ON DEMOCRATIC ACCOUNTABILITY IN EDUCATION

Derek Gottlieb
University of Northern Colorado

INTRODUCTION

The ongoing attempts to measure educational quality, to hold schools responsible for their performance against these measures, and to use the resulting data to design improvement initiatives have yet to yield the promised results. A universal definition of “good education” seems to elude our grasp. Despite broad public support for “accountability” in education, actual efforts to hold educators accountable for their quality over the past 15 years have tended to inspire rebellion from the very publics that these efforts are meant to serve.

In what follows, I attribute these persistent failures of accountability policy to two closely related but distinct assumptions about educational quality, the objects that policy seek to measure, to evaluate, and to act upon. The first of these pertains to the ontology of educational quality; the second pertains to the politics involved in measuring, judging, and improving educational processes. Taken together, these assumptions would leave the tasks of defining and measuring educational quality, and of prescribing interventions in educational practice, to be undertaken by experts in social and data science, as mandated by federal and state statute.

In seeking a new way forward, one that takes seriously popular demands for robust educational accountability, I will show that the ontological and political assumptions undergirding current accountability policies represent putatively hard-headed and objective defenses against the relativistic specter of what MacIntyre calls “emotivism”—the specter of “evaluative judgments” that are “nothing but the expressions of preference, expressions of attitude or feeling.” Crucially, the dilemma between the relativism that policymakers fear

---


and the objectivism to which they unsuccessfully turn is faulty. A certain version of Ordinary Language Philosophy (henceforth OLP), one derived from Wittgenstein’s middle and later works, and one associated with such exegetes as Stanley Cavell, Hubert Dreyfus, Naomi Scheman, Linda Zerilli, Toril Moi, and Simon Glendinning, among others, offers a way out of the false choice between a publicness conceived as objective universality and a privacy conceived as subjective and unshareable particularity. Following this path leads to a truly democratic form of educational accountability.

THE ONTOLOGICAL PROBLEM

What kind of a thing is “good teaching”? In advocating for certain measurement tools, proficiency levels, and the like, policymakers tend to define educational quality according to its effects. Raj Chetty’s 2011 and 2014 studies are the paradigm cases of this approach. Working backward from certain facts about individuals’ lifetime earnings, college attendance, etc., to the test scores of those individuals as schoolchildren, to the teachers responsible for those students’ education, and ultimately to the atomized practices of the teachers themselves, this method of determining the elements of educational quality and pedagogical practice renders this quality, and these practices, universal and abstract, singular.

But educational quality is not ontologically singular or reducible to a straightforward definition, however derived. Important evidence for the ontological multiplicity of “good teaching” comes from such unlikely sources as Nicholas Kristof. Kristof, defending Chetty’s approach to defining educational quality in The New York Times against skeptics, narrates the story of Alabama judge Olly Neal, whose life was transformed by the intervention of an exceptional teacher, Mrs. Grady. Kristof expects this narrative to underscore the importance of excellent teachers and so to support the idea that robust

Routledge, 2014), 1–4, similarly departs from this position, associated with G.E. Moore, for the similar reason that “emotivism” seems to deny something important about the reality of the value concepts that we use.

5 Stanley Cavell, Must We Mean What We Say? (Cambridge: Cambridge University Press, 1969), 261, describes this sense of privacy not only in the familiar mode of others being sealed out, but also in terms of the subject’s being sealed in.


7 As in Doug Lemov, Teach Like a Champion: the 49 Techniques that Put Students on the Path to College (San Francisco: Jossey-Bass, 2010).

accountability metrics of the kind Chetty proposes will help improve education wholesale.

As careful analysis shows, however, those accountability measures would not have been able to detect Mrs. Grady’s quality at all.9 The narrative Kristof relates certainly does underscore the importance of excellent teachers, but it reveals (a) that it is possible to grasp and to communicate such quality without recourse to the metrics for which Kristof advocates and (b) that if Chetty’s measures, or any other set of measures, are taken to absolutely define “educational quality,” those definitions will exclude certain cases of excellent teaching, and this is precisely because questions of educational quality are irreducible to value-free facts but are always bound up with one of several overlapping visions of the good.

In the run-up to ESSA’s passage, the assumption of ontological singularity was massaged by the inclusion of “multiple measures” of school or teacher quality, which purported to take account of these overlapping visions of the good. But this multiplicity was largely illusory. ESSA, for instance, requires the inclusion of a non-academic measure in states’ overall school rating systems, and that was supposed to introduce something like value plurality. Instead, a vast majority of states have chosen to use “student attendance” for this purpose because (a) it requires no additional resources since schools were tracking that data anyway, and (b) it correlates so powerfully with test scores.10

However, as Schneider has found with respect to evaluation,11 and as Lisa Delpit and Pedro Noguera have found with respect to practice,12 what it means to say that a school or a teacher is doing well depends profoundly upon who is asking, and why. Different stakeholders in different communities look to their schools to accomplish different ends. The kind of value pluralism at issue here cannot be settled by appealing to a common element—such as academic achievement—that all stakeholders agree belongs to the notion of “quality education.” After all, to say that all stakeholders value academic achievement is not to say that they value it equally or for the same reasons.

This pluralistic dynamism lies at the heart of American educational life, and yet our accountability procedures address it with an unacceptable clumsiness. Certainly a large part of the reason has to do with the fact that accountability is a state power, and states do not excel at perceiving local nuance.13 Kwame Anthony Appiah puts it colorfully, labeling it the “Medusa

---

10 Derek Gottlieb and Jack Schneider, “Putting the Public Back into Public Accountability,” in *Phi Delta Kappan* 100, no. 3 (2018): 29–32.
11 Schneider, *Beyond Test Scores*.
syndrome”: “When the state gazes at us—with its identity cards, educational stipulations, and other instruments of recognition—it invariably fixes and rigidifies a phenomenon that is neither fixed nor rigid.” 14 When the alternative to the state’s Medusa gaze is conceivable only in terms of an anarchic melee among purely private preferences or perceptions, though, the appeal of a single measure to rule them all is easier to understand. This brings us to the political problem.

**The Political Problem**

Since the twin events of *Brown v. Board* and the launch of Sputnik in the 1950s, as well as the waves of reform initiatives that proliferated in their wakes, accountability policies have pursued two goals that seem often to pull in different directions: the imperative to provide all children equal access to high-quality education and the imperative to foster international competitiveness. But America’s decentralized education system, including highly variable processes for and incentives tied to educational evaluation, posed a threat to both of these goals. 15 In the 1940s, (justified) mistrust of local discretion in the Southern criminal justice system called for standardization and federal oversight; 16 and that initiative came on the heels of the 1935 Social Security Act’s standardization and federalization of the nation’s poor laws. 17 When demands for justice and need to project national power required systematic reform in education, then, standardization and federalization were the most obvious strategies.

Reformers advocating for greater standardization in educational processes regularly framed matters in the same terms that civil rights activists had wielded before them: the lack of federal oversight enabled a feudal, backwards, and patchwork set of structures to proliferate, and this needed overhauling in the name of justice. Arne Duncan’s references to education reform as “the civil rights issue of our generation” highlights the link. 18

In the popular imagination, disinterested and enlightened standardization pursues the general good, while a multiplicity of local governance structures fosters superstitious prejudice and injustice. The solution to the possibility of discretion’s going awry is to preclude the intervention of discretion altogether.

However, by the same token, the quality of the justice ensured by an objective power mechanically following certain procedures depends on the content of those principles and procedures. The stories of welfare and criminal justice reform over the course of the 20th century are the stories of

---

mechanistically minded administrative structures, built with every intention of *precluding* the commission of indignities against marginalized populations, having their prime directives reoriented toward *redoubling* those indignities.\footnote{Tani, *States of Dependency*; Murakawa, *The First Civil Right*; Heather Schoenfeld, *Building the Prison State* (Chicago: University of Chicago Press, 2018).}

The notion that objectivity, distance, and standardization necessarily align with substantive justice ought to be examined with a jaundiced eye.

It is precisely this jaundiced perspective that we should adopt with respect to education accountability policies since NCLB. That law mandated universal testing and transparent reporting of school performance, disaggregated by various historically marginalized subgroups. The noble intention was to ensure that schools would not provide differential educational experiences to, say, their white and their Black students; sanctions would follow if schools failed to make adequate yearly progress (AYP) with any subgroup over a certain number of years. Ideally, the law’s incentivizing power would force schools to bestow upon their Black students the very same “just and loving gaze” they already shone upon their white students.\footnote{Iris Murdoch, qtd. in Toril Moi, *Revolution of the Ordinary* (Chicago: University of Chicago Press, 2017), 227.}

However, with schools’ white students in comparatively little danger of failing to make AYP, and Black students in comparatively greater danger, schools found themselves incentivized to relentlessly monitor the progress of their Black students and to narrow their curricula to the tested domains.\footnote{Daniel Koretz, *Measuring Up* (Cambridge: Harvard University Press, 2008).}

Rather than an equal practice of loving attentiveness, the pedagogical and administrative dynamics endured by marginalized children were something closer to what Simone Browne calls “the surveillance of blackness.”\footnote{Simone Browne, *Dark Matters* (Durham, Duke University Press, 2015).}

These two phenomena—loving attention and oppressive surveillance—are obviously opposites in the most important ways. One is clearly preferable to the other as a matter of both practice and policy, and they are not particularly difficult to distinguish at close range. However, when characterized in abstract terms, as in both NCLB and ESSA, surveillance and attention so closely resemble one another that policymakers cannot reliably tell which one a school is doing.

The purely objective perspective cannot get a grip on the nuances of particular educational situations and so are prone to misjudging them.\footnote{Kim Cook, officially judged ineffective despite being honored as her school’s teacher of the year, and Patrick Boyko, similarly discrepant, represent two such discordant instances. Valerie Strauss, “Lawsuit: Stop Evaluating Teachers on Test Scores of Students They Never Taught,” *Washington Post* (Apr. 15, 2013); Lisa Gartner, “Confused by Florida’s Teacher Scoring? So Are Top Teachers,” *Tampa Bay Times* (2014) https://bit.ly/2E3Zfl0.} And precisely because this objective perspective takes itself to be dealing with facts
rather than values, any local disagreement with accountability procedures’ output is taken to express simple ignorance about what educational quality is.\textsuperscript{24} This dynamic not only enables the persistence of harmful forms of injustice and inaccuracy in education, but also forecloses the political space in which to negotiate the values and the knowledge our educational system is entrusted with inculcating.

The ontological and political problems are therefore mutually reinforcing. Since the (educational) good is ultimately assumed to be universal and abstract, epistemologically available via certain objective procedures, no politics is necessary at all. Or, rather, the activity of politics extends no further than calculation. Evaluating educational experiences through procedurally objective means not only conceals and discounts the deep value pluralism in conceptions of educational quality but also forecloses the political space in which this pluralism might arise and further eviscerates the political praxis of living out, rather than simply avoiding, our society’s pluralism. OLP’s very different view of the way in which people share forms of life, including the objects and projects in which their “Being is implicated,”\textsuperscript{25} helps avoid the dilemma between the world conceived as either universally available or as a pure product of private perception.

**JUDGMENT AND EVALUATION**

Contemporary education reform, welfare reform, and criminal justice reform pose the same kind of problem for politics. Historically speaking, a problem of bad judgment was to be solved by forgoing the requirement for judgment altogether—by substituting facts for values, roughly.

But judgment is not, in this picture, actually eliminated; it is only hidden from view, exercised now with respect to the rules and definitions that make particular cases what they are rather than to the cases themselves. MacIntyre associates this sort of tradeoff with analytical philosophy’s approach to moral reasoning: “Questions of ends are questions of values, and on values reason is silent; conflict between rival values cannot be rationally settled.”\textsuperscript{26}

This vision of the purpose of judgment, as adjudicating conflict between rival positions, is too restrictive in Linda Zerilli’s view. On Zerilli’s

\textsuperscript{24} This happens with considerable frequency. Education reporter Peg Tyre, for instance, calls for “unbiased, accessible information about what solid research tells us what works best in schools”—according to the singular definition, that is—in order to make parental choice initiatives valid. This is necessary to Tyre because “sometimes the parents don’t even know that the ‘choice’ they’re making is a bad one.” Peg Tyre, “Putting Parents in Charge,” *New York Times* (Sept. 17, 2011), SR 8. For Tyre, as for so many others, the notion that parents might legitimately value aspects of an educational experience not captured in official definitions is ruled out of court. Such parents are, by definition, wrong.


understanding, the focus on adjudication is deeply implicated in the elimination of a public space in which pluralism might thrive: adjudication becomes “a kind of theoretical obsession that might well lead us to misunderstand what is at stake in judging politically.” To the extent that the picture of things in MacIntyre’s quotation parrots the notion of Rawlsian or Habermasian “public reason,” for example, public reason asserts universal principles for deciding the validity of any moral or political claim, but it does so at the cost of meaningful discussions about those principles themselves. In order to adjudicate, these must be treated as transcendent. On the opposite side, a “noncognitivist” or “emotivist” position would locate the validity of a given claim in the individual and would trap it there, as it were: if it suffices as a reason for what I do, it sacrifices its power to influence you.

MacIntyre charts one route between the extremes of absolute objectivism and absolute subjectivism, but OLP, as Zerilli’s work draws upon it, charts another—a way not only of returning a praxis of judgment to the public sphere, but of centering it there.

OLP, as I use it, stresses both that “the ordinary”—what we do or what we say as a matter of unreflective, habitual practice—reveals the world as we inhabit it and also that this view is not improved, but is rather distorted by attempting to take up a theoretical perspective located outside of the world. The Kristof anecdote illustrates both elements perfectly. In narrating Olly Neal’s story of his transformation at Mrs. Grady’s hands, Kristof does indeed put excellent teaching on public display, and the fact that he expects the narrative to do this shows the sufficiency of an ordinary recitation of an example for this purpose. At the same time, Kristof intends the example to buttress the need for a theoretical definition of “good teaching;” but the fact that the example itself would fall outside of the theoretical technique that Kristof proclaims necessary for perceiving quality highlights the distorting potential of theory as such.

Zerilli connects the reduction of judgment to mere adjudication with just such a theoretical drive—an urge to posit an abstract standard of correctness that will work across all cases. Wittgenstein’s therapeutic aim, as taken up by Zerilli, has to do with helping us choose to trust ourselves, plural, rather than the “crystalline purity” of our theories, when these two things point in different directions. As Naomi Scheman reads him, “What [Wittgenstein] expects us to find is that justification is a practice we engage in with particular other people for particular reasons, to lay to rest particular worries—and that sometimes it

28 Gert Biesta, "Why “what works” won’t work: Evidence-based practice and the democratic deficit in educational research," Educational theory 57, no. 1 (2007): 1–22 similarly connects the objectivism inherent in strictly defining, for research purposes, educational goals to a certain threat to democratic practices, and I take myself to be in league with his work.
works.” As Scheman continues, “To say that at this point we are on the terrain not of reason but of politics is to invoke a distinction that does no work. . . . Anything we think might ground our practices—whether it be Reality or Reason—is just one more thing to argue about.” For Wittgensteinians of this ilk, a Rawlsian attempt to specify rules for reasonable discourse is itself a political act that cannot, except by unjustifiable fiat, be rendered immune to objection.

In educational accountability, such attempts at boundary-drawing have been patently unsuccessful and for a variety of reasons. I cited the examples of Kim Cook and Patrick Boyko earlier—award-winning teachers declared “ineffective” by the procedures and definitions of Florida’s evaluation protocols—and I was able to cite these examples because they appeared in newspaper reports covering dissatisfaction with these accountability mechanisms. As a matter of phenomenological fact, then, constructing a theoretical definition of educational quality does not put a stop to political conversations about the justice of those definitions or the judgments they render.

The attempt to immunize such judgments from public reproach is straightforwardly ignoble. When Cathy O’Neill sets out to illustrate the problem of the increasingly algorithmic management of our lives, her first example is of teacher evaluation. “Like gods,” she says, “these mathematical models [of teacher quality] were opaque. . . . Their verdicts, even when wrong or harmful, were beyond dispute or appeal. . . . Instead of searching for truth, the score comes to embody it.” There is something morally wrong, she implies, something connected to the task of searching for truth, with the very attempt to render a principle or process “beyond dispute.”

This hint of moral turpitude reflects something important in the connection between imagining an objective definition of educational quality as necessary to adjudicating particular cases and the foreclosure of a crucial political space. To attempt such a foreclosure is to strip ourselves of an important capacity that we might put to good—even restorative—use. Against the “adjudication imperative,” Zerilli follows Arendt in asserting that “judging . . . must be a democratic world-building practice that creates and sustains human freedom and the common space in which shared objects of judgment can appear in the first place.” The notion that shared objects of judgment need a common space in which to appear maps the middle ground between an ineluctably unworkable objectivity and an inherently private subjectivism. The task of description in evaluation processes might access this space.

32 Zerilli, Democratic Theory of Judgment, xiii.
Wittgenstein says that philosophy’s job can never be “to reduce anything to anything, or to explain anything. Philosophy really is ‘purely descriptive.’” Wittgenstein elsewhere refers to this kind of description as “marshalling reminders for a particular purpose.” Understanding pure description as bringing shared objects of judgment into a common space requires that the act of description be tied to a particular need—that there be a reason for entering a description.

The presupposition of such a purpose, or a need for response, is an acknowledgment of the structural publicness of our lives with others—others whose needs are intelligible and demand responses. To acknowledge that we say things, or enter judgments, for particular reasons is to deny on a basic level the purifying effects of the fact-value distinction and own up to the fact that our utterances are always value-laden. Cavell notes that for any utterance to have a point, which an utterance must in order to be meaningful, it is to that extent “an expression of value, is something found worth uttering.” Or as he says elsewhere, “what can be communicated, say a fact, depends on our agreement in valuing, rather than the other way around.”

The necessary inclusion of values and of others, even in our propositional utterances, conjures what Arendt refers to as the “common world,” which we maintain partly through ongoing conversation about what we see and what we value. As Zerilli says, “politically speaking, it is not just a matter of recognizing something to be an objective fact but of recognizing, counting, or acknowledging this fact as meaningful for what we do or do not have in common.” Instead of looking away from others and toward some third thing (a metaphysical citizenship, say, or an explicit joining of wills) in virtue of which we are a community and do things in just this way—a form of life which just is, ultimately, shared—Zerilli’s view of things turns our gaze and attention to others with whom we take ourselves to be sharing a form of life. Owning up to the political responsibilities of judgment would mean to treat agreeing in judgments, agreeing in form of life, not as a fact from which to start, but as a task to be undertaken.

**Conclusion: Representative Democracy and Educational Accountability**

In the summer of 2009, Arne Duncan toured the country to drum up support for the various elements of his Race to the Top program, including the

---

idea that slipshod teacher evaluation structures were to blame for American schools’ underperformance, and that standardizing and rationalizing evaluation protocols represented the best solution. In a pattern typical of American policymaking over the 20th century, Duncan equated the absence of universal procedures and measures with the impossibility of judging educational quality.

The solution he proposed, however, as we have seen, would be a technocratic one—an attempt not at engaging a pluralistic, democratic citizenry in handling questions of educational quality, but rather at deflecting responsibility for acknowledging pluralistic value demands. This paper has shown that such attempts cannot actually escape the realm of politics, merely displacing, instead, the locus of political agency in ways detrimental both to the mutability and plurality of educational values and to the praxis of democracy itself.

Between the horns of requiring an absolute standard of judgment and an inability to judge at all lies a practice of evaluating education that would embrace rather than evade the positionality of a given school’s stakeholders: families, students, administrators, neighbors, and so on. The very fact that such people live together in a community and yet are differently positioned with respect to the conduct of a school indicates the possibility of what Arendt would recognize as genuinely objective judgment. Her view holds that objectivity is not to be achieved by seeking the famous view from nowhere; it is rather to be sought in the perspectives of differently positioned others. Cavell agrees: a convening of stakeholders “provide[s] an arena in which finality of judgment about actions is backed up by certainty, and in which certainty is essentially a matter of seeing.”

For Cavell, treating the practice of evaluation in this way—as inherently political insofar as it points out and names the meaningful aspects of a given school’s conduct—is necessary to a healthy democratic community. James C. Scott, whom I cited earlier, asserts that distant entities like the state simply do not see well when it comes to particulars; a better way to achieve insight, and therefore justice, is to make the Arendtian turn to others. This is the sense of objectivity in which “perspectives are corrigible not by something that is extraperspectival or neutral but by other perspectives.”

As Biesta shows, this is a meaningfully different notion of democratic practice than that espoused by Dewey or Guttmann. I would align it with Cavell’s sense of Emersonian “representativeness,” which he describes as a willingness to stand for others in reciprocal relation. He calls this “living as an

---

example of human partiality,” and he notes that “this is not a particular moral demand, but the condition of democratic morality; it is what that dimension of representativeness of democracy comes to which is not delegatable.”

43 Scott, *Seeing Like a State*, 318 contrasts “partisan knowledge” with “generic knowledge” in a way that echoes Cavell’s “partiality” here. “Partisan,” for Scott, indicates that the “holder of such knowledge typically has a passionate interest in a particular outcome.” Mentioning this echo has the effect of reading Cavell’s “living as an example of human partiality” as combining Arendt’s or Scheman’s sense of partiality as (human) incompleteness with Scott’s sense of “partisan” as indicating (human) attachment to a particular world.