

Can (and Should) Educational Research Be Value-Neutral?

Jonathan R. Dolle
Stanford University

INTRODUCTION

In August 2004, a *New York Times* front-page story headlined “Charter Schools Trail in Results, U.S. Data Reveals” described a new analysis of NAEP data conducted by researchers at the American Federation of Teachers (AFT). It began: “The first national comparison of test scores among children in charter schools and regular public schools shows charter school students often doing worse than comparable students in regular public schools.”¹ Prominent researchers and supporters of regular public schools lauded the study and criticized the Bush administration for burying the results.

But charter supporters were not to be outdone. Eight days later, thirty-one scholars signed a full-page advertisement rebutting aspects of the AFT study on methodological grounds, concluding that it “does not meet professional research standards. As a result, it tells us nothing about whether charter schools are succeeding.” A month later, Caroline Hoxby’s report, using methods similar to those criticized in the *New York Times* advertisement she cosigned, argued that charters actually outperform regular public schools.² Warring op-eds appeared in the *New York Times*, *New York Post*, and *Wall Street Journal*. The following year Martin Carnoy and colleagues published *The Charter School Dust-Up*, criticizing the research and policy claims of so-called “charter school zealots.”³ So the debate continues, with increasing data about charter performance not appearing to change the opinions of researchers on either side — at least so far.⁴

What’s really going on in policy debates like this one? Is political advocacy trumping the integrity of research? Can social science serve as a “neutral arbiter” in sensitive debates over social policy, or is it simply too laden with moral, social, and political values? Getting to the heart of such controversies is important because charter research is not a special case. Similar debates have occurred over class size reduction, vouchers, literacy, phonics, bilingual education, high stakes testing, the alternative credentialing of teachers, and many other educational policy domains.

The ongoing debate over charter school research illustrates a more general problem in the philosophy of science: what role (if any) should moral and political values play in research? Within philosophy of education, D.C. Phillips has vigorously defended the position that these values should play a limited role in scientific judgment and the justification of scientific claims. This essay examines Phillips’s account and argues that moral and political values, among others, can, do, and should play a more pervasive role in social science than he generally allows. For convenience, I call Phillips’s view the value-neutral (VN) account of scientific objectivity — a description he often uses, though his position should not be confused with other accounts of objectivity and scientific neutrality using the “value-neutral” label.⁵ Misunderstanding the role of moral and political values can have serious

consequences, stifling valuable discussion and debate. My goal here is to understand how Phillips's VN approach contributes to this misunderstanding.

Can, and should, educational research be VN in the way Phillips suggests? I approach this question in two parts. In the first section, I explain the VN view of scientific objectivity I take Phillips to hold. In the second section, I raise two criticisms of Phillips's position I do not think his account can handle. I conclude by discussing the importance of research frames and by briefly sketching what an alternative research norm might look like.

Before proceeding, let me highlight several points of agreement. First, Phillips and I agree that epistemic concerns — that is, concerns about the truth and justification of knowledge claims — are central to the enterprise of educational research and to science generally. We should be more confident that careful, systematic research conducted by honest, open-minded inquirers and vetted by a community of qualified peers gets at “the truth” (in a context-dependent sense) than does research that fails these criteria. The epistemic dimension of science justifies our belief that global warming really is happening, that evolution is more than a hypothesis, and that vaccines really can mitigate the spread of disease. And when science fails us, this failure can often be explained, at least in part, by a failure to meet epistemic criteria. Of course, understanding and defending particular epistemic criteria can be a difficult business, and this essay only hints at some of these difficulties.

Second, Phillips and I agree on the closely related point that moral and political values (among others) can and sometimes do corrupt the epistemic integrity of educational scholarship. If a researcher ignores evidence because it does not fit with her/his moral or political views, or fabricates data to fit a preferred model or theory, values are clearly playing an epistemically corrosive role in ways that undermine the ability of scholarship to arbitrate between competing claims. Where Phillips and I differ is in the necessary conditions for epistemically rigorous research.

Third, we agree that empirical research in education is both possible and exceedingly valuable when done well. I hesitate to say it is often done well, but there is no shortage of commendable examples. In his 1988 AERA presidential address, Richard Shavelson argued that one of the greatest contributions of educational research to policy and practice is “in constructing, challenging, or changing the way policymakers and practitioners think about problems.”⁶ Research on teaching and measures of school and teacher effectiveness are two prime examples he discusses, but the last two decades have brought many more. Claude Steele's work on stereotype threat — the fear that one's behavior or performance will confirm an existing stereotype of a group one identifies with — is changing the way achievement gaps are understood and standardized test results are interpreted.⁷ Tennessee's longitudinal study of the impact of class-size reduction on student achievement provided sufficient evidence for California to pursue a similar (if terribly implemented) policy.⁸ Perhaps most notably, Amartya Sen's work in welfare economics transformed the way economists think about development by prioritizing individual capabilities (or freedom) over aggregate utility, and influenced the educational

policies and programs of the United Nations, International Labor Organization, and World Bank.⁹

One final clarification: I use “science,” “research,” and “scholarship” to refer to one and the same thing — disciplined inquiry involving the collection of evidence and the mustering of argument, conducted in accordance with a set of scholarly norms. According to this definition, philosophy and history can make the same claims to knowledge production that psychology, economics, or particle physics can. This is a variation on Helen Longino’s argument for social knowledge. What these disciplines share (and undisciplined inquiry lacks) is a set of social practices conducive to knowledge production: venues for criticism, uptake of criticism, public standards for evidence, and tempered equality of intellectual authority.¹⁰ Naturally, different disciplines ask different questions, use different methods, and apply different standards of evidence. Which approach is best always depends on the question being asked or the problem being solved; questions of methodology are context dependent, but not subjective in the “anything goes” sense. My goal here is to sidestep unproductive debates about what counts as “science” or “scientifically-based” research.

With these caveats, I now turn to Phillips’s VN approach as a goal or norm of educational research.

PHILLIPS’S VALUE-NEUTRAL OBJECTIVITY

Phillips is a staunch defender of objective research. His account of VN social inquiry is one of four views about the relationship between values and science described in the final chapter of *The Expanded Social Scientist’s Bestiary*.¹¹ These views can be summarized roughly as follows (the labels and gloss are my own):

1. *Value-free science*. On this largely discredited view, values have no role to play. Science simply describes the empirical world, bracketing evaluative judgments. This “spectator view of knowledge” is often ascribed to logical positivists like Rudolf Carnap and A.J. Ayer, but was not universal within the Vienna Circle.¹² Few, if any, contemporary philosophers of science hold such an unqualified view.

2. *Science as politics*. On this view science is so value laden that its epistemic dimension is hard, if not impossible, to locate. Although Phillips attributes it to some feminists, it is not a feminist view. Elizabeth Anderson, Longino, and other prominent feminist philosophers of science reject science as politics as a general orientation, even if they are sympathetic to the more robust social analyses sometimes found in this work.¹³

3. *Value-laden science*. This view, which Phillips attributes to Michael Scriven and Richard Rudner, embraces the role of moral and political values in science while still holding a central place for epistemic criteria. This view sees no inherent tension between applying both epistemic and nonepistemic criteria in the evaluation of research judgments and the justification of research claims. When criteria are in tension, researchers

can reasonably debate advantages and drawbacks of prioritizing some criteria over others. I hold this view, and I believe it accommodates the challenges raised in the next section.

4. *Value-neutral science*. This is Phillips's position. On the VN view, understanding what makes research objective requires two distinctions. The first distinguishes truth- or knowledge-indicative values — sometimes called epistemic values — from other types of values. The second distinguishes aspects of science that are external to scientific reasoning with aspects that are internal.

With regard to the first distinction, the most uncontroversial example of an epistemic value may be empirical adequacy. When deciding between two competing hypotheses, we have reason to prefer the hypothesis that better predicts empirical outcomes — data that, under certain conditions, become evidence for the truth of a theory. So science is objective not because it is value-free, but because it privileges epistemic values over nonepistemic values in the evaluation of theories.

Consider the two central distinctions in Phillips's VN account. With regard to the second distinction, 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 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 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 identification of variables, the choice of research design, or the acceptance/rejection of hypotheses.

In practice, philosophers and methodologists acknowledge that nonepistemic values sometimes — perhaps often — impinge on processes of scientific reasoning. But, they argue, this does not undermine the value neutrality of science. Rather, keeping the internal aspects of science free from the interference of nonepistemic values should be understood as a regulative ideal, something epistemically responsible researchers strive for, even if they fall short. In this sense, the virtue of the VN approach is comparable to the virtue of rationality: researchers need not be perfectly rational for the research enterprise to get at truth; neither must their professional judgments be epistemically pure. As regulative ideals, value neutrality and rationality are goals of individual researchers that are enforced by the research community.

To sum up: on Phillips's VN account, objectivity is a function of epistemic purity. If a judgment is internal, nonepistemic values should hold no influence. But there is still a question of whether the VN approach applies to some or all social science. Call the *weak claim* the position that it should guide in a particular instance. It requires only the possibility that researchers can compartmentalize epistemic judgments, insulating them from nonepistemic values. Call the *strong claim* the position that the VN ideal should regulate all inquiry. So the strong claim entails the weak but not vice versa.

Phillips understands his argument as supporting the strong claim:

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.¹⁴

So a burden of Phillips's position is demonstrating that value-neutrality is not only possible, but that allowing nonepistemic values "to intrude internally" necessarily undermines the epistemic integrity of science.

Questions about the role of values in evaluating empirical claims, hypotheses, or theories are different from questions about what motivates the research generally. Fears about tenure and promotion, pride about one's work and legacy, envy of one's colleagues, greed for funding and recognition, and many other desires motivate 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. Wanting to do well may encourage a student to cheat, but it may also encourage study and effort to learn the material. The concern is only with how nonepistemic values do or do not affect the neutrality of reasoning processes and research judgments. The next section examines research judgments where segregating epistemic and nonepistemic value judgments is, I argue, unproductive and potentially dangerous.

TWO ARGUMENTS AGAINST VALUE NEUTRALITY AS A REGULATIVE IDEAL

Thus far, I have raised concerns about Phillips's arguments supporting the VN account as a regulative ideal for all social science. I now develop two arguments against the VN ideal and in favor of a value-laden conception of science. The first argument suggests that science is best understood as the search for significant truth, and debates about significance require moral, social, and political judgments. The second argument parallels the first but approaches from the opposite end: researchers have an obligation to protect the public from the foreseeable negative consequences of their work. This sometimes requires exercising epistemic caution when the social stakes of research are high and the degree of uncertainty is great.

First, science as the search for significant truth sometimes requires internal moral and political judgments. Science is not a purely epistemic enterprise, concerned solely with the discovery and justification of truth claims. Rather, the goal of science is what Philip Kitcher calls significant truth — truths that are meaningful, interesting, and important to human endeavors.¹⁵ The goal of significant truth (unlike the goal of run-of-the-mill truth) requires that competing research programs, explanatory theories, hypotheses, and background assumptions are evaluated on grounds that are not purely epistemic. Purely epistemic criteria cannot distinguish between the truly profound and the utterly mundane.

A common type of educational study investigates the impact of a particular curriculum or pedagogical approach on student learning, and thus must define some

measure or proxy for student learning. Here the difficulties begin: for what counts as a good measure depends not simply on epistemic criteria but also on what is worth knowing — a judgment legitimately influenced by moral, social, or political concerns.

A VN approach defender might offer several responses. First, it might be argued, important variables like learning or achievement are chosen external to the research process. They are inherent to the research question being posed and need not be part of a researcher's judgments or reasoning. But this response is inadequate for demonstrating that as a regulative ideal, moral, social, and political judgments about student learning should never be allowed to affect internal judgments (such as choice of research design or the interpretation of data). Establishing this stronger claim is incredibly difficult, for researchers often study programs for which the outcomes are undefined, at least initially. The researcher must identify or develop measures of learning that are not just operationalizable, but also meaningful, interesting, and useful. Even when desired outcomes are more robustly defined, researchers must still decide how to measure and assess them.

Second, accepting that judgments about student achievement must sometimes be made internal to the research process, it might be argued that these judgments need not be contaminated by moral or political values. This response goes as follows: measures and heuristics for student learning, while having moral, social, and political implications, need not be defended on these grounds. Instead, researchers can simply define their terms operationally, leaving it to philosophers and politicians to debate the "right" or "best" definition. This is fully compatible with researchers pointing out inconsistencies in academic debates, but it excludes methodological arguments about which measures, variables, or heuristics are better in the sense of being more meaningful, interesting, and useful. Yet this argument, like the previous one, begs an important question: even if some value-laden terms can be defined operationally without considering nonepistemic concerns, why as a regulative ideal should scientists bar such considerations from debates about the internal judgments of science? Perhaps such constraints work in areas of the natural sciences (though even here I am skeptical). But in the social sciences, where human well-being is often implicitly or explicitly the object of study, defining variables operationally without due consideration and argument about the moral, social, and political implications of particular definitions of well-being strikes me as both implausible and irresponsible.¹⁶

The VN account must thus provide a compelling explanation of why nonepistemic values should always be excluded when weighing the epistemic tradeoffs researchers inevitably make in developing a research design or evaluating competing hypotheses. For example, is it better to learn a little about a very consequential hypothesis or a lot about a less consequential hypothesis? This challenge becomes greater when the purpose of science is understood as the search for significant truth.

Second, uncertainty and social risk require scientists to exercise epistemic caution. Researchers also judge when to accept or reject a hypothesis or theory; these judgments, tentative as they may be, are often guided by disciplinary norms. For example, a very common measure for evaluating a finding is its p-value or statistical

significance. P-values indicate the probability that a particular finding occurred by chance in the random sampling procedure and therefore does not reflect a genuine finding about the population. In most sciences, statistical significance is benchmarked at three levels: 0.05, 0.01, and 0.001, representing progressively stronger degrees of confidence in the results. Statistical significance is one guide to crafting conclusions about, say, the efficacy of a particular mathematics curriculum at improving student learning of basic algebra. In education, significance levels of .05 or .01 are often accepted as confirming a hypothesis, though sometimes researchers apply more lenient standards.

But what level of certainty should researchers accept as sufficient? I believe disciplinary norms are inadequate justification. Instead, the strength of the evidence needed to justify accepting or rejecting a hypothesis can and should depend on what the likely (nonepistemic) consequences are of accepting a false hypothesis or rejecting a true hypothesis (in statistical language, type I and type II errors).¹⁷

Proponents of value-neutrality sometimes claim that moral, social, and political consequences are irrelevant to judgments about the truth of a theory, so scientists should remain unfettered by such concerns. But the fact that social consequences may have no bearing on the truth or falsity of a theory does not mean scientists should exclude such considerations from their professional judgments.¹⁸ The VN ideal appeals to the norm of free inquiry in science; yet almost all social scientists already accept methodological constraints on free inquiry in regard to the use of human subjects.¹⁹ Why exclude such considerations from the analysis of data or the reporting of research findings? Epistemic caution strikes me as not only justifiable but morally necessary.

So yet another burden for the VN account is a compelling explanation of why the risk of being wrong — the social consequences of accepting a false hypothesis or rejecting a true one — should play no role in researchers' epistemic judgments. I think Phillips's account cannot bear this burden.

A CONCLUDING REFLECTION ON VALUE-LADEN SCIENCE

The examples of significant truth and epistemic caution illustrate the important role problem framing plays in research.²⁰ How we analyze a problem depends on what we want to accomplish and what the inherent risks are. Recall Shavelson's observation that educational research can impact policymakers and practitioners by shifting their thinking about educational issues. These shifts often involve altering their analytic frameworks; he describes the advent of teacher cognition research as one such shift. Researchers had tried to investigate the traits of good teachers, but the work of Nate Gage and others shifted the focus to the behaviors and skills of teaching; later still, Shavelson's own work drew attention to teacher cognition, to "teachers' *decisions* about what to do, and when and how to do it [as] the basis of skilled teaching in a time- and context-bound classroom." Shavelson's explanation of these shifts is particularly salient:

Because teachers meet up with a multitude of unique situations for which they couldn't possibly be trained — and in which asking more questions would not "work" — I preferred to conceive of teaching as the exercise of professional decision making, that is, deciding how

to adapt their substantive and pedagogical knowledge to a time- and context-bound classroom.²¹

This is not a purely epistemic argument — it is practical, as well as moral and political. The problem with trait-based research is not that good teachers do not exhibit particular traits, it is that this knowledge does little to guide our practical understanding of what good teaching involves and how to help struggling teachers improve. The epistemic argument against trait-based teaching doesn't work without also arguing for a paradigm shift in what good teaching embodies. In retrospect this might appear to be a different research question or program, but clearly the field did not wake up one day and decide to work within a new framework — the cognitive approach was vigorously discussed and debated before gaining acceptance!

Within a framework, the VN account might appear innocuous enough: for a well-defined problem with predetermined measures of success and tools for measuring, a researcher can set to work without much reflection or judgment. But when the very problem, measures, tools, or impact is contested or unclear, it provides less guidance. The charter debate involves just such competing frameworks for comparing regular public schools and charter schools.

I have argued that value-neutral objectivity is an unproductive guide for educational inquiry, and I have suggested two ways non-epistemic values might influence key decisions researchers make. But apart from considerations of significant truth and epistemic caution, is there a general principle or regulative ideal that might replace Phillips's VN approach and the epistemic purity it entails? I do not commit to a particular alternative here, but let me briefly sketch one possibility: objectivity as openness to refutation. This view requires that a researcher not disregard evidence on the grounds that it does not fit with one's values.²² Evidence can only be excluded on epistemic grounds, but moral, social, and political values can play a pervasive role in all other judgments — including framing decisions about what counts as evidence. The precise role of nonepistemic values is left open, depending on the nature and context of the research question. Objectivity as openness is not neutral, in Phillips's sense, because moral, social, and political values (among others) are not excluded as reasons for selecting a research design, identifying particular variables as important, or even reasoning inferentially from theory and evidence to conclusion. When combined with other scientific values, this form of objectivity may be all that is required for objective social science.

In practice, value-laden science that makes moral, political, and practical judgments explicit encourages epistemic values (like honesty), elevates academic dialogue, and makes research objective in ways Phillips's VN account cannot.

1. Diana Jean Schemo, "Charter Schools Trail in Results, U.S. Data Reveals," *New York Times*, August 17, 2004.

2. Caroline M. Hoxby, *A Straightforward Comparison of Charter Schools and Regular Public Schools in the United States* (Cambridge, Mass.: National Bureau of Economic Research, 2004).

3. Martin Carnoy et al., *The Charter School Dust-Up: Examining the Evidence on Enrollment and Achievement* (New York: Teachers College Press, 2005).

4. See Jeffrey R. Henig, *Spin Cycle: How Research Is Used in Policy Debates: The Case of Charter Schools* (New York: Russell Sage Foundation, 2007).
5. See Heather Douglas, "The Irreducible Complexity of Objectivity," *Synthese* 138, no. 3 (2004); and Harvey Siegel, "Epistemological Diversity and Education Research: Much Ado About Nothing Much?" *Educational Researcher* 35, no. 2 (2006).
6. Richard J. Shavelson, "Contributions of Educational Research to Policy and Practice: Constructing, Challenging, Changing Cognition," *Educational Researcher* 17, no. 7 (1988).
7. Claude M. Steele and Joshua Aronson, "Stereotype Threat and the Intellectual Test Performance of African Americans," *Journal of Personality and Social Psychology* 69 (1995).
8. Edward Wexler et al., *California's Class Size Reduction: Implications for Equity, Practice, and Implementation* (Berkeley, Calif.: PACE-WestEd, 1998), http://www.wested.org/policy/pubs/full_text/class_size/main.htm.
9. Amartya Kumar Sen, *Development as Freedom* (Oxford: Oxford University Press, 2001); and Anthony B. Atkinson, "The Contributions of Amartya Sen to Welfare Economics," *Scandinavian Journal of Economics* 101, no. 2 (1999): 173.
10. Helen E. Longino, *The Fate of Knowledge* (Princeton, N.J.: Princeton University Press, 2002).
11. D.C. Phillips, *The Expanded Social Scientist's Bestiary: A Guide to Fabled Threats to, and Defenses of, Naturalistic Social Science* (Lanham, Md.: Rowman and Littlefield, 2000), 232–4.
12. See Nancy Cartwright, *Otto Neurath: Philosophy Between Science and Politics, Ideas in Context* (New York: Cambridge University Press, 1996).
13. Elizabeth Anderson, "Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce," *Hypatia* 19, no. 1 (2004); and Helen E. Longino, "Essential Tensions — Phase Two: Feminist, Philosophical, and Social Studies of Science," in *The Social Dimensions of Science*, ed. Ernan McMullin (Notre Dame, Ind.: University of Notre Dame Press, 1992).
14. D.C. Phillips and Nicholas C. Burbules, *Postpositivism and Educational Research* (Lanham, Md.: Rowman and Littlefield, 2000), 55.
15. See chapter 6 in Philip Kitcher, *Science, Truth, and Democracy* (Oxford: Oxford University Press, 2001).
16. On the value-ladenness of research on human well-being, see Elizabeth Anderson, *Value in Ethics and Economics* (Cambridge, Mass.: Harvard University Press, 1993).
17. I follow the argument in Heather Douglas, "Inductive Risk and Values in Science," *Philosophy of Science* 67, no. 4 (2000).
18. The qualification "may" is important; research findings can support or contradict established social categories and then affect human behavior, changing the categorized beliefs or behavior. Ian Hacking calls this the looping effects of human kinds. See Hacking, *The Social Construction of What?* (Cambridge, Mass.: Harvard University Press, 1999), 34.
19. For a critical examination of the ideal of free inquiry, see chapter 8 in Kitcher, *Science, Truth, and Democracy*.
20. On "framing" in debates about facts and values in science, see Catherine Z. Elgin, "The Relativity of Fact and the Objectivity of Value," in *Between the Absolute and the Arbitrary* (Ithaca, N.Y.: Cornell University Press, 1997).
21. Shavelson, "Contributions of Educational Research to Policy and Practice," 6–7 (emphasis in original).
22. Objectivity as "openness to refutation" is roughly the same as Douglas's notion of "detached objectivity." I use the former because "detachment" is easily confused with "value-free." See Douglas, "Irreducible Complexity of Objectivity."

This essay developed out of many conversations with Denis Phillips and Rich Shavelson on the nature of scientific research in education. I am deeply indebted to both; however, the views expressed here are my own.