THREE BARRIERS TO MORE USEFUL EDUCATIONAL RESEARCH

A DISSERTATION
SUBMITTED TO THE SCHOOL OF EDUCATION
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jonathan R. Dolle
July 2010
© 2010 by Jonathan Robert Dolle. All Rights Reserved.
Re-distributed by Stanford University under license with the author.

This work is licensed under a Creative Commons Attribution-Noncommercial 3.0 United States License.
http://creativecommons.org/licenses/by-nc/3.0/us/

This dissertation is online at: http://purl.stanford.edu/bd969pg3839
I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Eamonn Callan, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Denis Phillips, Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Helen Longino

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Richard Shavelson

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.
ABSTRACT

Many conceptual barriers stand in the way of more useful educational research. This dissertation examines three such barriers. The first intellectual barrier is an unproductive distinction between “scientifically based” research and other forms of educational inquiry—a distinction that, thanks to recent federal acts, now carries the weight of law. The second barrier is a failure to draw meaningful distinctions between “basic” and “applied” educational knowledge. Neither researchers nor federal funding agencies have outlined a clear definition of educational knowledge in these domains, nor have they offered a compelling account of how research aimed at one domain produces knowledge gains in the other, though a connection is frequently implied. And a third barrier is confusion about the relationship between what might be called the “positive” (or descriptive) and the “normative” (or prescriptive) dimensions of educational inquiry. Researchers often exhibit either too much comfort openly incorporating value judgments into research or too little, and the traditional distinction offers scant guidance to researchers wanting to directly engage practical educational problems.
# TABLE OF CONTENTS

**Abstract**

iv

**Chapter 1**

**Introduction: Three Barriers to the Advancement of Educational Inquiry**

1

1.1 Introduction 1

1.2 What Is Wrong with the U.S. Educational R&D Infrastructure? 3

   1.2.1 The Challenge of Developing Understanding for Use 6

   1.2.2 The Challenge of Producing Practically Useful Research 8

   1.2.3 Four Principles for Redesigning Educational R&D 10

1.3 Three Intellectual Barriers to a More Productive Educational R&D System 12

   1.3.1 Distinguishing “Scientifically Based” Research from Other Forms of Educational Inquiry 14

   1.3.2 Distinguishing “Basic” and “Applied” Educational Knowledge 16

   1.3.3 Distinguishing the Positive and the Normative Dimensions of Educational Inquiry 18

1.4 Chapter Summaries 20

**Chapter 2**

**The Problems of Educational Science and the Promise of Disciplined Inquiry** 24

2.1 Introduction 24

2.2 The Limits of Randomized Experiments as a Guide to Policy and Practice 26

   2.2.1 RFTs Have Limited Explanatory Power 30

   2.2.2 The Knowledge RFTs Provide Does Not (Usually) Match the Knowledge Policymakers Need 37

2.3 The Problems of Defining a Scientific Basis for Educational Research 42

   2.3.1 SRE: Popular Confusions 43

   2.3.2 SRE: Possible Tensions 46

2.4 Educational Research as Disciplined Inquiry 50

   2.4.1 Early Discussions of Education as Disciplined Inquiry 51

   2.4.2 Disciplined Inquiry as Social Knowledge 58

   2.4.3 Three Levels of Criticism 69

   2.4.4 The Potential Merits of Disciplined Inquiry 75
Chapter 3
Does Educational Inquiry Really Work in Pasteur’s Quadrant?
An Argument for Problem-Disciplined Inquiry

3.1 Introduction

3.2 Stokes’s Argument for “Use-Inspired” Basic Research
   3.2.1 Stokes’s Historical Argument
   3.2.2 Stokes’s Policy Argument

3.3 The Misapplication of Stokes’s Ideas in Educational Inquiry
   3.3.1 Misunderstanding “Use-Inspired” Basic Research
   3.3.2 Misunderstanding Stokes’s Policy Argument

3.4 Basic and Applied Knowledge in the Social Sciences and Education
   3.4.1 Historical Background
   3.4.2 Defining Basic Knowledge
   3.4.3 Defining Applied Knowledge

3.5 An Argument for Problem-Disciplined Educational Inquiry
   3.5.1 The Limits of Use-Inspired Basic Research in Education
   3.5.2 Problem-Disciplined Educational Inquiry
   3.5.3 Three Examples of Problem-Disciplined Educational Inquiry

3.6 Conclusion

Chapter 4
Practical Arguments in Educational Inquiry:
A Strategy for Linking Knowledge and Action

4.1 Introduction

4.2 Value Neutrality and Bias in Educational Inquiry
   4.2.1 The Value Neutral Thesis in Educational Inquiry
   4.2.2 How Do Values Bias Inquiry on the Value Neutral Account?

4.3 Linking Knowledge and Action through Practical Argument
   4.3.1 Validity Theory: A Brief Overview
   4.3.2 Validation as Practical Argument

References
You make a great, a very great mistake, if you think that psychology, being the science of the mind’s laws, is something from which you can deduce definite programmes and schemes and methods of instruction for immediate schoolroom use. Psychology is a science, and teaching is an art; and sciences never generate arts directly out of themselves. An intermediary inventive mind must make the application, by using its originality.¹

-William James (1892)

Some federal research agencies, by statute, are primarily about the business of basic research and the search for fundamental knowledge. The NSF, for example, has a mission “to promote the progress of science.” Other agencies, such as the Institute of Education Sciences, are primarily about practical action, solving real-world problems, and providing useful information to the public at large.²

-Russ Whitehurst, former IES Director (2003)

1.1 Introduction

The educational R&D infrastructure is broken, and three barriers stand in the way of fixing it. This dissertation explains how and why. The barriers I point to are manifest in the thinking of academics, policymakers, and educators. They are supported by disjointed systems of education governance and by a private industry that provides textbooks, instructional materials, technology, assessment instruments, and teacher development designed not to optimize student learning, but to maximize marketability. And they are perpetuated in the way educators and educational scholars are prepared, educational knowledge gets shared, and research money gets

¹ William James, Talks to Teachers on Psychology: And to Students on Some of Life’s Ideals (New York: Henry Holt and Company, 1906), 7-8.
distributed. Accordingly, understanding how intellectual barriers perpetuate these problems requires a basic understanding of the current educational R&D system and the reasons for its dysfunction.

Section 1.2 examines the system’s failure to address two fundamental challenges relating scholarly knowledge to the actions and decisions of educators. The first challenge is William James’s concern about the use of academic knowledge (in his case, psychological principles) in educational practice. This is the challenge of developing understanding for use in practice. The second challenge is Russ Whitehurst’s concern about the relevance of educational research to problems of schools and schooling. This is the challenge of producing practically useful research. The first challenge probes the educational limits of what research can do, the second what it means to do research that addresses pressing education problems.

These challenges will not go away. They are inherent to the development of professional educators and the allocation of scarce educational resources—two fundamentally difficult tasks. But what can change is the way the educational community rises to meet them. And it is this ability, I argue, that has been seriously hampered by the three intellectual barriers introduced in Section 1.3. (Each subsequent chapter develops an argument addressing one of these barriers.) The first intellectual barrier, I contend, is an unproductive distinction between “scientifically based” research and other forms of educational inquiry—a distinction that, thanks to recent federal acts, now carries the weight of law. A second barrier is a failure to draw meaningful distinctions between “basic” and “applied” educational knowledge. Neither researchers nor federal funding agencies have outlined a clear definition of educational knowledge in these domains, nor have they offered a compelling account of how research aimed at one domain produces knowledge gains in the other, though a connection is frequently implied. And a third barrier is confusion about the relationship between what might be called the “positive” (or descriptive) and the “normative” (or prescriptive) dimensions of educational inquiry. Researchers often exhibit either too much comfort openly incorporating value
judgments into research or too little, and the traditional distinction offers scant
guidance to researchers wanting to directly engage practical educational problems.
Finally, Section 1.4 summarizes the line of argument I pursue in the remaining three
chapters.

Beyond these practical concerns, this dissertation can also be read as a
modest contribution to social epistemology, a relatively new philosophical subfield
that examines how the social organization of inquiry affects the knowledge
produced. Although each intellectual barrier reflects a problem with how academics,
policymakers, and educators think about the relationship between research, policy,
and practice, the barriers are not strictly mental, since how we think about problems
often becomes manifest in the structure and culture of our institutions—in what we
might call our intellectual systems. These systems have a profound impact on the
organization of educational inquiry, from the problems researchers choose to pursue
and the methods they use, to the communities that scrutinize work and the
audiences that ultimately use research as evidence for decisions and actions.

But the first step towards solving a very hard problem is posing it as best one
can. Accordingly, before delving into the difficulties above, I offer a brief account of
the current state of educational R&D. With this background in hand, the challenges
and motivation for overcoming these three intellectual barriers to improving
educational R&D—and, ultimately, schools and schooling—will be clearer.

1.2 What is wrong with the U.S. educational R&D infrastructure?

On February 17, 2009, President Obama signed the American Recovery and
Reinvestment Act (ARRA), a bill worth approximately $787 billion in federal spending,
including $90.9 billion going to support school districts, Pell Grants, low-income
public schoolchildren, special education, Head Start, and childcare services. The act
marked the single largest federal investment in education ever made, with stimulus
spending equal to one and a half times federal discretionary spending. And yet, notably absent from ARRA’s education funding priorities was significant investment in research and development (R&D). Apart from $250 million for state data collection systems—amounting to just 0.27 percent of all education stimulus spending—virtually none of the funding will go to education R&D. Compare this to the ARRA’s $147.7 billion investment in health-care, where over 20 percent ($30 billion) will be spent on R&D, more than 70 times the comparable amount for education. The stimulus package also reflects a national disparity in R&D investment, with the health-care industry spending 5-15 percent annually on R&D compared to less than 1 percent annually in education.

This striking disparity in R&D investment should be a wake-up call for researchers. It reflects a lack of public confidence in the ability of education research to improve schooling. Almost 15% of all tax dollars spent to support K-12 public schools (second only to military spending), yet the U.S currently ranks 15th of 27 countries in reading literacy, 24th of 29 countries in mathematics, and 21st of 30 countries in science. Domestically, large educational achievement gaps between black and white students persist, despite decades of effort to eliminate them. Federal, state, and local policies for distributing education funds systematically offer

---

more money to higher-income students and wealthier school districts. Teachers with better professional training tend to leave the profession sooner than those with lesser training. The list goes on.

How are researchers addressing these and other educational problems? There is a sense in which educational research today is not much further along than it was over a century ago, at least when it comes to supporting more effective schools and schooling. When William James first delivered his “Talks to Teachers” in 1892, schools of education were in their infancy and focused more on training administrators than teachers. The preparation of teachers is an obvious conduit for bringing research to bear on educational practice, and this certainly continues today. But we still know surprisingly little about effective teacher preparation. Theories of teacher learning are relatively new, meaning “research-based” practices tend to focus on the effects of programmatic elements like strategic mentoring and professional development schools rather than on theories of teacher learning and professional development that might explain these effects and guide program design. And what we do know about effective teacher preparation has not been sufficiently connected to scholarship on educational policy and its effective implementation.

---

11 See for example Kevin Carey and Marguerite Roza, "School Funding’s Tragic Flaw," (Seattle, WA: Education Sector and the Center on Reinventing Public Education, 2008).
14 Certainly a lack of political will is also an important part of this disconnect. Yet even here educational research could be doing much more to make the policy case for educational change. Linda Darling-Hammond, for example, has recently detailed dozens of ways in which the U.S. educational system is failing to prepare teachers in accordance with what we know about effective teaching and effective teacher preparation. Linda Darling-Hammond, The Flat World and Education: How America's Commitment to Equity Will Determine Our Future (New York: Teachers College Press, 2010).
There are two fundamental reasons research is disconnected from practice in teacher preparation, and these same reasons can be extended to similar disconnects within educational policy and administration.

1.2.1 The challenge of developing understanding for use. The first reason is James’s: teaching involves artistry, and simply *knowing certain principles or theories* is very different from *understanding how to use them*. Researchers are skilled at producing rigorous scientific knowledge about what we should believe, but they are far less skilled at producing understanding about how we should act based on that knowledge. The philosopher Gilbert Ryle famously distinguished knowing *how* from knowing *that*.\(^{15}\) The challenge of practical understanding is really the problem of linking knowing that and knowing how. A scientist may grasp of the physics involved in skillful pool playing. An expert player may have the constitution and technique to play skillfully. The two kinds of knowledge are different, but they plausibly can be connected to work in mutual benefit, so knowledge how informs knowledge that and vice versa.

As James noted, part of the problem in education is that instructional methods cannot be derived from psychological principles. At best, he thought, “a science . . . lays down lines within which the rules of the art must fall, laws which the follower of the art must not transgress.”\(^{16}\) More recent research on teaching builds on James’s point about the relationship between psychological principles and instructional method, demonstrating that actual teaching practice also cannot be derived from an instructional method. Magdalene Lampert, for example, explains how that same approach, namely problem-based instructional methods, looks remarkably different in practice in different classrooms, for different students, at

---


\(^{16}\) James, *Talks to Teachers*, 8. I doubt these kinds of “necessary” laws actually exist in the social and behavioral sciences, a point I will consider in Chapter 3. But James’s larger point about the application of scientific knowledge still holds.
different points in time.\textsuperscript{17} And the problem of linking research and practice is further compounded by the fact that most preparation happens before teachers step into a school classroom—much less before they step into the classrooms that will be their first position after certification.

In the gap between educational research and its use, this is the challenge of developing \textit{understanding for use} in practice. As Carl Hempel observed, the concept of understanding is peculiar in that it has a pragmatic dimension. To use the concept “requires reference to the persons involved in the process of explaining” and this, consequently, also makes the concept relative.\textsuperscript{18} Many fifth grade math teachers know that their students have a hard time learning to divide fractions, but a smaller number understand this fact in a way that allows them to do much about it. Because most educational research (like research generally) does not have the goal of producing non-academic understanding, most research only becomes understanding when translated by journalists, consultants, intensive self-study, or the experts themselves. It is costly for the public to access research—in terms of time, energy, resources, etc—so it is usually much more efficient to sidestep understanding by relying on the testimony of experts.\textsuperscript{19} But skillful educational practice is a craft requiring more than knowledge that something is true. It also requires a particular kind of understanding: understanding for use in practice. To bridge the gap between knowledge and its use, strategies are needed for lowering the cost of translating scholarly knowledge into practical understanding. But there is also a need for research that distills practical understanding into more general knowledge, research that captures what Lee Shulman has called “the wisdom of practice.”\textsuperscript{20}

\textsuperscript{17} For an excellent, systematic examination of this problem see Magdalene Lampert, \textit{Teaching Problems and the Problems of Teaching} (New Haven: Yale University Press, 2001).
\textsuperscript{19} Russell Hardin, "If It Rained Knowledge," \textit{Philosophy of the Social Sciences} 33, no. 1 (2003).
1.2.2 The challenge of producing practically useful research. Beyond the artistry of the educator’s craft, there is a second reason research is often disconnected from practice, and this one is Whitehurst’s: *the research best suited to guiding the actions and decisions of educators and policymakers is often not the research educational scholars produce*. It is important to recognize that this is largely a modern challenge, made possible thanks to (i) a federal system for research and educational investment, (ii) an emerging capacity for large scale coordination of educational systems, and (iii) systems for the professional preparation of educators. In James’s day these were all non-existent or in their infancy, and the idea of large-scale coordination of educational R&D was nowhere in view.

This is the challenge of producing relevant research or, more precisely, *useful research*, since relevant research may not be particularly useful. Of course research can be useful in different ways for different purposes, and many factors beyond the control of researchers affect the usefulness of their findings. But researchers do have some control over the usefulness of work. One valuable strategy for producing useful research is stepping into the situation and mindset of those most likely to use it.

Again James’s thinking was prescient. In the preface to *Talks to Teachers* he wrote:

> My main desire has been to make them [teachers] conceive, and, if possible, reproduce sympathetically in their imagination, the mental life of their pupil as the sort of active unity which he himself feels it to be. *He* doesn’t chop himself into distinct processes and compartments; and it would have frustrated this deeper purpose of my book to make it look, when printed, like a Baedeker’s handbook of travel or a text-book of arithmetic.  

Replace “teachers” with “researchers” and “pupil” with “policymaker” or “educator,” and you have a Jamesian gloss on Whitehurst’s concern that educational researchers simply do not understand the needs of educators and policymakers. Early in his tenure as Director of the Institute of Education Sciences (IES)—he served from 2002 to 2008—Whitehurst even commissioned a series of interviews with local, state, and federal education administrators and policymakers to better understand their needs.

---

21 James, *Talks to Teachers*, iv.
needs. Yet throughout Whitehurst’s tenure, a gap between research and practice remained—a problem I discuss at length in Chapter 2.

The larger point is that many researchers have succumbed to the allure of a particular way of doing science that is often disconnected from educational practice. Consequently, they struggle to produce research insights that make their way into “the heart, head and hands of educators,” as Dewey once put it. If the history of educational research teaches us anything, educational historian Ellen Lagemann has argued, it is that applied research in the mode of Dewey’s laboratory school “lost” and psychological science in the mode of Charles Judd and Edward L. Thorndike “won.” By and large the gulf between research and practice became institutionalized in psychological experiments set apart from actual schools, teachers, and children. This gulf is felt today, not just by educators and policymakers, but by the many graduate students who enter schools of education with a strong desire to improve schools and schooling and then find themselves in courses learning research methodologies or psychological theories that seem disconnected from the problems facing educators. For these students as much as anyone, the gap between research and practice is palpable.

In the gap between educational research and its use, this is the challenge of producing practically useful research. But this challenge is bigger than educational research alone, or the organization of school governance in the U.S., or even a lack of resources. The larger problem is a faulty educational R&D infrastructure—that is, the lack of a coherent system connecting academic research; educational practice and administration; local, state, and federal policy; and the production of instructional

---

22 Unfortunately, Whitehurst conceived of the role of evidence and the needs of policymakers and educators more narrowly than research or these interviews warranted—an important problem I will examine at some length in Chapter 2. John Ralph et al., "Institute of Education Sciences Findings from Interviews with Education Policymakers," (Washington, DC: Institute of Education Sciences, 2003).
materials and technologies. The absence of such an infrastructure hinders more useful research from being produced and from entering the hands (and heads and hearts) of educators. The current system rewards researchers not for addressing educational problems, but for garnering research grants and publishing in academic journals—ideally of the more prestigious, disciplinary variety. The current system incentivizes publishers to design textbooks to cover as many state standards as possible, not based on pedagogical concerns or a demonstrated connection to student learning but with the goal of maximizing market potential. The current system works against, rather than in support of, a strong professional culture for educators. And the current system invests only a small fraction in educational R&D compared to other sectors of the economy.

1.2.3 Four principles for redesigning educational R&D. Despite these difficulties, it would be a mistake to abandon more ambitious attempts at reforming the educational R&D system. For one, research makes a valuable social contribution just by identifying new educational problems. The reason we can argue, with a high degree of certainty, that systematic disparities in funding, teacher quality, and student achievement exist is thanks to the systematic collection and analysis of educational data. There have also been some notable successes of research guiding policy and practice. Large-scale longitudinal studies have demonstrated the value of several major educational reforms. Two of the most famous are the Perry Preschool Study on the benefits of high quality preschool programs and the Tennessee study on the benefits of K-3 class size reduction. There have also been productive collaborations between researchers and school districts. One of the most well-known is the Consortium for Chicago School Research (CCSR), a former head of which, John Easton, is now the new director of IES. (Not coincidentally, the Chicago superintendent Easton worked with, Arne Duncan, is now Secretary of Education.)

28 Lagemann, An Elusive Science.
Furthermore, nascent efforts to fundamentally redesign the educational R&D infrastructure are currently underway. A handful of educational scholars have started down this road, including Tony Bryk, Linda Darling-Hammond, Allen Schoenfeld, and Mike Smith. And major reports by the National Academy of Education (NAE) and the National Research Council (NRC) have also recommended ways to restructure federal support for educational R&D. While the visions offered by these individuals and reports differ in their particularities, they share in common the recognition that the current educational R&D system is not meeting the needs of schools and schooling. Understanding the potential for research to help solve educational problems requires researchers to think creatively, on a large scale, and in ways that mark a departure from much of educational research’s past.

Four general principles can be found in proposals for redesigning the educational R&D infrastructure. The principles can be interpreted as ways of accounting for James’s observation about the inherent artistry of teaching and Whitehurst’s concerns about the relevance of research. The four principles are: (1) significant educational improvement will require significant educational investment, (2) most useful educational research takes a problem-centric, action-oriented approach, (3) large-scale, long-term collaboration and partnerships between researchers, educators, and policymakers are essential to ensuring effective implementation and lasting educational change, and (4) achieving educational


change at scale depends upon the production of reliable theories and general explanations. (Subsequent chapters expand on these themes.)

These four interrelated principles might be combined into a single mandate for transforming educational R&D: *significant investment in problem-centered, collaborative research aimed at solving educational problems while producing general knowledge*. Of course, not everyone agrees on what constitutes a significant investment, what problem-centered research and general knowledge means, or how collaborative efforts should be organized. Educational scholars will have to wrestle with these problems in the coming years, learning what they can from the past and from other fields. But it is reasonable to assume that the success of new educational R&D efforts will require successful implementation of these four intertwined ideas in some form. The efficient and effective use of greater educational R&D resources depends on research that works. What works depends on identifying and tackling the right problems. Successfully addressing problems requires the coordination and collaboration of many different educational stakeholders. And scaling this work requires theories and general explanations that facilitate the transferring of lessons learned across educational contexts.

1.3 Three intellectual barriers to a more productive educational R&D system

Sometimes the solutions to problems turn on how we think about them. Thinking differently rarely solves a problem *tout court*, but it can be an important step down the road to an eventual solution. Each of the intellectual barriers outlined in this section reflects a problem with how academics, policymakers, and educators often think about the relationship between research, policy, and practice. The barriers are also organizational, since how we think about problems can become manifest in the intellectual systems that shape our institutions. Consequently, if an intellectual shift is to take root, it must go hand in hand with shifts in the larger system. While I have less to say about systemic transformation than about how to
start thinking differently, I situate the three intellectual barriers within the context of existing educational institutions.

This section introduces three intellectual barriers to a well-functioning educational R&D system, locating each within a larger intellectual system; subsequent chapters will push the discussion further. Each barrier turns on confusions about the meaning and significance of a key conceptual distinction. The three intellectual barriers are: (1) confusion about “scientifically based” research and its relationship to other forms of educational inquiry, (2) confusion about the nature and relationship between basic and applied forms of educational knowledge, and (3) confusion about the acceptability of different value judgments in educational inquiry. Each is a particular expression of the challenges James and Whitehurst point to, and each works against the four principles for reforming educational R&D outlined above. They prevent the efficient and effective use of educational resources, undermine the foundation and incentives for genuine collaboration, direct researchers to the wrong kinds of problems, and detract from some of the most productive kinds of theory development.

One final preliminary about the dangers of turning productive distinctions into unfounded dichotomies—a concern that runs throughout my analysis of all three intellectual barriers. As John Dewey points out, sometimes a well-made distinction advances thinking about a problem, helping to illuminate an otherwise overlooked facet. But in other instances distinctions become reified dichotomies, especially when extended to contexts or used for purposes that have less to do with their original meaning and more to do with the cultures and institutions that have grown up around them. Thus, for the first barrier I suggest that too much is made of the distinction between scientifically based research and other forms of educational inquiry. For the second barrier I suggest the distinction between basic and applied knowledge in education is not properly understood. And for the third barrier I argue that the potentially useful distinction between the positive and normative dimensions of educational inquiry is not being made the right way. In each case my
concern is striking the right balance between the usefulness of a distinction for particular purposes and its unwarranted (and unproductive) application to other domains.

1.3.1 Distinguishing “scientifically based” research from other forms of educational inquiry. For more than a century educational scholars have debated whether or not a science of education is possible. Throughout these debates most (if not all) acknowledge, as James did, that there is a craft to teaching that prevents a completely scientific account. Some humanistic approaches to educational inquiry try and reflect this more artistic, aesthetic, or philosophical dimension, but the relationship of these forms of scholarship to more traditional scientific research in disciplines like psychology, sociology, and economics is often unclear. Over sharp distinctions between disciplines, research traditions, and methodological approaches have inhibited productive collaborations among researchers. Such distinctions have also limited the ability of researchers to see their work as part of a larger collective enterprise aimed at addressing pressing educational questions and problems that a particular method or discipline can only speak to incompletely. At the same time, this strong disciplinary grounding is often seen as essential to rigorous, systematic, high quality research.

Among educational scholars there is widespread disagreement about the merits of distinguishing science from non-science, where and how sharply to draw the distinction, and what role (if any) it should play in decisions to allocate research funding. These issues have taken on particular importance with the reauthorization of the Elementary and Secondary Education Act of 2001, popularly known as No Child Left Behind (NCLB), and with the passage of the Education Sciences Reform Act (ESRA) of 2002.31 NCLB required that all federally funded educational programs be “scientifically based”—a label used 119 times in the public law.32 Both acts offered definitions of what kind(s) of research count as “scientifically based” evidence for a

31 [Ref]
particular educational program, instructional method, or evaluation. ESRA even renamed the research arm of the Department of Education the “Institute of Education Sciences.”

Beyond NCLB and ESRA, there is also a larger trend towards “evidence-based” practice and policy. As in medicine, evidence-based practice in education has sought to align the actions and decisions of practitioners with the best available research. Resistance to evidence-based practice in teaching, like medicine, is often due to the impression that evidence can somehow replace professional judgment. And, in fact, some implementations of the evidence-based policies appear to attempt just that. Yet where evidence-based practice has been successful in both education and medicine, it has been carefully implemented as a complement to professional judgment, rather than as a replacement.\(^{33}\)

In this context, debate over what constitutes “science” and what counts as “evidence” for policy and practice takes on great importance. Some argue that only randomized experiments should carry the weight of scientific evidence in educational debates, while others argue for a broader conception of “scientifically based” research. Still others reject the usefulness of such general distinctions in the first place. This is the state of confusion many researchers, policymakers, and educators are faced with.

An important feature of this debate is the trend towards disciplinary specialization. A byproduct of highly specialized social science is an almost inevitable narrowing of perspective. Writing in the middle of the twentieth century, Edward Shils observed that classic social theorists from Aristotle and Plato to Machiavelli and Hegel “were all involved in the consideration of the fundamental problems of policy from the point of view of the man who had to exercise power and to make practical

decisions.” But increasing specialization of the scientific disciplines in the late nineteenth and early twentieth century led to the subdivision of human studies into many separate spheres. More holistic approaches to social inquiry faded in favor of increasingly specialized academic disciplines. With specialization came increased rigor but also diminished political relevance. Problems were sliced and reframed to facilitate generalizations and causal claims at the price of more holistic assessments that could more clearly guide actions and decisions. To use Nancy Cartwright’s helpful distinction, research that “vouched” for conclusions was often abandoned in favor of research that “clinched” conclusions, but little attention was paid to what was lost in the methodological evolution.

In education, specialization initially took on a distinctly psychological character, despite the efforts of some renowned scholars—most notably John Dewey—to keep education an interdisciplinary field firmly rooted in practice. (More recently, economics has also had its ascendance.) My concern is that focusing on what makes educational inquiry scientific detracts from the more important discussion of what makes research a source of credible and productive evidence. Without a system for encouraging productive collaboration around pressing educational problems, the prospects for an improved educational R&D infrastructure are dim. We must better understand the nature and limits of the intellectual contributions of science and disciplinary research to efficiently and effectively marshal research funding to address educational problems. This is the first intellectual barrier to reorganizing educational R&D, which I examine in detail in Chapter 2.

1.3.2 Distinguishing “basic” and “applied” educational knowledge. A second intellectual barrier to the restructuring of educational R&D follows naturally from the first. What is the relationship between basic educational knowledge (and the aim of advancing science) and applied educational knowledge (and the aim of addressing

---

educational problems)? The distinction is often glossed over in educational circles, but it matters for at least two reasons. First, as Whitehurst’s quote at the beginning of the chapter suggests, the federal system for funding research tends to be organized around these conceptual categories. If the goal of restructuring the educational R&D infrastructure is to better address the problems of schools and schooling, then it is important to investigate whether the distinction between basic and applied research makes sense and, if so, to understand the contributions of each. Second, researchers often speak loosely as if educational inquiry regularly contributes to basic and applied knowledge simultaneously, without probing how this works. If true, it is not clear exactly how this simultaneous pursuit happens or how well the labels fit the wide array of research found in the social sciences and education.

The relationship between basic and applied knowledge can be visualized as a simple Venn diagram consisting of two overlapping circles (see Figure 1-1). In one circle is knowledge researchers would generally agree is basic. In the other circle is knowledge researchers would generally agree is applied, and in the overlap is knowledge considered both basic and applied (that is, basic knowledge that is directly applicable to practical problems). While educational scholars frequently refer to basic and applied research, few examine what qualifies knowledge for each domain, much less how knowledge migrates across domains or manages to sit in both simultaneously. It is also unclear what bearing this distinction should have on the organization and funding of educational R&D. For example, some assume that unfettered scientific inquiry is a prerequisite for the production of basic knowledge, and that such inquiries will, over the long run, pay important, socially valuable returns back to society in the form of knowledge and technology that can be applied to pressing social problems. In terms of Figure 1-1, knowledge in the basic domain will either eventually be put to use directly (moving into the overlap) or will create a foundation on which applied knowledge later develops.
What constitutes basic research or knowledge in education and the social sciences? What constitutes applied research or knowledge? Have investments in basic social science research been translated into long-term benefits to schools and schooling? Is it possible that basic knowledge could come out of applied inquiry? How does knowledge in any form get translated into understanding for use? These questions are persistent sources of confusion in educational inquiry, and scholars have yet to tackle them head-on. This is the second intellectual barrier to reorganizing educational R&D, which I consider at length in Chapter 3.

1.3.3 Distinguishing the positive and the normative dimensions of educational inquiry. A third intellectual barrier to the restructuring of educational R&D is evident in the discomfort many researchers feel when trying to navigate what might be called the “positive” and “normative” dimensions of educational inquiry. The positive dimension reflects the view of the researcher dispassionately describing and explaining educational phenomena. This is the tradition of the researcher qua scientist. In contrast, the normative dimension of educational inquiry explicitly engages in evaluation and prescription, offering guidance about what decisions ought to be and what actions ought to be taken. This is the tradition of the researcher qua social critic, advocate, or engineer. There is an impulse to keep these two legitimate

activities entirely separate, lest the extra-scientific values involved in researchers’ prescriptions corrupt the accuracy of their descriptions, biasing the results. One concern is that the normative activity can and will drive the positive activity to predetermined ends, corrupting the representation of evidence and undermining objectivity.\textsuperscript{37} Another concern is that even if such corruption does not occur, there can be an equally problematic perception of bias that undermines the credibility of even the highest quality research.

These concerns about protecting the objectivity of research, on the one hand, and avoiding the perception of bias, on the other, have had the effect of driving talk of values and value judgments out of some domains of empirical educational inquiry. This is problematic for three reasons. First, researchers and philosophers generally agree that science is not (and cannot be) literally value-free. Science involves various norms or standards, and these represent particular values—variously referred to as scientific values, epistemic values, cognitive values, or constitutive values. When researchers judge the sufficiency of evidence, for example, they make a value judgment—a judgment about how much evidence is necessary to warrant a particular scientific claim.\textsuperscript{38} Despite their differences, scientific values are assumed to have a common aim: guiding researchers toward knowledge (and away from errors) about the world. Second, even non-epistemic (extra-scientific) values can play an instrumentally epistemic role.\textsuperscript{39} Fears about tenure and promotion, pride about one’s research program and legacy, envy of one’s colleagues, greed for greater funding and public recognition, and even less noble desires motivate much research. But nothing about these values qua motivations need undermine the epistemic merits of research, any more than a student’s desire to do well on a test implies cheating.

Taking a course and wanting to do well may encourage a student to cheat, but it may

\textsuperscript{37} There is less concern the other direction; in fact, some believe there are good reasons to wish our value judgments were more informed by scientific evidence.


also encourage the student to study hard and learn the material. The concern is not whether researchers are motivated by non-epistemic values, but how such values might affect the neutrality of reasoning processes and research judgments.

And third, some of the things we want to know about seem to be entangled with or embody human values. Teaching, for example, is a practice that people can be more or less effective at. When researchers study teaching, they must offer some account of what constitutes effective teaching, and this account will necessarily reflect certain human values. The same will be true of research on learning, segregation, development, conflict resolution, etc. In contrast, it is relatively easy to talk about atoms, gravitational waves, or mitochondrial DNA without offering obviously value-laden definitions. Put differently, it is natural to talk in evaluative terms about improved learning, worsening segregation, sustainable development, and better conflict resolution. Generally speaking, people want improved learning, they are concerned about worsening segregation, they promote sustainable development and seek out better approaches to conflict resolution. This is one sense in which these concepts seem tied up with human values. In comparison, atoms, solar radiation, and mitochondrial DNA seem much more neutral. They are not good or bad, better or worse, improved or worsened. Solar radiation might be good or bad for particular purposes: reading, gardening, skin cancer, etc. But it can be studied without having a particular purpose in mind. This is less obviously true of teaching, learning, segregation, etc.

Confusion about how educational researchers should deal with these complications is the third intellectual barrier to reorganizing educational R&D, which I consider at length in Chapter 4.

### 1.4 Chapter summaries

In the remainder of this dissertation I develop a set of arguments for overcoming these three intellectual barriers. The following chapter summaries focus on the broad arc of my argument in response to the intellectual barriers just posed.
Chapter Two – The Problems of Educational Science and the Promise of Disciplined Inquiry. This chapter examines two debates that often get unhelpfully conflated: the debate over what constitutes a scientific basis and the debate over what constitutes quality in educational research. My contention is that the first debate is less important methodologically (if not politically) and gets in the way of the second debate, which has practical, methodological, and political importance. Rather than focusing on what makes education research scientific, I argue attention should be on what makes research credible and productive. To this end, I parse concerns about research quality and usefulness from concerns about what constitutes science in the context of recent legislative debates over educational inquiry.

I recommend the more encompassing concept of disciplined inquiry as better capturing the conditions necessary for quality research. The notion of disciplined inquiry I propose is based on the organization and norms of scholarly communities. It is a more productive conception of educational research, I argue, because it encompasses the key epistemic elements of scientific research without unnecessarily excluding other legitimate forms of scholarship. This facilitates interdisciplinary research that can better align with an array of important educational questions and problems.

Chapter Three – Does Educational Inquiry Really Work in Pasteur’s Quadrant? The Case for Problem-Disciplined Inquiry. This chapter takes up an idea gaining greater traction in educational circles that research can be simultaneously basic and applied. The argument, put forward by Donald Stokes, has two parts: (1) a historical critique of the U.S.’s approach to science policy—a system that, Stokes argues, mistakenly dichotomizes basic and applied research, and (2) a proposal for restructuring federal funding of scientific research in the United States around what he calls “use-inspired” basic research.40 Stokes’s first point is widely cited by

educational researchers for a variety of purposes, while his second—more important—point has been largely ignored. Many educational scholars misinterpret Stokes on three fronts: (i) they misunderstand what Stokes means by “use-inspired” basic research, (ii) they ignore (or fail to appreciate) his larger point about federal research priorities, and (iii) they apply Stokes’s insights to educational inquiry without considering important differences between social inquiry (and its application) and research in the natural and biological sciences (and its application)—the latter being the focus of Stokes’s argument.

After explaining Stokes’s main points and examining how educational researchers often misunderstand or misuse his argument, I reassess the promise of Stokes’s ideas for any redesign the educational R&D infrastructure. I conclude that when properly understood, Stokes’s ideas are a helpful but ultimately inadequate intellectual framework for thinking about the conduct and funding of educational R&D. Rather than focusing on “use-inspired” basic research as the centerpiece of a federal educational R&D strategy, the emphasis should be on problem-disciplined research aimed at general explanation and understanding for use in practice.

Chapter Four – Practical Arguments in Educational Inquiry: A Strategy for Linking Knowledge and Action. This chapter considers a particular objection to problem-disciplined research: by encouraging engagement between scientific and non-scientific forms of inquiry, especially in the context of educational problem-solving, moral and political values are liable to corrupt the objectivity of the research and its ability to arbitrate between competing empirical claims. In short, the only kind of normativity empirical educational inquiry should be concerned with is epistemic normativity—that is, the evidence people should accept as credible. Policymakers and educators may choose to interpret the data as supporting or challenging a particular moral or political view, but such judgments should be left to the research consumers, not the research producers.

I challenge this contention, which I characterize as an argument for value-neutral inquiry. The argument proceeds as follows. Non-epistemic values are widely
accepted as legitimate influences on the research questions scholars pursue. Yet question choice directly influences the available alternatives and existing evidence by concentrating research on particular problems (and particular framings of problems) as opposed to others. But by helping direct the accumulation of research findings, non-epistemic values are also influencing the empirical adequacy of existing theories. As an alternative to the exclusion of extra-scientific values from scholarly discourse, I propose an extension of Michael Kane’s argument-based approach to validation41 to educational problem-solving more generally. Like earlier work by Stephen Toulmin,42 Samuel Messick,43 and Gary Fenstermacher,44 Kane’s approach is developed around the concept of a practical argument that lends itself to explicit incorporation of extra-scientific premises. I argue well-developed practical arguments have four potential benefits: (i) they can submit the logic of practical action to critical public scrutiny, (ii) they can locate a productive role for non-epistemic values without undermining objectivity, (iii) they can direct inquiry to pressing educational problems, and (iv) they can facilitate understanding for use in practice.

CHAPTER 2
The Problems of Educational Science and
The Promise of Disciplined Inquiry

Our first assumption is that governments, private foundations, and others are interested at times in estimating the relative effect of new social and education [programs]. Put another way, we assume that the public is interested in answering the questions, “What works better, for whom, and for how long? And how do we know?” Contemporary governmental emphasis on evidence-based policy illustrates this theme, as does some private foundation efforts.¹


To compress one’s education research into forms that fit any one discipline may corrupt the inquiry by distorting its questions into forms that more readily fit a particular disciplinary template. Education research must be disciplined if it strives to be credible and useful, but what standards or rules should it follow?²

-Lee Shulman (1999)

2.1 Introduction

Early in 2000, during President George W. Bush’s first year in office, a seismic shift began in national education policy, the reverberations of which are still being felt. The drafting and eventual passage of the No Child Left Behind (NCLB) act changed the educational landscape in the United States by making federal funding conditional on standardized tests taken by all students in particular grades. Less obvious to many—even most educational researchers—was that this legislation also signaled a dramatic change in the criteria for determining what educational programs are eligible for federal support. According to NCLB, only those programs and

instructional methods supported by “scientifically based” research would receive federal funding. Subsequent legislation also dictated that funding “scientifically based” educational research would be a top priority for the Department of Education’s research arm, now known as the Institute of Education Sciences (IES).³ With this legislation competing conceptions of science—something typically of most interest to methodologists, philosophers, and sociologists—became a focal point for debate over federal education policy and the distribution of billions of dollars in aid.

This chapter examines what it means to do “scientifically based” educational research and how different conceptions of scholarly inquiry can meet (or fail to meet) the needs of policymakers and educators. I agree with Robert Boruch’s suggestion that educational stakeholders want to know “what works.” To this end, educational inquiry should pursue research that attempts to address the problems of schools and schooling. However, there are good reasons to question the ability of strictly scientific research to rise to this challenge. This is especially true when science is narrowly defined, as it was in NCLB and as implemented by IES. In this regard, I share Lee Shulman’s concern that despite the many virtues of scientific methods and scholarly disciplines, they offer at best only partial answers—and in practice, imply only partial-truths—for the problems of schools and schooling. This can be more dangerous than having no knowledge at all when it blinds policymakers and (somewhat less frequently) educators to the practical complexities hidden in background assumptions, careful controls, and average effects of rigorous research. At the same time, appealing to educational researchers to address “real educational problems” by doing research that speaks more directly to policy and practice is not particularly helpful. In effect, this is what proponents of “what works” research have advocated, and I argue it hasn’t worked.

This chapter is the first part of a three part argument. In what follows I examine two debates that often get unhelpfully conflated: the debate over what constitutes a scientific basis in educational research and the debate over what constitutes quality in educational research. My contention is that the first debate is less important methodologically (if not politically) and gets in the way of the second debate, which has practical, methodological, and political importance. The next two sections (2.2 and 2.3) parse concerns about research quality and usefulness from concerns about what constitutes scientific research in the context of recent legislative debates. I conclude that defining scientifically based research and delineating it from other forms of educational inquiry has contributed little to discussions of quality and usefulness.

Rather than focusing on what makes education research scientific, I argue attention should be on what makes research credible and productive. To this end, I recommend the more encompassing concept of disciplined inquiry as better capturing the conditions necessary for quality educational research. The final two sections (2.4 and 2.5) propose and defend a conception of disciplined inquiry based on the organization and norms of scholarly communities. By encompassing key epistemic elements of scientific research without unnecessarily excluding other legitimate forms of scholarship, disciplined inquiry opens the door to interdisciplinary research without making the epistemic sacrifices that concern many educational advocates of scientifically based research. Looking ahead, Chapter 3 extends the concept of disciplined inquiry to research targeting practically important educational questions and problems. And Chapter 4 considers the normative problems that arise when researchers aim to influence the actions and decisions of educators rather than simply describing and theorizing educational phenomena.

2.2 The limits of randomized experiments as a guide to policy and practice

Recent legislative requirements that educational practice and policy be “scientifically based” were born of two frustrations. The first had to do with
uncertainty about what practices and policies educators should pursue. From local teachers and principals up through the highest levels of state and federal government, uncertainty reigned about which policies, programs, and instructional practices were most likely to improve student outcomes. Education research, it seemed, provided little useful guidance for addressing the problems of educators and policymakers. The right kinds of research simply weren’t available, and what research was available often wasn’t in the right form to guide policy decisions effectively. The second frustration had to do with concern about the general quality of educational research. While thousands of faculty at hundreds schools of education regularly produce scholarship, little appeared to significantly advance the understanding of the educational field or contribute to more basic disciplinary knowledge. In some quarters, the harshest critics charged that shoddy scholarship regularly passed muster, implying that the field itself was in danger of losing its scholarly credibility.⁴

When Russ Whitehurst assumed the directorship of IES in 2002, he saw the promotion of randomized experiments as a solution to both frustrations. For him, as for similarly-minded colleagues, the central educational question researchers should be tackling was Boruch’s: “What works best, for whom, under what circumstances?” and the gold-standard for answering “What works?” was the randomized field trail. Not coincidentally, this fit the narrow conception of “scientifically based” research defined in NCLB. In a 2003 presentation to the American Educational Research Association (AERA), he informed a room full of educational scholars that “the primary focus for the Institute will be on work that has high consideration of use, that is practical, that is applied, that is relevant to practitioners and policy makers.”⁵ He went on to argue that IES’s support for educational research would focus on

rigorously conducted applied research with correspondingly limited theoretical aspirations.

To appreciate the substance of this argument, it is important to understand what constitutes a randomized field trial (RFT). An RFT is an experiment where “subjects” (students, teachers, schools) are randomly assigned to either a “control” group or a “treatment” group. If the samples are large enough and properly maintained over time, the theory goes, any observed differences between the two groups can be attributed to the treatment.\(^6\) Thus, the treatment can be said to have caused the result. Such findings, Whitehurst believed, should weigh heavily (if not definitively) in the decisions of policy elites, though he acknowledged that RFTs are not suitable for all problems—for example, the development of assessment instruments. Other research methods, including “ethnographies, case studies, surveys, and correlational analyses” can play a supporting role when trying to make sense of RFT results, but Whitehurst left no question where he stood on the prerogative of educational research at the end of his 2003 presentation:

The people on the front lines of education do not want research minutia, or post-modern musings, or philosophy, or theory, or advocacy, or opinions from education researchers . . . [They] 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 a way to cut through the opinion and advocacy with evidence. They feel they aren’t getting that.\(^7\)

This vision was subsequently embodied in IES’s What Works Clearinghouse (WWC), a federally established center for evaluating and aggregating the “best” educational research, narrowly defined as RFT evaluations.\(^8\)

---

\(^6\) Of course, the conditional “properly maintained” is a catch-all, much like “all else equal,” that often obscures more than it illuminates. In practice, proper maintenance includes everything from monitoring the attrition and other confounding factors to ensuring fidelity of implementation.

\(^7\) Whitehurst, “The Institute of Education Sciences: New Wine, New Bottles.”

The creation of the WWC in 2002 involved facing up to the reality that relatively few RFTs had been conducted in education. Like Whitehurst, some prominent educational researchers, political scientists, and economists bemoaned the fact that educational programs were not being systematically evaluated with experiments, so it was impossible to say that outcomes were actually being caused by the interventions. In comparison, many other fields regularly evaluated interventions with RFTs, including medicine, public health, agriculture, and criminology. What few RFT evaluations of educational programs did exist were of mixed quality and scattered throughout the literature. The idea of producing a comprehensive and accessible synthesis of this work in even a single content area was daunting, but with significant support from IES, the WWC set out to produce a series of reports across a range of content areas. It was believed that doing so would create a reliable, up-to-date resource for policymakers and administrators wanting to know “what works” in education, as well as guide decisions about where IES’s should invest in future RFT evaluations. Reports have now been produced (or are in production) for eight topic areas including adolescent literacy, beginning reading, dropout prevention, early childhood education, elementary school math, English language learners, middle school math, and character education.

Given the central importance of RFTs in recent federal policy and the contention that they represent the most scientifically rigorous and practically useful method for evaluating educational programs, it is worth carefully considering what RFTs can offer in terms of evidence and how this evidence aligns with the needs of policymakers and educators. Section 2.2.1 takes up the first task, examining the logic of RFTs and their contribution to educational science. I find that although RFTs can be a powerful research method, their explanatory power is limited to a narrow set of

---

9 One of the most outspoken advocates for RFTs in educational research is Tom Cook, whose arguments undoubtedly had an impact on the formation of the WWC. See for example Thomas D. Cook, "Randomized Experiments in Educational Policy Research: A Critical Examination of the Reasons the Educational Evaluation Community Has Offered for Not Doing Them," *Educational Evaluation and Policy Analysis* 24, no. 3 (2002).

cases and even then depends on a groundwork laid by many other forms of inquiry. Therefore they can make, at best, only modest contributions to the aspirations of educational science. Section 2.2.2 takes up the second task, examining the relationship between the knowledge RFTs produce and the needs of policymakers and educators. Here I argue a mismatch exists between the evidence RFTs can provide and the needs and constraints of policy makers. By failing to appreciate these constraints, the WWC has been rendered less useful than it might be.

2.2.1 RFTs have limited explanatory power. Under the right conditions, well-designed RFTs can attribute effect(s) to a single, well-defined cause, but they also have limited explanatory power. This is an important and complicated claim, and several key concepts need unpacking, especially “cause” and “explanation.” It turns out RFTs employ an idiosyncratic understanding of causation, limiting their explanatory usefulness. To understand the idiosyncrasies, a more traditional account of cause and explanation in the social sciences is a helpful place to start.

Begin with a simplified account that has intuitive appeal. A straightforward understanding of “cause” is as a link in a causal chain: A caused D because A caused B which caused C which caused D. Causal chains can be mediated by relatively few links (as when one billiard ball hits another) or by many links (as when a rise in surface pressure over the Indian Ocean produces severe weather in Los Angeles). In this sense, to know that A caused D is to explain D in terms of A.

How do we know A caused D—that is, how can such a claim be warranted? The more parts of the causal chain we can spell out, the more complete the explanation. And the more evidence warranting each link, the more confidence we can have that A really did cause D. Ruling out plausible alternative explanations can also increases our confidence that A was the cause and not some other variable X. If

---

A happens to be a law or general theory (e.g., the ideal gas law, Kepler’s laws), and A explains more observations or events than A’, then A is said to have greater explanatory power than A’.

But this picture hides important simplifications. Three are worthy of special note, since they help account for the explanatory limitations of RFTs. First, decisions about how far along the causal chain one goes when explaining an event in the social sciences are often determined by pragmatic considerations, not by science. Bobby failed his math exam because he did not study. This is a straightforward causal explanation. But Bobby did not study because he fell asleep, and he fell asleep because his parents had a loud fight the night before that kept him awake and left him sleep deprived. So failing the exam can be explained by not studying, by falling asleep, and by his parents having a fight, and the causal chain could easily go back further. When researchers offer an interpretation of their findings, they are making a judgment about what links in the chain are most deserving of attention. In most cases these decisions could have been made quite differently, leading to a different (though, as in Bobby’s case, not necessarily incompatible) interpretation.

Second, events usually have multiple causes, not just one. We can say the player’s steady hand caused the billiard ball to go into the corner pocket, but her keen eyesight, consistent stroke, and supreme confidence were equally important influences. Absent any of these factors she might have missed the shot, so all caused the result in the sense that each contributed to the result, even if none by itself was sufficient. Even environmental factors like the slant of the table or the unevenness of the felt could have played a causal role in the final outcome. Like decisions about how far back to go in a causal chain, decisions about what to treat as causes and what to treat as background conditions are not strictly scientific; they also have a pragmatic dimension. Typically those conditions we can control (like concentration)

---

Francis Schrag has discussed this point and its implications for educational research. I will return to this problem in Chapter 4. Francis Schrag, “Values in Educational Inquiry,” American Journal of Education 97, no. 2 (1989).
are treated as causes, while those we have no control over (like the unevenness of the felt) are treated as background conditions.

Third, it is somewhat strange to talk about nonevents as causes, since nonevents are, by definition, things that didn’t happen, and there are an infinite number of things that do not happen (and, therefore, an infinite number of nonevent causes). In Bobby’s case, to say that not studying caused him to fail the math test is really to assert what is sometimes called a counterfactual conditional (or counterfactual for short): if Bobby had studied, he would have passed the math test. But this claim requires an assumption that all else would have been equal (or more weakly put, assumes that no mitigating circumstances occur). Perhaps if Bobby had studied instead of sleeping, he would have fallen asleep in class and still failed the test. Similarly, it is difficult to explain outcomes that did not happen. If the billiard player misses the shot, what explains the miss? There might be an obvious explanation (a mischievous friend’s taunting broke her concentration), but this kind of counterfactual is hard to assess when many other plausible mitigating circumstances also exist (the shot was very difficult, the lighting was bad, the table was poorly constructed). Many more plausible causal explanations exist for a missed shot than for a made shot. So while nonevents can still be useful in explanations, they also present special difficulties.

All three complexities help us understand the explanatory limits of RFTs, which have different explanatory goals than many other social science methodologies. Rather than probing the cause(s) of a particular phenomenon (e.g., altruistic behavior, political conflict), RFTs look for what Paul Holland calls the effects of a cause. In this research design “causes are only those things that could, in principle, be treatments in experiments.” RFTs are not designed to probe backwards, investigating the complex causal chains producing a particular outcome. Instead, they probe forwards, using random assignment to create two (or more)

---

equivalent groups tracking the effects of a single difference between them (the treatment). This allows RFTs to attribute any between-group outcome differences to the treatment, which may or may not be well understood. But probing forward is often just connecting $A$ (the treatment) to $D$ (the measured outcome(s)) without examining the mechanism in between. Thus, they are usually ill equipped to explain why a particular intervention did or did not cause the desired results, what changes to the intervention might improve outcomes, or how the intervention would fare in contexts different from those in which it was conducted.

These features of RFTs limit their ability to contribute to our theoretical understanding of social phenomena. RFTs need not be conducted without any guidance from causal theories, as advocates have sometimes acknowledged. Tom Cook, for example, writes that those designing experiments should exhibit “greater sensitivity to identifying moderating and mediating processes, and thus building into experiments the sampling and measurement particulars that such sensitivity requires. No more black box experiments.”\textsuperscript{14} Yet in a substantially similar essay published the same year he writes that “random assignment does not require well-specified program theories, or good management, or standard implementation, or treatments that are totally faithful to program theory, even though these features definitely make evaluation much easier. Experiments primarily protect against bias in causal estimates.”\textsuperscript{15} Regardless of how these two statements are reconciled, the contribution of RFTs to the causal explanation of educational phenomena is clearly limited. This would be less concerning if educational RFTs regularly uncovered significant results with medium to large effect sizes that could be reliably replicated in a meaningful array of contexts. Yet, for the most part, RFT evaluations of educational programs have produced limited demonstrated success identifying high impact educational programs.

\textsuperscript{14}Cook, “Randomized Experiments in Educational Policy Research,” 181.
Non-results are particularly hard to interpret for the reason discussed above: it is often harder to know why something doesn’t work than why something is working. If your car is running normally, lots of parts must be working (at least approximately) as they were intended to work. Of course, things can go wrong and problems can go undetected, but most of the time cars run reliably well. In contrast, if your car won’t start, myriad things might be wrong with it. Suddenly a puzzle exists—as Dewey might put it, a “problematic situation”—that needs an explanation and possibly a solution.

These explanatory and predictive limitations of RFTs have led some to conclude that experiments are useful only under a very limited set of circumstances. Murnane and Nelson, for example, argue that RFT evaluations only make sense if five conditions are met:

1. The treatment is well-defined. If this is not the case, it is difficult to make inferences from evaluation results.
2. The treatment is relatively easy to implement. Poor implementation is a common explanation for finding of ‘no effects’.
3. The effects of the treatment are evident in a relatively brief period of time. Selective attrition that undermines the validity of the [RFT] design becomes more severe over time.
4. The effects of the treatment do not vary among a great many subgroups of the intended population. The greater the number of interaction effects, the larger must be the experiment to identify these effects.
5. Feedback effects are modest. The presence of significant feedback effects means that the consequences of ‘going to scale’ with the intervention might be very different from the consequences of the [RFT] evaluation.\(^{16}\)

Relatively few educational interventions meet all five conditions, which may help explain why most of the educational RFTs examined by the WWC have small effect sizes. If, for example, an effective intervention is hard to implement, only a subset of

those in the treatment group may experience the program as it was intended. Others may receive what Ed Haertel has described as a “lethal mutation” of a program.\textsuperscript{17}

Interestingly, in the case of beginning reading programs, the WWC’s survey of research on beginning reading interventions showed much more promising results than middle school mathematics, with several reading programs demonstrating effect sizes of more than ten percentile points across each of four reading domains. One program in particular, Reading Recovery, had strong evidence of effect sizes a full thirty-two percentile points above the control for general reading achievement.\textsuperscript{18} Thus, Reading Recovery (RR) appears to be at least a partial vindication of the WWC approach. What made RR different from other programs? According to one of the authors of a WWC-reviewed study, RR fits Murnane and Nelson’s criteria for randomized experiments quite well.\textsuperscript{19} Furthermore, the evaluation of RR by Pinnell et al actually compared the program to several different conditions in an attempt to discern what features of the program made the biggest difference, including: “(a) a treatment modeled on RR provided by teachers trained in a shortened program, (b) a one-on-one skills practice model, and (c) a group treatment taught by trained RR teachers.”\textsuperscript{20} As the authors explained:

Criticisms of research on remedial programs indicate that traditional program evaluations are flawed because they have not attempted to discover why some programs seem to achieve positive outcomes. In this research, our intent was to go beyond a rigorous test of program effectiveness in order to examine program characteristics that contributed to positive results.\textsuperscript{21} Thus, the successful evaluation of RR went beyond the application of an RFT. The intervention itself was well-constructed, had previously been tested in a wide range of contexts, the Murnane and Nelson criteria were met, and the evaluation itself was

\textsuperscript{18} \url{http://ies.ed.gov/ncee/wwc/reports/beginning_reading/topic/tabfig.asp}
\textsuperscript{19} Personal communication from Tony Bryk, February 23, 2010.
\textsuperscript{21} Ibid.: 32.
designed to probe more deeply into the program theory—specifically, into what features of the program were critical to producing the desired long-term outcomes.

This discussion suggests an *epistemic* limitation of the RFT research design. Under the right circumstances and assuming certain background assumptions hold, they can offer strong evidence for a narrow knowledge claim. Nancy Cartwright captures the tradeoff between RCTs and other methods that offer less definitive evidence for a broader range of conclusions (and applications) in her distinction between methods that *clinch* a conclusion and methods that *vouch* for a conclusion. Clinching methods require many assumptions, but if the assumptions hold true, the conclusions follow deductively from the premises. Yet such methods are concomitantly narrow in scope. As Cartwright notes:

> The assumptions necessary for their successful application will have to be extremely restrictive, and they can take only a very specialized type of evidence as input and special forms of conclusion as output. That is because it takes strong premises to deduce interesting conclusions and strong premises tend not to be widely true.\(^{22}\)

In contrast, vouching methods cannot definitively prove a conclusion because “there are no general good practical ‘logics’ of non-deductive confirmation, especially ones that make sense for the great variety of methods we use to provide warrant.”\(^{23}\) In short, the logic of induction is necessarily less certain than the logic of deduction. We might firmly and rationally believe that all ravens are black, but no matter how many black ravens one observes there is always the chance a white one will pop up.

Notably, when educational RFTs are used as evidence for a program theory, the evidence is vouching rather than clinching. RFTs can only clinch inferences that a specific cause (the treatment) produced specific effect(s) (the outcome). This accounts for the limited explanatory power of RFTs: because they are ill-suited to probing mechanisms or confirming theories, they usually cannot explain why particular results obtained. This epistemic limitation entails corresponding limits for

\(^{22}\) Cartwright, "Are Rcts the Gold Standard?," 12.
\(^{23}\) Ibid.
RFT’s practical use as a guide to educational policy and practice. It also raises a second serious concern about using RFTs as a core strategy for linking research and policy.

2.2.2 The knowledge RFTs provide does not (usually) match the knowledge policymakers need. A second reason RFTs can be a particularly poor guide for policy and practice is that the knowledge a particular study provides does not usually match the needs of policymakers and educators. Of course, this is true of much educational research, not just RFTs—a concern I take up in Section 2.3. But for now the focus remains on RFTs which, despite the epistemic limitations discussed above, have been widely billed as the most useful form of applied educational research. Recall that Whitehurst’s vision for IES, and his motivation for creating the WWC, was to bring research to bear on the problems of policymakers and educators. But when policymakers ask “What works?” what they often mean is “What will work for me, in my context and given my constraints? And, by the way, what is the most cost-effective way to go about implementing the policy without sacrificing too much political capital?” Of course, in point of fact, policymakers do not always mean this. Policy elites often ignore the myriad problems associated with policy context and implementation, and sometimes they use research for political ends to justify a policy they have already decided to take. But wise and responsible decision makers should be asking these questions if they care whether or not their policies will actually produce the desired results.\(^{24}\) And responsible, policy-oriented researchers should help those they seek to influence better understand their situation, ask the right kinds of questions, and make use of the best available research.

One reason for the gap between research answers and policy questions was just discussed. Because RFTs often have limited explanatory power, it is difficult to know why a particular intervention did or did not work. And without such knowledge, it can be nearly impossible to know under what conditions similar results are likely to be achieved. Social scientists commonly call this a problem of external validity. In

\(^{24}\) Smith and Smith make this point forcefully in Smith and Smith, "Research in the Policy Process."

37
simplest terms, external validity is the warrant for thinking that what happened in
the study is applicable or “generalizable” to other contexts.

Another, related reason for the gap is the limited ability of decision makers to
interpret research findings. Just as William James observed that instructional
methods cannot be directly derived from psychological principles, policy decisions
cannot be derived from the findings of RFTs. A good case in point is class size
reduction in California. The California legislature passed class size reduction
legislation in July 1996. Although the measure was voluntary, there were strong
financial incentives for districts to reduce the number of students in K-3 classes to a
maximum of twenty. By the end of the 1998-99 school year 98.5% of eligible school
districts and 92% of eligible students were participating in the program. As several
evaluators of the California policy noted:

The strong political support for [class size reduction] in California was based
on the belief that reducing class size would produce significant improvement
in student achievement. This belief, in turn, was based on positive results of a
class-size reduction experiment in Tennessee. . . .Students who participated in
reduced-size classes in the [Tennessee] program during primary grades made
statistically significant achievement gains in all subject areas tested. The
achievement gains were equal for boys and girls. Also important from the
perspective of some California legislators, the achievement gains were largest
for minority students and for students attending inner-city schools.25

The Tennessee study the evaluators mention is considered one of the most
successful RFT evaluations because of its scale, the quality of its design, and the
policy implications of its findings. In addition, the study prompted further
investigation into why class size reduction in Tennessee had the impact it did on
student achievement.26 Nevertheless, while Tennessee’s class size reduction policy is
widely considered a success, the California experience has been quite different.
Among the key differences, California neglected to make sure a sufficient number of

25 Brian Stecher et al., "Class-Size Reduction in California: A Story of Hope, Promise, and Unintended
26 Jeremy D. Finn, Gina M. Pannozzo, and Charles M. Achilles, "The "Why's" Of Class Size: Student
classrooms were available to house the increased number of (smaller) classes, that an adequate number of well-qualified teachers were available, and that these teachers had the necessary professional development. As with the lesson of “evidence-based” practice discussed in the introduction, no matter how rigorously and well-tested a reform might be, no amount of evidence is a substitute for human judgment.

Marshall Smith and Matthew Smith argue that policymakers often need three different kinds of theory to get a policy right: (i) a theory of action to explain how a policy is supposed to work, (ii) a theory of context to explain what factors are likely to affect the theory of action in a particular setting, supporting or impeding the desired outcomes, and (iii) a theory of implementation to explain how to implement a policy in such a way that it is likely to work as planned. Generally speaking, RFTs contribute little evidence to these theories. (Recall Whitehurst’s claim that policymakers and educators have too much theory when what they really want is rigorous, objective evidence.) The WWC was designed around the operating principle that theories are not nearly as important as “facts” about effectiveness. This constraint significantly limits the usefulness of the WWC’s work.

To see how the avoidance of theory can undermine the ability of researchers to interpret the results of RFTs, consider the first topic area the WWC reviewed, middle school mathematics. The following results were reported:

We looked at 361 studies. Of these, 203 appeared to be studies of practices or other interventions that did not qualify for our review. Of the 158 remaining studies, 21 studies of 7 curricula met our evidence standards, 4 without reservations and 17 with reservations.

---

27 Stecher et al., "Class-Size Reduction in California."
28 There are, however, well documented cases where human judgment falters on even the simplest, most routine tasks. Under certain, well-specified conditions, checklists can be a valuable cognitive tool. For an excellent discussion of checklists and their uses, see Atul Gawande, The Checklist Manifesto: How to Get Things Right, 1st ed. (New York, N.Y.: Metropolitan Books, 2009).
29 Smith and Smith, "Research in the Policy Process."
30 These numbers actually include a subsequent update. http://ies.ed.gov/ncee/wwc/reports/middle_math/topic/tabfig.asp
In total, seven different mathematics interventions were studied, two of which had “strong evidence of a positive effect with no overriding contrary evidence”—one with an average of six percentile points of improvement, the other with eight. While this effect size is not what those who conceived of the WWC hoped for, an even more important question is what was learned from these studies. Is the correct inference that virtually nothing works? Or is something else going on that explains these poor results?

To understand this interpretive problem and its implications for the work of the WWC, consider several concerns raised by Alan Schoenfeld, professor of mathematics education at the University of California, Berkeley and a senior content advisor for the first review of studies of middle school mathematics curricula. Early on Schoenfeld wrote a briefing document for the WWC explaining some of the difficulties its staff was likely to face in their review. The most important problems, he argued, were long-standing disagreements about the nature of mathematical competency and their assessment. Traditional curricula focus on core concepts and procedures, while newer “reform” or “standards-based” curricula also included:

- Being able to employ a range of problem-solving strategies; being able to reason effectively using mathematical ideas and to communicate one’s reasoning effectively, orally and in writing; being able to make effective use of various resources, including the knowledge and time at one’s disposal; and having a productive set of beliefs and dispositions about the nature of mathematical enterprise.³¹

Unfortunately, most assessments of middle school mathematics curricula still focus solely on core concepts and procedures. This puts reform curricula at a distinct disadvantage in such comparative assessments:

[F]ocusing on issues of mathematical representation and problem solving takes time. The cost of that focus is that the newer curricula allow less time for mastery of basic skills. The argument made by advocates of the traditional

---

curriculum is that basic knowledge is a prerequisite for applications, and that
students will be seriously hampered by their lack of foundational skills. The
argument made by advocates of reform is that skills will develop in more
robust fashion if they are developed in meaningful problem solving
contexts. Consequently, he concluded that “it is not at all clear how much of the extant
literature can provide the information necessary to make meaningful comparisons of
mathematics curricula; nor is it clear what kinds of information can result from
syntheses of those studies.” At a minimum, he argued, it is essential that
mathematics curriculum assessments are themselves analyzed for content, so that
when the WWC reports results and aggregates findings we will know what kind(s) of
knowledge are actually being captured.

Schoenfeld’s briefing document was submitted to the WWC as the
mathematics background and content section for the literature review protocol. But
when the WWC submitted the draft protocol to IES, it was returned with instructions
to remove the background essay. The WWC promised to publish versions of the essay
on two subsequent occasions, but in each case ended up breaking its commitment to
make the ideas public. Ultimately Schoenfeld choose to resign from the committee
rather than be complicit in what he felt was, at best, academic censorship. This is
important background, but for the present discussion I bracket these concerns about
ideological bias, serious as they are. My interest is the importance of good theory in
the interpretation of results. Ultimately an RFT and its interpretation are only as good
as the theories and background assumptions upon which it is built. How the results of
an RFT should be interpreted depends on whether or not key assumptions held true.
In the case of middle school mathematics, the decision to lump many different
mathematics achievement outcomes into a single domain, rather than disaggregate
them, means studies will not be evaluated by the WWC based on whether tests cover

32 Ibid.: 16.
33 Ibid.: 17.
the spectrum of mathematical content that contemporary research suggests is essential.

As Schoenfeld notes, lumping different measures together presents a serious danger of reporting false negatives. Imagine a study comparing two curricula: a traditional curriculum emphasizing content knowledge and procedures (the control group), and a reform curriculum that also emphasizes problem-solving (the treatment group). If the study only uses a test that assesses content knowledge and procedure, it cannot tell us which curriculum does a better job teaching problem-solving. Inevitably, the quality of the WWC’s review depends on some theory about the key elements of mathematical proficiency. While aspects of the WWC are laudable—particularly if carried out with greater transparency and integrity in the future than in the past—the organization is unlikely to be able to provide all the knowledge policymakers need to make informed decisions.

The preceding discussion highlighted two limitations of RFTs, one epistemic and one practical. The epistemic limitation points to the logic of RFTs and the correspondingly limited evidence they could produce. The practical limitation pointed to the misalignment between the evidence RFTs can produce and the questions of policy elites. The next section examines a much broader proposal for what counts as credible, productive research.

2.3 The problems of defining a scientific basis for educational research

The narrow definition of what counts as “scientifically based” research in NCLB (essentially RFTs) and the concern that pending legislation would further constrain the ability of IES to fund a more balanced research portfolio prompted action on behalf of the educational research community. Most influentially, a report was produced by the National Research Council taking up the question of what it

---

34 Ibid.: 18-19.
means to do “scientifically based” educational research.\textsuperscript{35} The final report, \textit{Scientific Research in Education (SRE)}, concluded that many different research designs (not just RFTs) have a scientific basis. The authors went on to argue that any study can be considered “scientifically based” if it meets six principles:

- Principle 1: Pose Significant Questions That Can Be Investigated Empirically
- Principle 2: Link Research to Relevant Theory
- Principle 3: Use Methods That Permit Direct Investigation of the Question
- Principle 4: Provide a Coherent and Explicit Chain of Reasoning
- Principle 5: Replicate and Generalize Across Studies
- Principle 6: Disclose Research to Encourage Professional Scrutiny and Critique\textsuperscript{36}

Despite the authors’ good intentions, the publication of \textit{SRE} caused quite a stir among the community of educational researchers. This section considers several common confusions surrounding \textit{SRE} (2.3.1), and then argues that parallel practical and epistemic limitations that beset RFTs also apply to \textit{SRE}’s principles (2.3.2).

\textbf{2.3.1 SRE: Popular confusions.} Unfortunately, concerns about the report’s principles led to some confusion and fear among educational scholars. First, many of those reading the report did not understand the political context in which it was written. They interpreted the report’s authors as attempting to impose scientific principles on educational inquiry, rather than as an emergency effort to reinterpret a federal policy that had already written the requirement of “scientifically based” research into law.\textsuperscript{37}

A second confusion, closely related to the first, was that many scholars interpreted the report as promoting the idea that all educational inquiry should be assessed based on degrees of scientific merit. Specifically, it was unclear whether or not the scientific principles outlined in the report were supposed to be principles for

\textsuperscript{35} One goal of the report was to shape the final wording of the Education Sciences Reform Act (ESRA), which it ultimately succeeded in doing. Though in practice, the administration chose to fund randomized experiments to the exclusion of other forms of educational inquiry.

\textsuperscript{36} National Research Council, "Scientific Research in Education."

\textsuperscript{37} For a good discussion of the larger political context in which \textit{SRE} was written, see Eisenhart and Towne, "Contestation and Change in National Policy On "Scientifically Based" Education Research."
assessing the quality of educational inquiry generally. For example, the president of the National Academy of Sciences wrote in the preface that “the report shows that, within the diverse field of education, researchers who often disagree along philosophical and methodological lines nonetheless share much common ground about the definition and pursuit of quality.”38 This implied an understanding of the principles as general quality standards for educational research. But near the end of the report’s introduction, the committee writes that:

Although science is often perceived as embodying a concise, unified view of research, the history of scientific inquiry attests to the fact that there is no one method or process that unambiguously defines science. The committee has therefore taken an inclusive view of “the science of education” or “the educational sciences” in its work. This broad view, however, should not be misinterpreted to suggest “anything goes.” Indeed, the primary purpose of this report is to provide guidance for what constitutes rigorous scientific research in education. Thus, we identify a set of principles that apply to physical and social science research and to science-based education research. In conjunction with a set of features that characterize education, these principles help define the domain of scientific research in education, roughly delineating what is in the domain and what is not.39

In this passage the principles are invoked for two different purposes: (1) as quality standards for rigorous inquiry (the first emphasized section) and (2) as a definition of scientifically based educational research (the second emphasized section).

For some educational scholars, the conclusion that SRE only intended to address scientific inquiry led to a third confusion: what is the relationship of scientifically based research to other forms of educational inquiry? The report says very little about educational inquiry in general, within which scientific approaches (thus defined) make a distinct contribution. This makes the report’s target harder to decipher. How, for example, can we know whether or not the principles are an appropriate standard by which to assess a particular study, answer a particular

39 Emphasis added. Ibid., 24.
research question, or address a particular problem? There are two points where the report briefly discusses this larger picture of educational inquiry and the distinct contributions of scientific research. The first appeared shortly after the excerpt above:

Finally, and critically, the committee believes that scientific research in education is a form of scholarship that can uniquely contribute to understanding and improving education, especially when integrated with other approaches to studying human endeavors. For example, historical, philosophical, and literary scholarship can and should inform important questions of purpose and direction in education. Education is influenced by human ideals, ideologies, and judgments of value, and these things need to be subjected to rigorous—scientific and otherwise—examination.40

In this passage the focus is delineating science from not-science for the purposes of distinguishing a particularly valuable form of knowledge, not as a standard for research quality. The authors point to other worthwhile forms of inquiry and suggest that other value dimensions of education “need to be subjected to rigorous—scientific and otherwise—examination.” A second passage (relegated to a footnote on page 131) offers further explication:

Although our focus is on scientific research, we believe that the federal government should also fund related activities, such as development and demonstration work, a statistics function, a national library, other forms of educational scholarship (e.g., history and philosophy), and a research dissemination and implementation structure. Scientific research is related to, and often depends on, these functions. Indeed, we believe that it is the integration of scientific knowledge with insights from the humanities and other scholarly pursuits that will ultimately yield the most powerful understanding of education. However, we do not address them explicitly in this report because they are outside the scope of the committee’s charge to focus on scientific research in education.41

This passage suggests yet another, more general, standard for judging research: its ability to produce new insights and a powerful understanding of education. This

40 Ibid., 26.
41 Ibid., 131n.
standard has to do with the fruitfulness or relevance of educational research, a
criterion that might be related to the empirical adequacy of research but also goes
beyond it. This larger conception of educational inquiry emphasizes integration and
the potential continuity (and interdependence) of different scholarly endeavors. Of
course, no quality standards are discussed for these other forms of inquiry, nor is it
clear who will do the integrating or how integration might work.

2.3.2 SRE: Possible tensions. Based on this brief sketch, we can now consider
whether the distinction between scientifically based research and other forms of
educational inquiry implied by SRE helps or hinders the pursuit of credible,
productive research. From the political standpoint of influencing legislation, SRE’s
argument was successful.\(^{42}\) But if the goal is for funds to be efficiently and effectively
invested, the distinction fares less well. Deciding whether to invest more in
scientifically based research depends on some account of the aims of science and the
value scientific knowledge (thus defined) to the solving of educational problems. The
report offers only a vague and general account of the goals of science. The principles
are also a likely hurdle to interdisciplinary collaboration, since many educational
researchers do not see their work as fitting within the principles. Equally important,
those who do see their work as scientifically based may not want to risk that status
by incorporating forms of inquiry considered by some to be less rigorous. Thus, for
example, we find AERA setting up separate standards for reporting empirical and
humanities-based research in the association’s journals.\(^{43}\) Similarly, because the
report defines science as limited to the direct investigation (Principle 3) of empirical
questions (Principle 1) that can be replicated and generalized across studies (Principle
5), this mode of educational inquiry may not be up to the task of addressing
educational questions or problems that only lend themselves to indirect

\(^{42}\) In practice, the broader definition had little impact on what projects ultimately received funding.
\(^{43}\) American Educational Research Association, "Standards for Reporting on Empirical Social Science
Reporting on Humanities-Oriented Research in Aera Publications," *Educational Researcher* 38, no. 6
(2009).
investigation, limited replication or generalization, or that are less empirical (at least in the sense that scientists typically mean). Many important educational problems conceivably fit this description.

One underlying problem is tautological: the principles only apply to scientific research in education, but we only know a particular inquiry is scientific because it fits the principles. Thus, the principles are supposed to serve as both a general definition of scientific inquiry and as the standard by which such inquiry should be assessed. Imagine the confusion that would ensue if “automobile” was defined in terms of what it means to be a good automobile. Webster’s dictionary defines an automobile as “usually a four-wheeled automotive [self-propelled] vehicle designed for passenger transportation.” While this definition has a built-in purpose (passenger transportation), it excludes almost everything people care about when judging a car: cost, fuel efficiency, safety, capacity, handling, power, luxury equipment, etc. Adding a variable like cost to the definition of an automobile makes no sense, because the reasons cost is important depend on the purchaser and his or her intended use. When, in 2009, Congress implemented the “Cash for Clunkers” program, people bought old, run-down cars entirely for their trade in value towards the purchase of a new car. The cheaper the better: no other feature mattered, as long as the car could make it to the dealership. Conversely, some people buy cars precisely because they are expensive. When few people can afford something, its primary value can be as a symbol of social status. So here are two cases where people care a lot about the cost of a car, but because they care about cost for entirely different reasons (both only tangentially related to transportation), one wants the car to be as cheap as possible, the other as expensive as possible. Building some more definitive statement about cost into the definition of automobile-hood wouldn’t bring clarity to the situation, it only brings confusion.

So one problem with building quality standards into a definition is that they assume commonality of purpose where none may exist. When developing evaluative

criteria, purpose is paramount. One way to overcome this problem is to separate the essential, non-negotiable aspects of a concept from an analysis of what makes the concept good for a particular purpose. In the context of SRE, this argues in favor of a strategy that only captures the essential, distinguishing features of science from other forms of scholarly inquiry. To build in more—to try and include standards that are likely to vary across reasonably different contexts for reasonably different purposes—will have the (presumably) unintended side effect of excluding research in particular contexts or for particular purposes. Thus, the central challenge is defining the cognitive aim common to all scientific inquiry that distinguishes it from other forms of inquiry. Any general principles should then be only those absolutely essential to achieving these aims. This is what makes the narrowing criteria in some of the principles odd, like the requirement in Principle 5 that studies are replicable and generalizable. The problem with this requirement is that it excludes some forms of research from science without offering an epistemic rationale for either criterion. For example, lots of educational research relies on test scores. The analysis of test scores should certainly be replicable, but the data collection itself—the administration of the test—is often not replicable in principle or in practice. Similarly, some studies are not concerned with generalizability. For example, sometimes a complete data set is available, so the question is how to analyze and interpret the data, not how to generalize from it.

Replicability and generalizability are not principles of science so much as standards of evidence for some forms of scientific inquiry, but after the paragraph quoted above, the authors go on to suggest a more fundamental standard for evaluating research:

We argue that education research, like research in the social, biological, and physical realms, faces—as a final “court of appeal”—the test of conceptual and empirical adequacy over time. An educational hypothesis or conjecture must be judged in light of the best array of relevant qualitative or quantitative
data that can be garnered. If a hypothesis is insulated from such testing, then it cannot be considered as falling within the ambit of science.\textsuperscript{45}

Empirical adequacy—that is, the fit with relevant evidence—is the extent to which a given claim or theory fits the available evidence. Conceptual adequacy is presumably an assessment of fit with other theories, but this fit only matters to the extent these theories are themselves empirically adequate. This criterion opens the door to other forms of educational inquiry that could be empirically adequate without necessarily fitting SRE’s more narrowly defined principles. Is replicability a criterion for empirical adequacy? If so, what about a study needs to be replicable? Must it be replicable in practice or only in principle, and if only the later what does “in principle” mean? It remains unclear how far this and other criteria, as well as the ultimate standard of empirical adequacy, extend within educational inquiry. I return to question in the next section.

At root, my disagreement with the core ideas of the NRC report is not with its political goal but with its methodological implications. I disagree with readings of the report that suggest something methodologically or epistemically important rests on drawing a line between scientifically based research and other forms of educational inquiry, even if the line is only drawn in principle and not in practice. The line in principle, it seems to me, is part of the same intellectual system that results in more rigid lines being drawn in practice. This is not an argument against the merits of scientifically based educational research which, by most any definition, has made important contributions to educational policy and practice.

Rather, my concern is that criteria for scientifically based research are being suggested without a clear articulation of when and why they are important. Furthermore, it is unclear whether the key ideas underpinning the recommendations of SRE are really distinctive of science as opposed to all systematic scholarly inquiry. The next section argues for a more general conception of educational inquiry that

takes quality, credibility, and alignment with purpose as central epistemic desiderata. I call scholarship that meets these conditions disciplined inquiry.

2.4 Educational research as disciplined inquiry

The goal of this section is to suggest another way of thinking about research quality, where quality is understood in terms of credibility and productivity rather than scientific status. The approach I adopt, which I call disciplined inquiry, builds on the well-established idea that science is an essentially social enterprise. In considering the interdisciplinary study of education from this perspective, I mean to include disciplines often (though not always) excluded from the sciences. I have in mind not just philosophy, history, and anthropology but also hybrid approaches like curriculum studies and policy analysis. As a field of study, rather than a single discipline, education lends itself particularly well to this form of analysis. My argument is not that all educational inquiry should share the same goals, adhere to the same standards of evidence, or even be organized in a similar manner. Rather, I contend that disciplined inquiry involves a particular set of community norms capable of many different instantiations. By offering a set of epistemically efficacious norms, along with an insistence on greater clarity about the shared aims of inquiry, disciplined inquiry avoids the tensions of SRE. Building on several key ideas from the social dimensions of scientific knowledge, especially the work of Helen Longino, I attempt to clarify what the pursuit of disciplined educational inquiry entails.

The extension of arguments and ideas from the philosophy of science to educational inquiry will be imperfect and my argument incomplete, since there are significant differences not just between the natural sciences and the social sciences, but between the social sciences and the humanistic disciplines. A thorough defense would require a dissertation unto itself, and the issue that prompted this study—an argument for research that better addresses educational problems—would fall by the wayside. Instead, my goal is to suggest the plausibility of extending several key ideas from the social epistemology of scientific practice to scholarly inquiry generally.
I am under no illusion that my account is a definitive, knock-down argument and defense.

In Section 2.4.1, I offer a brief summary of two early conceptions of disciplined inquiry in education that point to a tradition worth building upon. Section 2.4.2 offers an account of a disciplining community, arguing that such a community requires several important norms to support the critical, discursive practices that promote more effective knowledge production. Then in Section 2.4.3 I consider in more detail the three levels at which critical discursive interaction occur. While all are important, and every discipline periodically engages in each, different disciplines vary in the manner, emphasis, and frequency of discourse spent at each level. These differences are important, because they help to explain how, for example, the work of one disciplinary community can be an epistemic resource for another community, as well as pointing to the epistemic challenges of interdisciplinary research. Finally, Section 2.4.4 sums up what I take to be the central advantage of disciplined inquiry over more circumscribed conceptions of science: the opportunity to align critical discursive practices and standards with a broad range of cognitive aims. At root, disciplined inquiry is a form of collective problem-solving that produces public accounts of the problem(s) and epistemically acceptable explanation(s) at which it aims. I conclude by explaining how this approach can better address the two challenges introduced in Chapter 1: promoting understanding for use in practice and producing more useful research.

2.4.1 Early discussions of education as disciplined inquiry. In 1969, Lee Cronbach and Patrick Suppes edited a National Academy of Education sponsored report, *Research for Tomorrow's Schools: Disciplined Inquiry for Education (RTS).*46 The forward describes the state of educational inquiry at that time: “Currently, confusion about the nature of research in education and its potential contributions to education and schooling seriously interferes with constructive efforts to create and

---

use relevant knowledge.” Thus, the goal of RTS was to offer a “more adequate understanding” of educational inquiry, and the report was addressed primarily to educational decision-makers with the hope of better informing those who use research about the scholarly process. RTS’s authors believed an improved understanding would help policymakers use educational knowledge more effectively and promote future investment in educational research. As they wrote:

There are misconceptions about the nature of scholarly work and its relation to educational practice, and these misconceptions divert effort into unprofitable channels. The course of inquiry is shaped by the ways in which resources are allocated, research institutions administered, and educational innovations adapted. These forces encourage one set of studies rather than another, and so determine whether the results will be transient and minimal, or fundamental and far-reaching.

Influencing and improving the work of educational researchers was only a secondary goal, since the authors believed these changes could only happen organically, from the ground up: “To discover and propagate new mechanisms for conducting research and new research styles will require changed attitudes within the academic community. These must grow; they cannot be legislated.”

The report’s educational context and policy audience shaped its form and content, limiting the depth of its conceptual framing of disciplined inquiry. Consequently, disciplined inquiry as presented in RTS is more assorted and aphoristic than systematic and detailed. The authors write that disciplined inquiry differs “from other sources of opinion and belief,” in that it “is conducted and reported in such a way that the argument can be painstakingly examined.” It does not simply “depend for its appeal on the eloquence of the writer or on any surface plausibility.” Different disciplines have different goals, ask different questions, and apply different standards

47 Ibid., vii.
48 Ibid.
49 Ibid., xvi.
50 Ibid., 7.
51 Ibid., 13.
that “separate sound argument from incomplete or questionable argument.”\textsuperscript{52}

Nonetheless “disciplined inquiry does not necessarily follow well-established, formal procedures” of a single discipline, nor are all its conclusions true since “each investigation is limited by its methods, and the consensus of the best informed members of a discipline is limited by the state of the art.”\textsuperscript{53} But at its core disciplined inquiry involves a particular set of dispositions and norms:

The success of academic men [sic] in breaking old intellectual molds and inventing fresh concepts results from the fact that they value the process of inquiry at least as much as they value its fruits. They are trained in specialized techniques of observation and analysis. Instruments to refine the judgment of the observer, statistical models to weed out chance effects, mathematical models, canons of documentation, and formal criteria of acceptable definition, constitute the technology of inquiry. But far more fundamental to disciplined inquiry is its central attitude, which places a premium on objectivity and evidential test.\textsuperscript{54}

Over the course of the report, four recurring themes are developed: (1) an argument for casting a wide net that includes many different forms of research as disciplined inquiry, (2) an insistence on the distinct value of scholarly knowledge over and above other possible sources of information, (3) a contention that the epistemic value of scholarly work rests (at least partly) on the dispositions and norms of research communities, and (4) a recognition that all knowledge is fallible and partial rather than certain and complete. In this regard RTS shared some commonalities with SRE, though the former obviously did not use these contentions to define educational “science”—a term largely absent from the report. The essential attitude described in RTS resembles several of SRE’s principles for scientifically based research, including the focus on collecting empirical data, careful analysis, and the subjecting of methods and conclusions to careful scrutiny and critique.

\textsuperscript{52} Ibid., 15.
\textsuperscript{53} Ibid., 16-17.
\textsuperscript{54} Ibid., 18.
Beyond RTS’s general conception of disciplined educational inquiry, a central organizing idea in the report is a distinction between “conclusion-oriented” and “decision-oriented” research. The main difference is audience: conclusion-oriented research speaks to some segment of the scholarly community, whereas decision-oriented research is for a particular decision maker and often comes in the form of a commissioned study. Because these categories map roughly (if only roughly) onto notions of basic and applied inquiry, I save my discussion of these categories for Chapter 3. For present purposes I note only that RTS conceives of both conclusion-oriented and decision-oriented research as amenable to disciplined inquiry. No argument is offered for placing more emphasis or resources into one mode of inquiry rather than another, just a recognition that the purposes of educational inquiry often differ depending on the context and audience.

Although RTS does not offer detailed recommendations for reorganizing educational research around disciplined inquiry, other scholars have refined and extended the concept within the field of educational research. In particular, Lee Shulman has made extensive use of the notion of disciplined educational inquiry, reorienting RTS’s framing to the needs of 21st century schooling.

Shulman recommends a “disciplined eclectic” approach to educational inquiry that builds on a range of scholarship from education and the social sciences. He attributes the eclectic label and related ideas to the work of Schwab (a scientist and educational philosopher), Robert Merton (a sociologist), and Lee Cronbach (a psychologist). In a series of influential articles and book chapters published over the course of several decades, Shulman’s conception of disciplined inquiry was addressed

55 Ibid., 20.
reservations about the ability of traditional academic disciplines to address problems of schools and schooling. With RTS, he agreed that “What is important about disciplined inquiry is that its data, arguments, and reasoning be capable of withstanding careful scrutiny by another member of the scientific community.” In brief, disciplined eclecticism requires a broader methodological understanding and openness than traditional disciplinary research, but its distinguishing feature is the disciplined application of research methods:

What distinguishes research from other forms of human discourse is the application of research methods. When we conduct educational research we make the claim that there is method to our madness. Educational research methods are forms of disciplined inquiry. They are disciplined in that they follow sets of rules and principles for pursuing investigations. They are also disciplined in another sense. They have emerged from underlying social and natural science disciplines which have well-developed canons of discovery and verification for making and testing truth claims in their fields. Education itself is not a discipline, but rather a field of study on which we bring to bear the various forms of disciplined inquiry.

And a precondition for disciplined inquiry is an understanding of the problem:

We must avoid becoming educational researchers slavishly committed to some particular method. The image of the little boy who has just received a hammer for a birthday present and suddenly finds that the entire world looks to him like a variety of nails, is too painfully familiar to be tolerated. We must first understand our problem, and decide what questions we are asking, then select the mode of disciplined inquiry most appropriate to those questions.

When the educational problems being investigated require the resources of many different disciplines or methodologies, the practice of critically “disciplining” scholarship remains essential:

Undisciplined eclecticism is no virtue when compared to carefully conducted research within a particular research program’s tradition. Indeed, it is probably worse. Nevertheless, we need not give up on the notion of research

---

59 Ibid.: 12.
60 Emphasis original. Ibid.
programs conducted in the spirit of disciplined eclecticism. A new generation of educational scholars is being prepared who are truly research methodologists, that is, capable of employing alternative approaches to problems as they are formulated, rather than the orthodox research methodologists of an earlier generation. Moreover, the development of research centers and institutes in which representatives of distinctly different research programs and traditions can work collaboratively shows promise for the development of healthy new hybrid programs. It may be that in many cases individual studies cannot be pursued jointly; the canons of each research program must be allowed to function, thus to discipline the inquiry as it is pursued. But when investigators have learned to speak each other’s languages, to comprehend the terms in which other programs’ research questions are couched, then processes of deliberation over findings can yield the hybrid understandings not possible when members of individual research programs dwell in intellectual ghettos of their own construction.\(^\text{61}\)

In cross-disciplinary settings, disciplined inquiry cannot mean inquiry aligned with a particular discipline or set of methods. Instead, he suggests yet a third, possibly more fundamental, understanding of “discipline:"

> Discipline refers to the management of impulse and the control of intellectual caprice. While flexibility and creativity are certainly to be valued in the attack on educational problems, scholarship must somehow be disciplined by some principles. Schwab’s solution to the dilemma is the creation of cross-disciplinary deliberation groups, collection of scholars representing a cross-section of disciplinary perspectives selected because of their bearing upon the topic or problem at hand.\(^\text{62}\)

We thus have three distinct senses in which educational inquiry is disciplined, on Shulman’s account: (i) it draws upon theories and knowledge found in the disciplines, (ii) it applies the methods of the disciplines to educational problems, and (iii) it requires a disciplined disposition.

Although Shulman’s writings are reasonably well-aligned with the key ideas of RTS, he emphasizes two themes largely absent from this earlier work, taking

---

61 — — — , "Paradigms and Research Programs in the Study of Teaching," 33.
disciplined inquiry in a somewhat different direction. First, whereas RTS remained agnostic with respect to the division of cognitive labor aimed at extending disciplinary knowledge versus addressing educational problems, Shulman consistently argued for eclecticism. This eclecticism is recommended not just for the field of educational research but also for individual researchers, whom he believed should be trained in at least two forms of research methodology. Doing so would help educational inquirers avoid the “hammer problem.” Recently he has taken the argument a step further, arguing that “The education of scholars would be significantly improved if we acknowledged that the Ph.D. is a form of professional education.”

Second, Shulman explicitly links the virtues of the disciplined eclectic researchers to their ability to function effectively within cross-disciplinary teams aimed at addressing educational problems. Thus, problem-definition and the value of mustering educational inquirers with the knowledge, methodological skill, and proper disposition to address educational problems are of great importance.

Despite the consistent recognition in RTS and Shulman’s writings that disciplined inquiry is a fundamentally social enterprise, this theme stays largely in the background and is not subject to much scrutiny when it comes to the actual practice of educational research. Shulman, for example, offers a detailed analysis of the training of educational scholars, but stops short of describing what happens when disciplined eclectic researchers finally get together at the table. In the few instances when it is brought to the fore, relatively little is said about the nature of the interactions between researchers, despite the apparent centrality of these interactions to the quality of the inquiry. My formulation of disciplined inquiry differs from these accounts in foregrounding the essential social norms for communities of educational inquirers. Accordingly, discussion of the knowledge, skills, and


64 Shulman, "The Practical and the Eclectic."
dispositions of individuals takes a back seat to discussions of the critical discursive practices that I argue characterize all forms of disciplined educational inquiry.

2.4.2 Disciplined inquiry as social knowledge. Across the academy there is a general recognition that the production of scholarly knowledge is a social enterprise. In this regard, NCLB, SRE, and some of their harshest critics are in loose agreement. What this means for the organization, content, and practice of educational inquiry has considerably less consensus. For social minimalists, the recognition that educational inquiry is a social practice largely amounts to acknowledging two points: researchers necessarily build on the work of others and the principal arbiters of scholarly merit are one’s academic peers. The latter idea, typically embodied in peer review processes, usually implies that published scholarly work or the selection of grant recipients should be done based on the anonymous critical review of the scholarly work or grant applications of one’s academic peers. Two successor reports to SRE produced by the NRC focused on strategies for strengthening peer review in educational inquiry as the primary means for improving the credibility of research. But for others, the insight that scholarly inquiry is social goes much deeper. On this stronger view, often associated with Thomas Kuhn’s well-known (and often misunderstood) work in the history and philosophy of science, research is social in a paradigmatic sense. Research on this account encompasses what Kuhn referred to as a disciplinary matrix of symbolic generalizations, metaphysical presumptions, values, and exemplars, all of which are socially constituted.

While the notion of disciplined inquiry I recommend falls in between the minimalist and strong views, in most respects my argument favors the stronger account. This section briefly introduces several key philosophical thinkers on the

---

social dimensions of knowledge production before turning to Longino, on whose work I develop a more explicitly social conception of disciplined inquiry.

Charles Peirce was one of the earliest philosophers to discuss the social dimensions of scholarly inquiry. He argued that “the opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth, and the object represented in this opinion is the real.”68 Such investigation is necessarily a project of a community of inquirers since:

We individually cannot reasonably hope to attain the ultimate philosophy which we pursue; we can only seek it, therefore, for the *community* of philosophers. Hence, if disciplined and candid minds carefully examine a theory and refuse to accept it, this ought to create doubts in the mind of the author of the theory.69

More than a half-century later, Karl Popper’s influential falsificationist program shared some resemblance to Peirce’s view.70 Both understood the fallibility of knowledge and the value of critical communities of inquirers. Yet, unlike Peirce, Popper’s approach focused on the problem of demarcating science from non-science. Popper accepted that not all knowledge was scientific, but he believed there was distinct social value in the sciences as a tool for social problem solving. On this view, what distinguished scientific inquiry was the “prohibitive” nature of scientific theories: certain observations were necessarily incompatible with a theory being true. While no amount of data could ever prove the truth of an empirical theory, observations that were incompatible with the theory could falsify it. Thus, the epistemic role of critical communities of inquirers was their efforts to falsify the empirical theories of their peers.

---


70 This is not surprising given that Popper describes Peirce as “one of the greatest philosophers of all time” Karl R. Popper, *Objective Knowledge: An Evolutionary Approach*, Rev. ed. (New York: Clarendon Press, 1979), 212.
Thomas Kuhn and Imre Lakatos also had fundamentally social conceptions of scientific inquiry, but they focused more on the structure and norms of scientific communities than either Peirce or Popper. Notably, neither Kuhn nor Lakatos was optimistic about the application of their ideas to the social sciences. Kuhn’s disciplinary matrix, for example, only works as a framework for understanding what he calls *normal science*, where there is widespread agreement about the aims of research and the standards of evidence. During periods of normal science (in contrast to periods of revolutionary science), researchers work within the existing paradigm or disciplinary matrix to solve modest puzzles. The problem with conceiving of the social sciences and humanities as fitting the “normal science” concept is the frequency with which these disciplines question their own aims and standards. As Kuhn wrote near the end of his career:

> In the natural sciences the practice of research does occasionally produce new paradigms, new ways of understanding nature, of reading its texts. But the people responsible for those changes were not looking for them. The reinterpretation that resulted from their work was involuntary, often the work of the next generation. The people responsible typically failed to recognize the nature of what they had done. Contrast that pattern with the one normal to philosopher Charles Taylor’s social sciences. In the latter, new and deeper interpretations are the recognized object of the game.

Of course some social science disciplines, notably psychology and economics, may fit the natural science model of normal science reasonably well. But humanistic forms of inquiry like philosophy and other hermeneutically inclined disciplines have reflexivity or reinterpretation baked-in as a cognitive aim of the discipline. One of the great philosophical questions since Plato has been “What is philosophy?” When Alfred North Whitehead wrote two thousand years later that “the safest general characterization of the European philosophical tradition is that it consists of a series

---

71 This has not prevented social scientists and critical theorists from applying both Kuhn and Lakatos’s ideas to the study of social phenomena.  
of footnotes to Plato” he meant, quite analogously, that the trajectory of western philosophy was largely systematic reinterpretations of “the wealth of general ideas scattered through” his writings.  

Lakatos was skeptical about the possibility of a social “science” for very different reasons than Kuhn. His methodology of scientific research programs was an attempt to offer a rational, continuous account of the advancement of science. Whereas Kuhn treated significant scientific change much like religious conversion, Lakatos understood the defensive stance of researchers to anomalous (potentially falsifying) data as perfectly rational. Research programs, on Lakatos’s view, consist of a positive heuristic that tell researchers what paths to pursue and negative heuristic that indicate paths to avoid. The negative heuristic encompasses the “hard core” of a research program—those core theories, principles, and propositions that are fundamental to the research program. Conversely, the positive heuristic includes a broad set of directives about the kinds of research questions and problems that should be pursued, including a “protective belt of auxiliary hypothesis” that protects the core from refutation. On this account, a defining feature of a successful research program is its ability to continually account for more and more anomalies by building out auxiliary hypotheses; Lakatos calls the result a progressive problem-shift. Lakatos was skeptical about the ability of the social sciences to produce such shifts because he believed social research tended to be either theory-rich and prediction poor (e.g., Marxism, Freudism) or predictively precise and theory-poor (e.g., psychology in the 1960s). In either case, Lakatos had difficulty identifying a coherent research program that matched his methodology of hard cores and protective belts.

---

75 Ibid., 88.
Before turning to a more promising approach to disciplined inquiry, a brief explanation of why I do not believe either Kuhn’s or Lakatos’s concerns about the social sciences are devastating to their knowledge-productive capacities. Kuhn’s skepticism had to do with the ability of certain social science and humanities disciplines to operate in a “normal science” mode. But this has less to do with their general knowledge-productive capacities and more to do with their ability to engage in the “normal science” enterprise. Many, including and especially Lakatos, have rightly criticized Kuhn for treating significant scientific changes like religious conversions. If all disciplines and fields allow, at times, hermeneutic reinterpretation, the difference is really just a matter of degree. Obviously there is a point—one which some forms of postmodern and poststructural scholarship approach—where self-critical reflection and reinterpretation prohibits collective knowledge or understanding. This is the point where some forms of scholarship diverge from disciplined inquiry. (More on this below.)

Lakatos’s skepticism turned on concerns about the empirical adequacy and predictive power of social theories. Without reasonably robust theory, the progressive problem-shifts that incrementally advance science are impossible because of inadequate predictions or explanations. Yet social theorists have identified work that clearly falls between these extremes. Notably, Robert Merton’s account of middle range sociological theories points to a form of social theory that has both predictive and theoretical content.76 And in the four decades since the writings of Merton and Lakatos, many more examples of middle range social theories have emerged in psychology, sociology, economics, and political science. Second, Lakatos’s notion of progressive problem-shifts only makes sense in the context of shared cognitive aims—that is, research programs that are trying to account for the same observations. But in the middle range theories of many social science research programs, cognitive aims diverge—that is, they often are interested in explaining and

predicting different aspects of social phenomena, sometimes for different purposes. More often than directly competing research programs, theory evolves to better conform to and explain the phenomena that are its object. Such aims are presumably reflected in a research program’s positive and negative heuristics, but given the emergence of new puzzles or problems even these might be reinterpreted, leading a research program in new directions. So while more hermeneutical modes of inquiry might soften parts of Lakatos’s hard core (and possibly harden parts of his protective belt), there is no need to abandon the insight that some theories are more important than others to a research program’s success. It might simply be the case that what those parts are tends to change over time in the social sciences more than in the natural and biological sciences.

Recently, Helen Longino has argued for a social conception of scientific knowledge she calls contextual empiricism. In the remainder of this section I argue that three aspects of Longino’s account are epistemically and practically important features of disciplined inquiry: a focus on the conformation of a hypothesis or theory to its object (a variation on the scientific value of empirical adequacy), the community norms essential to critical discursive practices, and the centrality of shared cognitive aims to the productive interaction of communities of inquirers. I focus only on those features of contextual empiricism I intend to apply to disciplined inquiry.

Longino’s account addresses a number of problems in epistemology and philosophy of science, but the most important is the underdetermination of theories by evidence. The problem of underdetermination has traditionally been of great interest (and concern) to philosophers, but it has received comparably less attention from empirical researchers. More familiar to researchers is the problem of induction,

---


78 Contextual empiricism is offered as a theory of scientific knowledge, but I extend the concept to all empirical inquiry—a category into which I lump not just history and anthropology but also (somewhat controversially) philosophy and literary studies. My reasons for doing so are discussed in Section 2.5.
which can be concisely stated as the problem of warranting generalizations from particulars. This is the problem pointed to in Section 2.2.1, where no matter how many black ravens I observe, I can never be certain that the generalization “all ravens are black” is, in fact, true. The problem of induction is a particular formulation of the more general problem of underdetermination. In this example, the claim “all ravens are black” is underdetermined by the available evidence. But here is another, more problematic example of underdetermination: imagine an astrophysicist with a hypothesis about the mechanics of planetary motion. To test the hypothesis, the astrophysicist makes some predictions by combining the hypothesis with some accepted background assumptions that imply particular observations. Then the observations are made, but they fail to conform to the predicted observations. Should the astrophysicist infer that the hypothesis is wrong, that one or more background assumptions are wrong, or that the observations (perhaps due to malfunctioning or miscalibrated equipment) were wrong?\(^{79}\)

At first glance the problem seems innocuous enough, since new observations can usually be made and background assumptions checked. But on deeper reflection the problem actually goes to the heart of scientific inquiry, to the connections between observation and language, evidence and hypothesis. As Longino argues:

> The background assumptions that fill that gap [between evidence and hypothesis], then, include substantive and methodological hypotheses that, from one point of view, form the framework or proximate intellectual context within which inquiry is pursued and, from another, structure the domain within which inquiry is pursued. These hypotheses are most often not articulated but presumed by the scientists relying on them. They facilitate the reasoning between what is known and what is hypothesized. From a traditional perspective this raises major problems for justification: if the justification of hypotheses requires assumptions, then how are these assumptions in turn justified? And how is it possible to screen out biasing factors such as individual idiosyncrasies, wishful thinking, values, social prejudices, ideologies, tacit metaphysics? In some cases evidence for

assumptions can be offered, but the same problem of underdetermination besets this evidential reasoning as well.\(^{80}\)

Another commonly referenced implication of the underdetermination problem is that for any given body of data, there are many (actually, an infinite number) of theories or hypotheses that are compatible with this data but incompatible with one another.

Contextual empiricism offers a partial solution to the underdetermination problem by recognizing that: (i) “purely logic constraints cannot compel” acceptance of a particular hypothesis or theory and (ii) inquirers are situated within a “network of relationships—among other individuals, social systems, natural objects, and natural processes” that can serve as a resource for closing the gap left by logic.\(^{81}\)

Thus, a claim or theory can be warranted or “epistemically acceptable” but also contingent on the background assumptions shared by the community of inquirers. Contextual empiricism thus avoids privileging some untested (or uncriticized) background assumptions over others. Knowledge is localized in the sense that it only makes sense against an implicit set of background assumptions. Longino’s solution is partial because it does not eliminate underdetermination so much as offer an account of how epistemically acceptable knowledge is possible with particular communities of inquirers. Shared aims and background assumptions constrain and direct cognitive labor by guiding inquirers to some questions or problems and away from others.

A second problem contextual empiricism addresses is the idea that truth is (or should be) a key evaluative concept in scientific inquiry. Two aspects of truth make its application to science problematic. First, truth is generally considered a characteristic of propositions, but much scientific inquiry involves non-linguistic data like images. It is relatively easy to assess the statement “Jon went to the grocery store today” as true or false (though, notably, the truth value of this same statement

\(^{80}\) Longino, *The Fate of Knowledge*, 127.

\(^{81}\) Ibid., 128.
may be different tomorrow.) But it makes much less sense to talk of the truth or falsity of a visual representation, like a map of California. What does it mean for a map to be “truth”? Second, truth is a binary concept: it is clear what is meant if a proposition is true or false, but it’s much harder to make sense of “more” or “less” true propositions. If two people are arguing about the population of San Francisco, and one claims the population is 500,000 people while the other claims the population is 800,000 when (say) the actual population is 812,891, it seems strange to dismiss both claims as false: one was clearly much closer to being correct than the other. But that requires an additional concept, something like “approximate” truth—Popper used the term “verisimilitude”—as a way to talk about propositional statements that fall short of absolute truth but are, nevertheless, much better representations than other statements.

Rather than evaluating knowledge claims as (approximately) “truth” or “false,” Longino introduces the much more useful concept of conformation as the relationship between some content and its object. Thus we can say that one map (some content) better conforms to the State of California (its object) than another without being in the strange position of describing a map as true or false. Similarly, we can say that the content of “800,000 people” better conforms to its object, the population of San Francisco, than “500,000 people,” even though neither gets the number precisely right (some might reasonably question whether there is a precisely correct answer to ambiguously stated questions.) We can even talk about how well a literary interpretation conforms to the text and the intent of its author, even if no single “true” interpretation exists.

The concept of conformation helps makes sense of contextual empiricism’s account of knowledge as a critical discursive practice amongst a community of inquirers. To produce knowledge claims that are epistemically warranted, such practices must meet four criteria:
1. **Venues for criticizing evidence, methods, assumptions, and reasoning.**  
“Criticism of research ought to be articulated in the same standard and public venues in which ‘original research’ is presented: journals, conferences, and so on. In addition, critical activities should be given the same weight or nearly the same weight as is given to ‘original research’: effective criticism that advances understanding being as valued as original research that opens up new domains for understanding.”

2. **Uptake of criticism.** “The community must not merely tolerate dissent, but its beliefs and theories must change over time in response to the critical discourse taking place.”

3. **Public standards.** “There must be publicly recognized standards by reference to which theories, hypotheses, and observational practices are evaluated and by appeal to which criticism is made relevant to the goals of the inquiring community. . . . [And] a community’s standards are themselves subordinated to its overall cognitive aims, which will be implicit in its practices even if not fully explicit.”

4. **Tempered intellectual equality.** “Communities must be characterized by equality of intellectual authority.” Meaning, more specifically, that “the social position or economic power of an individual or group in a community ought not determine who or what perspectives are taken seriously in that community. Where consensus exists, it must be the result not just of the exercise of political or economic power, or of the exclusion of dissenting perspectives, but a result of the critical dialogue in which all relevant perspectives are represented.”

Put more formally:

Some content A is *epistemically acceptable* in community C at time t if A is supported by data d evident to C at t in light of reasoning and background assumptions which have survived critical scrutiny from as many perspectives as are available to C at t, and C is characterized by venues for criticism, uptake of criticism, public standards, and tempered equality of intellectual authority.

---

82 Ibid., 129.
83 Ibid.
84 Ibid., 130.
85 Ibid., 131.
86 Ibid., 135.
Together, conformation and the norms for critical discursive practices address some of the problems of underdetermination and truth that have plagued previous accounts of scientific inquiry. But they also address the two tensions I point to in SRE: ambiguity about the aims of scientific inquiry and the extent of its social nature. The problem of aims is addressed in the third criterion requiring public standards for evaluating content. These standards, Longino notes, derive from what the community deems necessary for achieving its cognitive aims. Thus, the standards are problem or question relative, rather than absolute. This is similar to SRE’s contention that a research design should be matched to the problem or question it seeks to answer. The key difference is that for Longino, the standards are completely open, subject to better arguments and evidence. SRE’s standards take the form of predefined principles that fit some cognitive aims better than others.

Based on the preceding discussion, I propose that disciplined inquiry is, at root, a form of collective problem-solving that produces public accounts of the problem(s) and epistemically acceptable explanation(s) at which it aims. I adopt the features of contextual empiricism sketched above as a minimum standard for epistemic acceptability (e.g., venues, uptake, public standards, and tempered intellectual equality). In addition, the problem-solving criterion emphasizes the fact that inquiry is conducted with implicit or explicit cognitive aims in mind. All research is motivated by curiosity, even if that curiosity has not been translated into a well-formulated problem or question. A crucial step in solving difficult problems is clear articulation of the issue at hand, in fact, the early stages of some scholarly inquiry consists entirely of arguments about what the problem is. The publicity requirement emphasizes that disciplined inquiry is not simply problem-solving, collective or otherwise. It also involves the production of knowledge in the form of publicly accessible content, a criterion that makes critical discursive practice possible. It also is what allows the “detaching” of knowledge from particular agents or subjects, facilitating its transmission to those not involved in the production process. Another important feature of this account of disciplined inquiry is the recognition that critical
discursive interaction can occur at three different levels. The next section considers these levels in more detail, explaining each with illustrations from educational research.

2.4.3 Three levels of criticism. Disciplined inquiry as social knowledge roughly distinguishes three levels at which critical discourse can occur: normal criticism, criticism of cognitive standards, and criticism of cognitive aims. Because, as Kuhn noted, these distinctions may be clearer in fields like psychology and economics than in fields like philosophy or critical social theory, it is helpful to consider what criticism at each level looks like within education.\(^{87}\)

*Normal criticism.* Normal criticism is the most common criticism found in scholarly communities, especially in well-established ones. Loosely, this is criticism in the mode of Kuhnian normal science, where the focus is on the application of agreed upon standards to particular content (an article, grant application, presentation, etc). When peer reviewers evaluate an article based on a public set of academic standards, this is the normal criticism. The American Educational Research Association’s (AERA) standards for reporting empirical research, for example, are frequently referenced by referees of articles submitted to AERA journals. These standards include quality of problem formulation, design and logic of the study, sources of evidence, measurement and classification data, and analysis and interpretation of results.\(^{88}\) Of course, researchers often attempt (implicitly or explicitly) to apply standards like these to their own research in the course of inquiry. But normal criticism as part of a critical discursive practice involves their application by others. Kuhn believed normal criticism could also only occur from within a particular paradigm or disciplinary matrix, but this is not a requirement of disciplined inquiry (or of Longino’s contextual empiricism).\(^{89}\) Because standards of evidence are public, anyone can (in principle) apply a standard to some content regardless of

---

\(^{87}\) Kuhn, "The Natural and the Human Sciences."


\(^{89}\) Longino, *The Fate of Knowledge.*
whether or not they share the same background assumptions or cognitive aim. Of course different conclusions may be drawn as to how successfully the content meets the standard, but explaining resolving these differences would likely draw attention to conflicting background assumptions—just as it could within a paradigm—that could then be subjected to scrutiny.

Note that in some disciplines and fields of inquiry normal criticism is intertwined with criticism of cognitive standards and aims. This is true in many areas of educational inquiry; in fact, AERA published a separate set of standards for reporting humanities-based research that acknowledges the importance of what it labels “reflexive educational research.” Unsurprisingly, the standards for criticizing reflexive research and other forms of humanities-based research are even more general than the standards for reporting empirical research under the banners of: significance, methods, conceptualization, substantiation, coherence, quality of communication, and ethics. The content of these standards can be summed up as a call for scholars to ask interesting and important questions that get clearly and coherently addressed using appropriate methods and evidence. Thus, for reflexive modes of inquiry (as for many domains of humanistic scholarship), the standards for normal criticism amount to general standards for evidential reasoning found in any good argument.

**Criticism of standards.** Public standards of evidence are essential to scholarly inquiry, but these standards can themselves be subject to criticism. Two vivid examples of criticism of public standards were discussed in Sections 2.2 and 2.3 respectively, where debates over the definition of scientifically based research inevitably turned into debates over appropriate standards of evidence. It might reasonably be asked why any standards of evidence are necessary beyond vague reminders to muster appropriate evidence and argument.

---

The answer is that standards are often an indirect articulation of cognitive aims. In this regard the AERA standards I pointed to above are not especially informative, since the aims of educational researchers and the audiences for AERA journals are quite diverse. I imagine standards for publication in the journals *Science* or *Nature* would face similar difficulties, and it is one of the reasons good journals work hard to make sure the backgrounds of reviewers are well-matched to the articles they review. Often implicit or explicit in more specific standards are background assumptions about the domain(s) of knowledge (the questions and problems of the field). Catherin Elgin notes one important, often overlooked reason for this:

> We arrange for our standards to be met. We construct systems of categories that settle the conditions on the individuation of entities and their classification into kinds. Thus, for example, we devise a biological taxonomy according to which a dachshund is the same kind of thing as a Doberman, but a horse is a different kind of thing from a zebra.91

Although Elgin’s examples are standards for categorization, the point holds for other standards as well. The standards advocates of RFTs hold themselves to, for example, are standards they know RFTs can meet. This goes a long way towards explaining why, for example, classic texts on social experiments like Shadish, Cook, and Campbell’s *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* focus more on internal validity (a strength) than on external validity (a weakness).92 RFTs are often well suited to the former but poorly suited to the latter, but (as noted in Section 2.2) their advocates tend to prioritize the standards RFTs are best able to meet.

Two characteristics of the educational field frequently complicate agreement on common sets of standards: its *multidisciplinarity* and its *professional-orientation*. Education’s multidisciplinarity means many different scholars are trained in different

---


methodological traditions, pursuing different questions, and drawing on different concepts and theories as intellectual resources. Yet, with the notable exception of educational psychology, only a handful of faculty trained in any particular discipline may be found in a typical education school. Education’s professional-orientation—that is, its location in professional schools charged with preparing teachers, administrators, curriculum specialists, program evaluators, and policy analysts, as well as educational scholars—means that a parallel tension exists between the pursuit of more basic knowledge (and the goal of advancing science) and more applied knowledge (and the goal of practical use). This problem is the focus of Chapter 3, so I just want to explain briefly how this complicates the setting of standards of evidence. All research idealizes the educational context through key methodological assumptions (e.g., about the rationality of human actors, the consistency of educational contexts over time, the desirability of particular educational ends). Different standards take for granted different background assumptions about what is important and encourage different units of analysis. Some standards facilitate more fundamental knowledge, especially those that require the abstracting of particular educational phenomena to laboratory settings. Other standards facilitate application and use; for example, the *Harvard Business Review* regularly offers readers assessment “tool kits” they can use to improve themselves or their organization.93

Agreement on standards is further complicated when one’s disciplinary brethren outside of education or the grants supporting one’s research value certain kinds of knowledge, while parents, educators, and policymakers value other kinds of knowledge. Shared standards allow us to extend our understanding in particular knowledge domains by facilitating scholarly community, mutual understanding, and cooperation. But this disciplinary community can come at a price. As Lee Shulman notes in the quote that began this chapter:

To compress one’s education research into forms that fit any one discipline may corrupt the inquiry by distorting its questions into forms that more readily fit a particular disciplinary template. Education research must be disciplined if it strives to be credible and useful, but what standards or rules should it follow?\textsuperscript{94}

Standards of evidence designed to improve rigor and quality can also have a distorting effect when it comes to pursuing certain cognitive aims.

*Criticism of cognitive aims.* Given that scholarly inquiry is a collective enterprise, perhaps its most important feature is that participants have shared cognitive aims. Cognitive aims are simply goals about what knowledge gets produced. Without shared goals, it is hard to imagine agreement on standards or productive interactions. Imagine three friends who all want to vacation together in Florida, but they had different desires about whether to stop in South Beach on the way to Key West, whether to stay in a villa or a hotel, whether to travel for nine days to two weeks, etc. The aims of the friends are sufficiently aligned that productive deliberation can be had. On the other hand, if the three friends do not have reasonably well-aligned goals, deliberation becomes more difficult, if not impossible. If one friend (a homebody) really wants an excuse to stay home with his family, another (a socialite) wants to meet as many new people as possible, and a third (a surfer) values catching waves over spending time with others, it is unlikely they will be able to agree on Florida vacation plans.

The philosopher Larry Laudan has argued at length for the importance of cognitive aims as a guiding force in scientific inquiry.\textsuperscript{95} He makes two points of particular salience to educational inquiry. The first point is historical: the aims of science have changed over time. Although there is a tendency to think that science has always had the same aim—the pursuit of “truth” or “knowledge”—these generic labels disguise a diverse and evolving set of more specific cognitive aims addressed to

\textsuperscript{94} Shulman, "Professing Educational Scholarship," 161.
the problems and questions at particular point in history. Laudan’s second point has
to do with the justification of particular scientific aims. Philosophers like Karl Popper
and Karl Reichenbach argued that the goals or purposes of inquiry were outside the
domain of science. At most, the internal consistency of aims could be critiqued, but
the ultimate choice of aims was beyond the pale of science. Laudan takes a very
different view, which he calls a “reticulated model” of scientific justification. On this
model, he notes that aims can be criticized for at least two reasons external to the
aims themselves: “one may argue against a goal on the grounds (i) that it is utopian
or unrealizable or (ii) that it fails to accord with the values implicit in the communal
practices or judgments we endorse.”96 The first kind of criticism is about empirical
constraints, the second kind is about the consistency or alignment of one’s aims with
the research enterprise one seems to be engaging in.

It is important to note that a diversity of cognitive aims is both epistemically
and practically desirable. Educational historian David Tyack beautifully illustrates this
point in “Ways of Seeing: an Essay on the History of Compulsory Schooling.” In the
article, Tyack examines compulsory schooling from different political, organization,
and economic perspectives, arguing that each lends something unique to our
historical understanding. But he rejects “simple additive eclecticism” since “some
interpretations do fit certain times and places better than others.” Perhaps “more
fundamentally, the models deal with social reality on quite different levels: the
individual or the family, the ethnocultural group, the large organization, and the
structure of political or economic power in the society as a whole.”97 Rather than
offering a single correct perspective, different historical accounts with different ends
in view offer a more complex and therefore (perhaps) more complete picture of
compulsory schooling. If, as Ellen Lagemann has argued, one of the most important
reasons for studying the past is “to gain vantage on the present,” having a wealth of

96 Ibid., 50.
interpretations to draw upon also offers the practical benefit of bringing greater illumination to the present.  

At the same time, shared cognitive aims are essential to getting collaborative inquiries off the ground. Agreement about cognitive aims is always a matter of degree, but clarity becomes more important as disciplinary diversity increases. With shared cognitive aims comes the possibility of shared standards of evidence and norms of argument. In educational inquiry, discussion of cognitive aims has usually taken a back seat to debates over the acceptability of various methods or research traditions. Thus, qualitative research has often been contrasted with quantitative research, experimental designs with observational designs, critical approaches with postpositivist approaches, and scientific methods with other forms of educational inquiry. The most common interpretation of these debates is that different traditions recommend fundamentally different approaches to answering educational questions and addressing educational problems. An alternative interpretation is that different traditions frequently ask different questions and pursue different problems. So although they may have fundamentally different cognitive aims, they need not be fundamentally incompatible.

2.4.4 The potential merits of disciplined inquiry. The previous subsections sketched a framework for thinking about debates over what constitutes credible and productive educational inquiry. In this final subsection, I want briefly to summarize what I take to be the central merit of disciplined educational inquiry over conceptions of educational science: the improved alignment and reflexivity of research in the service of a broad range of cognitive aims.

---

102 See for example National Research Council, "Scientific Research in Education."
I argued in Sections 2.2 and 2.3 that two kinds of misalignment can undermine research credibility. The first kind of misalignment is internal to the research itself and involves a mismatch between the methods and organization of a research community and the community’s implicit or explicit cognitive aims. Many feminist critiques of scientific practices, for example, can be understood as arguments for precisely this kind of misalignment.103 The second kind of misalignment is a mismatch between the supply of research and the demand for findings. This is the failure of research knowledge to match to the knowledge needs of policymakers and educators. I argued that two popular conceptions of educational research—as RFTs and as “scientifically based”—worked against such alignment. The problem in both approaches was an implicit devaluing of research that could help narrow the gap between the knowledge research produces and the needs of research consumers. The two misalignments can be related when the cognitive aims of a research community are implicitly or explicitly understood to have practical consequences. If a research community believes it is producing practically useful knowledge, but reflective research consumers point out that the knowledge being produced is not meeting their needs, the research is misaligned in both senses.

Scientists sometimes object that the second misalignment—the failure of research to meet the needs of policymakers—should have no bearing on the credibility of research, since relevance has nothing to do with epistemic merit. Truth and significance are different questions, and science is primarily concerned with the pursuit of truth, not social significance. As long as researchers are producing well-warranted knowledge, research should be considered credible. This is correct if credibility is understood in strictly epistemic terms. But many researchers also believe that their work is of significant social value and therefore deserving of public investment. This is an argument for the extrinsic value of research: scholarly inquiry helps society accomplish socially important goals. When research loses its credibility with policymakers and the public, it is often has to do with the credibility of research

to address socially important goals. Of course, research can also lose credibility for epistemic reasons, most notably when the public perceives it as beholden to particular ideological interests that interfere with its objectivity. But the public is not usually well-situated to assess epistemic credibility in this sense—though this does not prevent them from rendering judgments about the likely origins of the universe, the credibility of evolutionary theory, or the strength of the evidence for anthropogenic climate change.\textsuperscript{104}

In contrast with epistemic credibility, policymakers and the public have more direct access to the social significance of research. In the case of education, the failure of policymakers to recommend significant investment in educational R&D has as much to do with failures in social significance credibility as with the epistemic merit of the research. And for non-experts, perception of social significance can be understandably linked to perceptions of trustworthiness. In addition, the ability of research programs to produce socially significant results is often taken to be an indicator of epistemic warrant. When research influences the daily lives of people—something advances in engineering, information technology, and medical research have all accomplished—public perception of the credibility of research in these fields goes up.\textsuperscript{105} This helps explain the lack of public confidence suggested in Chapter 1.

There is a sense in which interdisciplinary research has the potential to further improve the likelihood of both senses of alignment described above. This reflexivity might be conceptualized as “distributed metacognition” or perhaps as Chris Argyris’s “double-loop learning” in organizations.\textsuperscript{106} In either case the idea is


\textsuperscript{105}Of course, such assessments can also be profoundly misguided. Two fields where public trust often goes beyond what the research warrants are nutrition science and forensic science. In these instances, institutions like The National Academy of Sciences can play a particularly important role guiding public policy. See for example: National Research Council, \textit{Strengthening Forensic Science in the United States: A Path Forward} (Washington DC: National Academies Press, 2009), \textit{———, Improving Data to Analyze Food and Nutrition Policies} (Washington DC: National Academies Press, 2005).

the same, and it goes to the reflexivity suggested in Kuhn’s observations about the reinterpretive or hermeneutic dimensions of the humanities and some social sciences. By incorporating more reflexive methodological approaches into disciplined inquiry, more explicit attention is likely to be paid to the cognitive aims of research and the relationship between these aims, the standards of evidence, and the normal criticism that occurs. This increases the likelihood of alignment and therefore of more useful educational R&D.

The framework I have sketched argues that shared cognitive aims, shared standards, and a disciplining community are necessary for scholarly rigor, accountability, and objectivity. Different cognitive aims often require different standards of evidence as well as somewhat different systems for scholarly interaction. I also questioned the practical value of attempting to demarcate scientific educational research from other forms of disciplined inquiry. The features that make educational inquiry “scientific” according to the WWC (e.g., randomized experiments) or SRE (e.g., direct evidence, replication, generalization) are not essential to all instances of good science. Experiments cannot answer all scientific questions. Natural experiments are scientifically invaluable, but they do not usually lend themselves to replication. Black holes and mental states are not directly observable, but the indirect evidence is sufficient to allow for reliable inferences. Even the best-designed laboratory study might not generalize to other studies or to contexts outside the lab.
CHAPTER 3

Does Educational Research Really Work in Pasteur’s Quadrant?
An Argument for Problem-Disciplined Inquiry

The sources of educational science are any portions of ascertained knowledge that enter into the heart, head and hands of educators, and which, by entering in, render the performance of educational function more enlightened, more humane, more truly educational than it was before. But there is no way to discover what is “more truly educational” except by the continuation of the educational act itself. The discovery is never made; it is always making. It may conduce to immediate ease or momentary efficiency to seek an answer for questions outside of education, in some material which already has scientific prestige. But such seeking is an abdication, a surrender. In the end, it only lessens the chances that education in actual operation will provide the materials for an improved science. It arrests growth; it prevents the thinking that is the final source of all progress.¹

-John Dewey, The Sources of a Science of Education

3.1 Introduction

Some of the best educational scholars are intellectual omnivores, skilled at foraging the seeds of new educational ideas in the lessons of neighboring disciplines. Where even these researchers can fall short is in the careful cultivation of ideas to the point of bearing fruit, as Dewey writes, “in the heart, head and hands of educators.” The field of education, it seems, has rather arid soil: many ideas survive, sometimes for decades, but few truly flourish. This chapter takes up an idea gaining greater traction in educational circles that promises not just to bear fruit, but to transform the field. The idea, put forward by Donald Stokes, has two parts: (1) an historical critique of the U.S.’s approach to science policy—a system that, he argues, mistakenly dichotomizes basic and applied research, and (2) a proposal for restructuring the federal funding of scientific research in the United States around the aim of “use-inspired” basic research.

Stokes’s first point is widely cited by educational researchers for a variety of purposes, but his second—arguably more important—point is often ignored. In the decade since the publication his book, *Pasteur’s Quadrant: Basic Science and Technological Innovation*, many educational scholars and reports have cited this work, often in influential forums. A 1999 National Academy of Education (NAE) advisory report, for example, proposes a framework for “collaborative problem-solving research and development” that they attribute to Stokes, describing this model as a case of “use-inspired” basic research.\(^2\) Similarly, the 2002 National Research Council (NRC) report, *Scientific Research in Education (SRE)*, referenced Stokes repeatedly to explain the practical origins of many important research questions and to recommend a “use-inspired” educational research portfolio “including a mix of fundamental science and applied questions; projects with short-, mid-, and long-term horizons; and a variety of research types and methods.”\(^3\) Most recently, the call for proposals to the American Educational Research Association’s (AERA) 2010 annual meeting noted that “Educational research sits inside what Donald Stokes calls ‘Pasteur’s Quadrant,’ referring to the dual focus of building basic theory while simultaneously improving practice.”\(^4\) This sampling reflects the growing interest in using Stokes’s work to explain and guide educational inquiry.

Unfortunately, many educational scholars who reference Stokes misunderstand his argument. This chapter critically assesses the relevance of Stokes’s ideas for educational inquiry, proposing a variation that I call *problem-disciplined research*.\(^5\) In what follows, Section 3.2 explains Stokes’s main ideas, and Section 3.3 examines several prominent examples of misunderstanding or misapplying his ideas to education, especially his notion of “use-inspired basic research.” Interpretations of

---


\(^5\) I borrow the label from a personal conversation with Tony Bryk, but the account of problem-disciplined research I offer is my own.
Stokes have become a Rorschach test, of sorts, for how researchers understand the relationship between research and practice: scholars tend to see what they want to see, citing Stokes as evidence for their view. Section 3.4 takes a closer look at the historical backdrop relating social science and social policy, and I consider several plausible definitions of “basic” and “applied” research and knowledge in educational inquiry. Finally, Section 3.5 argues that the difficulty of translating sensible versions of Stokes’s ideas to the social domain reflects an intellectual barrier to educational inquiry. In place of “use-inspired basic research” I propose the concept of problem-disciplined inquiry. This approach aims at generating general knowledge and understanding for use in practice as strategies for narrowing the gap between research and educational problems.

This chapter builds directly on Chapter 2, which argued for conceiving of educational research not as science but as disciplined inquiry. The argument of that chapter was that disciplined inquiry, unlike most conceptions of science, offers a strategy for avoiding at least some concerns about narrowly focused disciplinary research. By focusing on the alignment of criticism with standards of evidence and cognitive aims, I argued that disciplined inquiry is more adaptable to the problems of schools and schooling while (potentially) avoiding some of the unhelpful divisions of traditional scholarly inquiry. Yet Chapter 2 focused primarily on the social norms and practices of inquiry that support the production of credible research. Relatively little was said about the feasibility or desirability of particular cognitive aims—practical, theoretical, or otherwise. In contrast, this chapter is about the cognitive aims of educational inquiry and their central importance shaping the research process. Basic, applied, and use-inspired research all imply different aims for inquiry, as does understanding for use. Better understanding these labels and their limits helps narrow the gap between educational inquiry and the problems of schools and schooling. To begin, consider what Donald Stokes’s framework has to offer.
3.2 Stokes’s argument for “use-inspired” basic research

Stokes’s book, *Pasteur’s Quadrant*, begins with a detailed historical analysis of post-war science policy in the U.S. Its stated goal is to offer “a more realistic view of the relationship between basic science and technological innovation to frame science and technology policies for a new century.” In what follows, Subsection 3.2.1 reviews Stokes’s historical argument, introducing his quadrant typology of research goals. Subsection 3.2.2 reviews Stokes’s policy argument, considering several templates for federal investment in research focused on societal needs. This summary provides a basis for critiquing some of the ways educational scholars use Stokes’s ideas in Section 3.3.

3.2.1 Stokes’s historical argument. The crux of Stokes’s argument is overcoming what he sees as an outmoded view about the relationship between basic science and technological innovation. The old view is directly traceable to Vannevar Bush—an engineer, entrepreneur, and science policy adviser to President Franklin D. Roosevelt. Bush famously led the charge for federally sponsored basic science after World War II, producing a widely influential report, *Science: The Endless Frontier*, that later became the intellectual framework for creating the National Science Foundation (NSF). This report also helped popularize what Stokes describes as the “linear” relationship between basic science and technology. According to Bush:

> There is a perverse law governing research: under the pressure for immediate results, and unless deliberate policies are set up to guard against this, applied research invariably drives out pure. The moral is clear: it is pure research which deserves and requires special protection and specially assured support.

His point was that because the aims and incentives of applied research are relatively clear, direct, and of “practical or commercial value,” over time these goals lead

---

researchers away from the pursuit of fundamental insights that are only possible in the absence of such short-term concerns.

Throughout the report Bush slips back and forth between the labels “pure,” “fundamental,” and “basic” research, but all fit roughly the same definition:

Pure [fundamental or basic] research is research without specific practical ends. It results in general knowledge and understanding of nature and its laws. This general knowledge provides the means of answering a large number of important practical problems, though it may not give a specific solution to any one of them. The pure scientist may not be at all interested in the practical applications of his work; yet the development of important new industries depends primarily on a continuing vigorous progress of pure science. ³

Three aspects of Bush’s definition are worth noting. First, the definition reiterates the essential tension he saw between the pursuit of basic and applied research. The intellectual freedom required for the broad advancement of science depends on insulating pure research from the narrowing goals of “specific practical ends.” Second, the definition makes clear that the aim of basic science is a particular kind of knowledge, that is, “general knowledge and understanding of nature and its laws.” And third, Bush’s definition explains the public rationale for pursuing basic research. His argument for federally supported pure research is not based on the intrinsic value of such knowledge, but on the Enlightenment view that advances in basic scientific understanding will later improve the human condition.³ The introduction to Science: The Endless Frontier, for example, begins with a series of section headings: “Scientific Progress is Essential” followed by sections “For the War Against Disease,” “For Our National Security,” and “And for the Public Welfare.” Arguments are offered for the significant social benefits that accrue to each thanks to investments in pure research, and subsequent chapters extend these themes.³³ Bush notes, for example, that “A nation which depends upon others for its new basic scientific knowledge will be slow

³ Ibid., 75.
³³ Stokes, Pasteur’s Quadrant, 100.
³³³ Bush, “Science: The Endless Frontier.”
in its industrial progress and weak in its competitive position in world trade, regardless of its mechanical skill.” In fact, throughout the entire 184 page report, pure research is always justified in terms of its long-term social benefits rather than its intrinsic value.

In contrast, Stokes distinguishes between pure research (which matches the above definition) and basic research, which refers more broadly to any research that aims to produce what Bush labeled “general knowledge and understanding of nature and its laws.” This distinction allows Stokes to mark out research goals from research processes. Thus, rather than seeing basic knowledge and practical application as two ends of a continuum, Stokes treats them as independent goals—goals that can sometimes overlap. The result is a two by two matrix designed around axes based on each goal: “Considerations of use?” and “Quest for fundamental understanding?” (See Figure 3-1). Each quadrant is illustrated with a major figure in the history of science and technology: Niels Bohr for pure basic research (and the single goal of contributing basic knowledge), Thomas Edison for pure applied research (and the single goal of practical use), and Louis Pasteur for what Stokes labels “use-inspired basic research” (and the dual goals of basic knowledge and practical use). Each figure was chosen as an archetype of a different approach to pursuing the aim along each axis. And in each case, Stokes emphasizes it is the ex ante aims of research that determine the quadrant rather than ex post outcomes. Bohr’s quest to model atomic structure aimed purely at fundamental understanding without regard for potential applications of this knowledge. In contrast, Edison’s quest to make commercial electric lighting widely available aimed purely at applied ends, with no serious regard for contributions to fundamental knowledge.

---

11 Ibid., 14.
12 Like Bush, Stokes uses the terms “knowledge” and “understanding” more or less interchangeably. As I suggested in Chapter 1 and return to in Section 3.5, understanding and knowledge can be very different goals. This distinction bears on how we think about improving the connection between educational research and educational practice.
13 Stokes, Pasteur’s Quadrant, 78.
Research is inspired by:

<table>
<thead>
<tr>
<th>Quest for fundamental understanding?</th>
<th>Considerations of use?</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pure basic research (Bohr)</td>
<td>Use-inspired basic research (Pasteur)</td>
</tr>
<tr>
<td>Pure applied research (Edison)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3-1: Stokes’s Quadrant Model of Scientific Research

And between these ostensibly opposing approaches, Pasteur reflects a third archetype, one that successfully combined applied concerns with the pursuit of fundamental knowledge. Stokes describes Pasteur’s approach as use-inspired research because he was driven to make both fundamental and applied contributions—a drive that placed him in a quadrant on par with Edison’s pursuit of application and Bohr’s pursuit of fundamental understanding. In this regard, Stokes’s “use-inspired” label for basic research in Pasteur’s quadrant is a misnomer, since Pasteur himself was equally concerned with use. As Stokes observed, “the mature Pasteur never did a study that was not applied, as he laid out a whole new branch of science.”

An excellent biography by Patrice Debré reinforces this point with a quote from Pasteur, noting his truly revolutionary view of applied science: “There is no such thing as a special category of science called applied science; there is science and there are its applications, which are related to one another as the fruit is related to the tree that has borne it.”

And he was especially passionate about the curing of disease, as Debré writes: “As far as Pasteur was concerned, disease was more than a field of observation. It was also a field of battle. He had understood the terrible consequences of microbial infections, but he was not a physician and his knowledge

---

14 Ibid., 13.
of physiology and pharmacology was limited.” Obsessed with the problem of vaccination, “Pasteur engaged in pioneering work, setting up long series of experiments designed to penetrate the secret of how vaccination worked. After many attempts that led nowhere, it was his research on chicken cholera that put Pasteur on the right track.” Thus, at least in Pasteur’s case, “inspiration” does not do full justice to his commitment to use in practice.

The quadrant typology offers two historical challenges to Bush’s view. First, it challenges the conditions necessary for the production of basic scientific knowledge. On Bush’s account, the pursuit of basic knowledge depends in large part on intellectual freedom, and this freedom is fundamentally incompatible with significant concerns about research use. This is why “pure,” “fundamental,” and “basic” research are one and the same for Bush. On Stokes’s view, basic knowledge can be the product of pure research, but it can also result from inquiry strongly motivated by practical concerns. Second, the typology challenges the relationship between basic science and technological innovation. For Bush, these activities are necessarily separate, with basic science producing the fundamental knowledge that over time becomes a resource for technological innovation. For Stokes, basic science and technological innovation can go hand in hand, with the pursuit of new technologies opening new frontiers for basic science. This was certainly true in Pasteur’s case, and modern examples from computer science to biotechnology also support the possibility of reciprocal benefit.

3.2.2 Stokes’s policy argument. What I have summarized thus far is only the first half of Stokes’s argument. The quadrant typology was not intended as a mere historical excise; Stokes uses it as the groundwork for a substantive critique of U.S. science policy. His analysis points out two intellectual flaws in the rationale for funding pure scientific research: the belief that applied goals undermine the pursuit of basic knowledge and the belief that basic knowledge supports applied research in

---

16 Ibid., 378.
17 Ibid.
a linear fashion. Both insights have significant policy implications. He notes, for example, that:

In the setting of American democracy, a broad awareness of how deeply modern science is inspired by societal need is more likely to renew the compact between science and government than is a generalized promise of technological return from pure science.¹⁸

The argument against emphasizing pure science is entirely pragmatic: the end of the cold war, the integration of the world economy, and the new realities of the federal budgetary process make continuing public support for pure research very uncertain. Reviewing survey research on public support for federal investment in science, Stokes notes that “this research has . . . shown how deeply the public values science not for what it is but for what it’s for. It is strikingly clear that the instrumental uses of science are the key to popular support.”¹⁹ Even Vannevar Bush understood this important point, which is why the vast majority of his report focuses on the Enlightenment argument that basic research, over the long-term, works to improve the human condition.²⁰ Ultimately, Stokes concludes, “the inspiration that basic research can draw from societal need strengthens its claim on public support in the policy community and from the public to which it responds.”²¹

Beyond public support, Stokes makes a second important observation: emphasizing use-inspired basic research in most fields may sustain or even improve basic research because it “strengthens the case for supporting the pure research on which the development of the field partly depends.”²² With a few exceptions (e.g., particle physics) a shift to use-inspired basic research strengthens the case for pure basic research as well:

As the emergence of goal-oriented basic research within a scientific field strengthens the case for public investment, it also strengthens the case for

---

¹⁸ Emphasis original. Stokes, Pasteur’s Quadrant, 111.
¹⁹ Ibid., 98.
²⁰ Bush, "Science: The Endless Frontier."
²¹ Stokes, Pasteur’s Quadrant, 99.
²² Ibid., 104.
public investment in the pure research that will enhance the capacity of the field as a whole to meet the societal goals on which it bears.\textsuperscript{23}

Yet a third argument for supporting use-inspired basic research challenges those who argue that it is impossible to know what basic research will turn out to be relevant. Although there is uncertainty in all scientific endeavors, Stokes argues that “the uncertainty as to who will capture the benefit in technology from new scientific knowledge is lessened when basic research is directly influenced by potential use.”\textsuperscript{24}

To contend that there is always uncertainty about the outcomes, relevance, and usefulness of research does not entail that attempts to promote research aimed at addressing particular social problems is pointless or counterproductive. It is uncertain what research will ultimately produce a cure for HIV, but who can doubt that significant public and private investment in R&D aimed at finding a cure increases its likelihood?

Given that use-inspired basic research has two distinct goals—contributing to basic knowledge and contributing social value—how are these goals best achieved? Stokes considers this policy question at two levels, which he refers to as microallocative and macroallocative problems, respectively.\textsuperscript{25} The microallocative problem is deciding what research to fund at the project level. The macroallocative problem is deciding what research to fund at aggregate levels including, for example, the setting of wholesale federal research priorities.

For Stokes, these allocation problems take on different characters at each level. If allocation decisions have two valences, social value and scientific merit, scientific expertise is particularly valuable at the micro-level where other scientists have the expertise to gage the scientific merit of particular projects. In contrast, judgments about social value are “much easier to grasp than the theoretical or technical elements.” Putting the asymmetry more pointedly:

\textsuperscript{23} Ibid.
\textsuperscript{24} Ibid., 106.
\textsuperscript{25} Ibid., 113.
The scientist and the lay observer are on vastly different footing in judging the scientific promise of next research steps in understanding the properties of hemoglobin. But the protein chemist and his or her funders are on much more equal footing in knowing the value of synthesizing blood free of HIV virus in a world suffering from AIDS.26

Stokes qualifies his view that scientists are reasonably good judges of social value in two respects. One, he notes that although judgments about social value are, in general, much less dependent on scientific expertise, “the analysis of social value may in some cases involve considerable technical rigor.” And two, “nonscientists may have an advantage in the analysis of complex moral and ethical issues that is sometimes required to judge the value of a potential application of scientific knowledge.”27 (I will return to these important qualifications in Section 3.5.)

Based on this analysis, Stokes draws several conclusions about the microallocation of resources at the project level. First, “a system for appraising scientific promise and social value at the project level should enlist the insight of the working scientist into the nature of the social goals on which his or her research bears.”28 No specific mechanism is recommended, but explicit, extended discussion of research interpretations and implications for policy and practice are one possible approach.29 Second, Stokes suggests that “it is unwise to use a system of allocation that separates these two judgments [social value and scientific merit].”30 Underlying this recommendation is a concern with “losing the creative insight of the bench scientist”31—a concern that is in tension with his earlier claim that, unlike scientific merit, social value can be judged relatively easily by the lay public. And third, because of the relative ease with which bench scientists can “comprehend and accommodate” incentives for societal need, they “may be built into a system of

---

26 Ibid., 114.
27 Ibid., 115.
28 Ibid., 116.
29 Chapter 4 proposes an interpretive framework for educational inquiry that fits this recommendation.
30 Stokes, Pasteur’s Quadrant, 116.
31 Ibid.
funding research at the retail level. In short, peer review of grants at the project level can help ensure the success of use-inspired basic research so long as social value is given explicit consideration and weight in the review process.

When we move to macro level allocation decisions, Stokes argues that the roles of scientists and the policy community are transformed. However, he is far less clear about what this transformation looks like in theory or practice. At the micro level, Stokes contended that allocation decisions could be adequately handled without input from non-scientists. But allocation decisions at the wholesale level require more formal mechanisms for greater public input on the relative social value of different research priorities. Similarly, scientists need new mechanisms for offering collective judgments about scientific merit, since judgments at the aggregate level often extend beyond the expertise of any individual. Three existing models for federal research investment are reviewed: (i) attempts to extend NSF’s research mission beyond Bohr’s quadrant of pure basic research, (ii) attempts to encourage smaller, mission-oriented research agencies that had traditionally operated within Edison’s quadrant to begin investing in use-inspired basic research, and (iii) models like the National Institutes of Health (NIH) which Stokes believes already centers its investments in Pasteur’s quadrant. Yet in introducing these models he largely abandons any discussion of their macro allocation processes, supplanting detailed analysis with more anecdotal discussions of their successes producing useful basic knowledge. He notes historical efforts to broaden the NSF’s mission to include, for example, the engineering sciences, but he does not analyze the process by which some use-inspired basic research funds are allocated. Similarly, several successful efforts to extend the work of mission-oriented agencies into strategic areas of use-inspired basic research are reviewed (e.g., the department of agriculture, the department of defense, the department of energy, NASA), but their respective allocation processes are not examined. Even in the case of NIH, discussion of the macroallocation process is thin.

Ibid., 117.
Nonetheless, Stokes does make several interesting general assessments of the NIH model as a research strategy. He notes, for example, that NIH’s approach, while centered on research in Pasteur’s quadrant, also includes pure basic and pure applied research—a fact that results in a sort of “schizophrenia” among staff and investigators:

The human significance of biomedical knowledge since the time of the Hippocrates placed the center of gravity of this research squarely in Pasteur’s quadrant. But the evolving strategy of NIH coupled this with substantial investments in pure research and purely applied research as well. What emerged was a comprehensive investment strategy, unique in America’s experience, of research investments that included all three patterns of research goals but was clearly centered on use-inspired basic science, an institutional strategy that has led at times to a kind of schizophrenia among both NIH staff and principal investigators. In policy circles, they are apt to emphasize their Pasteur’s quadrant role, whereas in academic research circles, where the ideal of pure inquiry still burns brightly, they are apt to emphasize their Bohr’s quadrant credentials.34 Stokes also thought the NIH model was an interesting contrast to use-inspired basic research sponsored by the mission agencies. “The outlook of these two groups of agencies,” he observed, is “fundamentally different.”35 Extending Abraham Kaplan’s famous law of the instrument metaphor36 that our tools shape our perceptions of problems,37 Stokes argues that NIH is the business of “making a better hammer”

33 Although I have not done a detailed review, there are reasons to question how much applied research NIH is engaged in. For example, a recent NRC evaluation of the Small Business Innovation Research (SBIR) Program—the only major cross-departmental grant program dedicated technology transfer into the applied arena—noted that “The NIH is focused on the pursuit of fundamental knowledge to extend healthy life and reduce the burdens of illness for the nation’s citizens. Most NIH programs generally do not seek to develop products and services for the marketplace. The SBIR program does.” National Research Council, An Assessment of the Small Business Innovation Research Program at the National Institutes of Health (Washington DC: National Academy Press, 2009), 3.
34 Stokes, Pasteur’s Quadrant, 135.
37 As paraphrased by Abraham Maslow: “It is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail.” Abraham H. Maslow, The Psychology of Science, [1st ed., The John Dewey Society Lectureship Series, (New York: Harper & Row, 1966), 15.
while mission agencies are traditionally in the position of “driving nails.”

Nevertheless,

These mission-related objectives (the nails) provide a distinct perspective for the basic scientists whose work they fund, a perspective that leads at times to strikingly original research. The value of this fresh angle of vision extends the list of “standard” advantages of a plural system of research support over a monolithic system.  

Although Stokes’s insight is not especially clear in this passage or the surrounding text, the import seems to be that use-inspired basic research within mission-driven agencies is a valuable source of innovation for basic research. He concludes that the NIH model is preferable to “ratchet[ing] upward from 50 to 60 percent the proportion of NSF’s budget that supports ‘strategic research.’” Because attracting the funding necessary to replicate this model would be quite difficult, the next-best option is “to build agendas of use-inspired basic science, funded by agencies across government, that bear on the nation’s needs.” The proposal attempts to coordinate research and improve its social value around NIH-like models for use-inspired basic research.

Looking to other areas, Stokes considers research on the environment a prime area for comprehensive, NIH-like investment because of its social importance and dependence on many different scientific disciplines. Although Stokes offers other compelling examples from the natural, biological, and medical sciences, examples from the social sciences are few and far between. Consequently, it is unclear how education or the social sciences fit Pasteur’s model. Stokes does cite the work of economists John Maynard Keynes and Sir Arthur Lewis (and points to the field of demography) as examples of social research in Pasteur’s Quadrant. This uncertainty has not stopped educational scholars from making frequent use of Stokes ideas for a

38 Stokes, Pasteur’s Quadrant, 136.
39 Ibid., 151.
40 Ibid., 17-18.
variety of purposes, some of which are at odds with one another. The next section explores the use of Stokes’s ideas in educational inquiry.

3.3 The misapplication of Stokes’s ideas in educational inquiry

The call for submissions to AERA’s 2010 Annual Meeting begins: “Education research sits inside what Donald Stokes calls ‘Pasteur’s Quadrant,’ referring to the dual focus on building basic theory while simultaneously improving practice.” This is an optimistic assessment, at the very least, and it is a questionable interpretation of Stokes. Much educational research does not contribute in obvious ways to our basic knowledge of educational phenomena or to the improvement of schools and schooling, much less to both simultaneously. In this section I consider two key issues with respect to the use of Stokes in educational inquiry. First, I am concerned with fidelity to Stokes’s intent. Sometimes educational inquirers use Stokes’s terminology in ways that are explicitly at odds with his argument. Second, I am concerned with overstretching the relevance of Stokes, who wrote almost exclusively about the natural, biological, and medical sciences, to the social sciences and education. Subsection 3.3.1 examines educational scholars’ use of Stokes’s typology, especially the notion of use-inspired basic research. Subsection 3.3.2, examines appeals to Stokes’s larger policy argument. Both illustrate difficulties in applying Stokes’s argument to the social sciences and education—areas Stokes gave relatively little attention. In Section 3.4 I probe the meaning of basic and applied knowledge in the social sciences more deeply.

3.3.1 Misunderstanding “use-inspired” basic research. Consider a recent article in Educational Researcher that makes use of Stokes. In it, Anthony E. Kelly and Robert K. Yin argue for the increased use of structured abstracts in education journals as a way of improving research use by policymakers, educators, and other

---

42 For an excellent history of educational inquiry discussing these struggles, see Lagemann, An Elusive Science.
researchers. The main idea is that the “vast bulk” of educational research is argumentative, in the sense that it “make[s] knowledge claims relative to some social goal.” The authors recommend revised and expanded standards for structured abstracts from those recommended by Mosteller, Nave, and Miech in an earlier article.

Setting aside the pros and cons of expanding the use of structured abstracts in educational inquiry, I want to focus on the point the authors attempt to support by appealing to Stokes: the idea that the “stage” of a research study should weigh in assessing the confidence we place in it. In this section, Kelly and Yin’s concern is how the research design is accounted for in the structured abstract. They argue that in addition to describing the design, researchers should also indicate the “stage of the research investigation.” Here is the relevant excerpt:

Drawing from Stokes (1997), we recommend that authors indicate from what stage of research their claims emanate: (a) basic or exploratory research, with the goal of advancing theory or discovering fundamental processes; (b) movement from informed theory to informed practice; (c) use-inspired basic research (from practice to theory); (d) technological development, including instrumentation, devices, new tests, metrics, and instructional software (Stokes noted that “in some cases [basic science] existed only in the technology,” p. 21); (e) methodological development, including data collection techniques, protocols, and argumentative grammars; (f) analytical development, including emerging statistical and qualitative frames; (g) synthetic literature review; and (h) synthetic-analytic literature review (challenging prior literature reviews in the same area of study). Each of these goals for research has different knowledge demands. Authors’ confidence in their written claims should be bolstered by a defensible choice of methodological warrant pertinent to the stage of investigation.

46 Ibid.: 136.
There are at least two serious problems with the way the authors use Stokes’s ideas in this excerpt. First, they structure their argument around the idea that research happens in clearly identifiable “stages” that are distinct from the research design itself. Moreover, they contend that the confidence we have in a study should depend upon the stage of the inquiry. Yet Stokes’s proposed typology was designed precisely as a way to break free of stage-like imagery, which he believed served only to reinforce Bush’s faulty linear model with research proceeding from basic research to applied research and the development and commercialization of new technologies. Notice too that despite talk of stages, the categories do not progress in any obvious stage-like fashion. When the authors refer to use-inspired basic research, to a stage in what progression are they referring? And what is the epistemic connection between a research “stage” and the confidence we should place in it? Stokes’s decoupling of basic and applied research goals alleviated the tension between the two, such that the pursuit of applied goals could be understood as independent of the epistemic merit of any knowledge the gets produced.  

Second, Kelly and Yin’s definition of use-inspired basic research as research that moves “from practice to theory” does not match Stokes’s notion. Use-inspired basic research in the mode of Pasteur is basic research aimed at addressing a practical problem. This is very different from deriving basic knowledge from practice. Practice-to-theory implies a form of reverse engineering, action science, or systematic naturalistic inquiry. Nothing in Stokes’s writings suggests this is what he means by use-inspired basic research. In all his examples, use-inspired basic research is basic research that has direct or indirect bearing on a particular practical problem.

47 Note too that discussion of research aims is another part of the structured abstract. Kelly and Yin use Stokes specifically to make the “stages” claim, not a more general assertion about research aims.  
and is conducted with explicit considerations of use in mind. Only certain practical problems lend themselves to this mode of inquiry, which often entails the extraction of a particular problem aspect into a controlled laboratory setting where causes can be probed more deeply. The practice-to-theory approach is a promising variation on Stokes’s typology, but it is a variation that diverges from his actual proposal. (Section 3.5 offers extended discussion of this variation, which I believe better suits much educational inquiry for reasons discussed in Section 3.4.) Also recall that part of Stokes’s rationale for a use-inspired basic research program was that it could provide some basis for supporting relatively “pure” basic research, since the capacity and technology produced from pure research could reasonably transfer to use-inspired work. This simply does not fit the practice-to-theory template Kelly and Yin suggest.

Educational historian David Labaree has also recently interpreted use-inspired basic research in a way that diverges from Stokes’s intent. In response to Jacquelien A. Bulterman-Bos’s article, “Will a Clinical Approach Make Education Research More Relevant for Practice?” Labaree defends pure basic research against the suggestion that researchers place greater emphasis on educational practice and relevance. Drawing on the work of Michael Polanyi, Bulterman-Bos argues that educational researchers (and their research) would benefit from greater engagement with the practice of teaching. Such clinical experience would help bridge rather large divides between the work of researchers and the work of teachers. Labaree expresses serious concerns with Bulterman-Bos’s proposal, particularly the idea that combining the roles of researcher and clinician into clinical educational researchers would somehow address the problem. Better, he suggests, “to acknowledge and honor the different zones of expertise and to promote a fruitful dialogue between practitioners in the two zones.” Thus an “allocation of functions between teachers and researchers seems fruitful for both professions, as long as the barrier is relatively low

49 Jacquelien A. Bulterman-Bos, "Will a Clinical Approach Make Education Research More Relevant for Practice?," *Educational Researcher* 37, no. 7 (2008).

and the conversation across the barrier is ongoing.”\footnote{Ibid.} (Notably, Labaree offers no evidence that either condition actually obtains in education. In my judgment, neither condition is obviously met in most educational inquiry.)

It’s strange that Labaree chose to use Stokes’s ideas to make his argument, since he argues \textit{against} the push for greater relevance in educational inquiry. For example, Labaree suggests that:

\begin{quote}
It is counterproductive to press education research—or, for that matter, any other form of research—to be relevant. One problem is that relevance is a tricky quality to define because it is easier to recognize in retrospect than in prospect. A related problem is that earnest efforts to make research more relevant can paradoxically make it useless or even harmful, by focusing on short-term results that are narrowly measured instead of on consequences with a longer horizon and broader scope.\footnote{Ibid.}
\end{quote}

These considerations more or less mirror Bush’s concerns. Nevertheless, Labaree is quick to point out, “to argue against the press for relevance is not to say that education research should be irrelevant.”\footnote{Ibid.} Here Stokes is brought in to explain how researchers may draw “inspiration” from real world problems without forcing relevance. After explaining the typology in brief, Labaree posits that:

\begin{quote}
Scholarly research justifies itself primarily by its contribution to theory, sometimes inspired by immediate social needs (like Pasteur) and sometimes not (like Bohr), and this applies to a professional field like education as much as to a scholarly discipline. If we as educational researchers fail to contribute to theory with our research, then we are less scholars than engineers or product developers. . . .The distinctive value of scholarly research dissipates quickly when it segues from being use-inspired to use-driven. And this is what happens in the press for relevance.\footnote{Ibid.}
\end{quote}

Two key points underpin Labaree’s argument in these brief passages: (i) the idea that scholarly inquiry may be inspired, but not driven, by practical concerns and (ii) the idea that applied research is not concerned with the development of theory.
and is therefore an inappropriate scholarly pursuit. Each is worth considering in more detail. The first idea directly conflicts with Stokes’s core argument and many of his examples, including and especially the work of Pasteur. The implication is that Edison’s applied research was use-driven while Pasteur’s research was use-inspired (but decidedly not driven). But Stokes never makes this distinction, and for good reason. Perhaps most obviously, Pasteur was a use-driven researcher who was also passionate about advancing scientific knowledge. This is precisely what made him the third archetype, on par with Bohr’s contributions to basic knowledge and Edison’s contributions to use. Stokes also argues explicitly and at length that social value should be a core criterion in evaluating which use-inspired basic research projects get funding. (Not for all research projects, it should be noted, just the use-inspired ones.) Such a criterion provides researchers with more than “inspiration” to pursue particular lines of research. In fact, virtually nothing drives research programs like outside financial support since without it, most of research programs Stokes discusses could not exist. And Stokes even went so far as to recommend that researchers do more to demonstrate the relevance of their work to the public good.

The second idea, that applied research is not concerned with the development of theory, also contradicts Stokes’s understanding, though the point is not widely discussed in the book and is obscured by Stokes’s discussion of Edison as the archetype for “considerations of use.” Early on in his discussions of the American research experience, Stokes writes:

The applied fields, while they seemed to repeat the separation of basic from applied science, have in fact provided an institutional home for research that is driven by the goals of understanding and use. Similarly, institutions outside the universities, with Bell Labs the prototype, provided a home for research melding these goals.\[55\]

The goal of advancing understanding in a research context is simply not possible without theory of some form or another, a point that is equally true of basic and

\[55\] Stokes, *Pasteur’s Quadrant*, 45.
applied research. The problem with defining applied research as unconcerned with
timey is that good theories are extremely useful in practice. This is unsurprising,
since a supposed strength of a good theory is its ability to facilitate generalization
and replication across contexts. If the applied problem one wants to address is
teaching a particular child how to read, it might be possible to address this problem
well enough through trial and error, without any theory. But if one wanted to tackle a
larger problem—say, making sure every child in the U.S. could read by age seven—it
would be impossible to accomplish this goal without a robust array of theories. We
would need good theories not just on how children learn to read, but also on how to
better train teachers and parents how to teach children to read. We might also need
theories about how to incentivize states to adopt the necessary policies, as well as
theories about how to implement them effectively in very different contexts. We
might even need theories about how to integrate diverse theories into a single,
comprehensive model. Some of these theories already exist, but not all of them.
Under these circumstances, robust theory development becomes an essential part of
applied research.

Of course, a theory coming out of applied concerns may be relatively local in
scope, narrow in purpose, or private in character—all important qualifications. But
theory development can be and often is a central activity in applied research and
development. Much modern applied research fits this mold, from the advent of the
compact fluorescent light bulb at General Electric to the sequencing of the human
genome by the Celera Corporation to the development of better search algorithms at
Google. Finance and marketing textbooks are filled with applied theories, as are
textbooks from other professions ranging from medicine to culinary science. In each
of these cases, the development of new theories was instrumental to solving
practical problems, not in tension with them.

Unfortunately, the role of theory in applied research is obscured in Stokes
larger analysis because of his discussion and portrayal of Edison as the template for
pure applied work. He writes of Edison, for example, that he “drove his research
team to complete the development of commercially marketable electric lighting, exemplified the applied investigator wholly uninterested in the deeper scientific implication of his discoveries.” Edison’s reputation for being atheoretical in his approach is especially evident in a lively description of his work by his assistant and subsequent rival, Nikola Tesla, the day after his death:

His method was inefficient in the extreme, for an immense ground had to be covered to get anything at all unless blind chance intervened and, at first, I was almost a sorry witness of his doings, knowing that just a little theory and calculation would have saved him 90% of the labour. But he had a veritable contempt for book learning and mathematical knowledge, trusting himself entirely to his inventor’s instinct and practical American sense.57

But there are difficulties with this interpretation of Edison. Paul Israel, author of a recent biography of Edison, notes that these interpretations are more caricature than reality:

Edison’s notebooks show him frequently interrupting more practical work to explore unexplained phenomena. Although he was always quick to seize on any possibilities for application, his primary interest was in gaining experimental knowledge and even producing an explanatory hypothesis.58

Unfortunately, Stokes does not consider this perspective on Edison’s research. Moreover, the larger question of what role theory plays in different forms of applied research is never addressed in any detail.

Labaree and Kelly and Yin are representative of some common misunderstandings of Stokes’s use-inspired basic research and related concepts in education. But neither really focused on policy recommendations for the reorganization of educational inquiry. The next section considers three different policy recommendations that draw heavily on Stokes without taking adequate stock of the policy side of his argument.

56 Ibid., 24.
3.3.2 Misunderstanding Stokes’s policy argument. Consider several problems with the way Stokes’s policy argument has been interpreted by policymakers and the groups that advise them. Russ Whitehurst, the former director of IES (discussed at length in Chapters 1 and 2), used Stokes to draw a lesson opposite that of Labaree. Whitehurst argues that if the history of educational research teaches us anything, it is the need for more applied research in the mode of Edison:

Without in any way diminishing the value of basic research, whether use-inspired or not, I want to argue for the importance of activities in Edison’s quadrant, particularly for topics in which there is a large distance between what the world needs and what realistically can be expected to flow from basic research, and for topics in which problem solutions are richly multivariate and contextual. Education is such an area: a field in which there is a gulf between the bench and the trench, and in which the trench is complicated by many players, settings, and circumstances. Choose what you consider to be the most exciting developments from basic research in Bohrs’ or Pasteur’s quadrants that are relevant to education. I’ll pick developments in cognitive neuroscience. Paint the rosiest scenario you dare for basic scientific progress in the topic you’ve chosen over the next 15 years. Then ask yourself what would need to be done to translate those imagined findings into applications that would have wide and powerful effects on education outcomes. I don’t know about you, but I’m not optimistic that the results of basic research, even if the findings are powerful, will flow directly and naturally into education. Goodness! Education hasn’t even incorporated into instruction what we know from basic research about the effects of massed versus distributed practice - and I learned about that in a psychology course I took in 1962. . . In summary, the Institute’s statutory mission, as well as the conceptual model I’ve just outlined, points the Institute toward applied research, Edison’s quadrant.59

Here again I think Stokes’s use of the adjective “use-inspired” facilitates the misperception that basic research cannot doggedly pursue pressing social problems in practically useful ways. This is evident in Labaree’s dismissal of problem-driven research in a scholarly setting, and it is equally evident in Whitehurst’s suggestion

that basic research has contributed virtually nothing to educational policy and practice in recent decades. Yet the point of Stokes’s quadrant typology was to show that those pursuing basic knowledge (like Pasteur) could be just as committed to social impact as those pursuing purely applied research.

It is worth considering whether Whitehurst’s skepticism is justified even if he is misinterpreting Stokes. This depends, in part, on how Whitehurst understands the goals of applied research. Unfortunately, echoes of an atheoretical approach (a caricature of Edison) are evident when Whitehurst writes that “the people on the front lines of education do not want research minutia, or post-modern musings, or philosophy, or theory, or advocacy, or opinions from education researchers.”60 In Chapter 2 I argued that this atheoretical orientation towards “what works” led to an uncritical embracing of randomized experiments. Here I want to note a different aspect of Whitehurst’s claim, an understandable sympathy with potential consumers of educational knowledge—the educators and policymakers, parents and reporters—who want clear and direct answers to what appear, at least on the face of it, as straightforward questions. Does class size reduction improve student achievement? Does an afterschool program narrow the achievement gap? Do alternative credentialing programs attract high caliber professionals by lowering barriers to entering teaching? The problem, of course, is that the answers to these questions are almost always “it depends.” This is unsatisfying for all stakeholders. But it also points to the potential value of good theory, not just in research but in practice. Theory can offer a cognitive framework for improved practice, a point I return to in Sections 3.4 and 3.5. (This is one important difference between the natural sciences and the social sciences.) But the larger point is that research efforts directed solely at Edison’s quadrant—at least as discussed by Stokes—will not lead to the kinds of practically useful theories necessary for widespread improvement of educational practice.

Stokes’s ideas are also present in the policy recommendations of Scientific Research in Education, the NRC report discussed in Chapter 2. Early on the authors

---

60 Ibid.
write that they “believe the distinction between basic and applied science has outlived its usefulness.”61 Their concern is that different research aims have taken on connotations about the rigor or quality of inquiry, when “what makes research scientific is not the motive for carrying it out, but the manner in which it is carried out.”62 A bit later they note more directly that “Stokes’s model clearly applies to research in education, where problems of practice and policy provide a rich source for important—and often highly fundamental in character—research questions.”63

Thus far the authors seem to argue, in line with Stokes, for educational investment in Pasteur’s quadrant. There is some underlying confusion about basic and applied knowledge, probably for the reasons discussed above. (Stokes gave no indication, for example, that the distinction between basic and applied research should be abandoned. But the authors real concern is placing epistemic weight on this difference—a point on which Stokes is silent.) Similarly the authors write that education’s “research portfolio should be use-inspired, including a mix of fundamental science and applied questions; projects with short-, mid-, and long-term horizons; and a variety of research types and methods.”64 Thus, “a program of research focused on [improving student achievement] might support short-term syntheses of what is known, mid-term evaluations of promising programs, and long-term studies of the acquisition and development of scientific competence.”65

And yet, rather inexplicably, when the authors turn to recommendations for reorganizing educational inquiry, their enthusiasm for comprehensive, use-inspired basic research diminishes. Specifically, they suggest that:

The research function of a federal education research agency should be organizationally separate from an educational improvement mission, leaving the latter to a parallel entity with its own budget. A measure of bureaucratic distance between these two functions is also desirable because it would be

62 Ibid.
63 Ibid., 58.
64 Ibid., 143.
65 Ibid.
difficult to develop a common culture in an education research agency given the appropriate differences between research and program administration.\textsuperscript{66} This is, in effect, a reiteration of Bush’s view, not Stokes’s, as recommendation cuts against what Stokes described as the productive “schizophrenia” among NIH staff and researchers, between the basic research and practical improvement rationales for their use-inspired work. And it was made under the banner of insulating “the agency from inappropriate political interference,” yet this framing seems to insulate the research function more from problems of education improvement than from political interference. Recall that under Whitehurst’s leadership, the What Works Clearinghouse (which aggregated research findings) was separate from the Department of Education’s funding of particular programs that (according to the WWC) were deemed scientifically based. But as described in Chapter 2, this did not insulate the WWC from political interference.\textsuperscript{67} The deeper problem is the concern that “a school improvement agenda can overwhelm the agency’s fiscal and intellectual capacity to focus on its core research mission.”\textsuperscript{68} This is an important practical difficulty, but it is not an epistemic argument against the possibility of use-inspired basic research; on the contrary, it is a compelling argument for greater federal investment.\textsuperscript{69}

Both Whitehurst and SRE seem largely to accept the status quo when it comes to the structure of federal R&D efforts, even if they disagree about what a properly balanced investment looks like. There is, however, at least one major report that has proposed a redesign of federal support for educational R&D. In 1999, the National Academy of Education (NAE) produced an advisory report for the Department of Education recommending a radical reorganization of educational R&D funding around Stokes’s framework. The report recommended that:

\begin{itemize}
  \item \textsuperscript{66} Ibid., 142.
  \item \textsuperscript{67} See for example Schoenfeld, "What Doesn't Work."
  \item \textsuperscript{68} National Research Council, "Scientific Research in Education," 141.
  \item \textsuperscript{69} The NIH Clinical Center, for example, is the nation’s largest hospital devoted to clinical research.
\end{itemize}
To progress toward an infrastructure that can support a more productive relation between research and educational practice, members of the study group recommend that the Department of Education initiate a long-term effort to develop support of collaborative educational research and improvement based on a new model of the relationship between these activities. The model, articulated by Donald Stokes in *Pasteur's Quadrant*, is fundamentally different from the model of research, development, dissemination, and evaluation (RDDE) as a mainly linear process. Instead, Stokes argued (with many supporting examples) that inquiry and invention can be categorized as low or high in their potential for advancing understanding of general explanatory principles and as low or high in their potential for betterment of a social problem.

The NAE study group recommends that OERI [now IES] begin an important new program of research in “Pasteur’s quadrant,” which should be focused explicitly on solving specific current problems of practice and at the same time should be accountable for developing and testing general principles of education that can be expected to apply broadly beyond the particular places in which the research is done.70

While I am generally in agreement with the recommendations of the NAE advisory report, my interest here is examining how accurately the NAE proposal reflects the use-inspired basic research model Stokes envisioned. Two potential differences stand out. First, the authors substitute “general explanatory principles” for basic knowledge. In places Stokes implies this conception, but it is not a particularly good contrast with applied knowledge and elsewhere he points to fundamental causes and the accumulation of knowledge as defining features. However, given the nature and constraints of the social sciences, this is a reasonable adaptation. (Section 3.4 discusses this issue at length.)

Second, the authors argue for research that is accountable for contributions to use. This too is a key idea discussed by Stokes that many overlook. However, the NAE report goes much further than Stokes when they describe how this accountability works:

---

The central idea in the study group’s proposal is to develop a system of support for projects in which professional researchers and professional educators share in the accountability for achieving success in improving educational practices and outcomes. These projects may also include program developers, curriculum specialists, or policy specialists. All the participants share in a commitment to and accountability for multiple outcomes of the work. . . . The recommendation’s novelty is in proposing that commitment to these goals and accountability for them should be shared by all of the participants in the project, rather than being separate responsibilities of distinct communities of educational practitioners, developers, and researchers working on separate projects.\footnote{Ibid., 9-10.}

This is indeed an important novelty in the committee’s recommendations, a novelty not present in Stokes’s writings. It is also an approach that differs from all three of Stokes models for moving more federal investment into Pasteur’s quadrant, and it leaves open the question of how “general explanatory principles” might result from this mode of inquiry.

Thus, at most, the NAE report elaborates on a Stokes-inspired template that only loosely resembles his actual policy proposals. But this is no criticism: as the next two sections make clear, there are important limitations to Stokes’s framework, especially as it might apply to educational inquiry.

### 3.4 Basic and applied knowledge in the social sciences and education

Notable differences between educational inquiry and the natural sciences have led scholars like David Berliner to label educational research the “hardest science of all,”\footnote{David C. Berliner, "Educational Research: The Hardest Science of All," \textit{Educational Researcher} 31, no. 8 (2002).} Ellen Lagemann to describe education as a field in search of “an elusive science,”\footnote{Lagemann, \textit{An Elusive Science}.} and David Labaree to suggest educational researchers must learn to live “with a lesser form of knowledge.”\footnote{David F. Labaree, "Educational Researchers: Living with a Lesser Form of Knowledge," \textit{Educational Researcher} 27, no. 8 (1998).} Each characterization acknowledges
differences between studying natural phenomena and studying educational phenomena. In this section I explain what some of these differences mean for the pursuit of use-inspired basic research in education. Many educational researchers aspire to the intellectual contributions of their disciplinary brethren in psychology, anthropology, sociology, economics, etc. Others take the opposite tack, turning away from the disciplines and pursing more applied lines of inquiry—work grounded in educational practice but often disconnected from social theory. Both approaches fall short in the quest to bring educational research to bear on pressing educational problems. This gap has to do, in part, with the different knowledge aims of basic and applied research, as well as with the different mechanisms available in the natural sciences, the social sciences, and education for translating knowledge into action.

Section 3.3 raised several concerns about the way Stokes’s ideas have been interpreted and translated into policy recommendations in education. Of course, imprudent readings of Pasteur’s Quadrant are one source of misunderstanding, but there are also ambiguities and tensions within Stokes’s framework that need clarification and explication. This section examines the aims of basic, applied, and use-inspired research more carefully, considering the challenges of applying each concept to the social sciences. Subsection 3.4.1 reviews several important debates over basic and applied research in the social sciences and the use of research in social policy. Subsections 3.4.2 and 3.4.3 propose definitions of basic and applied knowledge, considering the difficulties in applying these definitions to the social sciences and education.

3.4.1 Historical background. When the NSF was founded in 1950, its first director, Alan Waterman, shared Bush’s conviction that it should focus entirely on pure basic research. Notably, this emphasis was one reason some argued the social sciences should be excluded from the NSF. As Dael Wolfle notes in an early history of the NSF:
It was clear from the beginning that physical and biological sciences would be included [in the NSF’s funding priorities], but whether or not to include the social sciences was a major issue. Proponents contended that it was at least as important to advance the social sciences as the other branches of science and that inclusion would aid their development. Opposition was centered on one argument and one type of confusion. The argument ran that the social sciences were not so highly developed as other branches of science and that the ability of social scientists to solve important social problems was to be doubted. The confusion was between research on human problems and the practical control of human affairs. This confusion was illustrated in some of the debate when salesmen, legislators, and other practical manipulators of social affairs were identified as social scientists.75

Legislation ultimately gave NSF the freedom to decide whether or not to fund the social sciences, and some early grants did go to “anthropological and related sciences.”76 Wolfle points to two different sources of skepticism about basic social science research: a belief that the field was too immature to support or warrant significant basic research efforts and a belief that social scientists were hard to discern from those skilled in social affairs. More than a half century on the situation is quite different. Although the social sciences are still not funded on par with the natural, biological, or engineering sciences, they are now a recognized part of NSF’s research portfolio.77 Similarly, the social sciences are now well established within the academy, and at least two fields, economics and psychology, have large, well-established professional communities beyond the walls of academia. On most counts, the social sciences have flourished over the last half century.

75 Dael Wolfle, "National Science Foundation: The First Six Years," Science 126, no. 3269 (1957): 337. Wolfle was trained as an experimental psychologist but worked for many years on science policy, eventually becoming president of the American Association for the Advancement of Science.
76 Ibid.
77 Nonetheless, as recently as October, 2009, Senator Tom Coburn introduced legislation (Coburn Amendment 2631) to prohibit the National Science Foundation from “wasting federal research funding on political science projects. . .siphon[ing] resources away from research that promises greater scientific discoveries with real world benefits,” and he was critical of NSF support for social science generally.

And yet today federal investment in the social sciences remains largely uncoordinated. As the first head of NSF, Waterman shared the conviction that the NSF should not adopt a serious policymaking or coordination role “based on his experience with government research, as well as his contact with university and government scientists.” Instead, “he believed the foundation should focus on supporting research inappropriate for other agencies and should not be involved in evaluating the progress of other agencies.” Thus, one consequence of insulating basic science from practical concerns was, from the very beginning, a decentralized and largely uncoordinated approach to federal science policy (in general) and social science (in particular). Daniel Kleinman concludes in his 1995 study of post-war science policy that up until the beginning of the Clinton Administration there remained little centralization, coordination, or planning among different government departments and agencies. Although Stokes rejected centralization—especially in the form of a single “Department of Science”—the thrust of his policy argument calls for greater coordination among various federal agencies around important problems like energy and the environment. Such coordination would provide the framework for a portfolio of research centered on Pasteur’s quadrant. Yet coordinated investment has yet to happen in education.

One reason for the lack of coordination is longstanding uncertainty about the goals, prospects, and audience of educational inquiry, a concern that goes back to the founding of the first American schools of education. The inaugural issue of *Educational Review* in 1891 had several articles on the state of education theory and research at the time. Of particular note, philosopher Josiah Royce argued in an essay titled “Is There a Science of Education?” that “there is no universally valid science of pedagogy that is capable of any complete formulation, and of direct application to

---


individual pupils and teachers.” In fact, he continues, there will never be one because of the contingencies of particular educational contexts. Consequently, “the educator’s calling will be an art to whose beauty and complexity no science will be adequate.”

Nevertheless, Royce continues, educational inquiry should not be abandoned. Rather, a different relationship between educational science and educational practice is envisioned, one where the value of educational inquiry is more enlightening than instrumental:

To the educator we in effect say: “Work against the chaos of impulses, by using the very impulses themselves as the material for good order. In a word, organize.” Meanwhile, although the actual content of any attempted organization of life will be “historically determined,” and so imperfect and transient, relatively general accounts can be given of processes that do increase the orderliness of the life of the child.

For Royce (as for Wilhelm Dilthey, whose work he draws upon at length), the “general accounts” often take the form of “pedagogical rules.” These rules form the basis of a “scientific pedagogy.”

Far from telling the teacher finally and completely just what human nature is, and must be, and just what to do with it, will be limited to pointing out what does, on the whole tend toward good order and toward the organization of impulses into character. “This is the whole province of pedagogy,” as a general science. . . . Universal these rules would be, yet never universal in so far as they were precise guides in the concrete case. Aids they would be, but never substitutes for personal insight. In short, pedagogy, as a “science,” would be a good staff and a bad crutch.

Royce’s account resembles that of his colleague and contemporary, William James, who wrote similarly of the role of psychology in his *Talks to Teachers*. Royce nicely

---

81 Ibid.: 19.
82 Ibid.: 20.
83 James, *Talks to Teachers*. 

110
sums up the practical constraint of scientific theory when it comes to teaching: “Just what science abstracts from and ignores, just that you now most need to know.”

In the 1900s, the history of educational research reached another crucial decision point with John Dewey’s educational research program at the University of Chicago. This was all the more momentous for the fact that the fledgling social sciences, especially psychology and sociology, were also at a formative stage in their development. Dewey’s approach to educational inquiry is best captured by two facts about his short time at Chicago. First, much like Pasteur, Dewey was committed to pursuing socially relevant inquiry that engaged the problems of the day. This commitment was embodied in Dewey’s laboratory school, which served as a setting for observing otherwise abstract philosophical concepts up close and for attempting to put Dewey’s philosophical ideas into practice. Unlike traditional research psychology laboratories, the school offered an opportunity for naturalistic experimentation that could contribute both to our basic understanding of schools and schooling as well as “to create new standards and ideals and thus to lead to a gradual change in conditions.” Thus, educational historian Ellen Lagemann suggests that a “fundamental element” of Dewey’s research program was the idea “that educational research should serve as a testing ground and link between scientific and social innovation.” Second, Dewey was committed to crossing institutional boundaries. This was evident not just in the creation and operation of the Laboratory School, but also Dewey’s close relationship other faculty in psychology (e.g., George Herbert Mead, James R. Angell) philosophy (e.g., James H. Tufts, Addison W. Moore) and sociology (e.g., Albion W. Small, George E. Vincent).

Unfortunately, Dewey’s work at Chicago came to an abrupt end in 1904 when conflicts with the administration led Dewey to leave for Columbia University. While he had periodic contact with Teachers College, his primary home was the philosophy department and

---

84 Royce, "Is There a Science of Education?," 21.
86 Ibid.: 197-98.
87 Ibid.
his focus shifted from “reconstruction in education” to “reconstruction in philosophy.”

Dewey’s departure also signaled a new direction for educational research at the University of Chicago, where the psychologist Charles Hubbard Judd took over as chair of the education department. Judd shared much in common with his contemporary and former colleague Edward L. Thorndike, one of the most influential educational psychologists of the twentieth century. As Lagemann notes, “Whereas Dewey had hoped to find in education a scientific method for philosophy, Judd hoped to impose psychological assumptions and methods on the study of education in order to develop a distinct science of the field.” Although Judd was less well-known, he shared many similar perspectives with Thorndike. Judd also becoming a major professional and political force for a very different kind of educational inquiry—one that eschewed naturalistic inquiry in favor of more narrowly prescribed psychological studies in controlled settings. For Judd, as for Thorndike, schools were places for implementing laboratory findings rather than sites for educational discovery. Much of the history since Dewey left Chicago reflects Lagemann’s conclusion: in the rivalry for future of educational inquiry, “Edward L. Thorndike won and John Dewey lost.”

Interestingly, some of the most important recent contributions of research to educational policy and practice have been indirect. In his 1988 AERA presidential address, Rich Shavelson argued that the greatest contributions of research to education policy and practice are often “in constructing, challenging, or changing the way policymakers and practitioners think about problems.” Shavelson’s view reflects what Carol Weiss calls the “enlightenment” function of social science research. If this account is correct, “the major effect of research on policy may be the gradual sedimentation of insights, theories, concepts, and ways of looking at the

---

88 Quoting historian Harold Rugg’s description of Dewey’s professional transformation Ibid.: 204.
89 Ibid.: 205.
90 Ibid.: 184.
In contrast, rational decision-maker models like those proposed by Stokey and Zeckhauser and social engineering models like those proposed by Lindblom and Campbell describe much more direct influence for the social sciences in social policymaking. Note too that each of these models walks a fine line between describing the influence of social science research on policy (and methods in policymaking) and recommending a particular relationship. Weiss, for example, concludes that “at this point in their development, ‘enlightenment’ may be the wisest use of the social science.” This assessment may have as much to do with Weiss’s understanding of policymaking as with the state social science. She observes that:

On the research side, much of what goes by the name of social science knowledge is flawed, inconclusive, ambiguous, and contradicted by evidence from other studies. Many research conclusions are limited in scope or out of date. Ignoring such data may be a responsible stance for a decision maker to take. On the policy side, there are a host of competing claims for attention. The policymaking process is a political process, with the basic aim of reconciling interest in order to negotiate a consensus, not of implementing logic and truth. The value issues in policymaking cannot be settled by referring to research findings. As for the lack of fit between what decision makers want to know and what researchers can tell them—this is a chronic lament. The problematic factors are concreteness, specificity, representativeness, timeliness, and prediction of future conditions.

---

96 Weiss, "Research for Policy's Sake," 532-33.
Weiss’s enlightenment view is common among social scientists and educational researchers. It fits a straightforward template dividing cognitive labor between those who study problems of varying degrees of social relevance and those charged with making policy decisions or taking practical action.

On the face of it, the enlightenment view also seems to imply the very tension Stokes wanted to overcome, between making fundamental knowledge contributions and directly contributing to the solving of social problems. But where Bush wanted to promote basic science by protecting it from the short-sightedness of applied research, Weiss is concerned with promoting a well-functioning political process and buffering it from a narrowly instrumental understanding of social science. One view of the relationship between social science and policy, Weiss observes, is that:

Researchers are obliged to take off from policymakers’ specifications of what the problem is, what the goals are, and what alternative means are feasible for moving toward the goals. If the social scientist wishes to change the definition of the problem or broaden the scope of options, he is enjoined to win the policymakers’ assent to the formulation before he begins his research. To the extent that he departs from the goals and assumptions adhered to by policymakers, his research will be irrelevant to the “real world” and will go unheeded. This is the conventional wisdom: the social researcher whose work is to enter the policy sphere should research consensus with some important segment of policy actors on the basic value-orientation of his work. For maximum research utility, the researcher should accept the fundamental goals, priorities, and political constraints of the key decision-making group. He should be sensitive to feasibilities and stay within the narrow range of low-cost, low-change policy alternatives.97

But while Weiss accepts the division of labor between knowledge producers and political actors, the enlightenment model rejects the above account, which asserts that researchers must tailor their work to the needs, priorities, and value judgments of policymakers or practitioners. Thus, the distinction between enlightenment and engineering approaches is not analogous to the distinction between basic and

97 Ibid.: 544.
applied research. In fact, Weiss adopts the labels from the sociologist Morris Janowitz who (citing Edward Shils) notes that “the enlightenment model sees no strong distinction between basic and applied sociology.”

Rather, the intended distinction is between producing knowledge and taking action. The heart of Weiss’s argument is that the intellectual freedom of the enlightenment model can be extremely beneficial to policymakers even when the value commitments implicit or explicit in the research conflict with those of decision makers:

The enlightenment model of research use . . . does not consider value consensus a prerequisite for useful research. It sees a role for research as social criticism. It finds a place for research based on variant theoretical premises. It implies that research need not necessarily be geared to the operating feasibilities of today, but that research provides the intellectual background of concepts, orientations, and empirical generalizations that inform policy. As new concepts and data emerge, their gradual cumulative effect can be to change the conventions policymakers abide by and to reorder the goals and priorities of the practical policy world.

Weiss goes on to present empirical evidence that policymakers actually valued research that encouraged them to think differently about policy problems.

This historical background suggests closer consideration of this second distinction, the relationship between knowledge and action. In the next two subsections I consider different definitions of basic and applied knowledge. This builds to more detailed consideration of the relationship between knowledge and action in Section 3.5.

3.4.2 Defining basic knowledge. Before assessing the educational relevance of Stokes’s argument for “use-inspired” basic research, it is helpful to consider what constitutes basic knowledge (the presumed aim of basic research) and what

---


99 Weiss, "Research for Policy's Sake," 544.
constitutes applied knowledge (the presumed aim of applied research). The relationship, it turns out, is not as straightforward as it appears.

Recall the similarities between Bush’s and Stokes’s conceptions of basic research. Bush thought basic research “results in general knowledge and an understanding of nature and its laws.” Similarly, Stokes believed that:

The defining quality of basic research is that it seeks to widen the understanding of the phenomena of a scientific field. Although basic research has been defined in many ways and involves . . . extraordinarily varied steps . . . its essential, defining property is the contribution it seeks to make to the general explanatory body of knowledge within an area of science. In keeping with this conception, the Organization for Economic Cooperation and Development defines basic research as “experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts”

Stokes’s definition of basic research lumps together several distinct ideas: understanding, explanation, accumulation, and investigation of the “underlying foundation of phenomena.” Of these, the last two are most distinctive of basic research. Applied research involves degrees of understanding and explanation, since both are necessary for practical application. But applied research need not continue to probe more deeply or widely. (For example, it can rest content with finding that a treatment caused an effective without delving into how the treatment caused the effect.) Nor must applied research make its findings public, submitting to the scrutiny of the larger academic community. So when Stokes goes on to write that “the goal of basic research is, in a word, understanding and of applied research use,” I take him to mean public understanding that is sufficiently general and that can accumulate into a lasting body of knowledge.

---

101 Stokes, Pasteur's Quadrant, 7.
102 Ibid., 8.
At least two kinds of knowledge fit this notion of basic: knowledge of fundamental (causal) mechanisms and knowledge of general laws. Fundamental causal mechanisms are basic in a sense that general causal knowledge need not be. The key difference is how deeply a process is understood. For example, a pharmaceutical company can know a drug causes certain effects (good or bad) without knowing the biochemical pathways that actually produce the effects. The drug development process often begins with at least a rough theory of how a particular chemical might treat a particular disease. Over time, models are developed and refined that attempt to explain with progressively greater depth the connection between a drug and its effects. But sometimes drugs produce known effects when very little is known about the mechanism. For example, lithium has been used to treat depression for over a century and has been approved by the FDA since 1970, yet very little is known about the mechanism by which lithium affects depression. Simply knowing that a drug produces certain effects is causal knowledge akin to what a black box randomized experiment may tell us in education. But it would be strange to call this knowledge “basic” or “fundamental” without some understanding of the underlying mechanisms.

A potential problem with this definition is that it seems to imply a notion of “rock bottom” causal mechanisms. If researchers only dig deep enough, they will eventually discover fundamental causal mechanisms and make a “basic” discovery. On this conception, particle physics would be more basic than chemistry, chemistry more basic than biology, and so on. In practice, what scientists and their funders usually mean by basic research (in the causal sense) is much broader. Lee Cronbach summarizes Mackie’s account of causal knowledge this way:

Progress in causal knowledge consists partly in arriving gradually at fuller formulations. As some of the gaps are reduced, the uncertainties represented

---

103 I intentionally use the term ‘knowledge’ here rather than ‘understanding.’ The importance of this distinction will be made clear in Section 3.5.
104 Research on the mechanism by which lithium affects the behavior of mice is still being investigated. See for example Galila Agam et al., "Knockout Mice in Understanding the Mechanism of Action of Lithium," Biochemical Society Transactions 37, no. 5 (2009).
by X become correspondingly less. . . . The approach, then, is a start on the long road of “progressive localization of a cause” (Mackie, 1974, p. 73).\footnote{Lee J. Cronbach, \textit{Designing Evaluations of Educational and Social Programs}, 1st ed. (San Francisco: Jossey-Bass, 1982), 139-40.}

This captures a plausible criterion for basic science in Stuke’s sense: a research program is basic if it pursues the “progressive localization of a cause.” Note that in this context, localization refers to the precision with which a cause is identified and not to causal mechanisms that only operate “locally” rather than generally. (Of course, under certain circumstances increased precision may lead to local causal explanations rather than general ones, but this was not Mackie’s point.) On this view, the knowledge coming out of such a research program is basic if it pushes the frontier of what we know about a particular causal mechanism. The genius of Pasteur’s swan neck flask experiment, for example, was the elegance with which it isolated a plausible cause of contamination: airborne microorganisms. Subsequent work in microbiology has continued to localize the cause(s) of particular cellular processes, investigating the function each plays. This research is basic not because it has hit bottom, but because it continues to dig deeper. In principle, research on any natural phenomenon can be basic if the goal is progressively deeper (more precisely localized) causal knowledge. What this looks like in practice depends on the phenomenon being investigated, but the probing of causes is basic science because it is general (causal understanding of a type, not just a particular instance) and because it lays the groundwork for subsequent investigations.

Laws of nature are less well understood by scientists and philosophers, though the concept remains widely associated with basic science. Daniel Little summarizes the natural science view of laws and their role in scientific explanation this way:

The successes of the natural sciences have given natural scientists confidence that natural systems operate in accordance with a strict set of laws, that these laws may be given precise mathematical formulation, that they derive from the underlying real properties of constituent physical entities, and, finally,
that these facts entail that the future behavior of physical systems is in principle (though perhaps not in practice) predictable. And this conception of the nature of physical systems in turn gave rise to a paradigm of scientific explanation: to explain a phenomenon is to derive the explanandum from a set of general laws and a description of the initial conditions of the system. Prediction and explanation go hand in hand, and both depend on the availability of empirically supportable general laws.¹⁰⁶

On this account natural laws are not just true generalizations, they also “govern,” meaning the generalizations are not true by accident. Consider this example by Bas C. van Fraassen comparing two very two similar generalizations:

1. All solid spheres of enriched uranium (U235) have a diameter of less than one mile.
2. All solid spheres of gold (Au) have a diameter of less than one mile.¹⁰⁷

The first generalization, van Fraassen allows, is a putative law of nature. The critical mass of uranium means a nuclear chain reaction would occur before a solid sphere of uranium a mile wide was ever produced. The second generalization seems to be an accidental or incidental fact of nature. A mile wide solid gold sphere will probably never exist, but for reasons that have little to do with laws of nature.

Note that unlike causal understanding, where the knowledge is “basic” because of progressively deeper probing (localization of a cause), a law of nature is basic not because it is fundamental (in a “rock bottom” sense) or because it is universal (like the generalization about gold spheres), but because it is necessary.

When a law of nature is accepted as part of an explanation, it is this necessity that does the work. To try and explain the absence of solid gold spheres greater than a mile wide by appealing to the generalization that all gold solid spheres must have a diameter of less than a mile is not persuasive because the fact is not necessary. We know it is not necessary in the case of gold (but is necessary in the case of uranium) because of the larger body of scientific knowledge we have accumulated. Causal

¹⁰⁶ Little, "On the Scope and Limits of Generalizations in the Social Sciences."
knowledge helps us probe which generalizations might be natural laws and which are spatiotemporally constrained generalizations.\textsuperscript{108}

Yet, to really warrant the claim that a proposition is a law of nature, we often have to know something about causation—something that helps explain why the law is necessary. Consequently, it is harder to make sense of natural laws in biology or the social sciences, where causality is harder to probe. One important reason for this, as Alexander Rosenberg has argued in the case of biology, is that some sciences are fundamentally historical:

Its [biology’s] subject matter is almost exclusively what has happened to biological systems on this planet over the past 3.5 billion years of its almost 5-billion-year history. From molecular biology to paleobiology, and including all the compartments of the subject in between, the explanatory problems of biology are all those of explaining either particular events (e.g., the Cretaceous extinction, the effectiveness of antiretrovirals for HIV-A) or historical patterns of a longer or shorter scale (the 1:1 male:female ratio among sexually reproducing species, the persistence of sickle-cell disease for 1,000 years in west Africa). By contrast, a nonhistorical science—chemistry or physics—is one whose kind of terms, laws, and theories do not make essential reference to particular places or times in the history of the universe.\textsuperscript{109}

In Chapter 2 I discussed causal explanations at some length, likening it to identifying links in a causal chain. Rosenberg’s point is that eventually, if we try and probe far enough back along biology’s causal chain, we are bound to have to deal with questions about why particular structures evolved. In this sense, spatiotemporal constraints are built into biology, since these constraints help explain why certain molecular components survived and others did not. But if there is an element of historical contingency built into the explanation of some biological facts, it is hard to envision many “necessary” biological laws. In fact, Rosenberg takes the law of natural

\textsuperscript{108} Some philosophers of science, including van Fraassen, Ronald Giere, and Stephen Mumford, do not believe any natural laws exist in the technical sense that scientists and philosophers intend. I’m inclined to agree, but this technical debate over the realism of natural laws is tangential to my argument about the social sciences.

selection to be the only biological law, since it alone can explain the historical trajectory of biological systems. Hence, Dobzhansky's dictum: “Nothing in biology makes sense except in the light of evolution.”

Natural laws and causal explanations in the social sciences face additional difficulties because of human cognition and intentionality. Since social phenomena have biological, cognitive, and environmental dimensions, social generalizations tend to have lots of ceteris paribus clauses. Consequently, social scientists do not typically pursue natural laws in the above sense. Instead, they focus on the progressive refining of generalizations. That is, they attempt to identify more general regularities requiring progressively weaker ceteris paribus clauses and that, ideally, fit better with other existing generalizations. But because the element of necessity is missing, it is strange to think of this knowledge as basic, particularly when broader generalizations are developed at the price of predictive or explanatory power.

Practically speaking, it matters a lot whether a generalization is necessarily or accidentally true. Education and the rest of the social sciences have struggled to offer reliable generalizations not just because social phenomena are complex, but because social phenomena are continually changing. This recognition explains why, as Cronbach put it, social generalizations “decay:”

At one time a conclusion describes the existing situation well, at a later time it accounts for rather little variance, and ultimately it is valid only as history. The half-life of an empirical proposition may be great or small. The more open a system, the shorter the half-life of relations within it are likely to be.

This often makes it difficult, if not impossible, decouple social phenomena from their context, and it helps explain the difficulty of translating carefully controlled laboratory studies of psychological phenomena into real world applications.

---

Consequently, an important part of social theory is a clear articulation of the conditions under which the theory is expected to hold. Yet this constraint on social generalizations, combined with the limits of generating fundamental causal knowledge, explains why it is often much harder to distinguish basic and applied knowledge in the social sciences and education in terms of the kinds of knowledge each aims at.

**3.4.3 Defining applied knowledge.** What does applied research aim at that distinguishes it from basic research? Recall that for Bush, applied research was the bridge between basic knowledge and the more complete knowledge necessary to address practical problems. Stokes is much less clear about the nature of applied knowledge. This is partly a function of his quadrant typology (Figure 3-1), which had “basic knowledge” on the y-axis but “considerations of use”—not “applied knowledge”—on the x-axis. Consequently, Stokes focused on the actual application of Edison’s work rather than the applied knowledge it might have produced. Similarly, Stokes avoided addressing the nature of applied knowledge in Pasteur’s quadrant by defining this work as basic research of the use-inspired variety. This creates an analytical gap in Stokes’s framework between basic knowledge and other kinds of knowledge. This gap is partly to blame for inconsistent applications of his ideas, at least within educational inquiry.

Contrasting conceptions of “applied”—as a kind of knowledge versus the usefulness of knowledge—go to the heart of the difference between Bush and Stokes. By focusing on the commercialization of Edison’s research on the production and transmission of electricity, Stokes ignored the theoretical development that is central to much contemporary applied research. Modern commercial research labs are far more sophisticated than Edison’s in several important respects. First, few contemporary laboratory directors have the disdain for basic research that Stokes projected onto Edison. On the contrary, basic research often fuels technological

---

113 For a more balanced view of Edison’s approach to research and innovation, see Israel, *Edison: A Life of Invention.*
innovation, with many private labs now forming close partnerships with university researchers. If anything, the present concern is often that these public-private partnerships are too close for scholarly comfort.\textsuperscript{114} Second, the researchers and engineers now employed by labs in both sectors have far more extensive scientific training than employees in Edison’s day. Third, private research and development efforts compete with a speed and sophistication inconceivable in Edison’s day. To be competitive, modern research labs must use resources efficiently and expeditiously, and basic scientific knowledge serves as a valuable guide. Despite these differences, Stokes situates modern applied research in Edison’s quadrant, noting that “A great deal of modern research that belongs in this category is extremely sophisticated, although narrowly targeted on immediate applied goals.”\textsuperscript{115} The same characteristics that make Edison an archetype for “considerations of use” also make him a strange representative for contemporary applied research. Ultimately Stokes is not clear about either the purpose or meaning of applied inquiry, probably because “use” is not a specific goal but a collection of many disparate goals. It includes the development and commercialization of particular technologies, but it can also refer to research that supports interpretations, decisions, and actions in a wide range of extra-scientific contexts.

This ambiguity presents additional difficulties for the social sciences and educational research, which often want to understand something in a local environment, sometimes for a particular purpose. This introduces some of the historical contingency noted by both Rosenberg and Cronbach in the previous section. It also helps explain why Carol Weiss’s enlightenment model focuses not on differences between basic and applied research, but on the distinction between knowledge and action—research and “social engineering.” While Weiss’s framework may help educational scholars better understand how policymakers utilize research,

\textsuperscript{114} For an excellent account of this trend see Eyal Press and Jennifer Washburn, "The Kept University," \textit{The Atlantic Monthly}, March 2000.
\textsuperscript{115} Stokes, \textit{Pasteur’s Quadrant}, 74.
it provides little guidance to educational scholars who want to pursue research that better meets the needs of schools and schooling. As Israel Scheffler once wrote in regards to a curriculum’s relevance to a child’s life:

The theoretical problem, with relevance as with virtue, is to say in what it consists and why, thus specified, it ought to be pursued. Relevance is, in particular, not an absolute property; nothing is either relevant or irrelevant in and of itself. Relevant to what, how, and why?—that is the question. That is, at any rate, the question if the current demand for relevance is to be taken not merely as a fashionable slogan but as a serious educational doctrine.\(^{116}\)

But applied educational research requires more than educational relevance to be truly engaged with educational problems and oriented towards action. Most forms of research—from the most abstract philosophical reflections to the most sophisticated statistical analyses—can claim some relevance, making this a weak criterion for describing research as applied. In this loose sense research is relevant whenever a plausible link exists between a research conclusion and an educational problem. But relevant research is not useful unless actual educational actions or decisions might have been made differently barring such research. Of course there are lots of reasons research aimed at influencing actions and decisions falls short, and many of these reasons have nothing to do with the research itself. But the purpose of applied research is to try and be produce general explanatory knowledge that not just conforms to the evidence but is also useful.

So at a minimum, useful research requires a plausible link between a research conclusion and an educational decision. Many different actors make educational decisions: parents and children, teachers and administrators, policymakers and researchers. Some of these decisions are relatively trivial, while others direct the spending of billions of dollars and shape the opportunities available to tens of millions of people. Without a connection to educational decisions, research can still hold significant academic interest; however, at least in the near-term, it will not be

particularly useful to actors or stakeholders trying to improve educational outcomes. Accordingly, educational research is *more useful* when it influences actions and decisions that have comparatively greater impact on learners.

In the case of education, basic and applied knowledge exist largely in the form of general causal explanations at the frontier of what we know about a particular educational phenomenon (e.g., language acquisition, learning to read, effective teaching). These causal frontiers can change with the unit of analysis, the advancement of educational inquiry, and the evolving contexts in which inquiry is conducted. Given these sources of variation, educational inquiry tends to aim at general explanatory principles that are also delimited enough to be empirically testable (Robert Merton called this “middle-range” theory.) Many different factors can effect whether explanatory principles get labeled “basic” or “applied,” but the difference typically has more to do with the theory’s purpose or use rather than with an epistemic difference. This helps explain the tendency to deemphasize the distinction between basic and applied knowledge in the social sciences, with greater emphasis on the link between knowledge and practice. This is a distinguishing feature of problem-disciplined educational inquiry, the focus of the next section.

Before turning to the next section, here is a brief recap of the key distinctions developed in this section. The historical overview in Subsection 3.4.1 pointed to the uncertain relationship between education (and social science) and improving schools (and social policy). Two theories of how research influences practice were discussed. The first was a social engineering approach, where researchers are the “methodological servants,” to use Campbell’s memorable phrase, of the public (a public Campbell hoped was open to small scale experimentation and therefore an evolving approach to public policy). The second was what Carol Weiss and Morris Janowitz describe as an enlightenment approach, where the research is free to evolve and seep into the consciousness of policymakers and practitioners. In Subsection 3.4.2 I considered two definitions of basic knowledge: as the progressive

---

117 Merton, *On Sociological Theories of the Middle Range.*
localization of causes (increasingly robust knowledge of causal mechanisms) and as laws of nature (generalizations that are necessarily true). Both, I argued, had limited transferability to the social sciences and education, where robust causal mechanisms are few and far between, and where few (if any) necessary generalizations seem to hold true. Finally, in this subsection I pointed out that applied knowledge is a concept Stokes writes virtually nothing about, in contrast to the usefulness of much basic research which he discussed at length. This presents a serious hurdle to making sense of educational inquiry in Pasteur’s quadrant, since educational theories generally do not fit the template of traditional basic research. Instead, educational inquiry tends to aim at middle-range theories, constrained by the many contingencies of social phenomena, yet general enough to still have some explanatory power under certain circumstances. These theories are more useful when there is a plausible link between this knowledge and practical action. The next section suggests a strategy for strengthening this link.

3.5 An argument for problem-disciplined educational inquiry

This section makes the case for problem-disciplined educational inquiry as a promising approach to narrowing the gap between research and educational practice. Subsection 3.5.1 assesses the limits of use-inspired basic research as a central strategy for addressing educational problems. Subsection 3.5.2 introduces problem-disciplined inquiry as an alternative to research in Pasteur’s quadrant. I argue this approach can offer general knowledge while improving education policy and practice by aiming at the production of understanding for use. Finally, Subsection 3.5.3 offers three examples of problem-disciplined educational inquiry: design experiments, improvement science, and policy research.

3.5.1 The limits of use-inspired basic research in education. Given the accounts of basic and applied knowledge in Section 3.4, what are the prospects for a use-inspired basic research program in education and what could such a program reasonably be expected to accomplish? To address these questions, consider several
minimum requirements for a use-inspired basic research program. The requirements fall into two categories, one internal to inquiry and one external: requirements having to do with the current state of knowledge and requirements having to do with the social context of inquiry.

For the first category, a program must meet two internal requirements for basic social science research: there must be some established body of knowledge on which to build, and there must be a methodology for developing and assessing better general explanatory theories. The requirement that some established body of knowledge exists indicates a commitment to building upon/engaging with what has come before. There is no requirement that the research agree with this literature, it helps define the “frontier,” the cognitive domain, and the intellectual community through which knowledge is generated. The methodology requirement ensures that there is a strategy for advancing this frontier. Traditionally this had meant that the phenomenon under investigation lends itself to controlled experimentation, often in a laboratory setting. Controlled contexts facilitate systematic investigation of particular aspects of social phenomena that are often harder to isolate in more natural contexts. Alternately, some social science disciplines (e.g., economics) will employ strong assumptions in their analyses (e.g., rationality of preferences, perfect information) to facilitate theory development and testing in less controlled settings.

For the second category, a use-inspired basic research program requires a social context where particular external conditions are met. If, as I have argued, actual use is a weak constraint for research in Pasteur’s quadrant—that is, if mere inspiration and considerations of use are, in fact, sufficient—then the only external requirement is the existence of a social problem to which the research is plausibly relevant. The problem with this weak condition, especially in the social sciences, is that almost all research programs are plausibly relevant to some social problem. The same can even be said of much pure research in the natural sciences. For these reasons, classifying research solely on the basis of stated goals is insufficient unless these goals are accompanied with a plausible plan for achieving them. Much
contemporary scientific research, probably most, now claims to pursue both basic and applied ends. Large research grants frequently require principal investigators to support their funding requests with justifications demonstrating the possible applications of research knowledge. Under these circumstances there are strong incentives to claim both mantels. But if use-inspired research does not offer a plausible theory of action for how research findings will get used, it should be unsurprising if even the most promising findings lay dormant.

If the use requirement is strengthened, mere relevance is no longer sufficient and considerations of use become defining features of the problem. As research moves from relevance to usefulness, it also increasingly depends on the larger social context in which it takes place. This context helps frame the problem and determines whether or not external resources and mechanisms exist for acting on new knowledge. (Note that problem consensus is less important than problem clarity, since different research programs can pursue different problem framings, but a poorly framed problem will not readily lend itself to disciplined inquiry.) Context also determines whether or not an adequate social mechanism exists for addressing the problem. For some social problems the mechanism by which use-inspired knowledge gets translated into practice is straightforward. There are, for example, regulatory mechanisms and consumer protection agencies for putting into practice new knowledge that exposure to a household cleaner is hazardous to human health. But in other cases problems may emerge for which no existing infrastructure is entirely adequate. Global warming research is a good example, because the causes of global warming are diverse and politically contentious, the research is complex, the negative effects seem distant, and the responsibilities diffuse.

The absence or inadequacy of a mechanism for applying new knowledge can present researchers engaged in use-inspired research with a dilemma when the research context is inadequate to making good use of research findings. Pasteur’s research is revered because of its contributions to our basic biological understanding and to the curing of disease. But this dual contribution was only partly within
Pasteur’s control. Luck also played a role, as it does in much scientific inquiry, and Pasteur could easily have found himself in the position of many contemporary scientists who make a discovery and must decide whether to devote time and energy to applying or commercializing knowledge or to continuing the pursuit of fundamental causal mechanisms and general explanations. Stokes never offers an explanation of the link between fundamental causal understanding and use. A research program that shares both goals equally might not be typified by Pasteur if the context was different and the mechanisms for effective use were not in place. Implicit in Stokes’s argument for use-inspired basic research is a model for how basic knowledge gets applied in technical and medical fields. Fundamental causal knowledge is applied when it facilitates the production of useful products or practical control. Stokes assumes that by virtue of being use-inspired, basic knowledge is more likely to facilitate application. And by virtue of being basic, such research still contributes to our fundamental understanding of the world.

Concerns about the usefulness of research and the constraints of the immediate context must also be balanced with considerations about the benefits of inquiry with longer time-horizons. But longer time horizons do not require a strong distinction between basic and applied research, nor do they preclude strong considerations of use in practice. More salient to the problems of educational inquiry is the appropriate division of labor between the work of education schools and valuable disciplinary work in related fields. In the remainder of this subsection, I consider two examples that explore the challenges of educational inquiry and possible divisions of labor with related fields: research on effective teacher preparation and research on mitigating the effects of developmental and biological disruptions early in a child’s life.

*Research on effective teacher preparation.* Darling-Hammond, Bransford, and LePage, have developed recommendations for improving teacher preparation around four different bodies of educational literature (see Figure 3-2). Note that if all educational research fell into the base of the triangle—into basic research in the
learning sciences—it would be easy to understand Whitehurst’s skepticism about new basic knowledge getting quickly or directly translating into benefits for educational policy and practice. This knowledge can take a long time to produce, and even then requires additional effort and funding for translating the research into effective educational practices, developing strategies for influencing practice on a large scale, and implementing with sufficient fidelity and flexibility to work in a wide range of contexts. In some cases knowledge at the higher levels of the triangle might pay educational dividends more quickly than basic research in the learning sciences. And conceivably every step in this process, and every level in the triangle, could be the focus of use-inspired research aimed developing general explanations or principles.

![Figure 3-2: Research Bases Supporting Teacher Education Recommendations](image)

Research at all four levels falls within the traditional domain of educational inquiry. Research at the bottom of the triangle also currently goes on in psychology departments across the country, but the top three levels exist almost exclusively

---

within education schools. (This is not to say that very relevant research is not
happening outside education schools, for example, in sociology departments,
business schools, etc) Nevertheless, research in the top levels has not been well
funded, well coordinated, or well implemented in actual teacher preparation
programs. For example, much research on teacher education has developed
independently of research on teaching, though scholars are working to bridge this
gap. Pam Grossman and Morva McDonald argue for a new research agenda focused
on teacher preparation that includes (i) developing a clear analytical framework for
describing, analyzing, and improving teaching, (ii) expanding research beyond the
cognitive demands of teaching to include also its craft, affect, and relational
dimensions, and (iii) understanding how state mandates and local labor markets
influence teacher education and how programs can be more responsive to changing
conditions and policies.¹¹⁹

Thus, research on teaching has struggled to meet both the internal and
external criteria for use-inspired basic research. Internally, research on teacher
preparation is still in its infancy, lacking a common language, common problems, or a
promising, well-defined methodology for moving the field forward. Externally,
research on teacher preparation has not had many obvious mechanisms for
influencing teacher preparation programs, though this is beginning to change. If use-
inspired basic research is the core strategy for improving teacher preparation, I share
Whitehurst’s concern that another fifteen years could go by without any significant
impact on practice.

A second example of the limits of use-inspired basic research is new work
linking childhood poverty and adult health. A scientific consensus is emerging in the
fields of neuroscience and molecular biology that developmental and biological
disruptions in early childhood have lasting health consequences into adulthood.

¹¹⁹ Grossman and McDonald, "Back to the Future: Directions for Research in Teaching and Teacher
Education."
Recently the authors of a major review of this research explicitly linked the priority of new basic research to the priority of developing new intervention programs:

Much work remains to be done to elucidate the precise causal mechanisms that explain these linkages [between early adversity and subsequent health, learning, and behavior]. The identification of biomarkers of toxic stress and its physiological consequences offers particular promise as a source of short- and medium-term measures to assess the mediators of outcomes that require decades to confirm. In a parallel fashion, the design and implementation of new approaches to both the prevention and treatment of toxic stress and its consequences, beginning in the early childhood years, must be another key priority. For example, testing new community-based interventions or clinical treatments for preschoolers who have been abused or seriously neglected ought to be at least as high a research priority as conducting clinical trials of statins for school-aged children with elevated cholesterol levels. Focusing on access problems and differential treatment in the health care system is certainly important, but confronting the early childhood origins of disparities in physical and mental health may offer far greater return on investment.\(^\text{120}\)

The authors conclude that there is a clear and compelling need for research aimed at developing and implementing educational interventions. At the same time, they underscore a critical limitation of basic research: more fundamental knowledge about causal mechanisms does not necessarily tell us how to produce different effects. And yet, *the point of understanding the impact of early childhood influence on adult health outcomes is to be able to improve those outcomes*. Here again there is reason for skepticism that use-inspired basic research in education is sufficient to the task of tackling pressing educational problems. The usefulness of all research, basic or applied, depends on either eventual use or in helping advance other basic knowledge that eventually gets used.

Currently there are relatively few scholars attempting research that effectively bridges the gap between knowledge and practical use. The most obvious reason is that this research is exceedingly difficult and requires a skill set that many

educational researchers lack. It requires strong organizational skills to manage and scale up a project. It requires strong interpersonal skills to navigate complex relationships and coordinate large groups of people. It requires savvy political skills to persuade funders, partners, and other stakeholders to participate and remain involved when things inevitably go wrong. And it can require academic sacrifices for the chance at larger scale social benefit, since all of the above activities necessarily constrain the traditional activities of publishing and teaching.

Faced with these challenges, educational researchers and their funders have a number of choices. One possibility is to continue down the path Russ Whitehurst argued for, attempting to fill in a gap in applied research by focusing relentlessly on the needs of policymakers and practitioners. In Chapter 2 I argued that this approach is a mistake if it does not also include the robust development of general explanations and theories. Alternately, the educational community might follow a path akin to that recommended by Shirley Brice Heath and reinvigorate basic research programs in education:

Many journals and spokespersons in education. . . .seem to practice lexical avoidance behavior with regard to the adjective *basic*. It is time to bring the term back into the vocabulary of education research. Basic research in any field must stand apart in its fundamental procedures, standards, and audience from those on which journalism, advocacy, policy analysis, evaluation, and public service depend. This is not to say that it cannot and should not be of use to the latter groups. On the contrary, decisions, recommendations, and applications of these groups stand to carry more weight and merit when linked to basic research. When basic research figures in action, elements of consistency, fairness, and predictability increase for those who are pursuing careers that are embedded in practice, policy, service, and philanthropy. We need . . . to rechart old territories and explore new ones carefully, find new ways of preparing travelers to go there, and know that those who follow will
be the better surveyors, architects, and builders for the charting work of our basic research.\textsuperscript{121}

For reasons already discussed, I believe this approach is also unlikely to lead to the knowledge necessary to improve schools and schooling. Educational scholars have both a professional and civic responsibility to help address these larger educational problems. While this in no way precludes laboratory research or experiments or fundamental investigations into educational practice, it does call for a Stokes-like reorganization of educational R&D. The focus, I contend, should not be use-inspired basic research but problem-disciplined inquiry.

\textbf{3.5.2 Problem-disciplined educational inquiry.} In Chapter 2 I avoided discussions of basic or applied research, focusing instead on what makes educational inquiry credible and, to a lesser extent, productive. Credibility involved epistemic acceptability and depended on a community’s norms of inquiry. Productivity was understood in two ways. It could also be understood epistemically, as the result of alignment between a community’s cognitive aims, standards of evidence, and norms of criticism. Or it could be understood practically, as the result of alignment between the cognitive aims of a disciplining community and the needs of the public. My argument was that educational research is best conceived of as disciplined inquiry: a form of collective problem-solving that produces public accounts of the problem(s) and epistemically acceptable solution(s) at which it aims. In this section I argue that the cognitive aims of educational inquirers are most likely to match the needs of children, teachers, principals, and policymakers when the research is \textit{problem-disciplined}. And the aim of problem-disciplined educational inquiry is \textit{understanding for use} in practice. Both concepts need unpacking.

The concept of problem-disciplined inquiry is straightforward: it involves addressing a practical problem as the cognitive aim of inquiry. Whereas much (probably most) scholarly inquiry adopts a purely academic problem as its focus,

problem-disciplined educational inquiry adopts practical educational problems as its focus. Because the standards of evidence for problem-disciplined inquiry derive from the problem constraints, it is particularly important that the problem is explicitly stated and well-framed. At the same time, problem-framing is an important source of innovation, since new framings may lead to otherwise overlooked solutions.

Problem-disciplined inquiry differs from use-inspired basic research in two important respects. First, as discussed in Section 3.4, rather than categorizing knowledge as basic or applied, it aims at the development and refinement of general explanatory principles or theories. Second, whereas use is a weak constraint for work in Pasteur’s quadrant, it is a strong constraint for problem-disciplined research. The strong interest in research use also changes the character of the problems, which now must also involve consideration of plausible mechanisms for affecting social change. Yet a focus on research use does not entail that researchers become social engineers, at least in the sense Lindblom, Campbell, or Weiss envisioned.

The label “understanding” rather than “knowledge” is also intentional, and it goes to the strategy for narrowing the gap between knowledge and understanding, moving research from relevance to usefulness. First, understanding implies a purpose that knowledge often lacks. I can know that I got from point A to point B without understanding the fact. And what counts as understanding will depend on my purpose: Am I a lost tourist? An amnesiac trying to orient myself? A mechanical engineering trying to understand the new electric vehicle she’s test driving? Accordingly, understanding is also something one or more persons have, whereas knowledge often refers to the contents of a book or article. As academics well know, two articles can contain the same knowledge but one can be much better at producing understanding than the other. Which article results in better understanding can depend as much on who is doing the reading as how the article is written. In this sense understanding is specific to individuals in a way knowledge need not be. This distinction further illuminates the practical gulf that can exist between knowing that X causes Y and understanding how to use this knowledge.
Some domains of research are well-aligned with institutions that can put basic findings to use, so a basic discovery in microbiology can be translated into the mass production of a new vaccine. Because other research domains lack this institutional capacity, the prospects for use-inspired basic research are significantly diminished.

This focus on understanding for use draws researchers’ attention to the aim of social change. The goal is not to produce what Dewey famously described as “cold storage knowledge”—knowledge disconnected from experience and use. The goal is to produce knowledge about how to produce change, and this necessarily involves shifts in understanding. A short digression: consider for a moment an important difference between problem-disciplined inquiry in education and much innovative engineering research that could plausibly fall under the same banner. In many areas of engineering, knowledge becomes embodied in artifacts or technologies that do not require significant understanding on the part of the user to be quite useful. My iPhone, which I understand virtually nothing about, is one of my most useful possessions. A new injection molding technology and the latest microchip operate in clearly defined ways under clearly defined parameters. So long as the user knows the function and the parameters, she has all the understanding necessary for successful use in most contexts. In education, very few advances have such clearly defined functions or parameters. Consider the educational technology of standardized tests—specifically, the SAT and the ACT, which are widely used assess students and predict their college success. These tests are very carefully designed and are taken by over two million high school students a year. And yet they are regularly used in ways they were not intended; for example, to compare state-by-state performances.122 The parameters for successful use of educational knowledge and technology are often either so well-bounded as to be of limited use or so vaguely bounded as to

122 This is a mistake for a number of reasons, the most important of which the more students who take the test, the lower a state’s average score is likely to be. 
invite misinterpretation. So their effective, widespread use requires an understanding that traditional technological advances in the engineering sciences typically do not. This is why the goal of understanding for use is so important in the social sciences and education.

Research that pursues understanding for use prioritizes clear and direct contributions to the solving of practical problems rather than mere inspiration or considerations of use. Since addressing social problems often requires the coordination of many different actors and stakeholders, the audience for this research is necessarily much broader than academia. Researchers will need to collaborate with these stakeholders to understand the problems they face as well as possible solutions. Equally important, unlike many areas of research where the researchers themselves may be the first to discover something, the field of education has over a hundred thousand laboratories (our schools) with well over a million experts (our teachers), many of whom have tacit knowledge and cognitive skills that educational researchers do not know about or understand. This opens the door to a “practice to theory” model for producing general knowledge. (Bear in mind that research on teacher cognition did not really get underway until the 1970s.)

It might be objected that “research for use,” on my account, is no different than applied research. If what we really need is more applied educational inquiry, why not make this argument, rather than reconceiving use-inspired basic research? There are several important differences between problem-disciplined inquiry as I conceive of it and most applied research. First, my approach is explicitly concerned with solving social problems, while applied research can be much more parochial in outlook. In a similar vein, problem-disciplined inquiry is also concerned with the development of general explanations and theories as a core strategy for social problem-solving. Applied research need not adopt this goal. Second, much applied research happens behind closed doors in industry labs or consulting firms. In contrast, problem-disciplined research is committed to the public sharing of knowledge as a necessary step towards developing understanding for use.
Furthermore, as discussed at length in Chapter 2, the status of general explanations as knowledge—that is, as epistemically acceptable—depends on their publicity. This constraint also presents special ethical challenges. I cannot go into these difficulties here, but some concerns are addressed in Chapter 4. Third, problem-disciplined inquiry typically draws on three sources of expertise when tackling a problem: the scholarly literature, the wisdom of practitioners, and learning by doing. Because applied research is often more limited in scope and purpose, it need not draw on all these data sources.

Despite the field’s best efforts, structural differences in the pursuit of educational knowledge have limited the ability of researchers to uncover fundamental mechanisms and apply research findings as pervasively as the natural and biological sciences. One set of structural differences has to do with the different ways natural, social, and educational phenomena are constituted. Problems of complexity and indeterminacy, as well as definition and interpretation, set educational inquiry apart from the natural sciences. Another set of differences has to do with how respective types of knowledge get translated into use. The current US educational R&D system lacks much of the infrastructure necessary to put the best educational knowledge into practice. Furthermore, many important educational advances require widespread changes in popular understanding. My fear is that much educational research currently pursues knowledge that is incidental to understanding, gesturing at contributions to basic knowledge or practical use but successfully contributing to neither. The knowledge this research produces may be valued because it helps professional advancement or prestige within a small subfield, but it fails to seriously advance educational research or meet the needs of those wrestling most directly and personally with the problems of our educational system. Nonetheless, there have been some exciting new developments in educational inquiry in the last few decades that fit the template of problem-disciplined inquiry I have just sketch.
3.5.3 Three examples of problem-disciplined educational inquiry. The preceding discussion will benefit from several examples of problem-disciplined educational inquiry. In this subsection I take up three examples indicating a several different approaches to problem-disciplined inquiry: design studies, improvement science, and policy research. Each approach adopts a different unit of analysis and implies a somewhat different audience. Nevertheless, the approaches can be (and often are) complementary, if not interdependent. For each I explain the kinds of problems the approach is meant to address, the general knowledge or explanatory principles it can generate, and the way in which understanding for use is produced—sometimes as part of the general knowledge goals and sometimes independently.

Design studies. One genera of problem-disciplined research is the educational design experiment. The aim of a design study is creation of an instructional artifact, program, or environment. Educational design studies became popular in the last few decades thanks to the work of educational psychologists like Ann Brown, Allan Collins, and Erik De Corte who wanted, much like John Dewey, to study teaching and learning processes in real-world learning environments.\(^\text{123}\) Collins identifies at least seven differences between the traditional psychological methodology (first) and the emerging (and evolving) methodology of design experiments (second): (1) laboratory settings versus messy situations, (2) a single dependent variable versus multiple dependent variables, (3) controlling variables versus characterizing the situation, (4) fixed procedures versus flexible design revision, (5) social isolation versus social interaction, (6) testing hypotheses versus developing a profile, and (7) experimenter versus coparticipant design and analysis.\(^\text{124}\)

These features of design studies—messiness, variability, interaction, etc.—can enhance the ecological validity of design experiments, but this often comes at the price of general knowledge because of the challenges in identifying patterns across contexts. Consequently, good design research is incredibly difficult. Consider an example from Ann Brown’s work. Brown developed a technique known as reciprocal teaching in the context of a larger research program known as Fostering Communities of Learners (FCL). She spent several years honing the educational intervention, which was based on learning principles in cognitive science. In brief, the pedagogical approach involved “learning dialogues in which teachers and students [take] turns leading the discussion of segments of text.” As promising as the intervention initially seemed, large scale evaluation of FCL was not successful because the program underwent “lethal mutations”—changes that moved it away from core learning principles. Furthermore, the composition of these principles evolved over the life of the project, such that it was impossible assess program outcomes. Thus, the FCL evaluation indicates two challenges any design study must overcome to generate general knowledge and understanding for use of the sort needed for large scale change. First, experimentation must work towards a gradual honing or refining of an intervention, such that fidelity of implementation can be sustained. Without a move towards a more stable, well-defined program, it will be impossible to scale the program without a proliferation of lethal mutations. Second,

128 Ed Haertel led the evaluation of the FCL project. In a personal communication (October 19, 2005), he explained the evaluation was largely unsuccessful because the FCL principles evolved throughout the project. For a discussion of this and related challenges emerging from the FCL project, see Ginsburg, "The Mellon Literacy Project: What Does It Teach Us About Educational Research, Practice, and Sustainability?"
to also be a source of general knowledge, this refinement must include the articulation of new insights or the problematizing of well-established theories. Without this theoretical engagement, new and effective programs may result but not new general knowledge. (Per the discussion in Section 3.4, such knowledge need not be “fundamental” or “necessary,” it need only have general explanatory power that contributes to existing middle range theory.)

Having described an unsuccessful design study, consider some more promising approaches. Burkhardt and Schoenfeld, for example, have argued for a large-scale research agenda built around design experiments. They recommend a two stage “engineering” approach to instructional development. In the “alpha stage” an initial version of materials that work when taught by team members is produced, including a well-defined set of goals and a standards-based assessment. Once a promising set of instructional materials has been identified, research moves to a more structured “beta stage” where the intervention is scaled up to include 50-100 matched pairs of classrooms with varied characteristics and a randomly assigned treatment. Extensive data is gathered on curricular impact using standards-based assessments, as well as more limited instructional observations and curriculum-based assessments. Materials undergo further refinement based on analysis of beta stage data and are then ready for widespread distribution (in combination with detailed performance data for comparison purposes).

In the field of medicine, Atul Gawande has recently promoted the idea of carefully designed checklists for particular procedures as a way to reduce routine medical mistakes. While checklists may sound like a somewhat boring or antiquated idea, Gwande carefully studied their use in other industries from commercial aviation to mega construction projects. He also had a deep understanding of the practice of medicine, the medical profession, and the larger health care system in which medicine is practice. Noting that checklists had been

---

[129] Burkhardt and Schoenfeld, "Improving Educational Research."
used with great success to reduce central-line infection rates in Michigan hospitals, he wanted to see if he could design a surgery checklist that could similarly reduce the large number of surgical errors. Central to the success in Michigan was significant work strategizing about how to implement checklists in such a way that people appreciated their importance and used them consistently and appropriately.

Thus, while Gawande’s checklist is a good example of a design study, it also reflects an equally important challenge: making sure the checklist gets used the way it is intended. Like Anne Brown’s work on reciprocal teaching, it is not uncommon for checklists to undergo “lethal mutations” and not get used as they were intended. Even the best designed products, programs, or environments can fail without the right organizational culture. And yet, after many experiments and iterations, the checklist Gawande helped develop for the World Health Organization reduced the number of deaths and complications by more than a third.¹³¹

*Improvement science.* Another approach to problem-disciplined inquiry is research on organizational improvement.¹³² The goal of improvement science is to better understand the techniques, strategies, and theories that produce effective organizations. In a sense, improvement science is design research at the level of organizations. Lots of research on organizational improvement has been done in both the public and private sectors, but schools have not been studied as systematically. A notable exception is the work of the Consortium on Chicago School Research (CCSR), a research organization out of the University of Chicago that formed a partnership with the local schools in the wake of the city’s 1988 decentralization plan. CCSR worked with schools, the city, and community organizations and over the last two decades CCSR has made “important contributions to school reform, both through the findings and implications of specific research studies and more broadly by improving

---


¹³² See for example Argyris, Putnam, and Smith, *Action Science.*
the capacity of the district to use data, build effective strategies, and evaluate progress.”

Chicago’s decentralization plan was based on the theory that if local actors were given adequate resources and authority, they could solve local problems. The law also reshaped incentives (and sanctions) for the principals leading the schools. CCSR took on a capacity-building role working closely with local schools in their reform efforts. Thus:

The primary audiences for research findings were critical actors in reform: new principals, foundations and other organizations supporting change, and the broader civic community . . . [T]his expanded audience generated a new focus for research: research must speak to the central problem the practitioners and broader community were grappling with—what would it mean to judge the effectiveness of school improvement and create effective schools? In this context, the challenge for research was to find ways to inform the question about how to judge school improvement by bringing evidence to bear on the problem and providing critical frameworks for understanding the task.

With significant help from other research and reform organizations and the local school district, CCSR developed *The Essential Supports for School Improvement*, a conceptual framework that became the centerpiece for local school improvement planning. In addition to having a significant impact on Chicago schools, CCSR has drawn many important theoretical insights from their work. Perhaps most notably, CCSR’s research lead to the discovery that more than any other characteristic in schools, social trust was most important to creating an environment for school improvement.

---


134 Ibid., 2.


Policy research. Yet another approach to problem-disciplined inquiry is policy research. The goal of much policy research is to affect policy, though it is not often problem-disciplined in the ways it should be. The efforts of Russ Whitehurst and the WWC discussed in earlier sections and chapters are an interesting case, since some aspects of their efforts were clearly problem-driven. At the same time, the approach they adopted was not problem-disciplined in the sense of development general explanations or theories or in submitting their ideas or work to the critical scrutiny of their peers. (In a loose sense scrutiny did occur, but engagement and uptake were clearly lacking.)

A key difficulty in policy research, especially in education, is its potential complexity. As Smith and Smith write:

Whether a study is useful for policymaking depends on many things, including the nature of the knowledge produced in the study, how and in what context policymakers might use the research, and how it is packaged. “Basic” research in areas such as learning, teaching, motivation, and the nature of organizations is critical to the advancement of knowledge, but typically such research is not directly useful to policymakers, though it may eventually be part of a chain of evidence that leads to policy.

Similarly, a study that is useful at one level of government may not be useful at another. Studies designed to provide information about how to teach a specific, important concept in elementary mathematics will not be useful to policymakers in federal and state governments or even in most district offices, though they may be useful to teachers, principals, and publishers. Some research does not even indirectly influence policy because it lacks methodological or theoretical quality.137

Chapter 2 discussed the challenges of policy research in some detail, noting the importance not just of theories about policy action, but also theories of policy implementation and policy context. Rather than reviewing this work, I consider a very different approach that holds some promise for large scale problem-disciplined educational inquiry.

137 Smith and Smith, "Research in the Policy Process," 375-76.
Occasionally laboratory experiments lend themselves to larger scale implementation without the intensive, iterative processes characteristic of design studies or the organizational transformation involved in improvement research. One promising example is a relatively simple intervention built on over a decade of research on the psychological theory of stereotype threat. In brief, stereotype threat is the fear that one’s behavior or performance will confirm an existing stereotype of a group one identifies with. The intervention was a short 15 minute writing assignment targeted a racially mixed group of seventh grade students and was evaluated in two randomized double-blind field experiments. Students in the treatment group were asked to write about a value (or several values) that were important to them. The intervention “tested whether a self-affirmation intervention designed to lessen threat would enhance the academic achievement of negatively stereotyped minority students.” The results of the intervention were impressive: the average treatment effect for African Americans was 0.30 grade points, or roughly a 40 percent reduction in the achievement gap.

Several features of this study make it an interesting example of problem-disciplined research. First, it came out of extensive laboratory research on stereotype threat and related psychological phenomena. Second, the intervention design was short enough to be non-disruptive and simple enough to be administered without significant concerns about fidelity of implementation across contexts. Third, the intervention had a huge impact that persisted for over a year and was replicated in a subsequent study. Moreover, it was very inexpensive to conduct—particularly given the powerful influence it had on student achievement. A consequence of the intervention’s simplicity is that, at least in this case, little understanding on the part of students, teachers, or administrators was necessary for the intervention to be scalable. This makes the intervention fairly unique, but there are some reasons to be

---

optimistic that other interventions of this sort might be designed. In a review article, Dennis D. Embry and Anthony Bigland describe “evidence-based kernels” which they define as “fundamental units of behavioral influence that appear to underlie effective prevention and treatment for children, adults, and families.”

Each of these three examples is promising approach to problem-disciplined educational inquiry. Each focuses on a problem of schools or schooling and, drawing on the best available research, attempts to address the problem in a way that also generates general knowledge and understanding for use in practice.

### 3.6 Conclusion

Not all educational research need be problem-disciplined in the sense advocated in Sections 3.5. Like Stokes, SRE, and the NAE report, there is value in a balanced, pluralistic approach. But balance should not be misconstrued as giving equal emphasis to all approaches. Just as Stokes argued that use-inspired basic research was a rationale for continued public investment in federally sponsored research, educational scholars should consider the rationale for their claims on public resources. Why should state and federal government invest public monies in education school research, as distinct from disciplinary investments in psychology, sociology, economics, etc? Part of the answer to this question—the most important part, in my view—is that the scholarly inquiry going on in schools of education is making distinct contributions to addressing the problems of schools and schooling. Some of these contributions will necessarily come through models more akin to Weiss’s enlightenment view. But the problems of education are too great and too pressing for this to be the only approach, or even the primary one. Problem-disciplined approaches to educational inquiry are a better center of gravity for comprehensive investment.

---

CHAPTER 4
Practical Arguments in Educational Inquiry:
A Strategy for Linking Knowledge and Action

The ingenuity of man leaves us to despair of nothing within the laws of nature.  
-Thomas Jefferson

Not everything that can be counted counts, and not everything that counts can be counted.  
-Albert Einstein

Philosophers have hitherto only interpreted the world in various ways; the point is to change it.  
-Karl Marx, Eleventh Thesis on Feuerbach

4.1 Introduction

The expertise of educational scholars gives them unique insight into problems of schools and schooling, but it does not make them policy experts, skilled administrators, or effective teachers. Scholars may share these skills, but their connection to academic expertise is uncertain at best. The gap between scholarly knowledge and the skills and understanding necessary for effective practical action puts researchers in a difficult position when it comes to advising policymakers and practitioners. This is especially true when practical action is fraught with moral and political concerns traditionally excluded from scientific research. I contend this problem is an important intellectual barrier to productive use of educational research in practice.

In this chapter I recommend linking knowledge and action through the explicit use of practical arguments in disciplined educational inquiry. I argue such an account

1 Portions of this chapter were previously published in Dolle, "Can (and Should) Educational Research Be Value-Neutral?"
must meet at least two criteria. First, it must be able to explicitly incorporate non-epistemic values without biasing the research. I take up this challenge in Section 4.2, investigating what it means for research to be biased. Second, it must be able to publicly justify the non-epistemic values that are so incorporated. Doing so lends additional public accountability to inquiry that is more tightly connected to social action. In Section 4.3, I argue that properly formulated practical arguments can meet both criteria. Together, these conditions protect the credibility of inquiry while helping ensure that recommendations for practical action on behalf of the public are aligned with the public good.

4.2 Value neutrality and bias in educational inquiry.

Academic scholarship sometimes gets used as evidence or advice for practical actions and decisions. Many scholars wish it was used more often. Yet describing the world and prescribing action are generally taken to be fundamentally different activities. Knowledge production, it is sometimes argued, should be value-neutral. In contrast, recommendations for practical action are value-laden, since they involve judgments about what ought to be done. In this section I try to explain some of the complexities of value judgments in educational inquiry, especially when such inquiry aims to influence practical action. This sets the problem that I contend (in Section 4.3) practical arguments can help address.

Research aimed at accurate descriptions and predictions of the world is sometimes called “positive” inquiry. In contrast, research aimed at recommending what action(s) ought to be taken is often referred to as “normative inquiry.” The positive approach evokes the view of the researcher dispassionately describing and explaining educational phenomena. This is the tradition of the researcher qua scientist. In contrast, the normative approach explicitly engages inquirers in evaluation and prescription, offering guidance about what decisions ought to be

---

3 And in the case of problem-disciplined inquiry discussed in Chapter 3, the connection between academic knowledge and practical action is intentionally close.
made and what actions *ought* to be taken. This is the more eclectic tradition of the researcher *qua* social critic, policy adviser, or connoisseur. New educational scholars often feel ill-equipped to navigate between these approaches, as it a skill set not widely taught or promoted in academia. At the very least, doctoral students in education are not generally prepared with the skills.⁴ There is a strong impulse to keep positive and normative arguments separate, lest the values underpinning normative assessment corrupt the accuracy of description, biasing the results. One concern with intermingling positive and normative judgments is that the normative activity can and will drive the positive activity to predetermined ends, corrupting the representation of evidence and undermining objectivity.⁵

The next two subsections examine this concern and related attempts to guard against bias in educational inquiry. Subsection 4.2.1 summarizes the argument that educational inquiry can be protected from bias if it is regulated by the ideal of value-neutrality. Subsection 4.2.2 examines this thesis in more detail, arguing that a critical ambiguity exists in its scope. I conclude that the most concerning forms of bias have less to do with the particular intrusions of non-epistemic values and more to do with their publicity, scrutiny, and justification within a disciplined community of inquirers.

### 4.2.1 The value neutral thesis in educational inquiry

While different disciplines place more or less emphasis on the pursuit of positive or normative accounts of the world, distinguishing between positive and normative judgments within inquiry—observations and evaluations that presumably affect the character of the result—is more difficult. The argument that research should be value neutral (VN) has emerged as an alternative to the less subtle distinction between positive (value-free) and normative (value-laden) approaches. Various forms of the VN thesis

---

⁴ For a discussion of this and related challenges, see Labaree, "The Peculiar Problems of Preparing Educational Researchers."

⁵ There is less concern the other direction; in fact, some believe there are good reasons to wish our value judgments were more informed by scientific evidence.
have been formulated by scholars from Ernst Nagel\(^6\) and Karl Popper\(^7\) to Harold Kincaid\(^8\) and Hugh Lacey.\(^9\) For present purposes, I consider the VN account offered by Denis Phillips, which is more accessible and has had greater influence on educational inquiry.\(^10\)

Phillips explicated the concept of neutrality in terms of two distinctions.\(^11\) The first distinguishes truth- or knowledge-indicative values—sometimes called *epistemic values*—from other types of values. A relatively uncontroversial example of an epistemic value is empirical adequacy. When deciding between two competing theories, we have reason to prefer the theory that better fits the available data—data that, under certain conditions, become evidence for the truth of a theory. So on the VN account, science maintains its objectivity not by avoiding value judgments but by privileging epistemic values over non-epistemic values in the assessment of research claims.

A second distinction contrasts aspects of science *external* to scientific reasoning with those *internal* to scientific reasoning. Decisions about the choice of research problem, the application of research findings, and the ethical constraints on the use of human subjects are part of science, but they are decisions external to the process of scientific reasoning and therefore have no bearing on the truth or falsity of a particular theory. Moral, social, and political values, on this view, can be a perfectly acceptable part of the external features of science. But they should have nothing to do with judgments central to the internal aspects of scientific reasoning like the


\(^{7}\) Popper, *Objective Knowledge: An Evolutionary Approach*.


\(^{11}\) For an extended discussion of this position, see Chapter 13 of Phillips, *The Expanded Social Scientist’s Bestiary*. 
identification of variables, the choice of research design, the drawing of inferences, or the acceptance/rejection of hypotheses.

So the central tenet of one popular framing of neutrality is that moral and political values should never influence epistemic judgments (though they might reasonably influence epistemically productive cognitive behavior). As Phillips and Burbules put it:

The classic dispute about values—the dispute that has fired controversies for more than a century—is about whether or not external, nonepistemically relevant values (e.g., political or religious values, or values relating to one’s position of power in society or to one’s economic interests) legitimately and perhaps necessarily play a role in scientific research. In common with most if not all postpositivists we have made no concessions on this point but have maintained that a research field has been seduced if it allows such values to intrude internally. Research must be free of contamination from such epistemically irrelevant, external values.  

In short, on this account objectivity is a function of epistemic purity. If a judgment is internal, non-epistemic values should hold no influence. This reflects the view that a claim is not more likely to be true simply because it would be “good” or “moral” or “just” if it were true.

Finally, for concepts like teaching and learning that have an evaluative dimension, Phillips follows Francis Schrag in arguing that disagreements can be overcome by recasting the question. For example:

Suppose you deny that throwing spitballs, talking while the teacher is talking, and so on, constitute discourtesy, and I insist the characterization is apt. We still agree, however, that the students in the one school throw spitballs and the like, and in the other school they do not. If we are interested in the causes of the differential student behavior, in the two schools we can, therefore, easily reformulate the question in this way: Why do the students in one school throw spitballs, and so on, while those in the other school do not?  

12 Phillips and Burbules, Postpositivism and Educational Research, 55.  
The strategy is to focus on the reliably observable behavior (which both parties can presumably agree upon) rather than on integrated evaluative judgments about what this behavior means. In cases where evaluative judgments cannot be bracketed because observation criteria depend on the evaluative stance adopted (e.g., assessing students academic performance), such criteria might simply be stipulated (e.g., by a research funder), or they might be set descriptively according to prior public usage (e.g., NCLB’s definition of a “highly qualified teacher”). Regardless of the strategy adopted, the VN account makes clear that researchers should not interject their own evaluative judgments internal to the research study.

The appeal of the VN thesis is that it seems to deal neatly with these complications. It acknowledges the importance of epistemic values and segregates the influence of non-epistemic values without ruling them out entirely. And for value-laden concepts like teacher effectiveness, it safely factors the evaluative dimension from the empirical dimension. Nonetheless, on closer examination the scope of the VN thesis is ambiguous and its rationale unclear for excluding non-epistemic values in cases where two or more research judgments are equally defensible on epistemic grounds. I consider both concerns in the next section.

4.2.2 How do values bias inquiry on the value neutral account? Ambiguity about the scope of the VN thesis stems from two problems: difficulty in discerning where the external part of science stops and the internal part begins, and uncertainty about how to warrant key research judgments under conditions of epistemic indeterminacy—that is, when epistemic values alone do not dictate a single course of action and different choices lead to different conclusions. To probe these problems, I try and state the essence of the VN thesis more formally than in the previous section.

One possible statement of the thesis is as follows:

\[
\text{VN}_1: \quad \text{A claim’s truth should not be influenced by values.}
\]

\[14\] For a good discussion of the different criteria that can be used to determine educational definitions in public discourse, see Israel Scheffler, *The Language of Education* (Springfield, Ill.,: Thomas, 1960).
There are two problems with this formulation. First, it says nothing about researchers or the research process. If the VN thesis is to have some bearing on researchers and the research process, it must connect with these activities. Second, as Phillips notes, researchers and philosophers generally agree that science is not (and cannot be) literally value-free. At a minimum, research involves standards of evidence and norms of inquiry, and these represent particular values—variously referred to as scientific values, epistemic values, cognitive values, or constitutive values. These kinds of values are not viewed as problematic, since they (presumably) have the common goal of guiding researchers toward knowledge (and away from falsehoods) about the world.

Consider a second possible statement of the thesis:

\[ VN_2: \text{ The inference(s) a researcher draws about the truth of a claim should not be influenced by non-epistemic values.} \]

This formulation focuses on the activity of inferential reasoning and the distinction between epistemic and non-epistemic values. In the most general sense, inferential reasoning is reasoning from particulars to a general claim. In this context, epistemic values can apply to either the research claim (e.g., its empirical adequacy) or to the reasoning process (where they are sometimes referred to as cognitive values, e.g., honesty). The most uncontroversial interpretation of this claim is that a researcher’s personal values should not determine whether a claim is true or not. As Elisabeth Lloyd explains: “if one is personally invested in a particular belief or attached to a point of view, such inflexibilities could impede the free acquisition of knowledge and the correct representation of (independent) reality.”\(^\text{15}\) Heather Douglas calls this detached objectivity, which requires only that values should not supplant evidence.\(^\text{16}\)

But VN\(_2\) could be interpreted more broadly as the requirement that non-epistemic values have no influence on any part of an inferential reasoning process, which can involve more than relating evidence to a particular claim:


VN3: No part of a researcher’s inferential reasoning process about the truth of a claim, including interpretation and explication of the claim, should be influenced by non-epistemic values.

In cases where the link between the claim and the evidence is indirect and open to multiple interpretations, the process can be complex and involve steps that go beyond simply assessing evidence. In this way uncontrovertially internal aspects of the research process can grow outward, encompassing wider circles of research judgments and the research process. Imagine assessing the claim that ‘Laura is an effective teacher’ according to VN3. The process of reasoning from evidence about Laura’s teaching to the conclusion that she is an effective teacher requires substantial clarification about what it means to be an effective teacher. This assessment requires a well-developed theory of effective teaching, and it may also depend on what Laura is teaching and who her students are this quarter. According to VN3, all these judgments would also have to be made without the intrusion of non-epistemic values. This is potentially problematic because the concept of effective teaching is not one around which universal agreement exists, and epistemic criteria can be used to justify several different courses of action. To the extent that a rough consensus exists about some aspects of effective teaching (e.g., student understanding of content matter), disagreement remains about how to assess this understanding (e.g., standardized tests, student portfolios, performance assessments).

So how should an honest, epistemically concerned researcher adhering to VN3 resolve this indeterminacy? Consider three possibilities:

1. *The public standard.* Select some publicly determined standard for effective teaching (e.g., the state’s standard for annual yearly progress).
2. *The best evidence.* Review the literature on teacher effectiveness and select the one that seems to be best supported by the current research.
3. *The low-inference reframe.* Conclude that effective teaching is such a high-inference concept that it is not amenable to empirical investigation. Thus, the concept must be reduced to one or more lower-inference elements.
Note that each option can be defended on strictly epistemic grounds. Option (1) can be defended on grounds that since effective teaching is an evaluative concept entangled with non-epistemic values, researchers have no special authority in setting the social worth of various teaching skills. Thus, the criteria should come from the public, not the research community. Option (2) can be defended on grounds that the concept of effective teaching has significant empirical content, and some conceptions in the existing literature are more empirically adequate than other concepts—that is, they better conform to the existing evidence than other concepts. Option (3) can be defended on grounds that while effective teaching may have empirical content, this content can be separated from the high-inference evaluative concept of “effective teaching.” By being more precise and speaking only of the lower-inference concepts that are elements of effective teaching, the degree of extra-empirical content can be minimized while maximizing the $VN_3$ of the research.

As the VN thesis is developed by Phillips, it is unclear which of these three strategies is epistemically preferable or what criteria should be used for selecting one over another. Note that it is also unclear why non-epistemic values should be excluded from arguments for or against the choice of construct. The claim that a publicly funded study should adopt public criteria for teacher effectiveness is a non-epistemic value, but it seems like a reasonable consideration in trying to decide between the above options, even if it is not a sufficient consideration. Regardless of whether one agrees or disagrees with this value claim, it seems like a plausible consideration that should not be ruled out on grounds of $VN_3$.

The decisions involved in this example of inferential reasoning are not uncommon. Some of the phenomena we want to know about are entangled with or embody human values. Teaching is a practice that people can be more or less effective at, and when researchers study teaching, they must offer some account of what constitutes effective teaching. This account will necessarily reflect certain non-epistemic values over others, regardless of whether the researcher endorses those values. The same is true of research on learning, segregation, development, conflict
resolution, etc. In contrast, it is relatively easy to talk about atoms, gravitational waves, or mitochondrial DNA without offering such value-infused definitions. Put differently, it is natural to talk in evaluative terms about improved learning, worsening segregation, sustainable development, and better conflict resolution. Typically people want improved learning, they are concerned about worsening segregation, they promote sustainable development, and they seek out better approaches to conflict resolution. This is the sense in which these concepts seem tied up with human values. But atoms, solar radiation, and mitochondrial DNA seem much more neutral. In general, they are not good or bad, better or worse, improved or worsening. Solar radiation might be good or bad for particular purposes: reading, gardening, skin cancer, etc. But it can be studied without having any goal or purpose in mind. The same does not seem to be true of teaching, learning, etc, where what we mean is likely to depend on our purpose.

By ruling consideration of the morally and practically normative dimensions of educational phenomena out of bounds, VN₃ significantly constrains the kinds of questions researchers can investigate. According to VN₃, researchers can ask: given a particular criterion or set of criteria for student learning, how well does an educational program fair? But researchers cannot ask: what ought the goals of learning be (and how should these goals be assessed)? Similarly, researchers can ask: given a certain construct or specification of racial segregation, how does a school voucher program influence the composition of schools? But researchers cannot ask: what is the concept of segregation (and how should it be specified in different contexts, e.g., housing, schools)? Note the parenthetical constraint on developing assessments that would warrant inferences about the construct. This can involve further interpretation and explication of goals and, at least in the case of low-inference reframing, the requirement that goals are well-specified and translatable into low-inference observations that may be impossible for some goals or constructs. The difficulty is that when the evaluative dimensions of concepts like “learning” and “segregation” are factored from the directly observable dimensions, the result is a
gap between the operationalized research concept (the strictly observable dimension) and the underlying construct it is supposed to reflect. If these are constraints of VN, they would seem to substantially limit the innovative capacity of educational inquiry and social science more generally.

How should researchers deal with the epistemic and evaluative aspects of entangled concepts? With respect to the descriptive or “epistemic” dimension, Phillips writes that these judgments “are made within domains of activity or discourse, and they are judgments that stand or fall according to how well they can be justified in terms of the technical considerations internal to the relevant fields.”

For the evaluative dimension, several general approaches are possible corresponding to the above strategies. One possibility is Schrag’s recommendation to recast the research question so as to avoid the problem. This may work in some instances, but it is an example of methodology driving the choice of research question, and it pushes the problem down the road to the practical interpretation of research. Similarly, a researcher might stipulate a definition, for example, as one might do when choosing cut points for assigning grades. (Careful researchers will go a step further and do “specification checks” to demonstrate that the results remain the same (or nearly the same) if modest adjustments are made.) Here too the justification is merely deferred.

Yet another approach is to appeal to a widely accepted external criterion based on tradition, history, or ordinary language. When teachers justify setting an “A” grade at 90% because it is what is widely accepted by the administration, parents, or other teachers, they are effectively using this strategy. But note that in recommending researchers stipulate a definition or adopt prior public usage, the VN₃ account sees researchers as closer to social engineers or “methodological servants,” as Donald Campbell colorfully put it, rather than as the drivers of enlightenment envisioned by Carol Weiss. (Recall from Chapter 3 that for Weiss,

---

18 Campbell, "The Social Scientist as Methodological Servant of the Experimenting Society."
19 Weiss, "Research for Policy's Sake."
the enlightenment function of research involved the introduction of new value judgments as much as new theories or conceptual frameworks.) Rich Shavelson has argued that “the contribution of educational research most often lies in constructing, challenging, and changing how policymakers and practitioners think,” and the adopting of public usage or historical conceptions would seem to seriously limit the ability of research to make such contributions.

There are instances where all of the above approaches are insufficient. Consider the following example: suppose two researchers operationalize the construct of student learning differently (creating different assessments) for a middle school mathematics curriculum and, consequently, get different results when analyzing the outcomes of otherwise identical intervention studies. Many possible explanations could account for the different results, but one possible explanation is that the researchers were actually studying different constructs—that is, there was fundamental disagreement about what constitutes appropriate middle school mathematical knowledge. Recall the debate in Chapter 2 over the What Works Clearinghouse assessment of middle school mathematics curricula. Alan Schoenfeld’s concern was that insufficient data was being collected on the assessments used for evaluating different curricula, so it was impossible to tell whether the results of particular mathematics curriculum were due to the curriculum or were an artifact of the assessment. He was particularly concerned that aspects of conceptual understanding and problem solving ability were not being adequately assessed by the WWC. (In this case, conflicting evidence for middle school mathematics curricula was averaged together without probing differences in the underlying construct, losing potentially valuable information.)

How should such construct and theory disagreements be resolved? Appealing to empirical evidence should certainly be part of any approach, and the standards for

---


21 Schoenfeld, "What Doesn't Work."
middle school mathematics proficiency at the heart of the above debate certainly conform to much empirical research. But sometimes construct disagreement may turn on the conflicting value commitments of researchers. This does not mean that either assessment is biased in the sense of VN$_2$. Rather, this could be an expression of the underdetermination problem discussed in Chapter 2. Just as a researchers in Lakatos’s competing research programs are not being irrational or biased when they defend their theories from refutation by making adjustments to minor supporting theories, researches from competing theories of teacher effectiveness need not be biased in the ways they are inferring evidential support. It might, in fact, be the case that the research programs have subtly different cognitive aims—a difference that can be hard to pin down. Further investigation may bring more evidence to bear on the theories, and it may gradually uncover unwarranted background assumptions.

A major concern about VN3 is that it could inadvertently shut down productive discussion about how the cognitive aims and value commitment of researchers shapes inquiry. This would be especially unfortunately given that such conversations might actually improve the social relevance of research by both research programs, even if no agreement was reached. Furthermore, conceptual arguments of the sort more commonly found in history or philosophy might also deepen and improve such work in practically normative ways. Again, the capaciousness of VN$_3$ is in question: to what extent must the interpretation and explication of concepts within a research study occur without influence from non-epistemic values? Consider three examples of non-epistemic value judgments that would presumably be barred from consideration under VN$_3$:

- It is unfair to hold teachers accountable for factors that are external to schooling. Consequently, assessments of effective teaching should control for external variables (e.g., student SES).

---

• Equality of educational opportunity requires that every student receives a minimally adequate education. Consequently, teachers should be assessed primarily on their effectiveness teaching the students who struggle most.

• Teachers are professionals. Consequently, they should be assessed according to the content they aim to teach and the learning goals they set, based on their professional judgment and not on a single universal standard.

These non-epistemic values can do more than shape the preliminary research questions that scholars ask. They can also operate as the cognitive aim of a research program. Consequently, each of these value judgments could as a touchstone throughout the research process— influencing the choice of research design, the selection of variables, the type of data analysis, etc. None of these judgments necessarily biases the research in the sense of VN$_2$, since all of these decisions can be made independent of the inference from the accumulated evidence to the final conclusion. And it is the epistemic integrity of this inference that matters most. Values may shape the character of inquiry without driving researchers to predetermined conclusions (VN$_2$).

This raises a final difficulty with the framing of both VN$_2$ and VN$_3$: each implies that the most epistemically important unit of analysis is the reasoning process of the individual researcher. The problem with this focus is that non-epistemic values, like false background assumptions, can operate in epistemically corrosive ways outside a researcher’s awareness. No amount of intellectual fastidiousness can fully root out the many background assumptions (warranted and unwarranted) that influence the myriad professional judgments researcher make, yet these judgments rarely get scrutinized. (Phillips, it should be noted, is very clear that VN should be a norm of inquiry.) The argument of Chapter 2 was that the critical discursive practices of communities of inquirers did the epistemic work. On this account, different cognitive aims (e.g., consulting for the state governor, narrowing the racial achievement gap, improving teacher education, contributing to cutting edge theories in social
psychology) might argue in favor of approach (1), (2), or (3). In fact, this suggests a fourth strategy, the one I recommend:

(4) The best fit. Reflect on the research question and the cognitive aims of the research study (e.g., narrowing the racial achievement gap) and select (or develop) a conception of teacher effectiveness that best fits these aims.

On this account, all that’s needed on the part of individual researchers to prevent bias is participation in a disciplined community of educational inquirers. No need to offer a single prescription for weighing many different (sometimes competing) epistemic values.

This approach also addresses another potential difficulty that has not yet been raised: the epistemically productive instrumental role non-epistemic values can play in inquiry.²³ Fears about tenure and promotion, pride in one’s research program and legacy, envy of one’s colleagues, greed for greater funding and public recognition, and even less noble desires motivate much research. But nothing about these values qua motivations need undermine the epistemic merits of research, any more than a student’s desire to do well on a test implies cheating. Taking a course and wanting to do well may encourage a student to cheat, but it may also encourage the student to study hard and learn the material. As long as these values don’t result in motivated reasoning that undermines epistemic judgment, there is no problem for the VN thesis in any of the above formulations. However, to the extent that VN₃ constrains the kinds of research questions scholars can pursue, it can undermine epistemically valuable motivations for scholarly inquiry. Philip Kitcher, for example, has argued at length for the importance of non-epistemic incentives (like recognition and reward) for an epistemically well-organized research effort.²⁴ To the extent that VN₃ unnecessarily forecloses some of these research incentives, it may undermine instrumentally epistemic benefits.

---

²³ Firth, "Epistemic Merit, Intrinsic and Instrumental."
4.3 Linking knowledge and action through practical argument

When people put forward claims, they expect those claims to be taken seriously (unless, of course, they are joking or speaking hypothetically). Scholarly claims typically are given extra credibility, I argued in Chapter 2, because they have been deemed epistemically adequate by a disciplining community. Central to this process is the supporting of academic conclusions with evidence and argument. Section 4.2 considered how values—especially non-epistemic values—could influence those conclusions and their supporting arguments without undermining epistemic merit. This section considers practical arguments and their potential for warranting for practical action.

In brief, a practical argument is an argument for taking practical action. Following Stephen Toulmin’s formulation, practical arguments follow an informal logic (akin to legal reasoning) rather than the formal logic of logicians. For present purposes, the virtue of practical arguments are twofold. First, they allow for the explicit incorporation of practically normative premises (e.g., ‘education policy ought to close the achievement gap,’ ‘allow children to pace their own learning is a good way to honor their individuality’). Such premises can be combined with more straightforwardly empirical claims about the achievement gap and individually paced learning. And the conclusion of such arguments is a recommendation for practical action (e.g., adopt policy X, try pedagogical approach Y). As informal arguments, these conclusions are vouched for rather than clinched.

Second, practical arguments offer a strategy for making the logic of action explicit and public. This means it can be independently assessed by others. This could be a group of researchers, a community of practitioners, or the public at large. This allows for the public justification of practical actions and decisions, conceivably subjecting arguments to broader input and critique (if appropriate processes and venues were in place).

Recent work in an area known as validity theory, especially as developed in the field of educational measurement, offers helpful guidance about the validation of
practical arguments. Validity theory may seem a strange place to go looking for a strategy for narrowing the gap between research and practice. Nevertheless, I contend that validity theory holds promise for offering a relatively robust theory for evaluating practical arguments that connect knowledge and action. Furthermore, it does so in ways that accord with the conceptions of problem-disciplined inquiry and understanding for use sketched in Chapter 3. The concept of validity also has surprising broad appeal. Both promoters of randomized experiments\textsuperscript{25} and some of their strongest critics\textsuperscript{26} have couched their argument in terms of validity. (It is an open question whether, in fact, their conceptions of validity have any resemblance to one another, but at least there is a space for potentially productive dialogue.) And the \textit{SRE} authors write that “scientific inquiry emphasizes checking and validating individual findings and results.”\textsuperscript{27} Natural scientists, social scientists, and the humanities all make arguments, as a central part of the research process. In short, practical arguments are a strategy for linking knowledge and action in a way that facilitates some degree of public validation.\textsuperscript{28}

I recommend a conception of validity that draws on work in educational measurement—especially that of Lee Cronbach, Samuel Messick, and Michael Kane—and has several distinct virtues from its formulations in philosophy or experimental design. First, validity in educational measurement has an empirical component missing from the logic of deductively valid of philosophical arguments. Second, educational measurement has always been intimately connected to policy and practice. Good measurement is a difficult and expensive exercise, and those who invest in large scale assessments often want to be able to draw conclusions from such investments to guide actions and decisions. (Oddly, validity in experimental

---

\textsuperscript{25} Shadish, Cook, and Campbell, \textit{Experimental and Quasi-Experimental Designs for Generalized Causal Inference}.

\textsuperscript{26} Patti Lather, "Fertile Obsession: Validity after Poststructuralism," \textit{The Sociological Quarterly} 34, no. 4 (1993).

\textsuperscript{27} National Research Council, "Scientific Research in Education," 4.

\textsuperscript{28} Of course, due to limits in knowledge and expertise, everyone involved may not be able to assess every aspect of a practical argument.
design has been much more narrowly prescribed, focusing almost exclusively on particular causal inferences that may or may not be of any use to decision makers, a point discussed at some length in Chapter 2.)

In Subsection 4.3.1 I offer a brief overview of validity theory in educational measurement, contrasting this view with the more technical version found in experimental design. In Subsection 4.3.2 I discuss validity in practical arguments as formulated by Gary Fenstermacher and Michael Kane. Practical arguments offer a structure for linking epistemically normative and practically normative premises in ways that can be productive to both research and practice. The development and validation of practical arguments fits problem-disciplined inquiry aimed at understanding for use especially well. As Fenstermacher writes, “when it is argued that research has benefit for practice, the criterion of benefit should be the improvement of practical arguments in the minds of teachers and other practitioners.”

4.3.1 **Validity theory: a brief overview.** In education there is a large literature dedicated to understanding how empirical research warrants claims or inferences. This field, known as “validity theory,” comes primarily from two subfields: educational measurement, where researchers try to understand the meaning of test scores, and experimental design, where researchers try to identify the impact and effectiveness of different treatments on educationally relevant outcomes. Despite these relatively specialized origins, work in validity theory has surprisingly broad application to a wide range of interpretive problems in educational inquiry. This expanding scope is due, somewhat surprisingly, to the growing emphasis on educational assessment. As educators and policymakers have sought to use assessment data as a guide for policy and practice, more robust theories of appropriate and inappropriate use have become necessary. Likewise, improved knowledge about educational measurement has helped assessment designs better match interpretive needs for various uses.

---

29 Fenstermacher, ”Philosophy of Research on Teaching,” 44.
In this subsection I offer a brief overview of validity theory and its evolution in light of the demands of educational use. This history, I contend, is the history of a tradeoff between rigor and relevance. Often the most rigorous assessments—understood as those that lend themselves to interpretations that are the most certain—have limited scope and application. In contrast, more comprehensive assessments that measure complex constructs for a range of purposes make validation more difficult. Thus, the trend has been towards the validation of practical arguments—interpretation of test scores with an explicitly stated aim or purpose. Subsection 4.3.2 extends the concept of validity arguments to practical arguments more generally. Subsection 4.3.3 considers how practical arguments can be used to address the two problems outlined in Chapter 1: the pursuit of more practically useful educational research and the development of understanding for use in practice.

The concept of validity in measurement theory has evolved over the last five decades, and this evolution suggests a complicated relationship often involving difficult tradeoffs between the pursuit of rigor and relevance. The evolution came, in part, because researchers wanted to design tests and studies that encouraged well-warranted inferences and guarded against unwarranted inferences. Thus, debate over the warrant for knowledge claims also became a debate over the desirability of different knowledge claims and the dangers of facilitating mistaken inferences. Standardized tests are a classic example. Tests designed to measure student subject matter knowledge are being used for other purposes, for example: to shape curriculum, to compare state educational programs, and to evaluate teacher effectiveness. Interpretations of test scores for these purposes can not only mislead, they can produce feedback loops that undermine the ability of tests to measure what they were originally designed to measure. This phenomenon is sometimes known as Campbell’s Law: “The more any quantitative social indicator is used for social
decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.”

Validity became a significant research area in the early 1950s when the American Psychological Association (APA) was first developing standards for evaluating the adequacy of psychological tests. These standards attempted to address two problems. The first problem was how to validate indirect measures of psychological phenomena. A person’s height can be measured directly and is stable across most contexts. The speed with which someone can run a mile is more likely to vary over different conditions (e.g., if a runner is dehydrated or sick), but it can also be directly observed and measured at particular points in time. A construct like athletic ability, in contrast, cannot be directly measured. Assessing athletic ability involves what validity theorist Samuel Messick later described as an “integrated evaluative judgment” that must be inferred rather than measured directly. In the 1950s tests were being developed to assess all manner of psychological characteristics that could not be directly measured—personality, creativity, motivation, etc—and the APA wanted criteria for assessing the adequacy of these tests. The attempt to develop these criteria highlighted a second problem: a test could be valid for some purposes but not valid for others. Because the validity of a test depended, in part, on how it was used, validity theorists focused on the validity of inferences from test scores rather than the validity of the tests themselves. For example, SAT scores might be a good predictor of college-bound seniors’ GPA at the end of their first year but a poor predictor four year graduation rates, university


honors, or GRE scores. Rather than talking about the validity of the SAT test itself, it is inferences from SAT scores to particular conclusions that are more or less valid.

On the face of it, a claim’s consequences and social relevance appear to have no bearing on its validity. If validity assesses the warrant for a claim or interpretation, why would the potential use or consequences of a claim have anything to do with the likelihood that the claim is true? A psychological test either predicts an outcome or it does not; an outcome’s “goodness” or “badness” would seem to have no bearing on the matter. This is most clearly true when there is a readily available criterion for measuring an outcome. For example, a test of academic self-efficacy might be used to predict college persistence or retention. Over time, test scores could be compared directly to the predicted results, and inferences from scores to retention would become more or less justified based on this track record. While constructing a test that accurately predicts a criterion is a more or less mechanical task, significant judgment is still involved in selecting a good criterion. As Cronbach noted, the procedures for making decisions based on the criterion model are typically intended “to optimize some later ‘criterion’ performance.” Academic retention is only a good if it leads to desired ends like academic success and learning. A major limitation of the criterion model is the difficulty of selecting an adequate criterion—an important difficulty that is easily glossed over.

An alternative to the criterion model is the content model of validation. Here a criterion is validated “by establishing a rational link between the procedures used to generate the criterion scores and the proposed interpretation or use of the scores.” A ranking of the world’s fastest 400 meter runners is developed based on a sampling of their performances—that is, performances at international competitions.

---

33 There are good reasons to question the validity of these interpretations of SAT scores. For an excellent discussion on their poor ability to predict graduation rates, see Bowen, Chingos, and McPherson, Crossing the Finish Line: Completing College at America’s Public Universities.


36 Kane, "Validation," 19.
A student’s vocabulary is measured by testing their comprehension of a carefully chosen sample of words. In some instances a complete performance assessment is possible, eliminating the need for sampling, for example, when times tables knowledge or the names of the elements is assessed by asking students to perform all the content.

Despite these advantages, content models face some of the same challenges as criterion models. As with criteria, content-based analyses “rely on judgments about the relevance and representativeness of test tasks.”\textsuperscript{37} Content models are particularly problematic when they are used to make inferences about theoretical constructs or cognitive processes. As Cronbach noted, “Judgments about content validity should be restricted to the operational, externally observable side of testing. Judgments about the subject’s internal processes state hypotheses, and these require empirical construct validation.”\textsuperscript{38} Construct validity refers to the validity of inferences about things that cannot be observed directly. Early work in construct validity came out of personality assessments in psychological tests, where individual test items were designed to sample different aspects of the construct theory—the object of the test.

In validity theory’s early days, measurement theorists debated the priority and tradeoffs between these three validity types. According to this early thinking, validity in testing was conceptualized according to different testing aims. But because tests could support the validity of inferences for some aims and not for others, researchers debated the relationship between different validity types and the acceptability of sacrificing one type for another.\textsuperscript{39} Nevertheless, over time a consensus emerged among measurement theorists that one aspect of validity—construct validity—should be prioritized over others.\textsuperscript{40} This development will be

\textsuperscript{37} Ibid.
\textsuperscript{38} Emphasis original. Cronbach, "Test Validation."
\textsuperscript{40} Messick, "Validity.", Cronbach, "Test Validation."
discussed in Subsection 4.3.2. But before turning to contemporary accounts of validity theory and the validation of practical arguments, it is worth briefly considering the very different way in which validity has been used in experimental design.

Validity theory has a long history in the design of experimental and quasi-experimental studies in education and social science. In one of the most widely taught and cited books on the subject, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Shadish, Cook, and Campbell distinguish four interrelated types of validity: statistical conclusion validity, internal validity, construct validity, and external validity. Here is their concise definition of each:

**Statistical Conclusion Validity.** The validity of inferences about the correlation (covariation) between treatment and outcome.

**Internal Validity.** The validity of inferences about whether observed covariation between A (the presumed treatment) and B (the presumed outcome) reflects a causal relationship from A to B as those variables were manipulated or measured.

**Construct Validity.** The validity of inferences about the higher order constructs that represent sampling particulars.

**External Validity.** The validity of inferences about whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables.\(^{41}\)

Like validity in measurement, these validity types apply to inferences rather than specific research designs or methods. Unlike validity in measurement, validity in experimental research design focuses on a relatively narrow set of inferences, namely *causal inferences*, rather than the much broader concern about the *meaning of test scores*. This is most evident in the major questions they suggest “practicing researchers face when interpreting causal studies:”

1. How large and reliable is the covariation between the presumed cause and effect?

---

(2) Is the covariation causal, or would the same covariation have been obtained without the treatment?

(3) Which general constructs are involved in the persons, settings, treatments, and observations used in the experiment? and

(4) How generalizable is the locally embedded causal relationship over varied persons, treatments, observations, and settings?\(^{42}\)

These four questions are supposed to parallel the four validity types, and three of the four appear to line up reasonably well. However, as should be evident from the educational measurement discussion, the third question about “which general constructs are involved” seems to ask less of researchers than what construct validity, by their own definition, requires. Construct “involvement” sounds like a matter of definition, but construct validity requires a lot more than clearly defined terms. And the situation is even more complicated, as the authors go on to note:

Although these questions are often highly interrelated, it is worth treating them separately because the inferences drawn about them often occur independently and because the reasoning we use to construct each type of inference differs in important ways. In the end, however, readers should always remember that “A validity typology can greatly aid . . . design, but it does not substitute for critical analysis of the particular case or for logic” (Mark, 1986, p. 63).\(^{43}\)

As the Mark quotation suggests, the validity of causal inferences, like the validity of inferences in educational measurement, ultimate depends on a comprehensive validity argument rather than any particular typology.

The virtue of good experimental designs is the strength it lends to the internal validity of causal inferences. So it is unsurprising that Shadish, Cook, and Campbell tend to underscore the importance of internal validity while putting comparatively less emphasis on construct validity. Rigorous research designs that strengthen the internal validity of a study can also end up diminishing its external validity. The authors note the importance of external validity but are quick to emphasize that they

\(^{42}\) Ibid., 39.

\(^{43}\) Ibid.
disagree with their psychometric colleagues about the possibility of validating actions. Consider the following excerpts:

Inferences from completed studies to as-yet-unstudied applications are necessary to both science and society. During the last two decades of the 20th century, for example, researchers at the U.S. General Accounting Office’s Program Evaluation and Methodology Division frequently advised Congress about policy based on reviews of past studies that overlap only partially with the exact application that Congress has in mind.  

Whereas our understanding of validity is that *inferences* are the subject of validation, this definition suggests that *actions* are also subject to validation and that validation is actually evaluation. These extensions [sic] are far from our view.  

We strongly endorse the legitimacy of questions about the use of both tests and experiments. Although scientists have frequently avoided value questions in the mistaken belief that they cannot be studied scientifically or that science is value free, we cannot avoid values even if we try. The conduct of experiments involves values at every step, from question selection through the interpretation and reporting of results. Concerns about the uses to which experiments and their results are put and the value of the consequences of those uses is important . . . However, if validity is to retain its primary association with the truth of knowledge claims, then it is fundamentally impossible to validate an action because actions are not knowledge claims. Actions are more properly evaluated, not validated.  

Whereas Cronbach concluded that “Validation of a test or test use is evaluation,” Shadish, Cook, and Campbell reject this, concluding instead that “Validation is not evaluation; truth is not value.” Furthermore:

We need the distinction between truth and value because true inferences can be about bad things (the fact that smoking causes cancer does not make smoking or cancer good); and false inferences can lead to good things (the

---

44 Ibid., 84.
45 Ibid., 475.
46 Ibid., 476.
astrologer’s advice to Pices to “avoid alienating your coworkers today” may have nothing to do with heavenly bodies, but may still be good advice). Conflating truth and value can be actively harmful.49

But Cronbach’s point was that researchers need to be involved in helping those who use assessments understand their appropriate use. This alignment between an assessment and its proposed use can be subject to validation by analyzing the logic of the interpretation and its evidential support.

In sum, Shadish, Cook, and Campbell’s argument follows quite closely the logic of the VN₃ thesis discussed in the previous section. A clear implication of their argument is that it is impossible to validate an interpretation that depends on a value premise. In contrast, measurement theorists have generally been less skeptical about this concern, since both the construct validity concept and the valid interpretation of test scores for particular uses rely, at some level, on the validation of practically normative premises.

4.3.2 Validation as practical argument. In the third edition of Educational Measurement Samuel Messick defines validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment.”50 Note that Messick’s definition explicitly incorporates intended use as a key element of assessing the validity of inferences. Thus, just as the standards of evidence in disciplined inquiry depend (in part) on the cognitive aims of research, on Messick’s account the validity of an inference depends on both the cognitive and practical aims of the assessment. In this sense, validity can be said to have both a “consequential” and an “evidential” basis.51

In the fourth edition of Educational Measurement, Michael Kane writes in the new chapter on “validation” that:

49 Ibid.
51 Messick’s discussion of the evidential and consequential bases of validation is quite complex, and I do not go into the distinction here because the details are less relevant to educational inquiry generally.
To validate a proposed interpretation or use of test scores is to evaluate the rationale for this interpretation or use. The evidence needed for validation necessarily depends on the claims being made. Therefore, validation requires a clear statement of the proposed interpretations and uses.\footnote{Kane, "Validation," 23.}

On Kane’s account this “clear statement” is a practical argument that proposes an interpretation. This focus on explicit practical arguments is the primary difference from Messick’s approach to “validity.” In this regard the distinction between “validity” and “validation” is subtle but important. The main difference is that validity is a characteristic of an interpretation, whereas validation is a characterization of a process. As Kane writes:

The main advantage of the argument-based approach to validation is the guidance it provides in allocating research effort and in gauging progress in the validation effort. The kinds of validity evidence that are most relevant are those that support the main inferences and assumptions in the interpretive argument, particularly those that are problematic.\footnote{Ibid.}

The validation process for interpreting test scores involves two kinds of arguments: an interpretive argument and a validity argument.

An interpretive argument specifies the proposed interpretations and uses of test results by laying out the network of inferences and assumptions leading from the observed performances to the conclusions and decisions based on the performances. The validity argument provides an evaluation of the interpretive argument.\footnote{Emphasis original. Ibid.}

The interpretive argument offers an explicit account of the link between the observed performances (the assessment or test score) to claims based on these performances. For Kane, the inferences may be “if-then” rules, for example, “if an observed performance has certain characteristics, then the observed score should have a certain value, if an applicant’s score is above some cutscore, the applicant is admitted.”\footnote{Ibid.} The virtue of interpretive arguments is that they make “the reasoning
inherent in the proposed interpretations and uses explicit so that it can be better understood and evaluated.” The validity argument assesses the reasoning and evidential basis of the interpretive argument.

Although Kane’s account focuses on validation in educational measurement, the central idea can be generalized to other areas of educational inquiry as a tool for validating the implications of research findings for educational practice and policy. A more general formulation might be:

An interpretive argument specifies the proposed interpretations and uses of [research findings] by laying out the network of inferences and assumptions leading from the [findings] to the conclusions and decisions based on the performances.

One key difference is that in education measurement the validation process assesses the inferences that can be drawn from a tool (the assessment) that will be used to produce new data (in need of interpretation), but practical arguments in many areas of educational inquiry do not develop such tools (though some certainly do). Under these circumstances the “if-then” arguments could be applied to interpretations of the data for various purposes in different contexts. But even without such a contrivance, practical arguments could be used to accomplish what Donald Stokes suggested researchers and funders do more generally: create “a system for appraising scientific promise and social value at the project level [that enlists] the insight of the working scientist into the nature of the social goals on which his or her research bears.”

So practical arguments in education inquiry can be part of a research study—for example, as part of discussions about a study’s implications. But they can be proposed independently. There a sense in which this is what DC policy organizations like the Heritage Foundation, the American Enterprise Institute, the Center for American Progress, and the Brookings Institution do. They muster evidence for

56 Ibid.
57 Ibid.
58 Stokes, Pasteur’s Quadrant, 116.
particular practical arguments about the direction policy should go, attempting to use these arguments to sway representatives and their staffs toward particular policy actions. But such use has a major limitation: the practical arguments offered by these intermediary policy organizations are rarely subject to the kind of validation recommended by Kane. Put differently, these practical arguments are not scrutinized as part of disciplined educational inquiry. Consequently, a significant portion of the educational policy discussion in the U.S. happens with limited review or insight from the larger community of educational scholars. A recent study found, for example, that think tank reports were far more likely than university studies to be cited by major news outlets.59

This is one reason bringing practical arguments within disciplined educational inquiry is crucial to strengthening the connection between knowledge and action. Nonetheless, the building or strengthening of this connection, it should be noted, is not a one way street from knowledge to action. In fact, as Kane notes, “the main advantage of the argument-based approach to validation is the guidance it provides in allocating research effort and in gauging progress in the validation effort.”60 Practical arguments can not only help assess what the best evidence says about a particular problem, they also help identify practically important holes in the research base.

Yet another potential benefit of practical arguments, beyond directing educational inquiry to useful research, is their value promoting understanding for use. Gary Fenstermacher discusses this potential use, stating that

When it is argued that research has benefit for practice, the criterion of benefit should be the improvement of practical arguments in the minds of teachers and practitioners. As researchers cast about for interesting problems, they may seek problems that bear on the practical arguments in the minds of teachers, or they may ignore these arguments. The relevance of

60 Kane, "Validation," 23.
research for teaching practice can be understood as a matter of how directly the research relates to the practical arguments in the minds of teachers. Here’s an illustration of the kind of practical argument Fenstermacher has in mind:

1. It is extremely important for children to know how to read.
2. Children who do not know how to read are best begun with primers.
3. All nonreaders will proceed through the primers at the same rate (the importance of learning to read justifies this standardization).
4. The skills of reading are most likely to be mastered by choral reading of the primers, combined with random calling on individual students.
5. This is a group of nonreaders for whom I am the designated teacher.

ACTION: (I am distributing primers and preparing the class to respond in unison to me).

Imagine first that this practical argument came from a teacher in conversation with other teachers or with a researcher. Making this reasoning public could have a number of benefits, including making the teacher more conscious of his or her rationale for various pedagogical approaches and facilitating the sharing of these insights with other teachers. Furthermore, explicit articulation also facilitates criticism and critical reflection about the argument’s logic and evidentiary basis. This public dimension is very important, as it facilitates discussion, deliberation, and critical revisions in light of new evidence.

Organizational theorist Chris Argyris’s distinction between espoused theory and theory-in-use is especially relevant to this point and illustrates a problem with Fenstermacher’s framing. Espoused theory, on Argyris’s account, is the explanation people give of their own behavior. Theory-in-use is the logic that seems to guide and predict observed behavior. In Argyris’s many organizational studies, he found that individuals’ behavior often did not match their espoused theory. Furthermore,

---

61 Fenstermacher, "Philosophy of Research on Teaching," 44.
62 Ibid., 46.
disconnect was often something they were entirely unaware of. For practical arguments to have influence over teaching practice, the changing of these arguments in teachers’ and practitioners’ minds is insufficient if it does not also include this critical public dimension.

One final qualification: there is a large (and growing) body of philosophical literature on practical reasoning—a topic that clearly has some relationship to the role of practical arguments in decisions and actions. However, this literature has tended to focus on the internal reasoning processes in individual actions and decisions. My concern is with public arguments attempting to warrant particular decisions or actions and with the cognitive value of practical arguments as a pedagogical tool (of sorts). The practical reasoning literature becomes more relevant when delving into the epistemic and ethical dimensions of adopting a practical argument as a reason for action—an interesting issue I cannot address here.

---

64 See for example the anthology edited by Elijah Millgram, *Varieties of Practical Reasoning*, which does not include a single example of policy-oriented reasoning, though it does introduce a few examples of how external constraints like laws or social norms might figure into the practical reasoning of individuals. In all events, this literature takes us too far afield from the issues at hand. Elijah Millgram, ed., *Varieties of Practical Reasoning* (Cambridge, MA: MIT Press, 2001).
References


Cook, Thomas D. "Randomized Experiments in Educational Policy Research: A Critical Examination of the Reasons the Educational Evaluation Community Has


