economy began to take form in the mid-nineteenth century. Throughout the northeast, professional journals and societies arose dedicated to the study of educational issues. The first state boards of education were founded in states like New York, Massachusetts, and Connecticut, and in them were vested significant responsibilities for monitoring such things as expenditures, enrollments, and teacher quality. In 1840, after considerable lobbying, the federal government acted to create the first formal mechanism for gathering educational statistics.

Why did the federal government do what it did when it did it? A full answer would require nothing short of a complete account of the institutions, actors, and political constituencies of the nineteenth century. One would want to know the beliefs and prejudices of each individual, as well as the political constraints within which those beliefs were acted upon. Interesting though such an account may be, it may distract from the central questions this chapter seeks to answer—namely: (1) In what ways has the government intervened in the education research economy? (2) What justifications were given to support such interventions? These two questions set the stage for Chapter Four, in which I ask (3) Are these justifications ethically and epistemologically defensible?

Thus, this chapter aims at a somewhat more modest target than a full historical reconstruction. It restricts the examination to three basic themes: first, direct interventions in the education research economy; second, involvement in the production, consumption, and exchange of other forms of

knowledge (such as commerce and agriculture); and third, the growth of education and education research from 1840-1990s.

## The Federal Role in Education Research in the Nineteenth Century

These three themes are brought together by a common trend—the general growth of the federal government. The notion that federal intervention in education began in the nineteenth century is as common as it is false. Indeed, the federal role in education has been present nearly since the nation's founding. The Land Ordinances of 1785 and the Northwest Ordinance of 1787 established a precedent of federal allocations to education. The Northwest Ordinance of 1787 specifically required that 1/36 of the land for each new township be set aside for the maintenance of public schools, while the remaining portions be sold at public auction with the proceeds going to the federal government. Further, in order for state to be admitted into the union, it was required to have in the state constitution a clause recognizing the state's responsibility for education. Thus, the federal role in education was not new to the nineteenth century. The novelty is in the *mechanisms* by which it executed its role—that is, through the support of certain kinds of research on education.

### *Background to Federal Involvement in Research Economies*

Federal involvement in the political economy of research predates involvement in the production of *education* research. The oldest form of this research, dating back to Article I, Section 2:3 of the Constitution, authorized the collection of such population statistics as would insure the proper execution of governmental responsibilities. Specially, it called for a census to help determine the number of members in the House of Representatives. The Department of Treasury and Department of

Agriculture had also from a very early stage conducted this sort of fact-finding research as a means of improving their operations. Such activity can be seen as an early indication that the information needed to manage federal programs could not be reliably supported in the absence of federal intervention.

Early in the nineteenth century, a different kind of federally sponsored research began to emerge. This research aimed not simply at better coordinating federal programs, but at providing a general service to the nation. For instance, by 1816 the federal government had funded research to develop navigation technology and to map the Atlantic coast (National Advisory Committee on Education, 1931, pp. 378-379). By the 1840s, the Commissioner of the Patent Office was receiving a modest but steady stream of research funds to report to the House Agricultural Committee the latest developments in science. It was expected that these reports would both provide a "scientific basis for lawmaking" and disseminate information on "every new discovery in the science of agriculture…geology, chemistry, and botany." While the federal government did not fund this research, it acted as a source for synthesizing and disseminating the work of the "great experimenters…who kindly and generously communicated to the head of the office the results of their labors and experiments" (National Advisory Committee on Education, 1931, p. 379).

These examples suggest a gradual shift in the federal government's willingness to shape the research economies in various fields. The Patent Office's dissemination can be seen as an effort to facilitate the exchange of research goods and lower the cost of acquiring new information. As such, it can be seen as an intervention to increase the supply of available research. However, it has the potential to reverberate throughout the entire structure of the research economy. From the perspective of the research consumer, dissemination lowers the cost of acquiring research goods. From the perspective of the research producer, dissemination expands the pool of consumers. Rather than being limited to the audience of a scientific journal, the distribution of one's research goods is now

effectively subsidized by the government. Standard economic theory holds that this can lower the cost of production and therefore increase supply.

## Education Research in the Early and Mid-Nineteenth Century

By the 1830s, some form of free public education was available in most northeastern urban centers. Over the 1840s and 1850s, there was increased effort to extend free public schools to other areas and to bring consistency to those areas that already had it. The growth in education occurred at the same time as the proliferation of education research. Though educational psychology, which would come to dominate education research in the twentieth century, barely existed in the beginning and middle of the nineteenth century, there was no shortage of research on the topic of education. The rise of schools had been matched with a proliferation of textbooks, including works by Noah Webster (1785), Caleb Bingham (1785), Lindley Murray (1797), and William Holmes McGuffey (1836). The famous reading wars had begun in full earnest, with Samuel Worcester (1828) introducing the whole-word method as an alternative to the alphabet-syllable method (Cremin, 1980, p. 391).

What then can be said of the federal role in the growth of education and education research in the early nineteenth century? Whatever influence the federal government enjoyed was indirect. By creating mandates and incentives, the federal government may have contributed to the spread of education throughout the states and territories. This diffusion may have helped stir interest in the myriad difficulties involved in the tasks of educating citizens. How and what are students to be taught? How are teachers to be trained? How are schools and districts to be administered? Who ultimately has authority over curriculum and methods? Thus, though federal support for education research did not begin until 1840, it may be fair to locate the origins of federal influence on the political economy of education research much earlier. The Land Ordinance of 1785 and 1787, along with the conditions for

granting statehood to the several territories, helped bring questions about the provision and management of public education into the core of the nascent field of education research.

The professional publications of the period reflect the problems of the day. In a period in which there was no ready model upon which to base current efforts, early education researchers focused heavily on historical and comparative scholarship as a source of ideas. For instance, an early issue of *The American Journal of Education* (1826) included articles describing educational efforts underway in Boston, London, Buenos Aires, as well as descriptions of national practices in Germany, and France. Other articles focused on what states were doing to meet their constitutionally mandated duties, highlighting addresses by governors and state superintendents. By making education primarily a state matter, the federal government helped make the actions of agents of the states core subjects of study.

Education research was profoundly affected by the social and political crises that were brewing throughout the nineteenth century. Passionate debates over slavery and social justice were often entangled with more genteel disagreements over the fundamental relationship between the states and the federal government. Henry Barnard, one of the chief advocates for a larger federal role and the first Commissioner of the Department of Education, was not an abolitionist. Perhaps this is part of what made him a viable compromise candidate for the position. Though not all who favored an increased federal presence favored abolition, it is clear that those who opposed abolition looked at the federal government with suspicion.

Supporters of education research ran against strong resistance to an increased federal presence in education. Advocates for a large federal role drew upon research done on the education systems of countries such as Prussia and France, where the central ministry was responsible for overseeing what they saw to be a coherent and organized system of instruction (MacMullen, 1991, p. 245). In the time that such appeals were being made, most state departments of education were barely decades old.

New York established the first state board of education in 1812. Massachusetts followed twenty-five years later, appointing Horace Mann as the first secretary of the state board of education in 1837 (Lagemann, 2000, p. 6). Henry Barnard helped usher in similar legislation in Connecticut in 1839 (Travers, 1983, p. 20).

Reformers could thus draw on two sources to support their advocacy for a larger federal role in education research. The first was the fact that the federal government was already involved in the production and dissemination of research on other areas of national interest. The second was a small but vocal network of persons advocating for the expansion of education. The growing recognition of the importance of education is evidenced by the growth in the number of state boards of education, education journals, and schools. However, while education may have been seen as a national goal, it was by no means clear that the federal government should be involved in advancing that goal.

By the 1830s and 1840s, a number of groups were putting forth proposals for a federal agency responsible in some way for education. These plans varied tremendously in their visions of the appropriate federal role. Charles Brooks, the great advocate for normal schools, argued on behalf of a cabinet-level Minister of Public Instruction as well as a constitutional amendment that would make such legislation possible (MacMullen, 1991, p. 245). Another citizen group argued for the establishment of a department of agriculture and education to help prepare farm children for their future vocations. (Warren, 1974, p. 47). Robert Dale Owen, congressman from Indiana, argued throughout the mid-1840s to place an educational bureau and national normal school within the Smithsonian Institution.

In his official biography, Henry Barnard claims to have submitted a "rough outline" for a national role in education, beginning with the inclusion of education statistics on the 1840 decennial census. There were many persons who advocated the inclusion of education statistics on the census, and the evidence of Barnard's importance in bringing this about is in question (for differing representations, see

MacMullen, 1991, p. 244; Warren, 1974, p. 48). What is clear is that their advocacy was successful (National Advisory Committee on Education, 1931, p. 431). Information on illiteracy, schools, academies, and colleges was included on the 1840 decennial census (Smith, 1923, p. 1).

The movement to establish a national presence in education failed in the ensuing twenty years to rally around a single vision. In 1829, Representative Joseph Richardson argued that a House education committee should be formed and charged with spending surplus federal funds on the advancement of the cause. From 1838-1849, prominent educators, among them Charles Brooks, Henry Barnard, and Horace Mann, thought it would be sufficient to simply add to the census. Alexander Bache argued in 1840 against any direct federal role, favored instead a privately funded "national school agent" (MacMullen, 1991, p. 246).

By 1849, Horace Mann agued to an audience of education reformers that the ultimate end of education, especially in such perilous times, was national unity (Warren, 1974, p. 49). Within a year, Henry Barnard presented plans for a federal education agency. At the center of both plans was the belief that the deficiencies in education were largely responsible for social and political conflicts. The proper role of the federal government, in their view, was to assist the various states and localities in collecting and interpreting information on schooling. This would allow lawmakers and educators to better grasp deficiencies and possible means of addressing them.

Horace Mann and the American Association for the Advancement of Education petitioned congress through the 1850s and 1860s to create a bureau within the Department of Education which would be charged with just this task. After extensive lobbying by education boosters, a non-cabinet level Department of Education, housed within the Department of Interior, was authorized by Congress in 1867. The beginning of the act states that the Department shall be established:

> for the purpose of collecting such statistics and facts as shall show the condition and progress of education in the several States and Territories, and of diffusing such information respecting the organization and management of schools and school systems, and methods of teaching, as shall aid the people of the United States in the establishment and maintenance of efficient school systems, and otherwise promote the cause of education throughout the country.

It was ardently opposed by Southern whites and others who feared intrusion by federal agents. President Andrew Johnson looked upon the Department with suspicion. He was persuaded to approve it only after Senator James Dixon of Connecticut convinced him that the Department would simply be doing annually what the census does decennially (MacMullen, 1991, p. 258). By design, it had neither the stick nor the carrot to motivate states to fall in line with federal priorities.

After a bit of political wrangling, Barnard was chosen to be the first Commissioner of the new department. The choice proved to be inauspicious. Fraught by internal tensions and external suspicion, within a year the Department was reduced to an Office within the Department of Interior. Two years after that it was reduced to a Bureau. When considering the constraints of a limited mandate, inadequate funding, and poor leadership, it is remarkable how well the Bureau of Education still carried the novel function of collecting, analyzing, and disseminating nationally comparable education statistics. This began the process of laying the foundation for a public debate about the purpose and condition of education.

As the focus of this dissertation is on the kinds of research conducted, something should be said about the evolution of research questions, topics, methods, and designs. In his first report to Congress (1867-1868), Commissioner Barnard confessed that he had relied upon data from four principle sources (quoted in Smith, 1923, p. 19):

(1) Annual reports of school officers and systems, replies to special inquiries, and such information as could be gathered by visits of the commissioner and his agents in the field. (These were the principle sources.)

(2) Attendance at annual meetings and special gatherings of national, state, and local associations and other groups interested in education.

(3) Personal touch and individual correspondence with school officers.

(4) His own personal collection of books and pamphlets.

(5) The press.

However, the quality and quantity of data would soon improve considerably. In 1872, Congress appropriated funds for the employment of a full-time statistician (Smith, 1923, p. 20). During the middle and later portion of the 1870s, the Office of Education worked with the National Education Association to develop standards for reporting educational statistics. Among other things, they developed common forms for the reporting of data by state and local authorities, eliminating much of the eclectic reporting practices that had previously taken place.

In these early years and for a good while to follow, the primary mechanisms for government intervention in the education research economy were (1) to collect data which no other organization, agency, or individual would have access to otherwise and (2) to disseminate this information to those in charge of the supervision of education. This role is understandable and has been explained by others as reflecting the limited funding and mandate of the Bureau of Education (Lagemann, 2000; Smith, 1923; Warren, 1974).

This placed superintendents and policy-makers, rather than teachers and parents, as the primary producers and consumers research. These were the persons who tended to supply the information reported in the Annual Report of the Commissioner. They were also the persons most likely to receive copies of annual reports, circulars, and other publications. The implications of this were far-reaching—it helped create a permanent place in the research economy for topics related to management, administration, and allocation. These topics lend themselves to questions of inputs and outputs, the kinds of things which administrators can control, rather than processes, which are

ultimately the domain of teachers and students. The Office of Education helped stimulate both supply and demand for replicable, uniform, and comprehensive information about educational activities. As education research grew, it came to develop increasingly sophisticated designs and methods for addressing such questions. Steady improvements were made toward providing regular and accurate portrayals of educational activities across the various states and territories. In 1907, an external investigation headed by Edward Thorndike recommended an overhaul of the statistical capacities of the Office of Education. Among the changes was the establishment and expansion of a statistical division with a staff of nine persons. It also led to more comprehensive and uniform reporting standards for states in 1918-1919.

From its first Annual Report, the Office of Education seems to have been on an endless quest to figure out how best to disseminate information. Early on, the Department/Office/Bureau of Education adopted a plan for dissemination that would include monthly circulars, quarterly publications, educational documents and tracts, and annual reports. Barnard would also send reports to school officers, conference attendees, and those with whom he had personal acquaintance (Smith, 1923, p. 23). By 1870, the Bureau sent out 12,000 documents annually.

According to the Annual Reports of the Commissioner of Education, in the early years the demand for research products seemed to outpace the government's capacity to supply them. Congress failed to fully fund the printing needs of the Bureau, thus often requiring staff to compose lengthy letters rather than issue copies of reports to those who inquired (Smith, 1923, p. 24). It is easy to see why they were such valuable commodities—the range of information contained in them was unprecedented. They contained reports not just on the Bureau/Office's activities, but of progress across the nation, the latest updates on pedagogy, a list of publications of note, and reports on what other countries were doing in education. They touched on nearly every aspect of education, from the home

and church to school and college. They documented the latest advances in school architecture, nutrition plans, sanitation, and even furniture.

## The Federal Role in Education Research in the First Half of the Twentieth Century

In 1918, the Office began publishing *School Life,* a semi-monthly publication targeted for teachers and administrators. For the first time, a serious effort was made to position teachers as the consumers of education research, and thus suggest that research should be conducted which is of direct use to teachers and not just those who manage them. *School Life* contained updates on legislation, research, and a variety of education topics. Another regular series, the Teaches' Leaflet, appeared about the same time. It was more narrowly focused on issues of classroom instruction. For instance, Teachers' Leaflet #2, published in 1918, contained a number of examples on how to teach patriotism, a timely concern given the country's engagement in the most expansive war in human history.

These publications helped to round out what the 1920 Annual Report of the Commissioner of the Office of Education described as the Office's role as a "clearinghouse of information," a phrase that would gain increasing important over the next several following decades. The report states:

> …there are a few men and women in the United States and elsewhere whose opinions, because of their greater knowledge of the subject, are most valuable. This bureau tries to find for each subject who these persons are and to make lists of expert advisers whom it may consult and to whom it may refer others. It also undertakes, after correspondence and personal conference with these experts, to formulate the consensus of expert opinion (Smith, 1923, p. 66).

The Bureau used three basic sources of information: (1) questionnaires sent to "experts," (2) a review of the "most important current publications on education," and (3) special conferences organized around subjects of interests. Of course, it is not clear the degree to which these three sources overlapped. It seems that leading experts would be more likely to publish in the best journals, which would make them

more likely to attend high-profile conferences. Thus, this tactic may not have been incredibly successful at diversifying the producers and consumers of education research. However, it is unlikely that diversity was a major goal. Rather, these tactics can be seen as part of an overall strategy to influence the distribution of capital among producers and consumers.

According to a statement from the Commission of Education to the Secretary of the Interior in 1930, the years following World War I brought about "vigorous" agitation for a department of education, as well as "bitter" opposition. The Commissioner argued that a large amount of the debate was over the need, "of real fact on which to base a policy" (National Advisory Committee on Education, 1931, pp. 406-407). He imagined a department which would focus on research, and that this research would settle debates over the direction of education. In the Commissioner's view, those who opposed a federal department of education were opposed to having facts rather than politics decide policy matters.

In the Senate, a bill was introduced on June 4, 1929 (S. 1454) to better specify the activities of the Bureau of Education without suggesting any major change to its structure. It authorized the Bureau to conduct studies, investigations, and reports on a number of specified issues and any others the Secretary of Interior deemed necessary. On September 4, 1929, a competing bill (H.R. 10/S. 1586) was introduced. The bill called for the creation of a Department of Public Education. It called not just for the collection of statistics and facts, but for the conduct of "studies and investigations...and the results thereof to be disseminated" (National Advisory Committee on Education, 1931, p. 421). On December 11, 1929, a third, less detailed bill (H.R. 7249) was introduced, calling for the creation of a Department of Education. The bill specified that all educational activities would be consolidated within the Department. Further, it stated that the Department would collect statistics and facts in eleven specified fields of education (National Advisory Committee on Education, 1931, p. 413).

None of these bills would become law. Nor would the nearly one-hundred similar bills introduced over the next fifty years. They did, however, help stir debate over the possibility of expanding the federal role in education. In 1929, President Herbert Hoover commissioned the National Advisory Committee on Education to report on current federal involvement in education. Two years later, the Committee published *Federal Relations to Education* (National Advisory Committee on Education, 1931). Though it advocated largely for the preservation of the tradition of local control of education, it called for increased support in areas in which local and state governments struggled. One such area was educational research. The report states, "The Federal Government should render large intellectual assistance to the States in matters of education through scientific research, and the collection and dissemination of reliable information, particularly when the types of intellectual, scientific, and professional service needed cannot be provided by the States and the local communities" (National Advisory Committee on Education, 1931, p. 34). In providing such a service, the federal government could ensure that efforts to improve education rested on informed opinion as the basis of democratic debate.

The report argued that democratic governance requires the widespread dissemination of timely, accurate, and relevant information. In a representative passage it stated:

> Nothing enhances the free registration of intelligence and honest personal opinion on the part of political representatives more than impartial and widely disseminated information, secured through the scholarly services of a highly trained, scientific personnel, independent, impartial and constantly subject to the criticisms of the scientific world, in which they are by profession an inherent part (National Advisory Committee on Education, 1931, p. 73).

This information could be used to move past political conflicts by "revelations of pertinent facts established by scientific methods and presented in understandable terms" (National Advisory Committee on Education, 1931, p. 72). It argued that educational decisions should be based on "[f]acts scientifically established rather than unsupported opinion" (National Advisory Committee on Education,

1931, p. 75). The report further argued that whatever research can be conducted by the states or the private sector should be left to them. The federal government should only concern itself with the remaining portion which is of national interest.

In the 1930s, the Office of Education (OE) used a variety of different methods for gathering and analyzing data (Judd, 1939, pp. 82-83). From 1932-1936, OE published 334 studies of education. Eighty-eight studies (26 percent) involved conducting and analyzing survey data, frequently supplemented by "field work," which seems to mean having a person physically in the field recording information. Sixteen studies synthesized data from conferences. The largest number of studies (115) was "analytical," meaning that they reviewed the literature on some educational problem and attempted to provide a clearer and broader analysis of the situation. There were also 114 documentary and historical studies. Eighty-three bibliographical studies were conducted, gathering the literature on topics such as "the education of Negroes" (including pioneering work by Ambrose Caliver). Interestingly, only one "experimental" study was conducted during these four years. The authors attempted to study how to adjust behavioral problems in school children. The push for more experimental work would not gain traction until the George W. Bush administration.

## The Federal Role in Education Research in the 1950s

From the Great Depression through World War II, education research faded as a federal priority. While federal spending for research skyrocketed in other areas, the staff for the statistical division of the Office of Education was actually reduced by one-third (Vinovskis, 1998, p. 17). On this reduced budget, new projects were largely curtailed as OE focused on its core activity of collecting and publishing statistics.

Science and technology had played an essential role in World War II. At the conclusion of the war, Vannever Bush argued in *Science, the Endless Frontier* (1945; reprinted 1960) for a national foundation for the support of scientific research. In 1950, the National Science Foundation (NSF) was established. Though its initial budget was quite small, over time it assumed an important advisory role to the federal government. The NSF has since played a vital role in the shaping the relationship between the federal government and education research.

The federal role in research would grow throughout the 1950s. The Cooperative Research Act of 1954 (CRA) established the Cooperative Research Program (CRP), an initiative which allowed the government to outsource federal education research to universities and other organizations, rather than conduct it internally. The program was operated out of the Office of Education and overseen by the Research Advisory Committee (RAC), a group of researchers who were based outside of OE. The RAC clashed with those working within the OE over how best to spend the congressional appropriations, with RAC hoping to emphasize unsolicited, field initiated studies (Dershimer, 1976, p. 42) and the staff within OE hoping to capture a portion of the funds to further their own projects. Congress appropriated $1 million to the CRP in 1957, $2.3 million in 1958, and $2.7 million in 1959. As part of their effort to increase the rigor of education research, RAC sought and rewarded research by discipline based scholars, rather than educationists. In this effort they were successful; applicants from discipline-based scholars quadrupled from 1955 to 1963 (Dershimer, 1976, p. 56).

One of the more striking—and controversial—aspects of CRP was in its reliance on field-initiated projects. Until the 1950s, education research was supplied primarily by those working in education schools. For a variety of reasons, education schools were looked down upon by those working in the disciplines (Clifford & Guthrie, 1988; Lagemann, 2000). Ralph Tyler, chair of RAC in 1959, was convinced that the low quality of education research was due to the fact that the best and brightest minds were

not engaged in it. The only way to lure disciplinary scholars was to allow them a wide degree of latitude in their projects. He believed that given their expertise, they could be trusted to do good work. Tyler's main interest was to make sure that talented scientists would have the resources to conduct research and train young scientists (Dershimer, 1976, p. 47).

However, this attempt to redirect the flow of capital toward disciplinary scientists met hostile resistance. Consumers of education research—most notably those who used research as a tool for policy-making—made the predictable charge that the CRP was unfocussed and irrelevant to the needs of practitioners. The political might of education lobbying organizations—from unions to governor's associations—far outweighed the voice of academic scientists. Over the following decades, field initiated studies would come under increased scrutiny as part of a general suspicion of the quality and relevance of educational research. Thus, a major effort to give power to a select group of disciplinary-based producers ended largely in failure. The producers of education research would once again be those in the education schools who were seen as being more in touch with the needs of administrators and policy-makers.

During the second year of the CRP, congress passed the National Defense Education Act of 1958 (NDEA), which provided further support for education research by linking it to the economic and military needs of the United States. It allocated substantial funds for investigating ways to improve the quality of math and science education. This legislation set the stage for educational researchers to play a prominent role in shaping educational policy. NDEA brought two new streams of research funding to the OE. Title VI of the act called for research on effective methods for teaching foreign languages and cultures. Title VII called for OE to conduct and disseminate research on the use of technology in the classroom. In 1959, its first year, NDEA brought an additional $4 million of research funds to OE. In the three following years, this would increase to $6 million (Dershimer, 1976, p. 44).

However, not all research mandated by NDEA was assigned to the OE. The act assigned to the NSF the daunting task of fundamentally reforming science education. From 1956 to 1968, NSF would appropriate increasingly large amounts of money to support the development of science curricula. Following the passage of NDEA in 1958, appropriations for curriculum development increased from $835,372 to $6,030,325 (Lagemann, 2000, p. 169). This research would eventually extend to the development of an elementary school curriculum in history and social studies. "Man: A Course of Study," or MACOS, drew the ire of conservatives for its rather unvarnished interpretation of humankind. Though NSF funding for curriculum development was already declining, the controversy surrounding MACOS all but killed congressional enthusiasm for curriculum development. From 1969 to 1975, funding for curriculum development within NSF dropped by more than half.

## The Federal Role in Education Research in the 1960s

By the 1960s, there was growing criticism that dependence on unsolicited grants had led to excessive fragmentation of the research community. Small-scale, investigator-driven research was not yielding cumulative, replicable, and usable results. By 1961, there was talk within the Office of Education of creating large centers in which human and financial resources could be concentrated to the study of specific educational programs for an extended period of time. One of the chief architects of these centers, Francis "Fritz" Ianni, envisioned them as encompassing the full spectrum of scientific inquiry, from basic research to demonstration and dissemination (Dershimer, 1976, p. 59).

Francis Keppel, then Commissioner of Education, believed that these R&D centers could help move the research community past the model of the, "small, easily-managed project which focuses on miniature, obscure and non-controversial issues, which are seldom taken seriously by administrators or teachers" (Dershimer, 1976, p. 59). He sought to alter the incentive structure of education research,

believing that the enticement of being able to conduct 'big science' would attract the best minds to the study of educational problems. Keppel and Ianni eventually won congressional approval for ten of these centers.

In 1964, President Lyndon Johnson assembled thirteen committees to make recommendations regarding the federal government's role in creating the great society. The Task Force on Education, headed by John Gardner, emerged with two broad conclusions. First, that money alone would not fix education. Innovation was also necessary. Second, that there must be institutional mechanisms for the dissemination of innovations. The Gardner Report described the creation of national laboratories on a scale that would dwarf the educational R&D centers founded just a couple of years prior. Gardner imagined laboratories of the size and scope of the Atomic Energy Commission (Vinovskis, 1998, p. 11). Unlike existing R&D centers, they would place greater emphasis on developing and disseminating "educational innovations." These tasks would be executed in part by using experimental schools or pilot programs in regular schools, as well as training teachers (Dershimer, 1976, p. 66). Though Gardner and Keppell were able to get Congress to authorize the national R&D centers, bringing them into existence proved to be another story. Their ambitious vision for national laboratories was thwarted by Congress, which decided that rather than a small number of large laboratories, funding would be divided among twenty small regional labs.

Gardner's recommendations for research and development were incorporated into Title IV of the Elementary and Secondary Education Act of 1965 (ESEA). ESEA remains the single most important piece of education legislation in the United States, leading to an eight-fold increase in educational spending over a period of eight years, from $500 million in 1960 to $3.9 billion in 1967 (Vinovskis, 1998, p. 11). This brought fundamental changes to the relationship between the federal government and education research. One of the most significant changes is that it created within the Office of Education

the National Center for Educational Statistics (NCES). By the end of the 1960s, NCES had launched the National Assessment of Educational Progress (NAEP), a nationally representative assessment of academic achievement in the fourth, eighth, and twelfth grades. Though the form and content of NAEP examinations has changed over time (Jones, 1996), it remains one of the most important nationally representative data sets on student achievement.

With this increased spending came an explosion in the demand for program evaluations (Lagemann, 2000, p. 162). This heightened demand for a steady supply of research products capable of informing judgments on the effectiveness and efficiency of government spending. While it may be impossible to make an informed judgment as to what program evaluators would be doing if they were not evaluating programs, one may wonder whether their attention may have went to basic research and development, or perhaps even humanistic approaches to educational issues. This is not to suggest that the federal government's altering of the research economy was *bad* because it diverted attention from other questions, but merely to suggest that some trade-off may have taken place.

In addition to the push to shape research priorities, the Office of Education began to take a greater role is turning teachers into consumers of education research. If education was to be based on science, there needed to be a way of getting scientific research to those who could put it to use. As late as 1958, the dissemination efforts of the Office of Education were limited to encouraging researchers to report their findings at professional meetings, publishing a magazine, *School Life*, and issuing the occasional Teachers' Leaflets (Dershimer, 1976, p. 108). In 1965, The Educational Resources Information Center (ERIC) began operation as a clearinghouse for collecting, disseminating, and archiving "exemplary information and research" (Lagemann, 2000, p. 187). ERIC operated through twenty topic-specific clearinghouses which screened documents to be included in its database. It also archived all research conducted within the Office of Education. ERIC has continued to be extraordinarily successful at

collecting information. However, this information has been most vociferously consumed by those who produce it—researchers. Educators, having little time or interest to indulge in consuming research, have never made extensive use of it.

Perhaps the greatest piece of research to emerge from the 1960s resulted not from ESEA, but from the Civil Rights Act of 1964 (CRA), which mandated external studies of the effectiveness of federal education programs. One of these studies was the landmark *Equality of Educational Opportunity* (1966), a study led by James Coleman on the effects of differences in school resources on student achievement. To the surprise and chagrin of many in favor of equalizing school spending, Coleman's research team found that differences in school spending were not as strongly predictive of student achievement as differences in home resources. This finding would lead to decades of debate over the efficacy of federal programs in education. The Coleman Report was joined by a host of other evaluations which showed that the Great Society programs of the 1960s had not yet broken the "cycle of poverty" (Vinovskis, 1998, p. 13). It became increasingly clear that more research was needed to discover how to improve the chances that young people, especially those in poverty, could achieve educational success. Daniel Patrick Moynihan argued persuasively for a major shake-up to the organization of federal education research. He believed that such reorganization might lead to more rigorous and relevant research that could be put to use in the improvement of education.

## The Federal Role in Education Research in the 1970s

Upon his election to the presidency, Richard Nixon asked Moynihan and his assistant, Chester "Checker" Finn, to review the state of federally sponsored education research. Moynihan came back with the recommendation that Nixon go through with his campaign promise to create a National Institute for the Educational Future, later renamed the National Institute of Education (NIE) (Vinovskis,

2001, pp. 76-77). By 1972, despite significant opposition from the American Federation of Teachers as well as prominent researchers, they successfully won congressional approval for a National Institute of Education. However, Congress quickly slashed funds for NIE by more than one-third, as well as intervening in the establishment of research priorities (Vinovskis, 1998, p. 15). The reductions not only prevented NIE from funding new research, but required scaling back some current operations (Sproull, et al., 1978).

Though spending on research had fallen, by 1980 dissemination expenditures had tripled, comprising the single largest category of spending with the NIE. One of the new dissemination initiatives was the program for Research and Development Utilization (RDU), established in 1976 as yet another attempt to connect educators with scientific research. Meanwhile, development research plummeted in the 1970s, falling from $40.3 million in 1975 to $27.9 million in 1980. Though NAEP continued, the qualitative components were phased out in favor of the cheaper paper and pencil tests (Jones, 1996). One bright spot was in the initiation of the National Longitudinal Study (NLS-72), the first longitudinal study of education. Though many research programs were cut, the growth of longitudinal studies and national assessments helped swell funding for NCES from $1.9 million in 1970 to $13.9 million in 1978.

## The Federal Role in Education Research in the 1980s

In 1979, President Jimmy Carter made good on his campaign promise to create a cabinet level Department of Education (DOE). This was seen by conservatives as a giveaway to the teachers unions. Ronald Reagan was elected in 1980 having promised to dismantle the DOE. Though he was not able to fulfill his campaign promise, Reagan kept spending on education and education research to a minimum, with federal agencies having to fight to justify their appropriations. In real dollars, federal spending on

education dropped 14.4 percent during Reagan's first term (Vinovskis, 1998, p. 19). Edward Curran, Reagan's first Director of NIE, advocated for its abolition, citing as one reason, "NIE is based on the premise that education is a science….As a professional educator, I know that this premise is false" (quoted in Vinovskis, 2001, p. 102). The cuts to NIE were particularly deep, amounting to a 70 percent decrease in appropriations from 1981 to 1988 (Vinovskis, 1998, p. 20).

The publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983) emphasized a significant change in approaches to educational improvement. While the report generally found that the education system was failing miserably, it did not point to education research as either the cause or the cure for this failure. It presented the problems of education as moral and political, rather than scientific or technical. It was not that the country did not know what was wrong or how to fix it, but that it had grown complacent. This attitude was encouraged by William Bennett, Secretary of Education from 1985-1988. According to interviews conducted by Carl Kaestle, Bennett believed that the DOE should simply try "to get people to do what they know they ought to be doing." Another interviewee added that Bennett believed, "The solution lies in conviction, in character, not in understanding the cause of the phenomenon" (Kaestle, 1992, p. 28).

In this vein, Assistant Secretary of Education Chester "Checker" Finn, along with William Bennett, published *What Works: Research about Teaching and Learning* (1986). The DOE mounted an unprecedented dissemination campaign, eventually giving out over 500,000 free copies of the report to anyone who wanted it. The report issued some forty-one policy recommendations. It presented them as extensions of common sense and confirmed by research. The overall theme was that we already knew how to fix education without any increase in resources, but that we lacked the will to do so. David Berliner and Bruce Biddle (1995) later argued that the document was propaganda disguised as a research review. Later, Finn himself would admit that the report had minimal real impact.

# CHAPTER III

## FEDERAL INVOLVEMENT IN EDUCATION RESEARCH, 1990S-2009

This chapter examines federal legislation concerning education research from the 1990s through 2009. During this period, there have been four major pieces of legislation concerning education research: Goals 2000: Education America Act of 1994 (Goals), the Reading Excellence Act of 1999 (REA), No Child Left Behind of 2001 (NCLB), and the Education Sciences Reform Act of 2002 (ESRA, known earlier as H.R. 3801 and H.R. 4875). In addition, the National Research Council (NRC) issued a major report, titled *Scientific Research in Education* (*SRE,* 2002), which addressed the nature of scientific research in education and the federal government's role in supporting it.

This chapter seeks to describe comprehensively how these various interventions have conceptualized "science" and "education research." I argue that these conceptualizations can be seen as attempts to shift the distribution of capital throughout the political economy of education research, and through this shift affect the production, consumption, and exchange of education research.

### Goals 2000

On March 31, 1994, Congress passed Goals 2000: Educate American Act (1994). Section 902 of Goals 2000 reauthorized the Office of Educational Research and Improvement (OERI).[4] It states that the federal government, through the Office of Educational Research and Improvement, should make "a significant investment in attaining a deeper understanding of the processes of learning and schooling

[4] My account of Goals 2000 follows closely the one presented in Vinovskis (2001), p. 4.

and developing new ideas." It describes previous education failures as having resulted in part from a lack of, "a strong foundation of knowledge on which to design improvements."

The law brought about several important changes, including the creation of five new national research institutes. At least twenty percent of the funding at each institute was allocated for field-initiated research, while another one-third was set aside for the establishment of national research centers. It also established an Office of Reform Assistance and Dissemination (ORAD) to oversee the activities of the labs, as well as ERIC and the National Diffusion Network. Finally, it established a National Educational Research Policy and Priorities Board (NERPPB) to establish standards and monitor research funded by OERI (Vinovskis, 2001, p. 148).

NERPPB would be tasked with deciding what counts as quality research, and consequently which kinds of problems are most important, and by extension which kinds of people would be positioned at the core of the political economy of education research. Goals 2000 defined education research broadly, thus giving wide latitude to the board's work. It neither attempted to distinguish between scientific and nonscientific research, nor suggested a list of favored methodologies. Rather, the act delegated to the Assistant Secretary the duty to, "develop such standards as may be necessary to govern the conduct and evaluation of all research, development, and dissemination…to assure that such activities meet the highest standards of professional excellence." It states that the Assistant Secretary should develop these standards by consulting other federal agencies (such as the National Institutes of Health), research organizations, and members of the general public.

Goals 2000 demonstrated an unusual sensitivity to the role of the federal government in fostering inclusion in education research. It observed that minority researchers are "inadequately represented" in OERI and that the government should take actions to recruit, retain, and promote

(6) EDUCATIONAL RESEARCH.—The term ''educational research'' includes basic and applied research, inquiry with the purpose of applying tested knowledge gained to specific educational settings and problems, development, planning, surveys, assessments, evaluations, investigations, experiments, and demonstrations in the field of education and other fields relating to education.

Figure 1. Language defining education research in Goals 2000: Educate America Act.

qualified minority educational researchers. Further, Goals 2000 put the federal government in the role of an enabler rather than an enforcer. Its main role was to support the work of states through research, evaluation, and dissemination, including creating a "national treasure chest of research results, models, and materials at the disposal of the education decisionmakers of the United States."

As written, Goals 2000 does significantly less than subsequent laws to stimulate consumption for education research. For instance, it has no mandates for states to use research as a basis for choosing and evaluating curricula. Moreover, it makes virtually no claims as to which research the federal government expects people to produce or consume. The laws that followed would be far more specific.

## Reading Excellence Act

On October 21st, 1998, President Bill Clinton signed into law The 1999 Omnibus Bill (H.R. 4328) which included the Reading Excellence Act of 1999 (REA) (Sweet, 1998). This began a new era of government intervention. REA amends the sections of Elementary and Secondary Education Act of 1965 (ESEA) dealing with the way the federal government funds reading programs. The act's purpose was, "To improve the reading skills of students, and the instructional practices for current teaches (and, as appropriate, other instructional staff) who teach reading, through the use of findings from *scientifically*

*based reading research*" (emphasis added). Its purpose was to improve federally funded reading programs by conditioning their funding on scientific evidence of effectiveness.

It did so by limiting funding to "eligible providers," who were in turn defined as those whose product or service was based on "scientifically based reading research." It then went on to define scientifically based reading research as, "the application of rigorous, systematic, and objective procedures to obtain valid knowledge relevant to reading development, reading instruction, and reading difficulties." It is important to note that REA addressed the *consumers* rather than *producers* of education research. It told them which kinds of research they may use to justify receiving federal funds.

> (5) SCIENTIFICALLY BASED READING RESEARCH.—The term 'scientifically based reading research'—
>
> (A) means the application of rigorous, systematic, and objective procedures to obtain valid knowledge relevant to reading development, reading instruction, and reading difficulties; and
>
> (B) shall include research that—
>
> (i) employs systematic, empirical methods that draw on observation or experiment;
>
> (ii) involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;
>
> (iii) relies on measurements or observational methods that provide valid data across evaluators and observers and across multiple measurements and observations; and
>
> (iv) has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review.

Figure 2. Language defining scientifically based reading research in Reading Excellence Act (1999).

Its effect on education researchers was indirect; by changing the kinds of research which were in demand, it could thereby affect the kinds of research which was supplied.

It is worth pointing out the extent to which REA departed from previous legislation regarding the funding of federal programs. Consider Goals 2000, enacted just four years earlier. Both laws were meant to encourage the use of education research in the formation of state education policy. Under Goals 2000, the federal government did so by providing leadership and support. It specified increased investments in conducting and disseminating new knowledge relevant to state problems, without setting out prior criteria on how that research was to be judged. It did not even limit federal support to *scientific* research, but rather allowed for the possibility for other kinds as well. It left discretion in such matters to the NERPPB and Assistant Commissioner of OERI.

By contrast, with REA the federal government did not simply provide leadership or support. It mandated that educational providers base their practice on scientific education research in order to be eligible for funding. Further, it specified exactly what was meant by scientific education research. Both Goals 2000 and REA aimed to increase consumption for research. Goals 2000 did so passively by attempting to reduce the transaction costs associated with consuming it. REA did so more aggressively by demanding that federally funded reading programs come with a scientific seal of approval.

## H.R. 4875

The following year, on July 18, 2000, Rep. Michael Castle (R-DE) introduced the Scientifically Based Education Research, Statistics, Evaluation, and Information Act (H.R. 4875, 2000), hereto referred to as "The Castle Bill." Whereas REA had addressed the *consumers* of education research limiting the kinds of evidence recipients of federal funds could use to justify their programs, The Castle Bill addressed the *producers* of education research, or at least those producers who sought to have their research supported by federal funds. Perhaps reflecting these different purposes, The Castle Bill elaborated in much greater detail definitions of scientific research and related concepts.

The Castle Bill attempted, perhaps ambitiously, to articulate two sets of standards, one for scientifically based quantitative research and another for scientifically based qualitative research. The standards for quantitative research followed the general standards set out the previous year in REA, but with one important addition. They stipulated that such research was defined as "using experimental designs…or other designs to the extent such designs contain within-condition or across-condition controls." The Castle Bill reflected the view of experiments as the so-called "gold-standard" of scientific research, with all other research designs judged worthy to the extent that they approximate experimental designs.

The standards for scientific qualitative research were quite different. Qualitative research was defined as, "the systematic collection and analysis of data often associated with traditions of inquiry historically based in the humanities, such as narrative analysis." The Castle Bill listed a series of methods covered by the definition, including "participant observation, in-depth interviewing, and document collection." It further stated that such research may include, "case studies, ethnographies, multi-site case studies, and participatory action research." It identified the purpose of qualitative research as being, "to explore issues and hypothesis whose underlying dynamics and factors are not sufficiently well refined, understood, or amenable to experimental control to permit adequate study through quantitative research."

It is easy to see how qualitative researchers could read this as stating that their work was a fallback, the kind of thing one did as a last resort. The language suggested that if an issue *could* be approached with either qualitative or quantitative methods, quantitative would be preferred. The division of labor between quantitative and qualitative research became even clearer in the standards for "sound program evaluation." Quantitative research was said to be used to assess the programs *outcomes.* Qualitative research was used (in combination with quantitative research) to study program

(5) SCIENTIFICALLY BASED QUANTITATIVE RESEARCH STANDARDS.—The term "scientifically based quantitative research standards"

(A) means the application of rigorous, systemic, and objective procedures to obtain valid knowledge relevant to education activities and programs; and

(B) includes research that—

(i) employs systematic, empirical methods that draw on observation or experiment;

(ii) involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;

(iii) relies on measurements or observational methods that provide valid data across evaluators and observers and across multiple measurements and observations and across studies by the same or different investigators;

(iv) is evaluated using experimental designs in which individuals, entities, programs, or activities are assigned to different conditions with appropriate controls to evaluate the effects of the condition of interest through random assignment experiments, or other designs to the extent such designs contain within-condition or across-condition controls;

(v) ensure experimental studies are presented in sufficient detail and clarity to allow for replication, or at a minimum offer the opportunity to build systematically on its findings; and

(vi) has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review.

(6) SCIENTIFICALLY BASED QUALITATIVE RESEARCH STANDARDS.—The term "scientifically based qualitative research standards"—

(A) means the systematic collection and analysis of data often associated with traditions of inquiry historically based in the humanities, such as narrative analysis; and

(B) includes research that—

(i) uses some combination of participant observation, in-depth interviewing and document collection;

(ii) is intended to explore issues and hypotheses whose underlying dynamics and factors are not sufficiently well refined, understood, or amenable to experimental control to permit adequate study through quantitative research.

(iii) may include case studies, ethnographies, life histories, multi-site case studies, and participatory action research;

(iv) uses approaches to assess the experimental knowledge acquired to assure that the findings are scientifically valid and replicable; and

(v) has been accepted by a peer-review journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review

Figure 3. Definitions of scientific education research in H.R. 4875, the original "Castle Bill."

*implementation*. Both were best employed in "some form of a classical experimental design with random assignment," or, "when experimental designs are not feasible…the strongest possible quasi-experimental alternative."

(7) SOUND PROGRAM EVALUATION.—The term "sound program evaluation"—

(A) means program evaluation that—

(i) adheres to the highest possible standards of quality with respect to research design, statistical analysis and the dissemination of findings;

(ii) provides an adequate understanding of the programs being evaluated, and examine program implementation, program impacts and the relationships between these factors;

(iii) provides impact estimates that truly reflect what was caused by the program;

(iv) produces or leads to findings that are broadly generalizable; and

(v) uses valid and reliable measures to document program implementation and impacts;

(B) includes only those program evaluations that—

(i) use qualitative and quantitative methodologies that are judged by the social science and evaluation research communities to be of the highest quality.;

(ii) in order to study program impacts use, whenever possible, some form of a classical experimental design with random assignment, in order;

(iii) when experimental designs are not feasible in order to study program impacts, use the strongest possible quasi-experimental alternative, basing it on longitudinal data; and

(iv) in order to study program implementation, use a combination of qualitative and quantitative methods.

Figure 4. Definitions of sound program evaluation in H.R. 4875, the original "Castle Bill."

Eisenhart and Towne (2003) argue that one of the more noticeable features of the standards for qualitative research is that they contained no standards. This observation is not entirely correct. The

bill stated that such research must be "systematic," "replicable," and that it must be accepted by a peer-review panel or panel of independent experts. These may not provide much specific guidance—or worse, they may misguide—but they were not significantly vaguer than the standards for quantitative research.

It is important to distinguish between the push for randomized field trials and the push for quantitative methods. The bill clearly advocated for both, but the two are not necessarily connected. A qualitative researcher could randomly assign students to different tutoring programs and then conduct in depth interviews to assess their impact. Likewise, a quantitative researcher could tinker around at a single site, collecting various data in the hopes of having some interesting observation emerge. This relatively minor issue points to a larger problem with The Castle Bill—the identification of science with methodology. I return to this point in the next chapter.

The Castle Bill placed science at the core of education research, with quantitative research at the core of that core, and experiments at the core of that core. Unlike Goals 2000, federal support was limited to "scientific" research, even though the line between science and other fields of inquiry was rather unclear. Following REA, quantitative research is presented as a more matured version of qualitative research. Finally, in a novel addition, it inserted experimental designs to be the preferred method for conducting research, regardless of the kinds of questions one might be asking.

## No Child Left Behind

At the same time that The Castle Bill was moving through Congress, debates were under way concerning the reauthorization of ESEA, the bill that would be rebranded as the No Child Left Behind Act of 2001 (NCLB). Like REA, NCLB was intended to regulate which kinds of research can be used to justify

federally-funded educational programs. As written, it seemed that the kinds of research the government would allow to justify federal support were significantly more limited than the kinds of research the government would fund. Whereas The Castle Bill, a law addressing the funding of research, allowed for both quantitative and qualitative research, NCLB adopted the Castle Bill's definition (with slight modifications so as to allow quasi-experiments) of "scientific quantitative research" as the definition for scientific education research *as such* (NCLB, pp. 540-541).

(37) SCIENTIFICALLY BASED RESEARCH.—The term 'scientifically based research'—

(A) means research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities
and programs; and

(B) includes research that—

(i) employs systematic, empirical methods that draw on observation or experiment;

(ii) involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;

(iii) relies on measurements or observational methods that provide reliable and valid data across evaluators and observers, across multiple measurements and observations, and across studies by the same or different investigators;

(iv) is evaluated using experimental or quasiexperimental designs in which individuals, entities, programs, or activities are assigned to different conditions and with appropriate controls to evaluate the effects of the condition of interest, with a preference for random-assignment experiments, or other designs to the extent that those designs contain within-condition or across-condition controls;

(v) ensures that experimental studies are presented in sufficient detail and clarity to allow for replication or, at a minimum, offer the opportunity to build systematically on their findings; and

(vi) has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review.

Figure 5. Definition of scientifically based research in NCLB (2001).

Of all of the definitions of scientific research discussed in this chapter, the one adopted in NCLB was the most restrictive. The inconsistencies between these laws may be explained by comparing the different purposes served by REA and NCLB, on the one hand, and The Castle Bill (and later ESRA) on the other. The definitions seemed intended not to describe accurately the actual characteristics of science, but to serve the political purpose of using federal resources to concentrate capital on certain education and research programs. When it came to evaluating whether a certain educational program should continue to receive funding, the federal government was more selective in the kinds of evidence which it would admit. Thus, scientific research was described as being primarily aimed at making casual connections between interventions and outcomes. However, when it came to supporting the production of education research, the federal government aimed to develop a more diverse field of high-quality work. Thus, it sought a definition of scientific research which will allow it somewhat more latitude in allocating resources.

### *Scientific Research in Education* (*SRE*)

In the fall of 2000, the National Research Council formed the Committee on Scientific Principles for Educational Research. The Committee met for the first time in December 2000, at which point it was given presentations regarding the context of its report. In March 2001, the Committee hosted a workshop on "science, evidence, and inference in education" (National Research Council, 2002, p. ix). At the workshop, a number of presentations were made by researchers from universities, government, and private institutions. Following the workshop, the Committee spent a number of months collecting more information about how the DOE and other federal agencies supported scientific research. A draft of the Committee's report was sent for review to individuals chosen for their "diverse perspectives and technical expertise." The Committee considered the reviewers responses, made appropriate

amendments, and in early 2002 released its final report, *Scientific Research in Education* (referred to as *SRE*), a document which has become the centerpiece in debates over the nature and role of scientific research in education.

The Executive Summary states that the report was commissioned by the National Resource Council to address two concerns. First, legislation was currently being written which would define "what constitutes rigorous scientific methods for conducting education research" (National Research Council, 2002, p. 1). Second, they wanted to meet the "rising enthusiasm for evidence-based education policy and practice" (p. 1). The Committee investigated three questions (pp. 22-24):

1. What are the principles of scientific quality in education research?
2. How can a federal research agency promote and protect scientific quality in the education research it supports?
3. How can research-based knowledge in education accumulate?

It seems clear just by the questions the committee asked that it would suggest very different legislation than that included in REA or The Castle Bill. The first question foreshadowed the NRC committee's argument (1) that an examination of the nature and purpose of science should center on the *principles* of scientific inquiry, rather than the specific *methodologies* used in the pursuit of these principles, and (2) that "scientific inquiry is the same in all fields" (p. 2). The second question showed that the committee was focused on federal support of research, not in federal support of education. The third question anticipated the claim that "it is possible to describe the physical and social world scientifically so that, for example, multiple observers can agree on what they see" (p. 25). The authors rejected the notion that claims to knowledge have no rational basis and were merely expressions or exercises of social/political power.

The committee argued that "six guiding principles underlie all scientific inquiry, including education research" (p. 2). The principles were broad enough to reliably cover most of what one might consider to be scientific research. Importantly, the report argued that the principles extended to both

quantitative and qualitative inquiry, the distinction between the two being largely "outmoded"(p. 19). It cautioned against guidelines focused on certain methods or techniques, arguing that science advances by the self-regulating norms of a community of inquirers. It stated, "the history of scientific inquiry attests to the fact that there is no one method or process that unambiguously defines science" (p. 24). The federal role should focus on helping create a scientific culture among educational researchers, understood as a culture which adheres to the basic principles of scientific inquiry, rather than to certain tools or techniques. *SRE* rejected the notion of articulating separate guidelines for quantitative and qualitative research, insisting that scientific inquiry in all areas adheres to the same general norms.

Principles of Scientific Inquiry

1. Pose significant questions that can be investigated empirically.
2. Link research to relevant theory.
3. Use methods that permit direct investigation of the questions.
4. Provide a coherent and explicit chain of reasoning.
5. Replicate and generalize across studies.
6. Disclose research to encourage professional scrutiny and critique.

Design Principles for Fostering Science in a Federal Education Research Agency

1. Staff the agency with people skilled in science, leadership, and management.
2. Create structures to guide the research agenda, inform funding decisions, and monitor work.
3. Insulate the agency from inappropriate political interference.
4. Develop a focused and balanced portfolio of research that addresses short-, medium-, and long-term issues of importance to policy and practice.
5. Adequately fund the agency.
6. Invest in research infrastructure.

Figure 6. Principles concerning scientific research in education in *Scientific Research in Education* (2002).

On the view given in *SRE*, the guidelines set forth in Goals 2000, REA, The Castle Bill, and NCLB were both too broad and too narrow. They were overly narrow by confusing science with the

methodological tools used in some scientific inquiries. According to *SRE*, the guidelines missed the core of scientific activity, which is adherence to the basic principles of open and honest empirical inquiry. They were too broad because many studies which satisfied the methodological requirements may be seriously lacking in other components.

The report stated that it owed its existence at least in part to the contemporaneous debates surrounding The Castle Bill. Thus, it singled this bill out for criticism. The passage is worth quoting at length:

> [A]ttempting to boost the scientific basis of federally funded education research by mandating a list of "valid" scientific methods is a problematic strategy. The inclusion of a list of methods—regardless of how well they are applied in particular situations—erroneously assumes that science is mechanistic and thus can be prescribed. We have shown that science adheres to a set of common principles but its application depends greatly on the particulars of a given situation and the objects of inquiry. The definitions also make clear distinctions between quantitative and qualitative methods, implying that these two types of research approaches are fundamentally different; we argue the opposite. Furthermore, the use of definitions of methods as a tool for improvement fails to recognize the crucial role of theory and, as we emphasize, a strong, self-regulated, skeptical community of researchers that pushes the boundaries of knowledge (p. 130).

The reports claimed that in addition to scientific research, the federal government had a role in supporting humanities based research in fields such as history and philosophy. It stated, "it is the integrations of scientific knowledge with insights from the humanities and other scholarly pursuits that will ultimately yield the most powerful understanding of education" (p. 131). Thus, there should be some restraint against making it impossible for humanities and arts-based researchers to accumulate capital.

*SRE* was released two months after the signing of NCLB. Eisenhart and Towne (2003) argue that *SRE* was released too late to have a chance to influence NCLB. However, if one fully appreciates the different political purposes these definitions serve, it becomes plausible that *SRE* was not meant to influence NCLB. The questions driving *SRE* did not have to do with which kinds of evidence ought to be

allowed in the justification of federally funded education programs. Rather, *SRE* addressed how the federal government could play a positive role in supporting scientific education research.

*SRE* attempted to make a rather direct case for justifying federal intervention in the political economy of education research. The report opens, "No one would think of getting to the Moon or of wiping out a disease without research. Likewise, one cannot expect reform efforts in education to have significant effects without research-based knowledge to guide them" (National Research Council, 2002, p. 1). That is, the report argues strongly that the problems facing education are essentially due to a lack of knowledge, rather than some moral failure. The comparisons with aerospace and public health are telling. No matter one's level of care or commitment, it would simply be impossible to build a vehicle capable of escaping the earth's gravitational pull, or a vaccine capable of rendering a virus harmless, without significant investments in scientific research.

The report describes education as being "complex" and "changing," thus creating a situation in which educators "require new knowledge to reengineer schools in effective ways" (pp. 11-12). It continues, "To meet these demands, rigorous, sustained, scientific research in education is needed. In today's rapidly changing economic and technological environment, schooling cannot be improved by relying on folk wisdom about how students learn and how schools should be organized" (p. 12). However, the political economy of education research, as currently structured, is unlikely to produce such research. The report states, "In the absence of a federal leadership role, knowledge gained by one state or district that might be relevant to others would not likely be widely distributed" (p. 127). There is insufficient supply, stemming partly from a lack of public support, which is in turn partly attributed to problems of research quality, fragmentation, unrealistic expectations on behalf of policy-makers, and a lack of relevance to the needs of practitioners (p. 14).

Thus, in articulating how a federal agency can remedy this problem, the report focused on increasing support, staffing the agency with those who can help create a "scientific culture," crafting a balanced research agenda attentive to all stakeholders, and creating a mix of projects which are short-, medium-, and long-term. By making such investments, the federal government could encourage the production and consumption of research which is relevant to the needs of practitioners and rigorous enough to satisfy them. The report stated that a federal agency did not simply reflect the field it supported, but that the field it supported may come to reflect it. Indeed, this was the hope.

## ESRA

It is unclear to what extent *SRE* had any concrete effect on the Education Sciences Reform Act of 2002 (ESRA), the bill it was intended to address.[5] A comparison of the original Castle Bill and the bill that eventually passed as the Education Sciences Reform Act of 2002 reveals that there were indeed changes. ESRA dropped the narrow definitions recently inserted into NCLB. Margaret Eisenhart and Lisa Towne (2003) argue that the *SRE* was partially responsible for this shift. They describe the similarities between the NRC report and ESRA as follows:

> As in *SRE* (NRC, 2002), the ESRA (2002) definitions of scientifically based research that can receive federal education funding are relatively broad. There are neither different standards for quantitative and qualitative work, nor is it implied that qualitative research has no standards (i.e., is simply a list of methods) or is only preliminary. More importantly, in ESRA research is not confused with research methods; that is, its definitions are for research writ large rather than for methods only. Furthermore, ESRA no longer includes the requirement that studies always "test hypotheses," thereby opening up space for explanatory, descriptive, naturalistic, and hypothesis-generating studies critical to scientific research. ESRA also acknowledges that causal conclusions can be drawn from nonrandom assignment designs, and it revises the original Castle Bill language (H.R. 4875, 2000) associated with other causal methods from insisting on "controls" to the more appropriate call for the research to rule out competing

[5] Michael Feuer, Lisa Towne, and Richard Shavelson, three architects of *SRE,* argue that the effect was substantial. See Feuer, et al (2002), p. 56.

> explanations for observed differences….research cannot be prescribed a priori, but is nuanced according to the nature of the individual investigation (p. 34).

While Eisenhart and Towne do a fine job pointing to the changes from The Castle Bill (H.R. 4875) and the final passage of ESRA, the continuities between the two (and between NCLB and ESRA), are worth noting. Like NCLB, ESRA assumes that there are no significant differences in the standards for qualitative and quantitative research. While Eisenhart and Towne see this as a sign of progress, others (Baez & Boyles, 2009; St. Pierre, 2002) argue that is in fact an imposition of the standards of quantitative research upon qualitative inquirers. Elizabeth Adams St. Pierre (2002) goes so far as to accuse the report of endorsing "'terrorist' ideals of consensus"(p. 27). Beneath the rhetorical flourish of linking *SRE* with terrorism is a legitimate concern that federal policy might restrict inquiry. This is a concern shared by nearly all researchers.

On the issue of causality, ESRA states that scientific research makes "claims of causal relationship only in random assignment experiments or other designs (to the extent such designs substantially eliminate plausible competing explanations for the obtained results" (ESRA, p. 8). A similar clause in NCLB states that scientific research includes research:

> [U]sing experimental or quasi-experimental designs in which individuals, entities, programs, or activities are assigned to different conditions and with appropriate controls to evaluate the effects of the condition of interest, with a preference for random-assignment experiments, or other designs to the extent that those designs contain within-condition or across-condition controls (pp. 540-541).

Both passages clearly set out random assignment as the standard by which other designs are to be judged. The primary difference is that ESRA limits the preference for random assignment to investigations of *causal* questions, whereas NCLB does not make state any such limitation.

(18) SCIENTIFICALLY BASED RESEARCH STANDARDS.—

(A) The term ''scientifically based research standards'' means research standards that—

(i) apply rigorous, systematic, and objective methodology to obtain reliable and valid knowledge relevant to education activities and programs; and

(ii) present findings and make claims that are appropriate to and supported by the methods that have been employed.

(B) The term includes, appropriate to the research being conducted—

(i) employing systematic, empirical methods that draw on observation or experiment;

(ii) involving data analyses that are adequate to support the general findings;

(iii) relying on measurements or observational methods that provide reliable data;

(iv) making claims of causal relationships only in random assignment experiments or other designs (to the extent such designs substantially eliminate plausible competing explanations for the obtained results);

(v) ensuring that studies and methods are presented in sufficient detail and clarity to allow for replication or, at a minimum, to offer the opportunity to build systematically on the findings of the research;

(vi) obtaining acceptance by a peer-reviewed journal or approval by a panel of independent experts through a comparably rigorous, objective, and scientific review; and

(vii) using research designs and methods appropriate to the research question posed.

(19) SCIENTIFICALLY VALID EDUCATION EVALUATION.—The term ''scientifically valid education evaluation'' means an evaluation that—

(A) adheres to the highest possible standards of quality with respect to research design and statistical analysis;

(B) provides an adequate description of the programs evaluated and, to the extent possible, examines the relationship between program implementation and program impacts;

(C) provides an analysis of the results achieved by the program with respect to its projected effects;

(D) employs experimental designs using random assignment, when feasible, and other research methodologies that allow for the strongest possible causal inferences when random assignment is not feasible; and

(E) may study program implementation through a combination of scientifically valid and reliable methods.

Figure 7. Definitions of scientific research and evaluation in ESRA (2002).

## Implementation of ESRA

ESRA reorganized education research within the Department of Education, abolishing the Office of Educational Research and Improvement and replacing it with the Institute of Education Sciences (IES). Grover "Russ" Whitehurst was appointed to a six-year term as the first Director of IES. In coordination with the NERPPB, he would be responsible for interpreting the law and setting the priorities for research within IES. Over the course of six years, Whitehurst used this opportunity to bring greater attention to problems in which clear interventions were possible. Furthermore, he helped create a climate which would nurture and reward those who used randomized control field trials (RFTs) to investigate these problems. IES quickly gained the reputation as being intensely, and perhaps narrowly, focused on questions of causality, or "what works," and for insisting that the best and perhaps only valid evidence of such relationships can be found in RFTs.

This focus on causal questions was evident in the creation of the What Works Clearinghouse (WWC), an initiative intended to provide clear evidence regarding the causal effectiveness of various educational interventions. To be included in the WWC reviews of effectiveness, research must meet a list of methodological requirements. In essence, studies must either be RFTs or closely approximate the conditions of RFTs. Many have argued that these methodological requirements prevent serious criticism of the conceptual, theoretical, or normative assumptions behind the interventions (see for instance Baez & Boyles, 2009; Hyslop-Margison & Naseem, 2007; Lather, 2004; St. Pierre, 2002). Even among those who support the use of RFTs for the evaluation of causal questions, the degree of emphasis put on "what works" has drawn criticism (Feuer, 2006; Phillips, 2008). It seems clear that randomized trials and regression discontinuity are not the only reliable scientific sources of evidence of causal connections. A recent AERA/NSF report, *Estimating Causal Effects: Using Experimental and Observational Designs* (2007), argues that other worthy quantitative methods, such as fixed effects models, instrumental

variables, and propensity score matching, have been overlooked. The WWC has since somewhat liberalized its methodological requirements, allowing a wider range of research into its systematic reviews. In addition, IES has helped create another online clearinghouse, Doing What Works (DWW). DWW features video and other tutorials in an effort to make scientific research more user-friendly. The publications on DWW are not necessarily as methodologically stringent as those endorsed by the WWC, acting on the belief that practitioners cannot wait for the best possible evidence to arrive years down the road.

## Reviewing Fifteen Years of Federal Involvement in the Research Economy

This chapter began with two major questions. First, how has federal policy sought to influence the kinds of people engaged in the production of education research, as well as the questions they ask and the research methods and designs they use to answer them? Second, how has federal policy sought to influence the consumption for education research? That is, in what ways has the federal government attempted to shape the populations which demand research as well as the kinds of research these populations demand.

As the above discussion demonstrates, the government is actively intervening in both the production and consumption of education research. NCLB is a consumption-focused intervention, urging districts to base their decisions in scientific research. However, this intervention can be successful only if there is a steady production of such research. ESRA is the production-focused counterpoint to NCLB. It shapes, to a greater or lesser extent, the questions and methods which are worthy of federal funding.

The last fifteen years has brought attention on increasing the supply of rigorous and relevant education research. Once again, the government is turning to the hard sciences for guidance, though now the focus is on methods and designs rather than disciplinary orientation. Humanistic researchers and those who take a qualitative approach to the social sciences find themselves in a similar situation as the educationists of the 1960s.

One way the government has attempted to stimulate consumption is by lowering the barriers to access to research products. It has continued the longstanding attempt to construct various clearinghouses of education research through the Educational Resources Information Center (ERIC), What Works Clearinghouse (WWC), and most recently, Doing What Works (DWW). It has also continued to operate regional labs designed to help demonstrate and disseminate research to educators throughout the country. Beginning with the Reading Excellence Act and continuing with No Child Left Behind, the government has taken an additional approach to increasing demand for research by mandating that districts use educational strategies which are "scientifically-based." Furthermore, subsequent legislation has refined the meaning of scientifically-based research, thus making it clear which kinds of research ought to be used.

The pushing and pulling effect of legislation such as NCLB and ESRA may remedy the longstanding obstacle to the hope of making education a science. If history is any guide, there is little organic demand among educators to continuously consume education research as a way of improving practice. However, the federal government may be able to successfully generate demand by conditioning federal funding for educational programs on the rider that such programs have been scientifically proven to be effective. Superintendents and school boards who wish to be in compliance may therefore become active consumers of research.

The effects of federal involvement (or lack thereof) over the last fifteen years are unknown. Or more accurately, evidence of their effects is confined to anecdotal accounts given by various interested parties. Prior work has shown that policy implementation is a complicated matter (see for instance Weick, 1976). The intentions of lawmakers are often unclear, and even when they are transparent, they often confront a culture with multiple competing agendas. While it is clear that attempts to nudge the research economy to federal priorities have the possibility of producing unintended consequences, it is less clear whether those consequences will be good or bad. This chapter presented a detailed exposition of the conceptions of science contained in recent legislation. This helps to frame the more explicitly normative investigation that takes place in the next and final chapter, which examines the criticisms and defenses of these interventions, focusing on the ethical and epistemological arguments in support of the recent push for randomized control field trials.

# CHAPTER IV

# CRITICISM AND REPLIES

## Introduction

In the first chapter, I argued that the political economy of education research depended to a large extent on the free and unrelenting criticism of self-interested agents. If this economy is well-functioning, agents gain or lose capital as a result of their ability to withstand the scrutiny of a diverse range of critics. The second and third chapters illustrated some of the ways in which federal actors can attempt to manipulate the flow of capital in this economy, increasing the fungibility of certain credentials, affiliations, or disciplinary orientations. These attempts are not always, or even often, successful. Indeed, one might see the history of federal involvement as a series of mostly failed attempts to make the political economy of education research more responsive to its own priorities by changing the values attached to certain dimensions of research. Learning perhaps from the lessons of the past, recent federal interventions have reached deeper than ever before into the heart of the political economy of education research, going so far as to define the nature and uses of scientific inquiry in education.

These interventions have brought to light long-standing schisms in education research: activism versus technologism; clinical versus naturalistic; quantitative versus qualitative; scientific versus humanities and arts-based; basic versus applied. It may have even opened a new schism between innovation and evaluation. This fourth and final chapter explores the degree to which current federal policy does, in fact, seem designed to concentrate capital on scientists, quantitative researchers, and

those who conduct randomized control field trials. It then moves to the more difficult task of evaluating the arguments as to whether such a policy is justifiable.

## Core Argument for Focusing on RFTs

The strongest argument in favor of the increased federal focus on randomized control trials may be as follows: First, the federal government should support the production and consumption of education research only if that research is both *rigorous* and *relevant*. Second, only research which is scientific can be both relevant and rigorous (humanities and arts-based research lacks in one or both of these qualities). Thus, the federal government should support education research only if that research is scientific. However, even some scientific research (perhaps *most* scientific research) is not highly relevant to public concerns. The most *relevant* scientific research addresses *what works*, that is, causal relationships between distinct interventions and outcomes. The most *rigorous* scientific approach to causal questions is the randomized field trial. Therefore, the federal government should limit support to investigations of causal questions through the use of randomized control field trials.

## Four Claims

In order to access the strength of this argument, and thus the strength of the justification for current federal policy, I examine various challenges to the claims upon which it is based as well as the validity of the inferences drawn from these assumptions.

Specifically, I examine four basic claims which have drawn significant scholarly attention. The first claim is that the federal government should have a role in shaping the political economy of education research. That is, some have argued that either the nature of government or the nature of

education research mean that the former should not have a significant role in shaping the latter. The first claim is often defended by appeal to a second claim, namely, that the public needs to answers to the question of "what works" in education and that the federal government has a legitimate and necessary role in finding answers to these questions. The third claim is that given the need to concentrate on what works, the government should promote scientific rather than humanities and arts-based research, as they fail to provide relevant or rigorous answers to such questions. The fourth claim is that RFTs are the best or only scientific technique for establishing warranted answers to causal questions.

Claim 1: The federal government should have a role in shaping the political economy of education research.

I stated in the first chapter that federal interventions can differ importantly in kind and in degree. Regarding differences in kind, I suggested that we can usefully distinguish four areas of intervention: topics, questions, methods and designs, and dissemination. Regarding differences in degree, I stated that the level of influence can range from complete abstention to total control. It is still true that one might support the general statement that the federal government should exercise total control over some aspect of education research, yet still oppose the *specific* form this control might take. For instance, two people may agree that the federal government should strictly control which methods are used but disagree which methods should be dictated. This section addresses only the *general* arguments over the legitimacy or various kinds and degrees of federal intervention, leaving discussion of particular cases to the sections that follow.

Many have argued that either the nature of government or the nature of education research mean that the former should not have a significant role in shaping the latter (Baez & Boyles, 2009;

Hyslop-Margison & Naseem, 2007; National Commission on Excellence in Education, 1983; St. Pierre, 2002; Curran, as quoted in Vinovskis, 2001, p. 102). At least two arguments can be advanced on behalf of total abstention. First, one might argue that educational problems are impenetrable to scientific inquiry and are thus matters of moral or political debate rather than scientific or technical investigation (Baez & Boyles, 2009; Hyslop-Margison & Naseem, 2007). Second, one can argue that science requires absolute freedom over every dimension of the research process and that any attempt to impose external constraints on inquiry pervert integrity of scientific inquiry (it is not clear that anyone holds this position, but it may still warrant a preemptive rebuttal). Both arguments represent foundational attacks at the legitimacy of federal intervention in the political economy of education research. If either argument holds, then the more nuanced debates over which topics and methods are best will be rather unimportant.

*Scientific research in education is impossible.*

Let us then begin with the notion that scientific research in education is impossible, or at least that scientific research into the most relevant educational problems is impossible. The claim is sometimes made that teaching is a political or artistic activity. One can hardly deny the truth of such a statement. However, this in no way shows that scientific investigation into education is impossible. To make that claim, one would need to prove not simply that education is a political or artistic activity, but that political or artistic activity cannot be understood by means of scientific investigation (Dewey, 1984).

A series of proofs of the wrongheadedness of scientific inquiry in education are offered in Emery Hyslop-Margison and M. Ayaz Naseem's, *Scientism and Education: Empirical Research as Neo-Liberal Ideology* (2007). The authors are not skeptical of science *as such*, but rather are skeptical about the use of science to investigate human affairs. They argue that scientific methods are useful for investigating

"appropriate" subject matter, such as chemistry, physics, and biology. However, on their view humans cannot be understood in the "materialist" terms presumed by empirical education research. By "materialist," they mean the notion that all phenomena can be explained by appeal to observable, physical, entities.

Their first argument for the impossibility of scientific research in education proceeds as follows: The possibility of science depends on the assumption that events are causally determined. Thus, the possibility of a science pertaining to human actions must assume human actions are causally determined. However, the assumption that human actions are causally determined, "makes absolutely no sense in relation to our lived experience" (Hyslop-Margison & Naseem, 2007, p. 4). Therefore, human actions are not causally determined and are consequently beyond the scope of science.

This proof hangs to a great extent on the tenuous claim that our "lived experience" shows that it is not the case that "human actions are causally determined by antecedent forces that can be identified and manipulated" (p. 4). If our lived experience seems to indicate otherwise, then the proof fails. It may be true that people tend to experience themselves as exercising some measure of freedom. However, it is doubtful whether anyone has ever had a "lived experience" of being uncaused by antecedent forces. Being human requires one to choose under conditions of finitude and temporality. Decades of work in fields as diverse as sociology, cognitive science, and phenomenology continue to remind us that institutions, brains, and bodies all matter. To neglect the extent to which our actions are conditioned by antecedent forces is to forfeit whatever measure of freedom and capacity for control and redirection one may have potentially had.

The notion of a radically autonomous subject is phenomenologically unsupported. It certainly appears that my actions in the future are constrained (though perhaps not determined) by antecedent forces (including my prior choices). If I drink a pint of whiskey tonight, I am less able to choose to run a

marathon in under four hours tomorrow. If I have not yet been exposed to the alphabet, I cannot read *The New York Times.* William James (1978) made the point more directly, saying that there is always some pinch between the ideal and the real, and that some portion of the ideal is inevitably butchered. The relevant lesson is that the world both enables and constrains our choices. It is in understanding how it does so that social progress becomes more secure. Indeed, the authors' own insistence on social and political reform seem to assume such a view. If it were not for the assumption that conditions in the real world—such as the distribution of income—significantly influence the range of human possibility, there would be no reason to support the various political reforms they champion.

Hyslop-Margison and Naseem's second argument proceeds as follows: Scientific statements must be empirically testable. The statement "human actions are caused" is not testable. Therefore, any statement whose truth-value depends on the truth-value of the statement "human actions are caused" is non-scientific. The truth-value of statements about causes of human actions depends on the truth-value of the statement "human actions are caused." Therefore, statements about human action cannot be scientific.

This argument is a bit of a fallback from the first. That is, even if our experience *seems* to suggest that our actions are caused, one cannot make a testable statement about whether "deep down" they were really free or not. Such statements devolve into metaphysics, a branch of inquiry allegedly shunned by science. However, Hyslop-Margison and Naseem fail to prove that scientific inquiry requires any special conviction about what is happening "deep down." Or, put differently, scientific inquiry requires no more metaphysical commitment to this deterministic view that does any other activity which makes use of the past and present to help shape a future. We encourage the infant to reach for the bottle of milk rather than the bottle of bourbon on the assumption that what happens early in life might determine the range of possible actions in the future.

Their criticism is neither proof that the human sciences presume positivism or that positivism is necessarily self-refuting. The notion that the past matters can be endorsed by a range of non-positivists. Even within the positivist framework, the view might be tested without appeal to any thick metaphysical notions. If observations made at time $t_1$ in no way inform my understanding of what happens at $t_2$, then I may be warranted in my rejection of the assumption. If observations made at $t_1$ do inform my understanding of what happens at $t_2$, then I may be warranted to accept the assumption.

The third argument for the impossibility of scientific inquiry in education is a bit more difficult to discern. The authors offer more of a statement than an argument. Nonetheless, the following implied argument seems to lurk behind the statement: Scientific explanations are possible only if one can indentify clear and distinct causes. Even if human actions are caused, the process of causation is so complex that no scientific method could explain them. Thus, science explanations of human actions are impossible.

Scientists can concede that the world is indeed complex. The social and human dimensions of the world are especially complex. (Thus, one prominent researcher claims that education science is "the hardest science of all." See Berliner, 2002). For the better part of human history, this complexity has so impressed humans that few attempted to unravel it. Indeed, that the very attempt to study the world was long seen as being a sign of hubris is evident in the mythological accounts of Prometheus stealing fire from the gods and of Eve eating from the tree of knowledge. There are things, it has been said, that are rightfully beyond human understanding. The apostles of such views exhort others not to approach the world directly, but to passively accept things as they are.

There are also those who have interpreted complexity not as a cause for passivity but as the impetus for insight and investigation (for stirring examples, see Dewey, 1984; Dewey, 1986). Indeed, if the world were simple, scientific investigation into the connections between antecedents and

consequents might be unnecessary. Complexity gives science its *raison d'etre*. It seems rather odd to state *a priori* that education is so complex that nothing can help inform the decision to keep children in classrooms of 15, 20, or 50 students, or to see to it that teachers are educated through middle, secondary, or post-secondary school, or to have children spend 10, 20, or 200 minutes per day in physical activity.

This is not a defense of determinism but rather a statement that scientific explanation need not accept the determinism described by Hyslop-Margison and Naseem. It is one thing to say that educational consequences flow from myriad causes, or that causality is not simply linear. This is true, though trivial. It is an entirely different thing to imply that therefore the identification and manipulation of any causal agent or group of causal agents is impossible. This is demonstrably false. Through inquiry, we discover various fulcrums by which changes in social conditions can be leveraged. In the blooming, buzzing confusion of the world (to borrow again from James), careful observation can lead to testable hypothesis about why things happen the way they do, and how things can be changed.

*Scientific research in education is irrelevant.*

For reasons stated above, each of the three proofs of the impossibility of scientific research in education fail. This brings us to a second group of criticisms, namely that education research is irrelevant, immoral, or both. For present purposes, I refer to this as *skepticism concerning the relevance of empirical knowledge to educational problems.* Scientific research might be irrelevant if the true problems of education are matters of morals or politics and thus require changing the hearts of individuals or changing social and political institutions. This assumes that the answers have already been found and that all is needed is the conviction to act accordingly. Science, on this view, deals exclusively with means and has nothing to say about ends. It may follow from this that given a choice

between funding education research and funding education directly, some simple utilitarian calculus reveals that it may be immoral to fund education research, as it will produce less good than direct investment in education. The immorality of education research would then be conceived of consisting in lost opportunity costs. This assumes that there is a fixed amount of financial and intellectual capital to be divided between education research and education. If this is the case, any investment in education research might be, "virtually superfluous and an unnecessary drain on severely limited resources," as well as a distraction from, "the real issues of social class disparity and dwindling financial resources" (Hyslop-Margison & Naseem, 2007, p. 2). On this view, technical questions are easy and already understood. Such a view exudes confidence that we know exactly what is wrong (social class disparities), what causes it (a lack of political will), and how to fix it (changing the distribution of income).

There is a remarkable similarity between Hyslop-Margison and Naseem's opposition to education research and that of many conservative members of Ronald Reagan's Department of Education. Both seem to think that education research serves primarily as a way to smuggle reprehensive values into education while shielding these values from criticism. Just as Hyslop-Margison and Naseem assume that research is used to cloak conservative values, Edward Curran, who served as director of NIE while advocating for its abolition, reportedly suspected that social science was being used within NIE to justify a pro-government liberal agenda (Vinovskis, 2001). Both think that education research distracts us from, as Hyslop-Margison and Naseem state, "proper moral decision making" (pp. 5-6), or as William Bennett's views were described, getting "people to do what they know they ought to be doing" (Kaestle, 1992, p. 28). Hyslop-Margison and Naseem's claim that, "we already possess the required knowledge and resources" to fix our problems, but lack the "political and moral will to transform that knowledge into political, social, and ultimately, educational action" (pp. 5-6), is

remarkably similar to William Bennett's view that the "solution lies in conviction, in character, not in understanding the cause of the phenomenon" (Kaestle, 1992, p. 28).

Mark Twain once remarked, "It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so." It is difficult to reconcile acknowledgment of the complexity of human affairs with certitude about how best to improve them. This certitude warrants contempt once it refuses to open itself to the tribunal of experience. There is no *a priori* link between empirical research and "instrumental forms of thinking." Whatever evidence there is of some connection between the two must be found in experience. If John Dewey was correct to argue that we cannot fully understand the significance of our ends until we fully understand the means or instruments connected with bringing them about, then we should be hesitant to roundly denounce sustained analysis of instruments (Dewey, 1986, 1988). Education is ill-served not just by those who ignore ends in favor of discussion of means, but by those whose worship of some ideal precludes intelligent investigation into the material conditions for its realization.

It is difficult to imagine how such a position could be consistent with democratic processes. Presumably, those who believe that we know how to improve education should be forthcoming and specific in explaining how to do so. They should propose some intervention. It would then be fair to ask for evidence that this intervention—whether it be the nationalization of the means of production, or the abolishment of teacher tenure—will lead to the outcomes promised. It seems, to state it mildly, *morally hazardous*, to take actions which affect the public without providing publicly assessable evidence about the likely outcomes.

*Federal intervention should be limited to one or more of the following areas: topics, questions, methods and designs, dissemination.*

The two arguments for total federal abstention—based either on the *impossibility* of education science or on the *irrelevance* of education science—suffer major flaws. At least two positive reasons can be advanced in support of some sort of federal involvement. First, there is a certain amount of information necessary for the federal government to faithfully execute its duties. The various agencies within the federal government have long conducted research connected with its core duties, whether it be coastal navigation, economic forecasting, or census taking (see Chapter Two). The duties assigned to the various federal agencies necessitate the acquisition of information that may not be reliably available without federal intervention. Second, in addition to these basic duties, there are viable public benefits to be gained by improving our understanding of education. The federal government regularly makes investments in other areas of inquiry when there seems to be potential for public benefit, whether investing in the development of artists and scientists or conducting agricultural and epidemiological studies. Federal involvement is warranted whenever there is a shared problem, the solution to which is possible, but not guaranteed given the current political economy of research.

However, it is important to point out that there is always the danger of unintended consequences. Thus, it is just as important to decide *how* the government should intervene as it is to decide whether it *should*. The technical question is just as important as the normative question, and is indeed integral to it. Thus, I return to the typology of areas of intervention presented in the first chapter. It is on this vital question of *how* that much current debate turns. The federal government can intervene in the topics, questions, methods and designs, dissemination, and interpretation. In each area, the depth of intervention ranges from total abstention to total control. The extremes are

unattainable in practice, as even total abstention may be a subtle form of control, while total control is impossible to achieve.

Researchers who attach an especially high value to their autonomy may advocate for the *silent partner* model of federal intervention. That is, one might imagine a policy in which the government is the chief-check writer while the major decisions which shape the direction of inquiry are left to the self-policing community of researchers. The main benefit of the silent partner model is that it may insulate researchers from external political pressure. (It does little to resolve the sometimes bitter internal politics of research.) However, from the perspective of the federal government and perhaps the public at large, this benefit is also its main problem. The two main justifications for federal involvement in research—the effective execution of its core duties and the promotion of general welfare—require the continuous production of research on topics and questions which have been identified through proper deliberative channels as warranting public attention. The government must be able to explain why it operates the programs and conceives the public welfare as it does.

Thus, there seems to be little ground upon which to build an objection to *some* involvement by the federal government in shaping the topics and questions addressed within the political economy of education research. However, there are a number of important issues that should be considered. What is gained or lost by having either non-researchers or researchers be responsible for identifying topics of "public concern"? When is the public better served by allowing researchers to pursue their own curiosity rather than incentivizing them to focus on public demands? These questions have garnered more attention in recent years, and it is unlikely that their importance will soon recede. Further, it does not seem that any algorithmic answer will succeed. The flow of capital within the political economy of education research is somewhat volatile and mysterious, thus frustrating hopes of any simple calculation of utility.

Indeed, many of the arguments supporting recent federal interventions in the political economy of education research appeal directly to specific conditions. These arguments have seemed to favor a bureaucratic rather than field-initiated approach, not only to the definition of public problems, but also to the methods for studying them and the audience who will receive them. Specifically, IES has seemed to trend toward claiming that the areas of greatest public need are in examining "what works" questions, through the use of RFTs or close approximations, and controlling the flow of information by only disseminating studies which conform to these criteria.

I hope the above discussion has sufficiently argued that federal intervention in the flow of capital is not inherently nefarious, even when these interventions reach deep into the heart of the operation of research. That is, this dissertation does not present *general* objections to federal involvement in any aspect of the research process. In fact, it argues that most general objections fail. However, there are *specific* objections which warrant more attention. It is to these that I now turn.

Claim 2: Federal funding should concentrate on causal questions concerning "what works."

The last two decades have seen increased effort by the federal government to get researchers to focus on "what works." As early as 1986, William Bennett conducted a major campaign to collect and disseminate a pamphlet with the same title. As Walters (2009) chronicles, several national organizations later lent support to the what works movement. By 2002, there was a federal What Works Clearinghouse, along with a dramatic shift of funding toward causal questions. What justifies this particular emphasis on causal questions? What is gained and what is lost? What assumptions does it make?

To answer these questions, it may be helpful to return to the two justifications for federal involvement. First, the government has a role to play in the provision of mechanisms for public deliberation about public programs. Second, the government has a role to play in supporting research that addresses matters of general welfare. Given these two missions, it seems clear that the government must sponsor research which can help people deliberate over whether or not a public program is accomplishing the goals for which it was designed. That is, there must be continuous evaluation of whether it "works." Further, even in areas in which the federal government is not the provider of services, such as medicine, agriculture, and education, there is public benefit to be derived from knowing whether or not certain popular approaches actually achieve their stated goals. Thus, it seems that "what works" questions constitute an essential part of a federal research agenda.

However, overemphasizing "what works" questions may lead to negative unintended consequences. Perhaps ironically, excessive focus on "what works" may lead to the neglect of both the "what" and the "works." The "what" ostensibly refers to the intervention; the apparent simplicity of this statement belies the complexity of most interventions. Take for example the Tennessee STAR experiment in class size reduction. What led to increased test scores in some segments of the treatment group? One can imagine that decreased class size might be a condition which enabled one or more of a number of causes for increased test performance. Teachers may have been able to allocate more time to each student, or to adopt more inventive and time-consuming pedagogical strategies, or spend less time on administrative work like grading, or had fewer disciplinary problems. It is nearly impossible to provide useful answers to "what works" unless sufficient attention is given to studying the "what."

It is important to continuously question what it means for a program to "work." It is rare that any action has but one effect. One of the dangers of conceiving the main purpose of research as the search for "what works" is that it suggests a focus on a narrow range of effects. It may ignore questions

over whether the effect is desirable, or ignore the search for a better understanding of peripheral effects. A program may effectively raise student scores on a given test, but only by absorbing massive resources, or sacrificing public involvement, or giving up on the more liberal goals of education. Or, it may bring about a certain effect which not all people desire, such as certain political dispositions or attitudes toward sexual behavior.

One final potential negative consequence is that the push to evaluate whether something works may lead researchers to ignore the important question as to *why* it works. For instance, imagine that researchers randomly assign school districts to two different spending levels. The treatment groups receive $5,000 per pupil in addition to their previous spending levels, while control groups maintain their previous spending levels. Next imagine that student test scores rose in the treatment group by one standard deviation. One conclusion that might be drawn is that increased spending dramatically improves student test scores. But why? Perhaps they invested in lowering student/teacher ratios. But why would lower student/teacher ratios increase test scores? Phillips (2009) gives another compelling example:

> It might be a reliable finding that (on average) students learn more in high schools of medium size than they do in very small or very large schools, but this effect is mere magic if the mechanism producing the impact on learning is not known. It would be sad if the contemporary advocates of the "gold standard" turned out, in essence, to be supporters of the dark arts because of their refusal to countenance the search for underlying causal mechanisms as an important (even an indispensable) part of the scientific enterprise! (p. 183-184)

The further one pushes the question of why, the closer one can get to isolating the part of the intervention that is essential to the outcome of attention. When one isolates the mechanism, one can begin to isolate the desirable parts of the intervention from the more noxious elements.

Claim 3: Given the need to focus on "what works," the government should promote scientific rather than humanities and arts-based research.

The previous section addressed the question as to the place of "what works" questions in both public policy and in scientific inquiry. This section addresses the place of scientific vs. humanities and arts-based research in answering "what works" questions. IES policy holds that scientific research is best suited to answering what works questions. Humanities and arts-based research may be useful for other purposes, but they do not address questions concerning "what works." Thus, federal efforts should turn away from research originating the humanities and arts and toward research in the sciences. In his address to the American Education Research Association, "The Institute of Education Sciences: New Wine, New Bottles" (2003), Russ Whitehurst states:

> People on the front line of education do not want research minutia, or post-modern musings, or philosophy, or theory, or advocacy, or opinions from education researchers….The people on the front lines want to turn to education researchers for a dispassionate reading of methodologically rigorous research that is relevant to the problems they have to solve. They are surrounded by philosophy, and theory, and points of view. They want us, the research community, to provide them a way to cut through the opinion and advocacy with evidence.

Humanities and arts-based research is portrayed not only as unhelpful, but as obfuscatory and partisan. It is the job of education researchers to draw educators out of the confusion beset upon them by philosophers, theoreticians, and, worst-of-all, *post-modernists.* The job of a federal research agency, in Whitehurst's view, is to support research which helps educators and policy-makers cut through the jumble of values to arrive at solid facts.

Whitehurst rightly observes that there is a need for rigorous and relevant research. However, in order to achieve this, the sciences must be in unremitting critical contact with the humanities and arts. The analytic precision and depth of insight typical of philosophical scholarship might help prevent the

all-too-frequent confusion in the social sciences between technicality and rigor. Social scientists manipulate variables constructed in accordance with largely unexamined assumptions about race, gender, and class. Without sustained theoretical scrutiny, their models involve artifactual entities predicting other artifactual entities. The cheerful ahistorism behind so many educational interventions keeps them from constructing a science which learns from its mistakes (Lagemann, 2005). The distance from which social scientists must often work can result in dependent variables which are but pale imitations of objects of significant educational interest. The focus on obtaining answers to what educators have done can leave us unable to provide a compelling answer to questions about the nature and purpose of education. Rather than being obfuscatory, humanities and arts-based research is perhaps the last, best chance to conduct scientific research which is as conceptually and theoretically rigorous as it is technically complex.[6]

The Department of Education cannot expect scientific research to find and solve every educational problem. It may not be within the mission of the Institute of Education Sciences to fund and disseminate research in humanities and the arts. The committee responsible for *SRE* appealed to this limited jurisdiction in their choice to limit their discussion of what works questions to scientific approaches. However, the appropriate response then is to either fold IES into a broader research agency that can support the complete range of research necessary to answer what works questions or to create a separate entity devoted to humanities and arts-based research. If the public interest in supporting inquiry into assumptions of education research is as legitimate as the public interest in supporting inquiry into the empirical conclusions drawn from them, such support may seem justified.

[6] I have worked with the assumption that there is a useful distinction to be made between "scientific education research" and what I termed "humanities and arts-based research," both admittedly catch-all terms. Such a distinction can be useful, so long as one is cautious in the inferences one draws from it. For instance, it does not follow that the two have radically different ontological or epistemological commitments, or that the former consists of objective inquiry into facts and the latter subjective inquiry into values (Anderson, 1998; Lykins, in progress; Putnam, 2002). Nor does it follow that one is inherently better for examining "what works," while the other is better for examining what it should work toward.

DOE could operate the educational equivalents of NSF and NEH. One may speculate—and it is only speculation at this point—that a failure to do so may reflect the unfounded assumption that the humanities and arts simply do not offer much to those concerned with educational problems.

The institutional arrangements proposed above do not fully answer the practical question as to how much the federal government should prioritize support for scientific versus humanities and arts-based research on what works. This decision likely depends on the overall condition of the political economy of education research, or at least the perception of the overall condition of the political economy of education research. If potentially beneficial forms of research are not in sufficient supply, the federal government may concentrate its resources on stimulating production and consumption in those areas. (Of course, the question remains, "Beneficial to and according to whom?" Let us assume that the answer to this question is continuously contested via some sort of democratic processes.) Perhaps, as Whitehurst (2003)suggests, and Phillips (2009)comes close to seconding, there is already a healthy stream of rigorous and relevant humanities and arts-based research that is not at all dependent on federal resources.[7]

In conditions of such super-abundance (evidenced perhaps by recent AERA programs), a compensatory policy may be to devote *one-hundred percent* of federal resources to scientific education research. Rigorous scholarship in the humanities and arts may be less resource-intensive than research in the sciences. It may be possible for independent grant-making agencies, universities, other research organizations to continue to produce such research without federal support. Further, it might be the case that without federal support, many resource-intensive forms of scientific education research would come to a halt. In such a situation, the federal government might be wise to put its dollars where they can have the most impact. Perhaps these dollars could even draw some aspiring researchers away from

[7] Thomas Cook and Monique Payne (2002) make a similar statement about the robustness of every area of education research *other* than research on the efficacy of educational reforms.

the humanities and into the sciences, thus bringing about a better distribution of talent in education research.

This argument requires rather strenuous assumptions. But even if one assents to each assumption, the argument only goes so far. It may justify focusing federal research funds entirely on scientific education research, but it would not in itself justify actively suppressing humanities and arts-based research from the reviews posted in the What Works Clearinghouse. If the preceding analysis of the important of the humanities and arts to the study of education has merit, the federal government has grounds for ensuring that these forms of inquiry are not pushed from core of the political economy of education research. Current policy, if not designed to encourage this marginalization, does little to stop it.

Claim 4: Federal funding of causal questions should be limited to RFTs.

The next issue concerns not so much which traditions (science or the humanities and arts) are best suited for investigating causal connections, but which method and designs *within* the sciences are rigorous enough to yield reliable information concerning causal connections. That is, assuming that the sciences have something to contribute to the investigation of causal questions, which scientific approaches are best.

It may be easier to accept the notion that the federal government has a role to play in placing certain topics and questions at the center of the political economy of education research. It may also have a role to play in seeing that these topics are treated by scientists. However, the debate over whether and how the federal government ought to intervene in the choice of research methods and designs has emerged as perhaps the most contentious point in the debate over the federal role in

education research. Many education researchers argue that this approach is harmful both to the community of researchers and to those who stand to benefit from their work (see for instance Feuer, 2006; Howe, 2003; Howe, 2005; Phillips, 2006, 2009).

Over the last decade, federal policy has clearly indicated that, for causal questions, randomized control field trials (RFTs) are the "gold standard" by which other designs and methods are judged. In 2003, Russ Whitehurst made the point quite emphatically, stating, "Randomized trials are the only sure method for determining the effectiveness of education programs and practices."[8] In more recent years, the official position within IES has begun to soften. For instance, the Education Research Request for Applications for 2010 states, "Studies using random assignment to intervention and comparison conditions have the strongest internal validity for causal conclusions and thus are preferred whenever they are feasible" (p. 68). The only exceptions to this are (1) if "randomization is not possible" and (2) when "external validity of the quasi-experiment provides valuable information that is not obtainable from a randomized counterpart" (p. 68). In such cases, quasi-experiments are permissible, including, "regression-discontinuity designs," " instrumental variables," or "matched comparison groups designs," so long as they "substantially minimize selection bias or allow it to be modeled" (p. 68). Thus, it does not seem that RFTs are still seen as the "only sure method" for investigating causal claims.

There is still disagreement over whether this amounts to a wise evolution in public policy or dangerous backsliding. To investigate this disagreement, it may be useful to separate two claims: first, that RFTS are the "only" method for determining effectiveness; and second, that they are a "sure" method of determining effectiveness.

[8] Not all supporters of RFTs have put their case in such fundamentalist terms. For a more cautious endorsement, see for instance, see important essays by Boruch, de Moya, and Synder (2002) and Cook and Payne (2002).

*RFTs are only Method*

Causal investigation forms an important part of every branch of science. Thus, depending on the objects under investigation, different scientists have developed varying approaches to try to understand causality. The claim that RFTs are the only "sure" method for investigating causal claims can be defeated by any one of a number of simple existence proofs. As Michael Feurer (2006) indicates, "astrophysics research on the movement of the planets and origins of the universe does not require the establishment of control groups to test major, scientifically defensible hypotheses" (p. 62).[9] Astrophysics, geology, evolutionary biology, anatomy and physiology, and epidemiology have all made significant advances through techniques other than RFTs. The results furnished by these techniques seem to be at least as sure as the inferences drawn from the results of RFTs in education research.

The nascent claim behind the current push for RFTs seems to be that there are causal relationships between (1) research designs and methods and the reliability of information and (2) the reliability of information and educational outcomes. Curiously, these causal claims have not been submitted to experimentation. That is, there are no RFTs on the relationship between research designs and methods and the reliability of information or on the reliability of information and educational outcomes. If Whitehurst is correct that RFTs are the "only" way of demonstrating effectiveness, then it seems we have no evidence that RFTs are effective.

Backed into this rather uncomfortable conclusion, there seems to be only two options. First, one can concede that there is no evidence of the effectiveness of RFTs. Second, one can concede that RFTS are not the only way of demonstrating effectiveness. Either option leads to the abandonment of the position that RFTs are the "only" reliable evidence of effectiveness. The latter is preferable to the

[9] It might be worth noting that the physical sciences do not seem to present the same threats to external validity that are present in the human sciences. Experiment artificiality, while always a threat, seems especially pronounced in small experiments of large policies.

former, as it comports with the way causality is demonstrated throughout the sciences, in court cases, and in our everyday experience. In addition, it leaves a space open for a discussion of the role of RFTs in the overall quest to better understand the causes and conditions of educational effects.

*RFTS are a sure method*

To concede that RFTs are not the "only" source of evidence is logically equivalent to saying that conducting a RFT is not a necessary condition for making warranted causal claims. However, Whitehurst's description of RFTs as a "sure" way of establishing such warrants suggests not only necessity, but also sufficiency. That is, a well-designed and executed RFT fully satisfies the conditions for establishing causal connections. Whereas it may require *n* quasi-experiments to demonstrate causal efficacy, the same warrants can be generated by a single RFT.

The truth of this claim substantially depends upon what kinds of causal claims one hopes to make. Researchers commonly distinguish *internal* and *external* validity, with internal validity referring to the quality of inferences to the sample under investigation and external validity referring to the quality of inferences to another population. RFTs are commonly referred to as being highly effective at establishing internal validity and less effective at external validity.

These appraisals of RFTs often conflate "validity" with "truth." Roughly speaking, a conclusion is valid if when the premises are true, the conclusion must also be true. Soundness can be understood as validity plus truth. Take for instance the following invalid argument:

1. If the banks fail, then house prices will plummet.
2. House prices have plummeted.
3. Therefore, the banks failed.

The argument is invalid even though the conclusion is true. It is invalid because the conclusion does not follow from the premises (this fallacy is commonly referred to as affirming the antecedent). Contrast this to the following valid argument.

4. All humans are mortal.
5. Chad is a human.
6. Therefore, Chad is mortal.

Premises (4) and (5), if true, give logical support to the conclusion (6). Thus, (6) is a valid inference. If validity were our only concern, we might stop there. However, we also care about whether the argument is *sound*, that is, whether the premises are true. Consider the next argument:

7. All philosophers are decathletes.
8. Chad is a philosopher.
9. Therefore, Chad is a decathlete.

Just an in the first argument, premises (7) and (8), if true, give logical support to the conclusion (9). Thus, (9) is a valid inference. However, one should not accept premise (7), because it is false. An inference made from a false premise is not *sound*.

In his popular textbook, *Investigating the Social World: The Process and Practice of Research* (2006), Russell Schutt conflates truth with validity, stating, "We have reached the goal of validity when our statements or conclusions about this empirical reality are correct" (p. 19). Scientific inquiry does not aim simply to reach "correct" conclusions. Rather, it aims at reaching *warranted* conclusions. Conclusions are warranted when they follow from premises that are warranted. Premises are warranted when one has good reasons or evidence for believing them. The goal of validity is not to achieve correct conclusions, but to achieve warranted conclusions. In order to reach a "correct" conclusion, one must not only reach the goal of validity, but also the goal of soundness.

To sum up, all of the following combinations are possible. Both the premise and the conclusion can be false while the inference from premise to conclusion is valid. Similarly, both the premise and the

conclusion can be true while the inference from premise to conclusion is invalid. Most of what are described as threats to validity concern the truth of the premises, not whether the premises (if true) warrant the conclusion. For instance, Schutt distinguishes three kinds of validity, each with several subspecies (see Figure 8).

In all three instances, Schutt confuses two important aspects of social scientific research. First, the researcher needs to know the extent to which his or her premises are true. Second, the researcher needs to estimate the extent to which the truth of the premises warrants a given conclusion. The first case deals with soundness, the second with validity. As described, all three variants of validity are actually variants of soundness. Thus, they can be recast as in Figure 9. The revised typology emphasizes that *valid inferences* are in some respect more difficult to achieve than *true conclusions*. The former require evidence and justification. The latter do not. The goal is to reach true conclusions based on valid inferences.

This provides a useful way of evaluating the extent to which experimental designs yield "sure" conclusions concerning causality. A sure conclusion should follow from a valid and sound argument. What do RFTs contribute to the validity and soundness of arguments about causation? In the following section, I evaluate these three issues of internal validity, external validity, and measurement validity. In so doing, I argue that in addition to sample and cross-population generalizability, one other important form of external validity deserves attention in policy research, *similar-intervention validity*, or the extent to which a conclusion drawn about *n* permutations of an intervention warrant claims about other permutations of an intervention. Depending on the level of specificity with which one describes an intervention, similar-intervention validity can be classified as either a kind of content validity or a kind of external validity.

- *Measurement validity*: exists when a measure measures what think it measures (p. 117).
    - *Face validity*: exists when a measure "obviously pertains to the meaning of the concept being measured more than to other concepts" (p. 117).
    - *Content validity*: exists when a measure "covers the full range of the concept's meaning" (p. 118).
    - *Criterion validity*: exists "when the scores on one measure can be accurately compared to those obtained with a more direct or already validated measure of the same phenomenon (the criterion)" (p. 119).
        - *Concurrent validity*: exists "when a measure yields scores that are closely related to scores on a criterion measured at the same time" (p. 119).
        - *Predictive validity*: exists when a measure can be used to predict scores on some other measurement in the future.
    - *Construct validity*: exists when a measure can be shown to be "related to a variety of other measure as specified in a theory" (p. 120).
        - *Convergent validity:* exists when "one measure of a concept is associated with different types of measures of the same concept" (p. 120).
        - *Discriminant validity:* exists "when scores on the measure to be validated are compared with scores on measures of different but related concepts" (p. 120).
- *Generalizability*: exists when a conclusion holds true for the population, group, setting, or event that we say it does, given the conditions that we specify.
    - *Sample generalizability*: exists when a conclusion based on a sample, or subset, of a larger population holds true for that population.
    - *Cross-population generalizability (external validity):* exists when finding about one group, population, or setting hold true for other groups, populations, or settings.
- *Causal validity (internal validity)*: exists when a conclusion that A leads to or results in B is correct .

Figure 8: Schutt (2006) on the kinds of validity

- *Measurement validity*: To what extent do the measures for A, B, and the relationship between A and B warrant inferences about A, B, and the relationship between A and B.
    - *Face validity*: exists when one's claim that a measure measures what one thinks it measures is warranted by the obviousness that a measure pertains to the meaning of the concept being measured more than to other concepts.
    - *Content validity*: exists when one is warranted to believe that a measure covers the full range of the concept's meaning.
    - *Criterion validity*: exists when one's claim that a measure measures what one thinks it measures is warranted by the fact that the scores on one measure can be accurately compared to those obtained with a more direct or already validated measure of the same phenomenon (the criterion).
        - *Concurrent validity*: exists when one's claim that a measure measures what one thinks it measures is warranted by the fact that the measure yields scores that are closely related to scores on a criterion measured at the same time.
        - *Predictive validity*: exists when one's claim that a measure measures what one thinks it measures is warranted by the fact that the measure can be used to predict scores on some other measurement in the future.
    - *Construct validity*: exists when one's claim that a measure measures what one thinks it measures is warranted by the fact a measure can be shown to be related to a variety of other measure as specified in a theory.
        - *Convergent validity:* exists when one's claim that a measure measures what one thinks it measures is warranted by the fact that one measure of a concept is associated with different types of measures of the same concept.
        - *Discriminant validity:* exists when one's claim that a measure measures what one thinks it measures is warranted by the fact scores on the measure to be validated are compared with scores on measures of different but related concepts.
- *Generalizability*: To what extent does information about the relationship between A and B in sampled population warrant claims about A and B in the broader population or in some other population, given the conditions that we specify.
    - *Sample generalizability*: exists when a conclusion based on a sample, or subset, warrants a conclusion about a larger population from which the sample or subset is drawn.
    - *Cross-population generalizability (external validity):* exists when a conclusion based on one group, population, or setting warrants inferences about other groups, populations, or settings.

*Causal validity (internal validity)*: To what extent is one warranted to claim that within the sampled population, A caused B.

Figure 9: Revised conception of the kinds of validity

It is commonly held that there is a distinct trade-off between internal and external validity. In one sense, this is obviously true. Tightly controlled laboratory work helps guarantee that the only difference between treatment and control groups is the intervention. However, the less an experiment's conditions resemble conditions outside of the experiment, the less we are able to infer from its conclusions. To have the sort of "sure" evidence that is of most use to policy makers, one needs all three forms of validity.

*Description of RFT*

Before delving too deeply into claims made about RFTs, it may be helpful to describe an example of one particularly strong form, the Solomon Four-Group Design. In this design, a sample of individuals is drawn (either randomly or nonrandomly) from the general population of interest. Individuals are then randomly assigned to one of four groups. The first group undergoes a pretest, treatment, and posttest. The second group undergoes a pretest and a posttest. The third group undergoes treatment and then a posttest. The fourth group only undergoes a posttest. Allowing *R* to stand for randomization, *O* for observation (pretest or posttest), and *X* for intervention, the design can be symbolized as follows:

| | | | | |
|---|---|---|---|---|
| Experimental group 1: | R | O | X | O |
| Control group 1: | R | O | | O |
| Experimental group 2: | R | X | | O |
| Control group 2: | R | | | O |

This design provides a great deal of scaffolding for causal inferences. Because individuals are randomly assigned, pre-existing differences are less likely to exist across the groups. If such differences still linger (at least between Experimental group 1 and Control group 1), they will appear on the pretest and can

thus be accounted for. Further, the design allows the researcher to separate the effects of the pretest from the effects of the treatment. While the Solomon design is very resource intensive, it does partially mitigate the threats to internal validity that haunt quasi-experiments, thought such threats never completely disappear (Schutt, 2006, pp. 217-221).

RFTs can provide results which warrant claims such as, "within the sample population, A caused B." However, answering questions concerning internal validity is not equivalent to answering questions concerning "what works." In order to know "what works," we need to answer at least three other important questions. *First*, what is the relationship between (1) the measure of A and B and (2) A and B? *Second*, what is the relationship between the sample and the population(s) one hopes to draw inferences about? *Third*, what is the relationship between the intervention and other possible iterations of it? RFTS are not the best way of answering these three questions. Thus, in order to answer "what works" questions, the federal government should support a broader range of research methods and designs.

Regarding the first question, RFTs offer no particular insight into whether one's measures measure what one thinks they measure (Schutt, 2006, p. 217). This requires a number of other methodological approaches. For instance, imagine that a researcher wants to know whether self-reports of voting behavior measure actual voting behavior. There is no intervention to randomly assign, thus there is no RFT. In more complicated cases, such as attempting to find out whether a measure of citizenship measures actual citizenship, the task is even more complex. "Citizenship" is a significantly more difficult concept to operationalize and observe than "voting behavior." In addition to the research methods and designs originating in the sciences, such as task may require techniques from the humanities and arts.

Regarding the second question, Cook and Payne (2002, pp. 164-174) concede that RFTs typically involve a trade-off between external and internal validity, but argue that such a sacrifice is worth it. "Experiments place priority on unbiased answers to descriptive causal questions—internal validity. To draw such causal conclusions in the conditions of likely application—external validity is give lower priority" (p. 164). This might be true if policy-makers and scientists are concerned only about the group being studied, but this is seldom the case. More often, policy-makers want evidence not of what *worked*, but of what *works*. They want to know whether a program should be expanded or contracted and whether its effects will accumulate or wash out. In short, they want the researcher to be able to make a reasonable, evidence-based prediction. RFTs have a role to play in such a discussion, but it is not as provider of a guarantee of goods.

Cook and Payne respond that external validity is parasitic upon internal validity. The plausibility of broader generalizations can only be as valid as the validity of the claims from which they generalize. Thus, internal validity should be cared for first and foremost (Cook & Payne, 2002, p. 165). There is a degree of truth to this. However, logical priority does not necessarily warrant ignoring the latter. If so, then the fact that measurement validity is in some ways logically prior to internal validity would warrant focusing all resources on measurement validity.

Regarding the third question, careful investigation is required to understand how exactly a future intervention can capture the causal force of a past intervention. Educational interventions are seldom alike. Take for example randomized studies of teacher performance incentives. Assume that a random experiment conducted in Madison, Alabama, showed that the offer of a $10,000 performance bonus failed to result in any gains in student performance. What exactly would such a study teach us about the causal effects of performance incentives? It seems that the answer is that it would teach us precious little. All it would tell us is that this *particular* incentive—let us call it $i^1$—had no effect on this

particular measure. Perhaps the failure of $i^1$ is due to it being too small. Perhaps it is due to it being too large. Perhaps the incentive is not a large enough percentage of overall compensation. Perhaps it is due to it being linked to individual teachers, rather than the entire school. Perhaps it is linked to the wrong measure or the wrong outcome. Perhaps the effect was simply measured too early or too late. Perhaps it is due to it being too easy to achieve. Perhaps it is due to it being too hard to achieve. Perhaps it over-incentivized cooperation over individual effort, or individual effort over cooperation.

In itself, the study of $i^1$ does not give us "sure" evidence of the causal efficacy of performance incentives. Many popular education policy reforms, such as school choice, performance pay, and test-based accountability, exist in multiple forms. No amount of methodological sophistication can guarantee reliable inferences from the efficacy of one permutation of the reform to another. Thus, it is important not to overpromise what randomized studies can deliver to policy-makers and the public. Feurer (2006) argues that the notion that randomized trials provide a "gold standard" for research tends to "exaggerate the perceived accuracy of the findings" (pp. 63-64) It misleads policy-makers and the public into thinking that randomized studies offer "definitive answers to complex questions" (Feurer, 2006, pp. 63-64). Valid generalizations depend on reason, logic, and a wider range of empirical data. For instance, a relatively inexpensive teacher survey may reveal more about the likely consequences of incentive pay than any costly experiment.

Take for instance the much-studied Tennessee STAR experiment in class-size reduction. The results of the randomized trial suggested that class-size reduction caused modest gains in the test-scores of children in early grades. Boruch, De Moya, and Synder (2002) cite this study as evidence that "a single RFT can help to clarify the effect of a particular intervention against a backdrop of many nonrandomized trials" (p. 74) In fact, the experiment taught, at most, only that class-size reductions were responsible for increased test-scores *for these particular students.* It did not lend warrant to the

claim that class-size reductions are an effective way for raising achievement as such. This became clear when California implemented a state-wide policy of class-size reduction. The California program not only failed to increase student achievement, but may have been responsible for a substantial increase in the number of poorly qualified teachers in high-poverty schools, thus actually harming student performance.

*The Role of RFTs is finding What Works*

As indicated earlier, advocates of RFTs often point to the medical and biological sciences as proof of their superiority. However, the task of finding what works in education is significantly different than finding what works in medicine. The threats to measurement validity and external validity, as well as the amorphous state of many interventions are more problematic in education research than in medical research. However, this is not, as critics maintain, an argument against RFTs. Rather, it is an argument against assuming that an RFT by itself tells us what works. To understand what works, we need to both continue to develop better ways of measuring the things that are of most concern and work to understand the essential parts of an intervention and how they interact with different contexts.

*Quest for Certainty and Rationality*

What explains the exuberance surrounding RFTs? Why have some policy-makers and practitioners fingered RFTs as the source of evidence for what works? This attempt to locate an infallible source of knowledge is not new. Plato looked to the forms. Abrahamic traditions have looked to deities. Descartes turned to rational introspection. John Dewey (1988) described this as the Western tradition's "quest for certainty," or the attempt to escape the world of change and process and arrive at

a place of unshakable truth. More recently, Michael Feuer (2006) has drawn on developments in behavioral economics and cognitive science to argue that this quest for certainty is, given the constraints and complexity the real world imposes upon is, irrational. As executive director of the Division of Behavioral and Social Sciences and Education at the National Research Council (NRC), Feuer was one of the main architects of *Scientific Research in Education* (2002). Given the recalcitrance of cognitive limitations, it seldom makes sense to seek the *maximally* correct solution. To do so would require an algorithm for testing an indefinite number of possible solutions and comparing the utility of each one. Rather, a *procedurally rational* person instead seeks to *satisfice,* Herbert Simon's term for the seeking of a solution which is good-enough, though perhaps not optimal.

As an additional harm, exuberance around randomized trials often leads to exaggerations about the lack of an empirical base for educational spending. Randomized trials are expensive and difficult to conduct. Many developers of interventions may lack the capital to conduct such studies. Michael Petrelli (2006) argues that this might stifle innovative practices, as educators would be prevented from doing "whatever works" to help their students, instead relying on a prescribed list of what some RFT says about what works.

The excessive faith in RFTs may indicate a temperamental predisposition toward having a certain belief in a fact that is of dubious relevance, rather than to have uncertain belief in a fact that is of high relevance. Feuer does not think that such a temperament can be described as procedurally rational. However, following Kitcher (2001), it should be noted that science does not simply attempt to find truths, but relevant truths. The conclusions of science must eventually meet higher standards than internal validity to be relevant. They must also possess measurement validity. Except in cases in which the relevant question really does not go beyond the sample, they must also possess some degree of external validity. And finally, if a study is to ultimately tell us anything about the policy reforms

currently debated, such as issues in school choice, finance, accountability, and compensation, it must be structured in such a way to warrant inferences about similarly designed interventions. That is, it must postulate and test hypotheses about the causal mechanisms that make one permutation of an intervention work differently than another.

## Conclusion

People "on the front lines" of education do not need to know, "what worked," but "what works." Neither the sciences nor the humanities and arts are prepared to answer such a question by themselves. It is especially true that no one research method within the sciences can furnish such answers. Efforts to disseminate information on what works, such as the What Works Clearinghouse, need to take this into account.

In conclusion, this dissertation has argued that policies which concentrate capital on randomized control field trials of what works undermine the rigor and relevance of the political economy of education research. They isolate RFTs from the crisscrossing criticism furnished both by other scientific investigations and by investigations in the humanities and arts. However, contrary to some critics, this does not mean that the federal government should back away from taking a role in shaping the flow of capital among agents, or that it should cease to support RFTs. RFTs provide a valuable form of criticism of the claims of other research on what works.

There are a number of empirical questions the answers to which may cause one to rethink some of the normative issues discussed in this dissertation. Government's are seldom successful at micromanaging any economy. At best, government interventions can attempt to change the general

incentive structure and through this the way that agents perceive their own interests. Future work should investigate if and how nudges are effective. A partial list of relevant empirical questions would include: How have professional researchers responded to changes in federal policy and priorities concerning education research? Have these changes affected the topics, research questions, research designs, and analytic methods used? Has there been a shift in the kinds of studies submitted to and published in leading research journals and national conferences? Has there been a change in the paradigmatic or disciplinary orientation and methodological preferences of those drawn to education research? Has there been a shift in predoctoral training programs in education, either in the kinds of students admitted or in their research preparation they receive? Has there been a shift in the hiring practices of universities toward certain research orientations? How have other funding agencies (such as foundations) responded? Are doctoral dissertations in education and related social and behavioral science fields now fashioned after IES research criteria? Are there consumer effects? Do leaders of school practitioner organizations, teachers and administrators, perceive a change in research? What change, if any, do policy makers see as having taken place?

While this dissertation has argued that the goal of federal interventions should be to cultivate the sort of crisscrossing, unrelenting criticism that fosters rigorous and relevant research, it largely stops short of offering an appraisal of whether past attempts have done this or a prescription for how such a goal is best achieved in the future. These sorts of conclusions can be reached only by tying the sort of ethical and epistemological investigation undertaken in this dissertation with a sophisticated empirical strategy for evaluating the effectiveness of various federal policies. I hope that those pursuing such questions seek the unrelenting criticism upon which rigor and relevance depend. Moreover, I hope the political economy of education research is such that these sorts of investigations are possible.

REFERENCES

*American Journal of Education.* (1826). *1*(2).

Anderson, E. (1998). Pragmatism, science, and moral inquiry. In R. Fox & R. Westbrook (Eds.), *In face of the facts: moral inquiry in american scholarship* (pp. 10-39). New York: Cambridge University Press.

Atkinson, E. (2000). In defence of ideas, or why 'what works' is not enough. *British Journal of Sociology of Education, 21*(3), 317-330.

Baez, B., & Boyles, D. (2009). *The politics of inquiry: Education research and the "culture of science"*. Albany: SUNY Press.

Bennett, W. (1986). *What works: Research about teaching and learning*. Washington D.C.: U.S. Department of Education.

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

Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis : myths, fraud, and the attack on America's public schools*. Reading, Mass.: Addison-Wesley.

Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory, 57*(1), 1-22.

Bingham, C. (1785). The young lady's accidence; or, A short and easy introduction to English grammar; designed, principally, for the use of young learners, more especially those of the fair sex, though proper for either. Available from http://opac.newsbank.com/select/evans/18934

Boruch, R., de Moya, D., & Snyder, B. (2002). The Importance of randomized field trials in education and related areas. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: Randomized trials in education research* (pp. 50-79). Washington, D.C.: Brookings Institution Press.

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241-258). New York: Greenwood.

Bourdieu, P. (1991). The peculiar history of scientific reason. *Sociological Forum, 6*(1), 3-26.

Bush, V. (1960). Science, the endless frontier. A report to the President on a program for postwar scientific research. Available from http://name.umdl.umich.edu/HEB01131

Clifford, G. J., & Guthrie, J. W. (1988). *Ed school : a brief for professional education*. Chicago: University of Chicago Press.

Coleman, J. S. (1966). *Equality of educational opportunity*. Washington, D.C.: U. S. Department of Health, Education, and Welfare.

Cook, T., & Payne, M. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: randomized trials in education research* (pp. 150-178). Washington D.C.: Brookings Institution Press.

Cremin, L. A. (1980). *American education, the national experience, 1783-1876*. New York: Harper and Row.

Denzin, N., Lincoln, Y., & Giardina, M. (2006). Disciplining qualitative research. *International Journal of Qualitative Studies in Education, 19*(6), 769-782.

Dershimer, R. A. (1976). *The federal government and educational R&D*. Lexington: Lexington Books.

Dewey, J. (1984). The sources of a science of education. In J. A. Boydston (Ed.), *John Dewey: The later works, 1925-1953* (Vol. 5). Carbondale, IL: Southern Illinois University Press.

Dewey, J. (1986). Logic: The theory of inquiry. In J. A. Boydston (Ed.), *John Dewey: The later works, 1925-1953* (Vol. 12). Carbondale: Southern Illinois University Press.

Dewey, J. (1988). The quest for certainty. In J. A. Boydston (Ed.), *John Dewey: The later works, 1925-1953* (Vol. 4). Carbondale, IL: Southern Illinois University Press.

Diamond, A. (1986). What is a citation worth? *Journal of Human Resources, 21*, 200-215.

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

Feuer, M. (2006). *Moderating the debate: Rationality and the promise of American education*. Cambridge: Harvard University Press.

Feuer, M., Towne, L., & Shavelson, R. (2002). Scientific culture and educational research. *Educational Researcher, 31*(8), 4-14.

Goldhaber, D., & Brewer, D. (2008). What gets studied and why: Examining the incentives that drive education research. In F. Hess (Ed.), *When research matters: How scholarship influences education policy* (pp. 197-218). Cambridge: Harvard University Press.

Howe, K. (2003). *Closing methodological divides: Toward democratic educational research*. Boston: Kluwer Academic Publishers.

Howe, K. (2005). A question of education science: Experimetism vs. experimentalism. *Educational Theory, 55*(3), 307-321.

Hyslop-Margison, E., & Naseem, M. A. (2007). *Scientism and education: Empirical research as neo-liberal ideology*: Springer.

Institute for Education Sciences (2009). *Education research request for applications for 2010*. Available from http://ies.ed.gov/funding/pdf/2010_84305A.pdf

James, W. (1978). The moral philosopher and the moral life. In J. J. McDermott (Ed.), *Writings of William James*. Chicago: University of Chicago Press.

Johanningmeier, E. V., & Richardson, T. R. (2008). *Educational research, the national agenda, and educational reform : A history*. Charlotte, N.C.: IAP-Information Age Pub.

Jones, L. (1996). A history of the National Assessment of Educational Progress and some questions about its future. *Educational Researcher, 25*(7), 15-22.

Judd, C. (1939). *Research in the United States Office of Education*. Washington, D.C.: The Advisory Committee on Education.

Kaestle, C. (1992). *Everybody's been to fourth grade: An oral history of federal R&D in education.* A report to the National Research Council Committee on the Federal role in Education Research. Research Report 92-1.

Kaestle, C. (1993). The Awful Reputation of Education Research. *Educational Researcher, 22*(1), 23-31.

Kitcher, P. (2001). *Science, truth, and democracy*. Oxford ; New York: Oxford University Press.

Lagemann, E. C. (2000). *An elusive science: The troubling history of education research*. Chicago: The University of Chicago Press.

Lagemann, E. C. (2005). Does history matter in education research? A brief for the humanities in an age of science. *Harvard Educational Review, 75*(1), 9-24.

Lather, P. (2004). This IS your father's paradigm: Government intrusion and the case of qualitative research in education. *Qualitative Inquiry, 10*(1), 15-34.

Lykins, C. (in progress). *Social science and social progress: Facts, values, and empirical inquiry*. Nashville, TN: Unpublished Dissertation, Vanderbilt University.

MacMullen, E. N. (1991). *In the cause of true education : Henry Barnard & nineteenth-century school reform*. New Haven: Yale University Press.

McGuffey, W. H. (1836). *Eclectic first reader*. Cincinnati: Truman and Smith.

Merton, R. K. (1968). *Social theory and social structure*. New York: Free Press.

Murray, L. (1797). English exercises adapted to the Grammar lately published by L. Murray; ... Designed for the benefit of private learners, as well as for the use of schools. Eighteenth Century collections online. Available from http://galenet.galegroup.com/servlet/ECCO?c=1&stp=Author&ste=11&af=BN&ae=N006310&tiP

G=1&dd=0&dc=flc&docNum=CW110283606&vrsn=1.0&srchtp=a&d4=0.33&n=10&SU=0LRK&locID=nash87800

National Advisory Committee on Education (1931). *Federal relations to education*. Washington D.C.

National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform*. Washington D.C.

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

Petrilli, M. J. (2006). What works vs. whatever works. *Education Week, 25*(43), 36, 44.

Phillips, D. C. (2006). A guide for the perplexed: scientific educational research, methodolatry, and the gold versus platinum standards. *Educational Research Review, 1*(1), 15-26.

Phillips, D. C. (2008). A quixotic quest? Philosophical issues in assessing the quality of education research. In P. Walters, A. Lareau & S. Ranis (Eds.), *Educational research on trial: The search for rigor and the promotion of randomized studies*. New York: Routledge.

Putnam, H. (2002). *The collapse of the fact/value dichotomy*. Cambridge: Harvard University Press.

Ross, D. (1991). *The origins of American social science*. Cambridge: Cambridge University Press.

Schneider, B. (2009). Assessing quality in education journals. In P. Walters, A. Lareau & S. Ranis (Eds.), *Educational research on trial: The search for rigor and the promotion of randomized studies*. New York: Routledge.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. (Eds.). (2007). *Estimating causal effects: using experimental and observational designs : a think tank white paper*. Washington, DC: American Educational Research Association.

Schutt, R. K. (2006). *Investigating the social world: the process and practice of research* (3rd ed.). Thousand Oaks, Calif.: Pine Forge Press.

Smith, D. H. (1923). *The Bureau of education; its history, activities and organization*. Baltimore: The Johns Hopkins Press.

Sproull, L., Weiner, S., & Wolf, D. (1978). *Organizing an anarchy: belief, bureaucracy, and politics in the National Institute of Education*. Chicago: University of Chicago Press.

St. Pierre, E. I. (2002). "Science" rejects postmodernism. *Educational Researcher, 31*(8), 25-27.

Sweet, R. (1998). The Reading Excellence Act: A breakthrough for reading teacher training. *Right to Read Report, December*. Available from http://www.nrrf.org/essay_ReadingExcel.html

Thaler, R., & Sunstein, C. (2009). *Nudge: improving decisions about health, wealth, and happiness*. New York: Penguin Books.

Travers, R. M. W. (1983). *How research has changed American schools: a history from 1840 to the present*. Kalamazoo: Mythos Press.

Vinovskis, M. (1998). *Changing federal strategies for supporting educational research, development, and statistics. Background paper prepared for the National Educational Research Policy and Priorities Board.* Available from http://www.ed.gov/pubs/FedStrat/title.html

Vinovskis, M. (2001). *Revitalizing federal education research and development : improving the R&D centers, regional educational laboratories, and the "new" OERI*. Ann Arbor: University of Michigan Press.

Vinovskis, M. (2009). *From a nation at risk to No Child Left Behind : national education goals and the creation of federal education policy*. New York: Teachers College, Columbia University.

Walters, P. (2009). The politics of science: battles for scientific authority in the field of education research. In P. Walters, A. Lareau, & S. Ranis (Eds.), *Educational Research on Trial: Policy Reform and the Call for Scientific Rigor* (pp. 17-50). New York: Routledge.

Warren, D. R. (1974). *To enforce education: a history of the founding years of the United States Office of Education*. Detroit: Wayne State University Press.

Webster, N. (1785). A grammatical institute of the English language comprising, an easy, concise and systematic method of education; designed for the use of schools in America. Eighteenth Century collections online. Available from http://galenet.galegroup.com/servlet/ECCO?c=1&stp=Author&ste=11&af=BN&ae=W020405&tiPG=1&dd=0&dc=flc&docNum=CW112122794&vrsn=1.0&srchtp=a&d4=0.33&n=10&SU=0LRK&locID=nash87800

Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly, 21*(1), 1-19.

Whitehurst, G. (2003). The Institute of Education Sciences: New wine, new bottles. Available from http://165.224.220.253/rschstat/research/pubs/ies.html

Worcester, S. (1828). *A primer of the English language*. Boston: Hillard, Gray, Little, and Wilkins.