Theses and Dissertations

Spring 2013

Race to the top and the senses of good teaching

Derek Gottlieb
University of Iowa

RACE TO THE TOP AND THE SENSES OF GOOD TEACHING

by

Derek Gottlieb

ABSTRACT

Following up on the education reform initiatives of the 1990s and early 2000s, which are centered on the notion of accountability, Secretary of Education Arne Duncan's Race to the Top initiative strives to bring such accountability down to the level of the individual teacher through the use of advanced statistical parsing of student achievement data. Through the calculation of "teacher value-added," a given teacher's "effectiveness" can be measured and ranked, hence assigned a value. Duncan's speeches around the issue, and the assumptions visible in the studies of teacher quality and effectiveness that he and other reformers cite, suggest that at long last we as a society will be able to know and to communicate just who our best and our worst teachers are. Such an ability will allow us as a polity, on this view, to direct public funding much more efficiently than has heretofore been possible: armed with this new knowledge, we can reward the best teachers to ensure that they do not abandon the profession for higher-paying employment, and we can cull the worst teachers so as to replace them with more effective personnel.

The newfound ability to distinguish between good and bad teachers also has transformative implications for teacher training programs. By analyzing the practice of the highest quality teachers, one might discover "what works" in classrooms, the specific behaviors, skills, or mental states involved in highly effective teaching. Once discovered, these behaviors, skills, or mental states might then be given to pre-service teachers, which would dispense with what Duncan considers to be the overly theoretical and largely abstract curricula of current teacher education programs.

The problems that arise in attempts to implement these accountability measures for evaluation and training purposes are philosophical in nature, I contend. The method of investigation that follows is therefore best characterized as a conceptual analysis of Race to the Top's teacher-quality and achievement-data initiatives, comparing enacted policies to the Secretary of Education's public rhetoric employed to market those policies to the public. Taking the public rhetoric as an expression of the various needs to which the policies will be responsive, this thesis tests the coherence of the underlying assumptions about teaching and learning, and it assesses the conceptual fit between the needs visible in the rhetoric and the outcomes sought and measured according to the resulting policies.

The thesis finds that Duncan's public rhetoric expresses largely unproblematic needs, fears, or disquietudes around questions of teacher quality, but that the policies intended to answer those needs are entirely insufficient to the task, and in fact destructive in a number of ways. At issue is a misconception of teaching as a skillful endeavor, a mistaken idea about what teaching *is*. This thesis concludes that any adequate response to the needs expressed in Duncan's speeches will require a different view of teaching and learning. The thesis offers the basis for the formulation of a view better suited to answering the needs of the public, as expressed by the Secretary of Education.

Abstract Approved: _____
                    Thesis Supervisor

                    _____
                    Title and Department

                    _____
                    Date

RACE TO THE TOP AND THE SENSES OF GOOD TEACHING

by

Derek Gottlieb

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Teaching and Learning (Curriculum and Supervision)
in the Graduate College of
The University of Iowa

May 2013

Thesis Supervisor: Professor Peter S. Hlebowitsh

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Derek Gottlieb

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Teaching and Learning (Curriculum and Supervision) at the
May 2013 graduation.

Thesis Committee: _____
                Peter S. Hlebowitsh, Thesis Supervisor

                _____
                David Stern

                _____
                Carolyn Wanat

                _____
                Leslie Schrier

                _____
                Gregory Hamot

To, for, and on the behalf of my many teachers.

My wish to keep a number of strata of ideas moving in the same direction, seeming to require perpetual modulations as well as multiple themes, strikes [some] as an improvisation meant to link insights that, like punch lines, can be reached no other way, but the cost of which is the persistent danger of abusive obscurity.

Stanley Cavell
Little Did I Know

ACKNOWLEDGMENTS

TABLE OF CONTENTS

INTRODUCTION

In the early part of 2012, a number of media and governmental outlets trumpeted a new round of reform-minded educational initiatives and discoveries in the American landscape. On January 24th, President Obama heralded the importance of good teachers to a classroom's future economic prospects in the State of the Union address. On February 16th, The Secretary of Education, Arne Duncan, appeared on *The Daily Show* to promote the RESPECT initiative that the department had introduced on its website and with a press conference earlier in the same day. In two separate columns, both appearing prior to the State of the Union address, the Pulitzer Prize winning *New York Times* columnist Nicholas Kristof took up the ongoing discussion of education reform in the United States in the same terms, even gesturing to the same stakes, as both Obama and Duncan (Kristof, 2012a, 2012b). The columnist succinctly sketches the outline of the discussion, cites new research in the field (Chetty, Friedman, & Rockoff, 2011), and implores his readers to recognize the importance of good teaching. But what exactly *counts as* good teaching, as these various figures use the term? Combing through their tropes and studying the evidence they cite – and more to the point, the way in which they make use of the evidence that they cite – will help one to understand the view of good teaching under discussion here.

The amalgam of source material cited above also bears preliminary mention. In venturing into an examination of the nature and consequences of identifying something like "good teaching," one reaches, of necessity, beyond a subject-matter conceived in strictly academic or scholarly terms. The distinction to be made appears in witnessing the difference between the Chetty, et al article mentioned above, and the nature of the venues in which it is cited. The Chetty article is academic in nature. It seeks to explore the linkages between

high-performing teachers (defined here as the top tier of achievement-score lifters) and the socioeconomic outcomes of the students in those teachers' classes. Characteristic of such approaches is the appropriately strict and explicit definition of terms, the limited scope of the study (so as to achieve clarity on a matter quite specific), and most importantly a heightened consciousness of both of the foregoing points, which is visible in the researchers' warnings about "evaluating the policy implications of our findings" (Chetty, et al., 2011, p. 5), given that the conditions under which their data was generated would change substantially should such findings be used to drive policy. In other words, the researchers have drawn a strong correlation between "effective teaching," as they define it, and improved financial and social outcomes for the students of that teaching, as they likewise explicitly define it; they strongly caution that this ought not to be confused with having established a causal relationship between such generally undefined terms as "good teaching" and "a better future" – a lesson from Statistics 101.

The 2012 State of the Union address, the official press release and the *Daily Show* interview concerning the RESPECT program, as well as the two Kristof articles, however, are examples of *public* discourse, political speech. These uses of academics' work starkly reinforces the fact that while education may be an object of academic or scholarly attention, it is never *merely* academic or scholarly. It is a matter of practice, both in the conduct of teaching and learning itself and in the sphere of political and social maneuvering, as well. As an area of investigation, then, it would be both conceptually inaccurate and inherently delimited to confine one's field of vision to studies like Chetty's. The Chetty article demonstrates precisely what it seeks to demonstrate, and no more. In fact, the authors are explicit on the boundaries of their findings. The fact that commentators and policy-makers

draw conclusions beyond these boundaries and refer, explicitly or implicitly, to research like this in support of those conclusions represents the profound and consequential interdependence of these discursive fields. The application of findings drawn in severely controlled experimental contexts to the wider world of educational policy, as though these findings really *did* represent a kind of uncontestable fact about "good teaching" far beyond the scope that the researchers themselves claim for their conclusions, makes readily visible the apparent continuity of the concept of "education" that both discursive fields address, as well as the stark distinction in the terms of the two forms of address. This sharp distinction, in turn, reveals a certain blurriness surrounding the *concept* of "good teaching" that each field seeks to understand.

For this reason – that political and public discourse inherently bears upon an academic or scholarly research agenda and vice versa – the references to media outlets like *The Daily Show*, *The New York Times*, and others are not only warranted, but absolutely necessary. In the academic field of education, one's findings will be put to practical use in the public domain, and so the nature of that political or public praxis remains itself within the field of education as an object of study, particularly where the conceptual stability of "good teaching" is at stake.

The two Kristof articles highlight the problems inherent in establishing the conceptual stability of good teaching most clearly, and therefore provide the most auspicious starting point for the discussion.

The first opinion piece, which appeared on January 12th, makes the following statements about the importance of good teachers: "A great teacher is worth hundreds of thousands of dollars to each years' students, just in the extra income they will earn. . . .

Conversely, a very poor teacher has the has the same effect as a pupil missing 40 percent of the school year. We don't allow that kind of truancy, so it's not clear why we should put up with such poor teaching" (Kristof, 2012b). Later in the same piece, Kristof says "Our faltering education system may be the most important long-term threat to America's economy and national well-being, so it's frustrating that the presidential campaign is mostly ignoring the issue." And finally: "This latest study should elevate the issue on the national agenda, because it not only underscores the importance of education but also illuminates how we might improve schools. An essential answer: more good teachers. Or, to put it another way, fewer bad teachers. The obvious policy solution is more pay for good teachers, more dismissals for weak teachers."

Before segueing to Kristof's second article, which seeks perhaps to humanize the above picture, I would like to draw out clearly the terms in which Kristof characterizes the importance of good teaching and the urgency of developing policy that addresses the needs that he emphasizes. Firstly, Kristof posits an overarching axiological statement in declaring that our "faltering" educational system is crucial precisely to the degree that "America's economy and national well-being" are at stake. This statement functions in his argument as an indicator of urgency. Nothing less than our national economic health – and therefore our sovereignty and global influence – hangs in the balance.

Without venturing too far into or developing a full-fledged argument *against* such a statement here, it is at least worth mentioning that this statement makes two questionable assumptions. That our education system is "faltering," or failing, or decrepit, or whatever term of political art is most appropriate these days, is simply taken for granted. In pointing with the methodology of thick description to singular examples of the most awful conditions

in the public schools nationally – which almost always occur in high-minority, high-poverty, inner-city schools – one can easily find horrific instances of obvious failure. Then, in pointing to national and international data on high school and college completion rates, as well as international measures of achievement, one invites extrapolation of the horrific conditions to the general state of public schools, and *voila*, one sees systemic failure, a "faltering" *national* education system. The generalization from horrific individual cases, backed with data comparing our graduation rates and test scores with other countries', is not necessarily supported by the data. In questioning the assumption that our educational system is "faltering," I wish first of all to underscore *how deeply ingrained* the rhetoric of crisis is within the educational debate. Mainly, however, I wish to question the assumption of *systemic* failure, which would then call for *systemic* solutions.

The second assumption that Kristof makes has to do with the particular relation he imagines between our "faltering" public education system and our "economic and national well-being." On the one hand, it is easy to see how one might easily confuse this relation: the specific Chetty, et al. study he cites seems to promise – though its authors explicitly foreclose on this promise in the article itself – that better teaching causes more money in student lifetime (i.e. future) earnings; more future money causes higher future GDP; higher future GDP equals better future "economic and national well-being." That the relation is more complex than this simplistic rendering ought to be obvious, but, perhaps due to the deeply-ingrained crisis narrative to which I alluded above, the simplistic relation is almost always taken as truth.

In the 1960s, during the height of the space- and arms-races, the Kennedy administration lavished the fields of math and science with monetary and rhetorical attention,

in the same crisis terms (Berliner & Biddle, 1997, pp. 144-148). Crisis rhetoric aside, the United States convincingly won those races. But in 1983, *A Nation at Risk* came out, famously declaring that "If an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war" (Gardner, 1983), which certainly indicated that our education system was still mired in some sort of crisis. The Soviet Union collapsed during the years 1989-1991, and the United States moved into a position of unipolar dominance. In the 1990s, fueled by superlative levels of financial and technological innovation centered around Silicon Valley, the United States economy experienced a decade-long, unprecedented level of expansion. Conspicuously absent from reflections upon winning the Cold War and experiencing an economic windfall were accolades for the American education system; in fact, given the pervasive continuity of the crisis narrative, one might well conclude that the United States accomplished these feats in direct *spite* of its educational system's obvious flaws. The crisis narrative has remained untroubled by the United States' global dominance over the past half-century, even to the point of defeating its great political and economic rival. But today, with a global economy struggling to recover from a recession, politicians and policy-makers once again point to the United States education system as a cause of our economic problems, and thus the economic downturn becomes another fact that indicates the dire need for educational reform.

Against Kristof's assumptions that shape the contour of the debate, the foregoing discussion reveals that one of two things must hold: either the metrics that we take to indicate the systematic failure of our education system in fact indicate no such thing, or else quality education has far less to do with "economic and national well-being" than we assume. The

fact that the crisis narrative seems to function as a foundational truth immune to facts could

indicate either.[1] But in either case, what Kristof proposes we do in order to *simultaneously*

address the metrics that indicate systemic educational failure and our "economic and national

well-being" appears suddenly problematic.

The problematic nature of Kristof's assumptions manifests itself in considering the

relative possibility of addressing the future well-being of the nation by means of addressing

the indicators of our "faltering" educational system. Kristof's proposal for schools, which is

also Obama's and Duncan's, is "more good teachers. Or, to put it another way, fewer bad

teachers." Because he treats the problem in simple cost-benefit terms, the fact that good

teachers are correlated with certain economic gains for their students indicates to him a

glaringly self-evident consequence: "The obvious policy solution is more pay for good

teachers, more dismissals for weak teachers." In fact, Obama's State of the Union address

makes a nearly identical point:

> We know a good teacher can increase the lifetime income of a classroom by
> over $250,000…. Teachers matter. So instead of bashing them, or defending the
> status quo, let's offer schools a deal. Give them the resources to keep good teachers
> on the job, and reward the best ones. And in return, grant schools flexibility: to teach
> with creativity and passion; to stop teaching to the test and to replace teachers who
> just aren't helping kids learn (Obama, 2012).

---

[1] By "immune to facts," I do not mean to imply that the crisis narrative is unsupported by
any facts. I mean that no evidence of successful teaching or learning is capable of
overturning the systematic nature of the narrative of systematic failure. In the interview with
Jon Stewart, Arne Duncan said that, "We're trying to put a huge amount of resources behind
those places to take to scale what is working. I would say for every challenge we face in
education, the answers are out there. It's happening. The challenge is, it's not happening at

Both Kristof and Obama cite the Chetty study in attributing the economic gains to "good teachers," and in doing so they each tacitly smuggle in the study's rigorous definition of "good teaching," which has *only* to do with the amount of improvement students demonstrate in reading and math achievement scores. Importantly, however, neither Kristof nor Obama make explicit reference to this particular definition, instead using the concept of "good teaching" in its ordinary, everyday, non-technical sense. The State of the Union drives this point home as Obama appeals to the *everyday familiarity* of the experience of having a good teacher: "Every person in this chamber can point to a teacher who changed the trajectory of their lives"(Obama, 2012). The operative assumption, then, is that *there is no difference* between Chetty's technical definition of "good teaching" and the kind of "good teaching" so generally present and available that Obama can rely upon the experience of "*every person* in this chamber" (emphasis added) in order to make his point.

I would be remiss to let the quoted passage from the State of the Union address drift by without comment, and so before turning to Kristof's second opinion piece, I wish to highlight Obama's proposed "deal" with our schools, in which the schools will "reward the best" teachers and in exchange, they can "stop teaching to the test." Bearing in mind that a "good teacher" in both Kristof and Obama's discourse is identifiable in terms of his or her ability to raise student test scores, Obama's proposal to "reward the best ones" represents a straightforward call to increase the salaries of teachers whose students' scores climb most dramatically. In the entirely disingenuous sense that the *schools* will presumably no longer be penalized based on its students' achievement data, it is not technically false to claim that it is "the schools" that are "granted flexibility… to stop teaching to the test." But the notion that "teaching to the test" will disappear as a result of states' signing on to Obama's deal is

exceptionally unlikely. Under NCLB, Obama implies, teaching to the test occurs out of a natural urge to avoid federal sanctions. Under his proposal, however, *individual teacher compensation* will (somehow or other) be a function of achievement score growth, which, recall, is the criterion of "good teaching." Where under the earlier policy, schools were institutionally incentivized to encourage teachers to teach to the test – a practice Obama's speech marks as misguided – under his new plan, teachers themselves are *directly* incentivized to teach to the test. This isn't turning away from a dysfunctional practice; this is cutting out the middleman.

In order to turn to Kristof's second article, now, I wish to emphasize again the two senses of "good teaching" that occur in the State of the Union: on the one hand, there is the "good teaching" so familiar that the President can assume that "every person in this chamber" has the memory of such an experience; while on the other hand, there is a kind of "good teaching" that is knowable (only) by means of looking at gains in student achievement scores. The assumption, once more, is that these kinds of good teaching – one that reflects an explicit and technical definition, and one that reflects an experience of everyday school life – are (more or less) the same. In fairness to Kristof, Obama is the one who directly conflates the senses of "good teaching." Nothing in Kristof's first article refers to any ordinary, non-technical sense of teacher quality, remaining wedded, as it is, to the Chetty article's definition.

However, in Kristof's second article, the columnist does indeed risk conflating the two senses, and in particularly egregious ways. This egregiousness itself, and the fact that he cannot see the way he runs together different senses of teacher quality, indicates the level of difficulty at hand.

Kristof pens his second article on January 22nd, as a rejoinder to some feedback on his first article. He says, "skeptics of school reform wrote me to say: sure, a great teacher can make a difference in the right setting, but not with troubled, surly kids in a high-poverty environment. If you think that, or if you scoff at the statistics, then listen to Neal" (Kristof, 2012a). Perhaps now, a week and a half removed from his first article, he begins to allow his definition of "good teaching" to wander. Granted, at this point he is voicing the objections of "skeptics of school reform" when he says that "a great teacher can make a difference," but this no longer sounds quite like the "good teachers" that were the subject of Chetty's analysis. It is also important to note that Kristof here responds specifically to claims that socio-economic status and student temperament would confound the findings of VA analysis.[2] In his response, though, he opts not to return to the Chetty study that provided the impetus for the initial article – where such claims would be eminently and directly answerable – and decides to respond instead with the human narrative of Olly Neal. This decision itself speaks volumes.

Neal has the traits that Kristof requires in order to answer his critics. In terms of his temperament, Neal describes himself as "not a nice kid," as being "the only one who made [Mrs. Grady] cry." He "had a reputation." Neal was also poor, on which basis Kristof's critics objected to the commentator's championing of the reform agenda in question. As Kristof says, "Olly Neal was a poor black kid with an attitude. He was one of 13 brothers and sisters in a house with no electricity, and his father was a farmer with a second-grade

---

[2] It is, in some ways, too bad that Kristof decided to respond to these claims in particular. One of the *easiest* things to do, statistically speaking, is to partition out the effects of SES and other sociological factors, and Kristof might have simply pointed to the researchers' own algorithms in order to indicate that the VA metrics accounted for these traits.

education. Neal attended a small school for black children — this was in the segregated South — and was always mouthing off. He remembers reducing his English teacher, Mildred Grady, to tears."

Kristof here retells the story of a child with neither socio-economic status, nor "attitude," nor even race – "this was the segregated south" – on his side; and yet, it is Neal's story of having his life changed by a "great teacher" that for Kristof indicates that good teachers do indeed make an important difference in all cases, and thus we ought to adopt the reform agenda he lauded in the previous article.

I must summarize the story in order to bring out the salient points. In order to avoid the risk of distorting the tale, I will directly quote with some frequency. In 1957, "the fall of his senior year" of high school, Olly Neal "cut Mrs. Blakely's class" and "wandered in the library, set up by Grady, the teacher he had tormented." A book with a "risqué cover" "caught his eye," but he "didn't want word to get out to any of his classmates that he was reading a novel." "So I stole it," as Neal himself says. Upon returning to the library sometime later, he was surprised to find another book by the same author, and he stole that, too. "Four times this happened." As Kristof says, "His trajectory changed, and he later graduated to harder novels, including those by Albert Camus, and he turned to newspapers and magazines as well. He went to college and later to law school. In 1991, Neal was appointed the first black district prosecuting attorney in Arkansas. A few years later, he became a judge, and then an appellate court judge."

At a reunion some years later, "Grady stunned Neal" by revealing that she had known he was stealing books and that she had in fact been restocking the shelves with the author that he liked – "all in hopes of turning around a rude adolescent who had made her cry. She

paid for the books out of her own pocket." Kristof then asks, "How can one measure Grady's impact? Not only in Neal, but in the lives of those around him. His daughter, Karama, earned a doctorate in genetics, taught bioethics at Emory University, and now runs a community development program in Arkansas."

For Kristof, the "implication is that we need rigorous teacher evaluations, more pay for good teachers and more training and weeding-out of poor teachers. The need for more pay is simple. In the 1950s, outstanding women like Grady didn't have many alternatives, and they became teachers. Grady was black, so she didn't have many options other than teaching black children in a segregated school." Grady's impact, in Kristof's retelling, is difficult to measure – not only did she turn Neal's life around, but that intervention paved the way for Neal's daughter's success as well, a success which is, in turn, benefiting communities across the state. Grady is the model, here, of a great teacher.

It is precisely at this point that I would like to examine the way that Kristof returns to the "implication" he derives from this particular example. The example, first of all, is *without doubt* an example of an extraordinary teacher who achieved or enabled profound educational results even in a troubled, impoverished student. The question remains, however, as to whether this particular example supports the need for those "rigorous teacher evaluations," in defense of which Kristof offers the narrative in the first place.

A good test of whether there are in fact two different concepts of "good teaching" at work here would be to attempt to map Kristof's example onto the "rigorous teacher evaluations" he champions. For our current purposes, I assume that by "rigorous teacher evaluations," Kristof means teacher evaluations that are based on or otherwise tied to the teacher's ability to lift a student's scores on achievement tests. I trust that my assumption

about Kristof's notion of rigor is safe because of the words Kristof himself has used. Neal's story is offered, after all, as a response to those who "scoff at the statistics," – the statistics being the ones he offered in his previous article, presumably. The "good teaching" he has in mind, then, is specifically the "good teaching" visible by means of Chetty's definition. At issue in this exercise is the relative similarity of that definition to the everyday-familiar one that Obama has cited, and that Kristof's narrative also reveals. For Mrs. Grady is surely an example of our ordinary sense of great teaching – this everydayness stands out in the fact that no recourse to advanced metrics is required in order to see her excellence. But is Mrs. Grady an example of both? If so, we cannot be confident that any conceptual problem exists. If not, on the other hand, we must take that possibility into account.

Let us imagine that the Olly Neal story occurred under the regimen of "rigorous teacher evaluation" as Kristof demands, in the present day rather than in 1957. Would the rigorous evaluation agree with our judgment that Mrs. Grady is an excellent teacher? Would Mrs. Grady's excellence register by means of the metrics?

This is actually a very simple matter. One glaring fact leaps out before all the others: Olly Neal "cut Mrs. Blakely's class" in order to venture into Mrs. Grady's library in the first place. Mrs. Grady *had been* his teacher, but was no longer at the time of the manifestation of her teacher quality. While it is unclear whether Mrs. Blakely was Neal's *English* teacher – in which case *she* would be credited with lifting Neal's reading scores – under no circumstances whatever would any reading score gains have accrued to Mrs. Grady. Mrs. Grady's life-changing intervention on Olly Neal's behalf would not show up in any "rigorous" attempt to trace his achievement score gains back to their source.

In the second place, however, it is absolutely unclear that Neal would have been assessed to begin with, such that score gains would have existed at all. Neal was a high school senior; the law calls for yearly assessment between grades three and eight, and one assessment during high school. No state of which I am aware assesses seniors – as such assessments occur during the spring semester.

In the third place, it is not out of bounds to draw into question the likely size of any score gain. We know that Neal stole the first book "in the fall of his senior year," sometime after which, he absconded with books two through four. College applications are due in the late winter or early spring. When Kristof claims that Neal's "trajectory changed" and that "he went to college and later to law school," the commentator invites one to believe that prior to the first book-stealing, college and law school were not on Neal's agenda. That may be the case. But one's acceptance into college depends upon the demonstration of aptitude and accomplishment in a variety of realms and over the course of one's high school career. It is implausible to suppose, in other words, that Neal was *without* visible aptitude or academic skill prior to Mrs. Grady's intervention, even if he was a tough kid to handle in class. This is not to minimize Grady's impact on Neal's direction in life; it is rather to call into question the assumption that Neal would have performed terribly on standardized assessments of reading in previous years, such than any assessment following Mrs. Grady's intervention would have revealed substantial achievement *growth*, upon which Kristof's "rigorous teacher evaluations" would depend.

In the final analysis, it is impossible that the "rigorous teacher evaluations" that Kristof advocates would recognize Mrs. Grady's excellence in this case. I find it unlikely that a student of Neal's age would be assessed at all; and even if assessed, given the facts of

the case, it is not immediately reasonable to assume that Neal's change of trajectory would register in terms of any boom in his reading scores; but, most concretely, even granting the assessment *and* the score explosion, it is straightforwardly obvious that Mrs. Grady herself would miss out on the credit. It was *her* intervention, but she was not the one nominally responsible for teaching Neal how to read. Mrs. Grady is an example of excellence, but it is a kind of excellence that Chetty's definition excludes.

It seems reasonable to conclude on this basis that the public and policy discourse conflates two different concepts of "good teaching," erroneously treating them as identical to one another. Though Kristof holds up Mrs. Grady as the reason we need "rigorous teacher evaluation" in order to reward the good and weed out the bad, the very methods he defends would *fail to identify* Mrs. Grady's quality. And yet, it is simultaneously obvious that Mrs. Grady is an exceptional teacher. Two divergent senses of "good teaching" are thus in conflict here. Mrs. Grady's story appeals to an everyday sense of "good teaching," precisely that to which Obama gestures in the State of the Union address, as he asks the assembly to recall that "every person in this chamber" can remember an experience with such a teacher. The "rigorous evaluations" to which Kristof gestures, meanwhile, access a technical view of "good teaching" that fails to map onto the ordinary sense of the term. The assumption in applying advanced metrics to achievement data is that the potential discrepancy between the two senses can be eliminated with the proper application of statistical methodology. But it is no such matter.

Let us go further and charitably extend the idea of "rigor" to cover something like on-hand observational measures, since Arne Duncan has insisted on the appropriateness of using

"multiple measures" in determining teacher quality (Stewart, 2012). For the sake of

argument, let us also allow that Grady was Neal's teacher at the time of her intervention.

Kristof himself does not discuss the classroom-observation aspect of teacher

evaluation, but it, too, belongs to the Obama administration's reform agenda, and it is

factored into determinations of which states qualify for NCLB waivers and for Race to the

Top funding (DOE, 2011). The pairing of achievement data metrics with advanced

observational data is highlighted, in fact, in a recent newspaper article: "Spurred by the

requirements of the Obama administration's Race to the Top competition, Tennessee is one

of more than a dozen states overhauling their evaluation systems to increase the number of

classroom observations and to put more emphasis on standardized test scores" (Anderson,

2012). The sense of rigor in such evaluations derives from their standardized nature.

Observers look for specific characteristics and behaviors – best practices, one might say – in

a given teacher's planning and instruction. Tennessee offers a case in point, as detailed in the

aforementioned *New York Times* article:

> Each observation focuses on one or two of four areas: instruction,
> professionalism, classroom environment and planning. Afterward, the observer scores
> the teacher according to the state's detailed and computerized system. Instruction, for
> example, has 12 subcategories, including "motivating students" and "presenting
> instructional content." Motivating students, in turn, has subcategories like "regularly
> reinforces and rewards effort." In all, there are 116 subcategories (Anderson, 2012).

This particular description emphasizes the level of detail in the observation rubric,

which, while aimed at ensuring its rigor, simultaneously makes it unwieldy. But this

particular criticism misses the point in our case. Under which observational category, after

all, does one imagine "Allowing a student to steal a library book" to appear? Mrs. Grady's

decision to allow Neal to simply take the book was both *laudable* and *situation-specific*. It

was the right decision for that child in that place at that time, but as the rhetorical question makes obvious, it is simply not generalizable to *all* teachers in *all* circumstances for *all* children. Obviously, only generalizable traits can appear on an observation rubric, as one implicitly expects any teacher whatsoever to demonstrate or perform the trait in question. The on-site evaluation process would therefore *likewise* miss Mrs. Grady's excellence.

It becomes evident, then, that Mrs. Grady's action can serve as an example of great teaching not in *spite* of Neal's poverty or temperament (or anything else peculiar to him), but precisely *because* of them. Without taking Neal's particular history into account, one cannot see Grady's excellence. Without taking Neal's history into account, a neutral observer says, "I'm sorry, you let a kid steal a *book?*" Statistical methods of partitioning out race, socio-economic status, and other factors from the numbers themselves thus cut two ways, and it is important to record that fact. Stripping student characteristics away from achievement data may not in fact yield an unimpeded view of good teaching. It might instead obscure its visibility altogether.

In his second article, Kristof narrates the history of one great teacher's impact on an otherwise lost student. The teacher's greatness shows up clearly in Kristof's recounting, which he offers as a means of defending the main contention of his first article: we should pay good teachers more money and we should weed out the weak ones. Insofar as Kristof specifies methods of identifying good teachers, thus establishing a basis on which to decide whom to weed and whom to reward, the methods fail to accurately identify good teaching *even in the case Kristof cites* in defense of employing "rigorous teacher evaluation" to "improve national competitiveness, improving the ranks of teachers." He labors under the misapprehension that the technical definition of "good teaching" is, if not identical to, then a

reliable proxy for, the everyday sense of "good teaching" that we in fact wish to reward. The case of Olly Neal and Mrs. Grady profoundly undermines the compatibility of the two concepts at work in this discussion.

Whether or not to reward and punish on the basis of teacher quality remains beyond the scope of my concern here. Rather, I am struck by slipperiness of the concept of good teaching, the ways in which something so obvious that a simple narrative can effectively reveal it proves so onerous that it eludes or evades 116-point evaluation rubrics and the very best efforts of the economics field. The classroom-evaluation methods examine specific behaviors and practices on the one hand, and outcome measures on the other. If these means remain insufficient to demonstrate or secure the conceptual recognition of "good teaching" we wish to reward, on what basis is "good teaching" so immediately visible in Mrs. Grady's story? If not communicable in terms of specific practices or measurable outcomes, how else could our ordinary concept of "good teaching" become visible? If the aforementioned methods are as insufficient as they appear, how deeply has the conceptual confusion penetrated the popular and political discourse on the subject? How might an alternative approach arise, and what might it look like?

Mrs. Grady's example is so near to us that we need neither rubrics nor metrics to reveal its quality; yet it remains for all that beyond the very reach of our best rubrics and metrics. Perhaps our trouble in coping with teacher evaluation lies not within good teaching's mysteriousness, but rather in its profound familiarity.

This study will seek to explore the constructions of "good teaching" throughout the policies of the Obama administration, across the platforms of press releases, policy speeches, and state policies themselves; and it will likewise trace the contours of the discourse on

"good teaching" as it appears in commentary in national news media of all varieties. In exploring the shape of the conceptual distortion revealed in this Introduction, I will also develop and offer an alternative theoretical framework by means of which questions pertaining to teacher quality can safely – without, that is, requiring divergent concepts of good teaching or technical definitions – be posed.

Mrs. Grady did the right thing in Olly Neal's case, and it had a tremendous and lifelong effect. Teachers like her are worth retaining and rewarding. Our methods of identifying good teachers ought to be able to pick out Mrs. Grady's quality. I take her, then, as the test of any approach to questions of teacher evaluation: if it cannot capture Mrs. Grady's excellence, it cannot function as an indicator of teacher quality. Call it the Mrs. Grady test.

CHAPTER 1: KNOWING TEACHER QUALITY

In order to clarify the concepts of teacher quality circulating in the various discourses around American education under the Obama administration with an eye toward rendering them amenable to evaluation, augmentation, and correction, it will be necessary to explore the aforementioned discourses in some detail. In this introductory section, I wish first of all to make explicit the conceptual form of teacher quality as it appears in official policy, which includes for these purposes the Department of Education's Race to the Top application process, Secretary of Education Arne Duncan's speeches and interviews in regard to this initiative, and the substance and evaluation of various states' applications for these funds, insofar as they tailor state law and policy to the prevailing discourses on teacher quality. I do not assume, of course, that these various documents, policies, and speeches will reveal a single, self-consistent notion of teacher quality. Rather, my goal in this section is better characterized as drawing out and making visible the plurality of concepts of teacher quality that emerge in this discourse.

In approaching the material in question, I will trace the development of these ideas through their myriad forms, first in Department of Education press releases and Duncan's 2009 speeches to various groups, and then finally in their appearances in the form of state educational policies designed to win the federal money at stake. I will spend considerable time on both Duncan's policy speeches given during the summer of 2009 and on the notions contained therein as they later translate to state teacher-evaluation policies undertaken in attempts to win Race to the Top funding.

1.1 Introduction

In the summer of 2009, Arne Duncan delivered four speeches meant to lay the groundwork of the Obama administration's signal education policy. In each of these, and in fact through the four understood as a cycle of sorts, he announces the creation of what would become the Race to the Top Fund: as part of The American Recovery and Reinvestment Act of 2009, congress committed "$5 billion in competitive grant funding… to help advance these four reforms" (Duncan, 2009c). The premises underlying the reform initiative, the outline of the goals it hopes to achieve, and the skeletal structure of the eventual policies associated with the program are all visible in this early form.

One major element that must be confronted immediately, because it winds its way throughout Duncan's policy speeches and interviews from 2009 to early 2012, is the Obama administration's relationship to the educational policy of its governing predecessor: the No Child Left Behind Act (NCLB). Duncan opens his 2009 speech at the Governors Education Symposium with soaring rhetoric about the "genius of our system," visible in the fact that "the power to shape our future has, wisely, been distributed to the states instead of being confined to Washington" (Duncan, 2009c). A few moments later, he says, "When I was running Chicago's schools I knew that the federal government's role was to support our work—not to direct it or micro–manage it but [sic] to encourage, reward, and support the innovation and progress that were being made at the state and local levels." The Secretary is here at pains to distance the governmental efforts he will shortly announce from the perceived intrusiveness on states' autonomy of NCLB's stringent accountability measures, which remain highly contentious to this day (Gabriel, 2011). Duncan is careful to couch his

own initiatives in deferential terms, as "support" for "state and local governments, which are the real hothouses of innovation in America" (Duncan, 2009c).

The relation between local autonomy and federal control remains obscure, even in the context of this single introductory speech, however. Immediately after praising the distribution of educational power to the states, Duncan lauds Governor Jim Hunt for "call[ing] for common national standards when it wasn't politically popular" (Duncan, 2009c). Two sentences later, Duncan praises Governor Roy Romer for his role in the development of common standards, saying "we wouldn't have 46 states and three territories agreeing to adopt high common standards if it weren't for his hard work" (Duncan, 2009c). In adopting de facto national standards, of course, states voluntarily give over a large degree of their autonomy – which is to say, their autonomy consists of, and reaches its limit in, making the once-and-for-all commitment to a national educational agenda. I point this out early on merely to showcase the way in which Duncan is careful to characterize his own program as a bottom-up approach in contrast to NCLB, which has come to be viewed as top-down. It ought to be noted, however, that both programs involve making a certain infusion of federal money earmarked for education in general contingent upon individual states (autonomously) adopting policies designed to meet specific requirements established by the federal government. In order to receive the money promised under the original Elementary and Secondary Education Act (ESEA), its later incarnation as NCLB, and the Race to the Top program alike, states must conform their nominally autonomous education systems to meet federal standards. The basic relation of the federal government to state education agencies (SEAs) does not necessarily bear out the profound distinction Duncan wishes to draw in this speech, or in subsequent ones.

But the insistence on the "genius of our system" reveals a deep rhetorical commitment to the centrality of local control in the basic structure of Race to the Top, in light of which the fundamental aims of the program ought to understood.

The very name of Race to the Top reflects a dramatic turn away from NCLB accountability methods. NCLB requires the annual assessment of all students in grades 3-8 and once again in high school in order to evaluate whether or not students have made adequate yearly progress (AYP) toward achieving proficiency in the domains of reading, math, and – beginning in 2007-2008 – science. Implemented in the 2001-2002 school year, the law required that states develop a plan to move 100 percent of its students to proficiency in each of those domains by 2014. However, because the education of children is not a federal power specified in the Constitution and thus devolves to the states, as per the 10th Amendment, each state was empowered to define proficiency for itself. Some states seemed to avail themselves of the opportunity to ensure their federal funding by defining proficiency at a minimal level. In introducing the impetus for Race to the Top, Duncan in fact points to this very problem: the gap between the definition of proficiency utilized by the Nation's Report Card – the NAEP – and that which certain states use for compliance purposes can be "staggering" (Duncan, 2009b). This tendency to deliberately define down the concept of proficiency, as unintentionally incentivized by NCLB, came to be derisively termed the "race to the bottom." Duncan particularly decries this practice in the same breath as he announces the upcoming Race to the Top Program: "When states lower standards, they are lying to children and they are lying to parents" (Duncan, 2009b).

The Race to the Top initiative begins with encouraging – through, as in NCLB, the promise of additional federal money – uniformly higher standards. That is the major bully-

pulpit exhortation from the Obama administration's Department of Education from which all other demands – for example, as most directly concerns this project, policies regarding the improvement of the teacher corps – stem. Defining the height of the standards, and proposing means of achieving them, entails the more or less explicit establishment of the criteria by which progress in this domain will be visible. As regards the improvement of the national cadre of teachers, then, such improvement necessarily involves statements that reveal the working conceptions of what it means to teach and to learn.

<div align="center">1.2 Teacher Quality and Student Achievement Data</div>

In the second 2009 speech, addressed to the Governors Education Symposium, Duncan proposes the "four core areas" targeted for reform. These are: (1) "Robust data systems," (2) "teacher and principal quality," (3) "turning around our lowest performing schools," and (4) "higher standards and assessments" (Duncan, 2009c). Although Duncan posits these four areas as self-evidently distinct, a substantial common ground stands out immediately. In discussing issues of performance and quality – core areas (2) and (3) – Duncan implies a means by which high performance or quality might be distinguished from low. Core areas (1) and (4) hint at these means: robust data systems consisting of measures of student performance on assessments against "higher standards" (the criterion according to which a standard would be measured or made visible as either high or low remains unspecified).

Contrary to the rhetorical structure of this speech, then, teacher quality is not merely one of four areas addressed by the proposed educational reform; rather, every one of the core areas of reform bears directly upon teacher quality. In fact, ironically, the three other points are more revealing of the conceptual shape of teacher quality here than the explicit mention

of "teacher and principal quality" as an area of reform. By gesturing toward the intertwined needs for "higher standards and assessments" and "robust data systems," Duncan outlines the form in which teacher and principal quality will be knowable in the first place.

The linkages among robust data systems, standards and assessments, and teacher and principal quality are readily visible throughout Duncan's speeches. Speaking ostensibly about the data systems segment of this core drive, Duncan notes that reform "starts with robust data systems that track student achievement and teacher effectiveness." It is obvious, of course, that data must be data *of something*, but it is nevertheless worth pointing out that, far from being one of four elements on a reform agenda, teacher quality occupies a central position, even when Duncan explicitly talks about data as a separate issue. Duncan continues: "We need to do a much better job of tracking students from pre–K through college. Teachers need this data to better target instruction to students. Principals need to know which teachers are producing the biggest gains and which may need more help" (Duncan, 2009c). Data allows teachers more insight into their students' particular needs, he implies, but it also allows principals to better distribute their resources in the direction of teachers who "may need more help." The need for "robust data systems" is precisely the need for greater insight into an emerging and particular view of teaching and learning.

In his first speech of early June, 2009, Duncan in fact delves into detail about his conception of the data that pertain to the evaluation of teaching and learning (Duncan, 2009b), and, as it is the first of four policy speeches designed to articulate the policy initiatives the Obama administration proposes, it will be helpful to open with an exploration of the shape and purpose of "data" in Duncan's proposals.

Data plays a wide-ranging and multifaceted role in Duncan's proposals, and it will be crucial to keep one's eye on all of the aspects at once. In the policy speech geared specifically to data, he calls for "robust data systems" first of all in the following forms and spheres: he calls for "a longitudinal study of teachers and an international assessment of adult competencies"; he notes that the department of Education wants to know "how teachers are using data to drive instruction. Many teachers are hungering for data to inform what they do"; and he proclaims that "we will also ask whether the data around student achievement is linked to teacher effectiveness." The ultimate goal of these proposed uses emerges in two forms. Duncan says, "hopefully, someday, we can track children from preschool to high school and from high school to college and college to career. We must track high growth children in classrooms to their great teachers and great teachers to their schools of education." Lastly, Duncan avers that

> We can one day do a better job of understanding what makes great teachers tick, why they succeed, why they stay in the classroom and how others can be like them. Hopefully, we can track good programs to higher test scores to higher graduation rates. Hopefully, one day we can look a child in the eye at the age of eight or nine or 10 and say, "You are on track to be accepted and to succeed in a competitive university and, if you keep working hard, you will absolutely get there" (Duncan, 2009b).

In this speech alone, the "robust data systems" Duncan envisions will be used by teachers (for instructional purposes), principals (for teacher evaluation purposes), LEA and SEA administrators (for parental reporting and COE evaluation purposes) and academic researchers (for the purposes of "understanding what makes great teachers tick," among other things). The data in question will then *simultaneously* reveal (a) student achievement levels, including specific areas for improvement, such that teachers can tailor instruction accordingly; (b) teacher quality, which is explicitly apparent in the phenomenon of "high

growth children"; (c) teacher-education quality, which is recursively defined in terms of producing the teachers who produce "high growth children"; and, in a most utopian fashion, (d) the complete trajectory of a third- or fourth-grader's life. These four projects remain distinct from one another, though the temptation obviously exists to view them as a single entity, which Duncan's recursive logic makes clear. Duncan appropriately calls for *robust* data systems; given the number of tasks to which the data will be assigned, robustness is the operative term.

The obvious question to raise, at this point, concerns the *nature* of this robust data. Duncan never says explicitly *what* the data will consist *in* or *what counts* as data in his vision. But he does hint in one particular direction: student achievement scores. In introducing the need for increasing our use of educational data, Duncan says, "Our best teachers are using real-time data in ways that would have been unimaginable five years ago." He follows this sentence immediately with an explanation: "They need to know how well their students are performing" (Duncan, 2009b). This construction clearly refers to measures of *student* achievement.

Duncan proceeds to speak, as I have already quoted, about the links between student achievement measures and indices of teacher quality. A scant few moments apart, he posits two revealing positions. First he asserts that "we will also ask *whether* the data around student achievement is linked to teacher effectiveness" (emphasis added); a minute later, he objects to what he sees an union obstructionism in considering achievement data when he says that "to somehow suggest that we should not link student achievement and teacher effectiveness is like suggesting we judge a sports team without looking at the box score" (Duncan, 2009b).

The first statement presents the existence of a connection between student achievement and teacher effectiveness as a hypothesis requiring some sort of testing; the second stipulates it as a common-sense assumption. In thus linking the concept of teacher effectiveness to the concept of student achievement, which itself is measurable in terms of standardized tests, Duncan seems to run the risk of unwarrantedly conflating the concept of "raising test scores," in which test-preparation companies like Kaplan and Princeton Review specialize, with the concept of "teaching."

Yet Duncan's speech on data also reveals a certain public ambivalence about reducing an understanding of teaching to measures of student achievement. He speaks to that point in such statements as, "I absolutely respect the concerns of teachers that test scores should never be used solely to determine salaries" and "I also appreciate that growth models as they exist today are far less than perfect." But he rejects any iteration of the notion that, as he says, "since standardized tests are not perfect, eliminate testing until they are," calling such an idea "absolutely ridiculous." While acknowledging that these tests are "not perfect," Duncan still avers that "we need to monitor progress. We need to know what is and is not working and why" (Duncan, 2009b).

The danger of conflation is particularly acute here: "monitoring progress" is suddenly bound up with understanding "what is and is not working." Granting the sincerity of Duncan's concern about the imperfections of tests in their current forms, it may yet be possible to ascertain a ballpark estimate of a child's academic progress in the assessed domain(s), as long as that measure is understood as a rough estimate. But it hardly follows therefore that such measures reveal much at all about "what is and is not working," since any findings as to what is and is not working are derived from admittedly blurry measures of

achievement. Even less warranted would be specific instructional or personnel (or salary) decisions based on such flawed measures. Loosely "monitoring progress" through ad-hoc means necessitated by imperfect assessments thus seamlessly metamorphoses into making "data-driven decisions" based on acknowledged imperfections as a matter of policy. Despite the known imperfections of the measures, the assumption that one can use those measures to reliably distinguish "what is and is not working" remains somehow intact.

Several eminent scholars (D. Koretz, 2008; Ravitch, 2010; Rothstein, 2008) have emphasized, in particular, the "curricular distortion," as Rothstein terms it, that results from overemphasizing achievement measures in policy decisions. This criticism derives from the fact that the achievement measures in question, generally speaking, refer to the domains of reading and math alone, to the exclusion of the many other tasks that schools and teachers must accomplish. While the curricular distortion argument certainly and accurately impeaches the "robustness" of Duncan's notion of data, my analysis tends in a slightly different direction, and I would like to make this distinction explicit. For Rothstein, in particular, the major problem with conceiving of "data" primarily in terms of achievement measures has to do with *subject matter* exclusions: if reading and math become the sole measure of a student's learning, teachers of art, physical education, music, and so on will soon find themselves offering reading or math instruction, rather than offering students the subjects for which they were explicitly hired. Anecdotal evidence of this phenomenon abounds.[3]

---

[3] For example: "Since Mr. Shinlever knows his fate now depends on math and reading scores, he is making his classes more academic. After watching the documentary "Food, Inc." recently, his sophomores wrote essays. Similarly, in Chester County, a gym teacher recently spread playing cards around and had students run to find three that added to 14" (Anderson, 2012).

My analysis, once more, focuses specifically on the conceptual construal of "teaching" as a skillful activity in the policies under discussion, and as such I will for the most part attempt to avoid subject-matter discussions insofar as it is possible to do so. However, I, too, wish to question whether the data that Race to the Top utilizes in its policies evinces the robustness necessary to the myriad tasks it is assigned, particularly as regards the evaluation and study of teachers. If it proves to lack appropriate robustness, I propose to examine the specific conceptual terrain of "teaching" that achievement measures fail to address.

With an eye, then, toward considering the robustness of Duncan's data as it concerns the concept of teaching, let us return to some of Duncan's statements in the above speech. It seems doubtful that his limited conception of data could accomplish all of the tasks that he proposes for it, but that does not imply that it cannot accomplish any individual project. My purpose for the time being is to investigate the concept of teaching that *is* visible in terms of the data that Duncan proposes to use, and only later to ask whether that concept of teaching is appropriate or sufficient.

To return, then, to the salient passages of Duncan's first speech of June before transitioning to Duncan's second one, which focuses more directly on teacher evaluations, what would fundamentally need to be true of teaching in order for Duncan's view of data to succeed in the tasks he envisions for it?

Of data's utility for teachers, Duncan says,

> They need to know how well their students are performing. They want to know exactly what they need to do to teach and how to teach. It makes their job easier and ultimately much more rewarding. They aren't guessing or talking in generalities anymore. They feel as if they're starting to crack the code (Duncan, 2009b).

Data, in this passage, first of all allows teachers to know how their "students are performing." But it also, Duncan implies, allows them "to know exactly *what they need to do to teach and how to teach*" (emphasis added). That is a curious statement, and it will require some analysis, as it is not immediately clear how one could derive "how to teach" from examining student achievement data. Duncan provides a clue to his working assumptions in referring to data's enabling insights in terms of "crack[ing] the code," which seem to allow teachers to avoid "guessing." In contrast to "guessing" about what to do, teachers will presumably be acting with a given amount of certainty.

In this passage, the only one in this speech in which Duncan connects the use of data to the practice of teaching itself, some of his fundamental assumptions emerge. The code-cracking metaphor is hardly an idle figure of speech, but it is a complex one. In the first place, code-cracking implies the transmission of a message, which is not an uncommon way to characterize teaching, though it is a reductive and largely discredited one in both P-12 and higher education (Laurillar, 2002; Tishman, Jay, & Perkins, 1993; Trigwell, Prosser, & Waterhouse, 1999). In a transmission model, the teacher's task is to impart certain knowledge to the students, who then come into possession of said knowledge, and can thus be said to have *learned*. Given that this metaphor appears in Duncan's speech that focuses on student achievement data, these achievement measures would presumably demonstrate the amount of students' new knowledge, and thus the relative success of the "message's" "transmission." So, students who demonstrate the most growth on achievement measures have learned more and better than those whose scores indicate little growth. Recursively, the teachers of "high growth" students have done a better job of transmitting the message. The

higher rate of transmission, one might call it, is then attributed to the teacher's access to some sort of code.

The appearance of a "code" in this context is counterintuitive, however. The conveying of encoded messages, as with the Enigma machine in World War II, requires that both the sender and the receiver share access to the code. Duncan's metaphor indicates that the message's failure to get through – an absence of apparent student learning – is due to its encoded nature. Using data to guide instruction, as Duncan sets it up, somehow allows teachers to feel as though "they're starting to crack the code," which obviously constructs the problem for teachers in particular terms, to wit, discovering the means of breaking through to students. This construction implies, however, that if teachers need to possess a code that they so far lack in order to transmit knowledge to their charges, then students "have" – or, perhaps more faithfully to the metaphor's application here, "are imprisoned within" – this same code, which would otherwise make teachers' attempts at communication futile, gibberish. Students are not tasked, in this construction, with *un*scrambling a teacher's coded message; teachers are rather charged with properly scrambling the message in the first place so that it is immediately visible to students who are envisioned here as encountering the world ineluctably in terms of this code.

To pose the question again: What would fundamentally need to be true of teaching in order for this construction to hold? First of all, the very concept of teaching would have to be limited to the transmission model, wherein teachers are in possession of knowledge that students require, and the job consists in transporting that knowledge to the students' brains. Secondly, knowing "how to teach" would consist in the having of a particular "code" that would enable the successful transmission of knowledge from teacher to student. Such

knowing-how would therefore rest upon a knowledge *of* something abstract, a group of facts, a secret syntax. In other words, knowing-how would really be a specific kind of knowing-that, inasmuch as the "how" of good teaching is contained in the "that" of a universal code. Thirdly, as I just hinted, the "code" that allows and constitutes skillful teaching would have to universally shared among students. Duncan claims that data allows teachers to feel like "they're starting to crack *the* code" (emphasis added), which obviously implies the code's singularity; without the universal applicability of the code to all students, there is no sense in talking about "the" code at all.

The data so central to the Obama administration's Race to the Top program must be "robust," as Duncan says, in order to accomplish the tasks to which it assigned. The foregoing exploration suggests, however, that the notion of robust data rests more or less entirely upon student achievement scores. Parts 1.5 and 1.6 will explore more deeply the extent to which, whatever the "multiple measures" of teacher quality to which Duncan appeals, test score data alone functions as an ultimate criterion, but for now I leave it as a mere suggestion.

In Duncan's data-oriented speech of early June 2009, the role of data in the practice of teaching that he describes reveals a very particular view of the nature of teaching itself. Through delving into Duncan's preliminary remarks on the matter, the shape of this view of teaching has emerged in outline form, whose major contours involve an adherence to a transmission model of instruction, a concept of skillful classroom practice based upon something almost algorithmic in nature – the "what" of the code – , and a vision of students as fundamentally identical (on the levels that matter, at any rate) to one another, such that

methods that "work" to achieve test score gains in one place are assumed to "work" to the same or a similar degree in any other.

Given the fact that this view emerges from the context of a single speech leading up to the announcement of Race to the Top, it is entirely possible that later iterations of administration proposals and policies will revise or refine this view of teaching. It remains unclear at this point, for example, how data can reveal "what [teachers] need to do to teach and how to teach," as Duncan has claimed that it can, and a fleshing-out of that idea has the potential to clarify the operative concept of teaching in any number of directions. Exploring Duncan's next speech in this rhetorical cycle – a speech geared specifically to standards and assessments – ought to reveal a greater degree of conceptual nuance as regards the notion of what teaching most fundamentally is (Duncan, 2009c).

<div style="text-align:center">

1.3 Teacher Quality, Rigorous Assessments and Higher Standards

</div>

Before the Governors Education Symposium, one week after delivering his address on data, Duncan highlights issues around the means of producing the data in question, as well as the objects to which the data point, namely, teaching and learning.

Duncan notes near the opening of the speech that his primary emphasis lies on the need for high standards and rigorous assessments, and that he will deal with teacher and principal quality in another forum later on, but he wishes to "touch on it briefly" here nonetheless. In doing so, he both echoes points from the earlier speech and raises new ones as well.

Duncan first refers to the urgency of the problem of teacher quality, and then declares that we "have to fix our method of evaluating teachers, which is basically broken." He points

specifically to two problems: in the first place, "A recent report by the New Teacher Project shows that 99 percent of teachers are all rated the same," and in the second place, that "most teacher rating systems don't take student achievement into account" (Duncan, 2009c). The New Teacher Report is not unique in pointing to the flaw Duncan mentions: a recent article in the *New York Times* notes that in Chicago, "Despite low graduation rates, test scores and other measures of student performance in the district, more than 90 percent of its teachers are now rated excellent or superior" (Vevea, 2011). Duncan clearly points to a compelling problem when he declares current teacher evaluation systems broken.

However, the galling element for Duncan lies in the second point, that "most teacher rating systems don't take student achievement into account." He notes, as he did in his first speech, that some states in fact have laws on the books preventing the use of student achievement in teacher evaluation. What he has in mind in terms of the meaning of student achievement has already been indicated in his comments on the use of data, and he goes so far as to call the "firewalls" that some states have constructed between "student achievement" and the evaluation of teachers "not fair" to students and teachers alike, and what's more, "not honest." The obvious appropriateness of taking achievement data into account manifests itself in the rhetorical nature of the question he then poses: "How can you possibly talk about teacher quality without factoring in student achievement?" The structure of his argument here once more mirrors his approach in the first speech.

The refusal to use student achievement in teacher evaluations draws Duncan's ire here, and from his earlier comments, one can surmise that when he discusses "achievement," he means specifically student performance on assessments. In fact, Duncan addresses the

issue directly in his speech to the National Education Association a few weeks after the talk in question:

> I understand that tests are far from perfect and that it is unfair to reduce the complex, nuanced work of teaching to a simple multiple choice exam. Test scores alone should never drive evaluation, compensation, or tenure decisions. That would never make sense. But to remove student achievement entirely from evaluation is illogical and indefensible (Duncan, 2009a).

While Duncan once again allows himself some ambivalence about the use of "test scores" in teacher evaluations, my purpose in quoting the above passage is to tease out the sense of "student achievement" that Duncan thinks ought to be included in teacher evaluations. Of course, the above passage requires little in the way of teasing: Duncan directly equates "test scores" with "student achievement." Thus, when Duncan asserts that teacher evaluations need to take "student achievement" into account, whatever his reservations about relying too heavily upon a single measure, he clearly has in mind student performance on state assessments, the raw material that comprises his "robust data systems." Teacher quality, then, as constructed in the official rhetoric leading up to the announcement of the Race to the Top program, is visible – and thus knowable and evaluable – in terms of test-score data, with the important caveat that "scores alone should never drive evaluation." His notion of what else is required will emerge shortly.

Conceptually speaking, however, this vision of teacher quality is wholly determined by the assessments generating the data in question. Teachers needing more help are tellingly contrasted to those "producing the biggest gains" (Duncan, 2009c). Teachers failing to produce these gains need help; teachers who do produce big gains do not. The particular role of data in Duncan's proposal has explicitly to do with *knowing* good teachers from those who "may need more help." To word it more strongly, which is not unwarranted, given the

context: the data in Duncan's vision will do that work by itself – it is fully capable of drawing the distinction in question.

The nature of the assessments, then, rises for consideration at this point. The responsibility that achievement data bears in terms of sorting strong teaching from weak, good learning from bad, and so on, has been emphatically established. Thus, what is assessed, and how it is assessed claim a place of prominence in the discussion. In order for the assessments to serve the role they have been assigned – generating the data that will determine teacher quality, reveal student academic and career trajectory, guide instructional practice, and so on – these assessments must be valid.

Validity here ought to be understood in its technical psychometric sense. The assessments employed to gauge student progress and teacher quality, among other things, must evince good construct validity, which means both of the following things: that what the tests in fact measure must conceptually match what they purport to measure, and that what they purport to measure is adequate to the evaluation purposes assigned to this data. Duncan himself alludes to this necessity in discussing the insufficiency of "multiple choice tests" in this speech:

> We need tests that measure whether students are mastering complex materials and can apply their knowledge in ways that show that they are ready for college and careers. We need tests that go beyond multiple choice, and we know that these kinds of tests are expensive to develop (Duncan, 2009c).

The implication in this passage is that *because* multiple choice exams cannot adequately "measure whether students are mastering complex material" such that students would be "ready for college and careers," they do not have good construct validity. Although they purport to reveal something important about student learning, they do not adequately

represent what we really mean by "student learning," which, Duncan suggests, is visible in the concrete terms of students' readiness for college and careers.

I must pause so as to draw attention to the complexity of the concept of "student learning" as Duncan defines it above. A recent census report has shown that 30 percent of American adults ages 25 and older hold bachelor's degrees, which is an all-time high (Census, 2012). Obviously, the percentage of adults 25 and over that hold careers is vastly higher than 30 percent. Duncan's construction, which he and Obama both repeat more and more often over between 2009 and the present – most recently, on the day of this very page's composition, in urging South Carolina to adopt the Common Core standards he discusses in this 2009 speech (Duncan, 2012a) – seems to assume that college readiness and career readiness are the same thing, in that they are to be made visible by the same metric. But given the discrepancy between the percentages of American adults holding college degrees and those working careers, it follows that readiness to succeed in college is not necessarily identical to readiness to succeed in a career. The notion that these two constructs could be adequately measured on a single metric deserves more consideration. Is one type of readiness more essential than the other? If a teacher is adequately preparing students for careers, but not necessarily for college, will that teacher still have a claim to a high-quality rating? How is career-readiness defined, if it is different from college-readiness? The criterion according to which "student learning" will be gauged and assessed is "college and career" readiness. But what exactly is that, and how will the new assessments be able to address that?

"College and career readiness" never receives an explicit conceptual definition, but it continues to function rhetorically as the criterion of "student learning" that Race to the Top

works toward. In the absence of such an explicit definition of this criterion, one must derive

or excavate the working sense of the concept from its appearances and applications in

Duncan's discourse. Duncan's speech on assessments and standards begins to outline what

he takes "college and career readiness" to consist in, albeit in partial and limited form.

The limitations of this conceptual outline are directly related to the confused sense of

"standards" that appear in the talk. Praising the governors, Duncan says, "You've taken the

first step. Your stated goals are 'higher, clearer, and fewer' standards, and I absolutely

support your goals" (Duncan, 2009c). The higher, clearer, and fewer locution seems to

indicate that standards are conceived as both *levels of* achievement and as *areas for*

achievement, or in other words, as both a quality component (higher standards) and a content

component (clearer and fewer standards). This confusion persists throughout the speech. At

one point, Duncan says:

> Right now, standards are too broad, covering 35 to 40 topics per subject in
> each grade as opposed to 15 or 20 standards in many high–performing countries.
> Teachers scramble to cover everything—a little of this, a little of that, and not enough
> of what's really important. They can't dig deeper on a challenging subject that excites
> their students. And students can't master material when they are racing through it
> (Duncan, 2009c).

 In referring to standards as "too broad, covering 35 to 40 topics," Duncan obviously

refers to a content component. In calling for change in this regard throughout the speech, he

exhorts states to focus on "what's really important," "the most important things kids need to

know," and "the essential knowledge and skills our kids need," all of which imply (1) that

some, but not all, of what happens in school is essential to what he has in mind, and (2) that

one can know the important from the unimportant on the particular basis of students'

preparation "for college and the jobs of the future." The "college and career" readiness

criterion appears at multiple points as a determinant of "important" content: as Duncan says, "The standards must be tied to the end point of making sure students are ready to succeed in college or in the workplace" (Duncan, 2009c). Its importance as a criterion is therefore easy to establish in the realm of content: the concept of college and career readiness distinguishes what is "important" or "essential" to learn from what is apparently inessential.

But the sense of standards as an indicator of quality also weaves its way throughout this speech, and "college and career readiness" plays a criterial role here, as well. Duncan laments the current state of affairs in similar terms to those of his earlier speech, in which he called for "honest" data:

> For too long, we've been lying to kids. We tell them they're doing fine, give them good grades, and tell them they're proficient on state tests that aren't challenging. Then they get to college and they're put into remedial classes. Or they go into the workforce and find out that they don't have the skills they need to succeed. We need standards that will get them ready for the day after they graduate. That means they must be rigorous (Duncan, 2009c).

In this passage, by calling for "rigorous" standards and by referring once more to the corruption of "proficiency" under NCLB, Duncan more or less explicitly postulates a qualitative sense of standards. Additionally, in both claiming that too many students require remediation once they arrive at college – a datum relatively easy to measure – and that too many high school graduates "go into the workforce and find out that they don't have the skills they need to succeed" – a datum whose measurability is much less obvious – he reinforces the central role that "college and career readiness" plays in determining the level of quality he thinks states should mandate.

He moves to a grand combination of these two notions of "standards" as he makes his main point, and the answer to the question I posed earlier – what is college and career

readiness and how will it be assessed? – is more clearly answered here than at any other point in the speech:

> Today, our standards are too low and the results on international tests show it. Worse yet, we see the signals in the international economy as more and more engineers, doctors, and science and math Ph.D.s come from abroad. You must resist the temptation to make these standards too easy. Our children deserve to graduate from high school prepared for college and the jobs of the future. Your standards must be rigorous and they also must be tightly focused on the most important things students need to know.

An influx of foreign "engineers, doctors, and science and math Ph.D.s," combined with "the results on international tests" serve as evidence of both the inadequate level of our standards and the unnecessarily diffuse nature of the curricular content areas to which standards are currently applied. At the end of a passage raising the specter of high-level math- and science-related professionals from abroad taking the jobs that (implicitly) ought to be occupied by Americans, Duncan calls for standards that are both "rigorous" and "tightly focused on the most important things students need to know." This conflation communicates the sense that the content areas under discussion here, as well as the levels of achievement by which success or failure will be apparent, are those that would equip a person for a high-level math- or science-related career. At this point, a Ravitch or a Rothstein would gesture toward the numerous things we expect schools to do outside of preparing kids to be doctors, but that, once again, is not my aim.

My focus remains on the concept of teaching as constructed in these discourses pertaining to Race to the Top. The international tests to which Duncan refers, and which he in fact mentions by name in his first speech (Duncan, 2009b) are tests of math, science and reading. Although the problems inherent in comparing such results internationally are well-documented (Ercikan & Koh, 2005; Ercikan & McCreith, 2002; Sireci, 1997), these results

nevertheless influence the policies and popular discourse not only of the United States, but of other countries as well, even ones with which we are often unfavorably compared (Brügelmann, 2011; Fehr, 2010). In sum, then, by appealing to these particular measures – as well as high school and college graduation rates – Duncan establishes particular criteria in particular places. The dual senses of "standards" here, which indicate both achievement levels and content areas, serve to indicate the nature of student learning that good teachers will address, as well as the expectations of quality that questions of student learning will address in reciprocal fashion.

The "limiting" effect in Duncan's call to "limit standards to the essential knowledge and skills our kids need so teacher can focus in depth on the most important things kids should know" involves by necessity narrowing what it is that we (officially) expect teachers to accomplish. "College and career readiness," insofar as it surfaces in minimally concrete form in Duncan's speech, is revealed by certain proxy measures: standardized test scores, most evidently; and also the percentage of our "engineers, doctors, and math and science Ph.D.s" that emerge from our own education system. Given the careers that Duncan chooses to highlight and the international tests to which he refers, college and career readiness will be established in terms of improved scores in math, particularly, but likely also in reading. Should college and career readiness consist in certain levels of math and reading scores, and should all teachers be evaluated according to their contributions to students' readiness for college and careers, it follows that teacher quality will be accessible in terms of students' growth in measures of math and reading.

Perhaps this appears too simple of a verdict, too pat, particularly given the numerous instances in the speeches on data, assessments, and teacher quality in which Duncan

explicitly insists that teaching is far too "nuanced" a practice to be reduced to a test score alone. In fact, in the assessment speech in particular, Duncan highlights the pilot evaluation program that he and others developed in Chicago and hammered out with the union: "It was based on classroom observation, whole school performance, and individual classroom performance, measured in part by growth in student learning" (Duncan, 2009c). "Whole school performance" and "individual classroom performance" are most certainly represented by test scores, as Duncan himself implies, at least in the latter case. Classroom observation proves to be the only place in which any sort of skillful practice not evident in score growth might reveal itself.

The apparently inevitable recourse to achievement data in the grounding of the concept of "learning" that "teaching" ought to impact is particularly important to point out at this juncture precisely because in each of the three speeches under consideration here, as has already been noted, Duncan makes a special point of acknowledging that the "complex, nuanced work of teaching can never be reduced to a simple multiple choice test," while stating in the same breath that "to remove student achievement entirely from evaluation is illogical and indefensible" (Duncan, 2009a). Duncan is at pains to make explicit the limitations of test scores in talking about student learning. The fallibility of these measures is on particular display in the way in which NCLB unwittingly encouraged states to make "proficiency" into an empty signifier, unmoored from any real-world application. Duncan's "college and career readiness" criterion functions as a more or less explicit correction to the previous error, re-anchoring the concept of "student achievement" to the future goals of college and career.

It remains unclear at this point, however, whether Duncan's notion of "college and career readiness" as an indicator of "student achievement" and therefore of "teacher performance" is itself anything more than standardized test scores. After all, in calling for narrower content domains and higher achievement standards against which to *measure* students, he seems to imply that "college and career readiness" is visible, ultimately, in test data. College and career readiness will be defined in terms of these standards. Students will be tested against these standards. Ergo, college and career readiness will be revealed in achievement data. If the concept of "good teaching" is to exceed the reduction to test data in a way for which Duncan seems to advocate, the "classroom observation" that Duncan includes will play a crucial role, as it appears to blunt the potential objection that teacher quality is defined merely in terms of test scores.

I wish to point this out now primarily because it will come up again in a later section. The rationale for *not* overemphasizing the profile of achievement data in the evaluation of teachers has everything to do with the common-sense notion that some aspects of excellent teaching will not show up in such data. Other measures, particularly the classroom observations of the sort in Duncan's Chicago pilot program, provide a means of determining whether or not the qualities to which achievement data cannot speak are present in the teacher's practice.

Because the assumption that undergirds the insufficiency of using achievement data alone for evaluation purposes has to do with the ability of classroom observation to *add* something to what the achievement data can reveal, it stands to reason that the data gleaned from observations ought to demonstrate, perhaps, *some* correlation with the what the achievement data reveals, but that the correlation ought not to be perfect. If the correlation

between observational data and achievement data *were* perfect, the evaluation system would merely be measuring the same thing twice, making one or the other method superfluous. The utility of evaluation methodologies that go beyond the use of achievement data must evince a certain amount of divergence from what achievement data can demonstrate. Duncan implies repeatedly that "good teaching" is too "complex" and "nuanced" for perfect representation in student achievement data. Whatever it is that he takes to be the excess in good teaching beyond such representation will therefore be the portion uncorrelated with achievement data.

The concept of construct validity reemerges at this point, as we return to Duncan's speech on standards and assessments. Teacher quality – my principal concern – is at least partially knowable in terms of student achievement to such an obvious degree that to omit achievement from teacher evaluation is "illogical and indefensible." Student achievement is determined in Duncan's proposals according to "rigorous" standards, which are reflected both in terms of specific content areas and specific quality levels. These "rigorous" standards are derived from the real-world notion of "college and career readiness." But college and career readiness, which drive the determinations of the achievement levels and the content areas in and according to which students and teachers will either succeed or fail, has been only sketchily defined. The hints that Duncan offers in 2009 point merely to international competitiveness on standardized tests in math and literacy and the proportion of our "engineers, doctors, and math and science Ph.D.s" that come from abroad.

It is not too much to concede that the results of international tests and the percentage of superlatively skilled foreign workers coming from abroad might adequately impugn our educational system's ability to prepare 100 percent of our children for college (and for graduate and professional school, apparently, as the only careers Duncan names in his speech

require substantially more schooling than an undergraduate education), but these measures are certainly insufficient to impugn our educational system's ability to prepare our children for all careers as such, since the vast majority of American adults, as revealed in the recent census report already cited, have a career despite lacking a college degree. If the "rigorous" assessments and standards that Duncan encourages states to adopt are expected to ultimately speak to a given student's "college and career readiness," then it is unclear that standards geared toward producing doctors, engineers, and Ph.D.s in the fields of math and science would necessarily reveal all instances of career readiness (unless it is somehow assumed that the skillfulness required to be a master chef, for example, is some subset of the qualifications required to be a college mathematics professor, for instance).

The issue of construct validity is a pregnant one for the Race to the Top program precisely insofar as "college and career readiness" strikes one as an entirely unobjectionable goal for American schooling. It is further unobjectionable to factor something like "whether a teacher is preparing students for college or career" into the evaluation of teachers. But in rendering the concept of college and career readiness operational, in translating it into content standards and achievement levels, one risks a great deal of conceptual slippage, precisely inasmuch as the requirements for college readiness and career readiness are supposed to be identical. Furthermore, if teaching is (partially) conceived as "preparing kids for college and career" by means of teaching to these operationalized definitions of readiness that appear in achievement and content standards, then it is exceedingly difficult to see how one avoids the temptation to evaluate teachers on the preponderant basis of student test data, where those assessments are aligned with the standards derived from the (limited) definition of college and career readiness.

Duncan proposes that school ought to prepare students for college and career, and that sounds perfectly sensible. Duncan proposes that teaching is too complex and nuanced to be reduced to a test score, which likewise sounds perfectly reasonable. The way in which teacher evaluation proposals fill in the gap between what test data can reveal and what else goes into preparing students for the real world therefore will determine whether or not Duncan's view of teacher quality is both (or either) self-consistent and/or sufficiently accurate and robust. Duncan's final two speeches leading to the announcement of the Race to the Top program speak to his conception of teacher quality much more directly than his first two, particularly the speech he delivers to the National Education Association on July 2nd (Duncan, 2009a), which pertains specifically to reforming the teaching profession, though his speech of June 22nd (Duncan, 2009e), which deals with turning around America's lowest-performing schools, offers some insightful comments about teaching and teacher preparation, as well. Carrying the preliminary hints and conceptual outlines of "good teaching" that Duncan has already offered, one looks to the NEA speech in particular to fill out the picture of teacher quality that he has so far merely drafted.

<u>1.4 Taking What Works to Scale: Teacher Practice as Rule-Following</u>

Since the June 22nd speech – third in the cycle – merely glances at teacher quality, while the final speech of early July is expressly geared toward goals for the teaching profession, this investigation would do best to take the speeches in chronological order, fleshing out as much as possible the existing conceptual portrait of quality teaching with Duncan's remarks on the bottom five percent of America's schools before turning to the most

complete depiction of teacher quality available in the run-up to the announcement of Race to the Top.

The existing depiction of good teaching so far refers to at least two loci of visibility, if one can call it that; one is visible in terms of its effects on students – student achievement data – and one is visible in classroom observations. The existing depiction of good teaching, that is to say, has specified two particular sites that will reveal teacher quality without saying anything concrete about what teacher quality will consist of at either site, though in the realm of student achievement data, one can reasonably assume that teacher quality will be directly related to improvement in test scores.

In his speech on turning around the bottom five percent of schools, Duncan never in fact details the particular criteria by which "the bottom five percent" are knowable as such. The way he talks about successful turnaround projects, though, indicates the criteria at work here:

> Sherman Elementary saw a five-point jump in the percentage of students meeting standards in the first year. Harvard reduced absences by five days per student in the first year. And Orr High School saw a 15-point jump in attendance in its first year (Duncan, 2009e).

In these examples, he refers both to student attendance and to the "meeting" of (qualitative) standards, which would be visible in achievement data. But he ventures outside of the strictly or obviously measurable, as well, in this speech, citing anecdotes of Chicago parents who

> talk about their kids 'looking forward to school for the first time,' coming home and 'talking about their teachers.' They say it's 'a totally different atmosphere' even though it's the same schools with the same kids and the same socioeconomic conditions (Duncan, 2009e).

The atmospheric comments deserve further attention, but first I must point out the way this anecdotal evidence highlights a popular and ready-made argument against Duncan's proposals. Such opponents of Duncan's – like those who complained to Kristof after the publication of the first editorial cited in the Introduction – sometimes appear to paint a picture of socioeconomic determinism, or at least that seems to be the picture against which Duncan argues here. Duncan's anecdotal evidence displays the test-score and attendance improvement of children but simultaneously shines the spotlight on the unchanged socioeconomic circumstances of the same students. The unspoken conclusion is one that Duncan makes explicit in his next speech, when he says, "We believe every child can learn and every school can succeed" (Duncan, 2009a).

The use of anecdotal evidence itself is remarkable in this speech because it helps to concretize that which Duncan resolutely avoided making explicit in his first two speeches, namely, the aspects of teacher and school quality that lie beside the achievement data. "Lie beside," in the previous sentence, merely indicates that while such anecdotal evidence may or may not correlate in a particular way with achievement data, this relative degree of correlation appears simply beyond the scope of concern as Duncan cites this evidence. The anecdotes speak for themselves, in other words. Examining the type of evidence he cites and the way he expects it to function in other contexts reveals important aspects of his approach to educational quality in general, both at the level of the school and at the level of the teacher.

Duncan returns to the use of atmospheric comments as he refers to specific indicators of good schools, both in traditional public and in charter public schools:

> We know what success looks like. I see it the moment I enter a school. It's clean, orderly, the staff is positive and welcoming, and the kids and the classroom are the focus. I see award-winning school work on the walls. I see discipline and

enthusiasm in the children. I see parents engaged and teachers collaborating on instruction (Duncan, 2009e).

Absent from these indicators is any mention of the achievement data that has appeared to play such a prominent role so far. In this instance, Duncan discusses "what success looks like." In terms of school-wide building characteristics, he points to cleanliness, orderliness, and the presence of "award-winning school work" hanging on the walls; in terms of student affect, he points to "discipline" and "enthusiasm"; in terms of the adults in the building, he cites a "positive and welcoming" atmosphere, an atmosphere where "kids and the classroom are the focus," and one that seems to encourage parent engagement and teacher collaboration. Obviously, certain of these elements would require greater specificity in order to be operationalized into a "rigorous" evaluation system of the sort Duncan would desire – what does "positive and welcoming" consist of, exactly? And what opposes "kids and the classroom" as a school focus? – but what stands out immediately in this description is Duncan's assertion that he can see success "the moment [he] enter[s] a school." The immediate visibility of success inherent in the claim that "we know what success looks like" signifies the ordinary, everyday experience of judging schools, whether by laypeople, experts, or politicians.

The conceptual development here, in which Duncan explicitly references an immediately apparent aspect of a school's quality, will serve as an important bookmark to which to return on occasion. In his call for "robust data systems," Duncan's approach presents a view of learning – of school and teacher quality – that is fundamentally *mysterious*, such that the state-level use of achievement data to sort good teachers from bad teachers, and high-performing from low-performing schools, is required. He points to

California, one of the states whose laws create a "firewall" between student achievement and

teacher evaluation, as an example of the outright *need* for such data-driven accountability:

> In California, they have 300,000 teachers. If you took the top 10 percent, they
> have 30,000 of the best teachers in the world. If you took the bottom 10 percent, they
> have 30,000 teachers that should probably find another profession, yet no one in
> California can tell you which teacher is in which category. Something is wrong with
> that picture (Duncan, 2009b).

Where, precisely, is the need for such data that the California example is meant to

make evident? Since apparently "we know what success looks like," what specific need does

the additional achievement data fulfill? Says Duncan, "no one in California can tell you

which teacher is in which category." Does this statement imply that without the use of data,

no one can *know* which teachers are in the top 30,000 and which are in the bottom? Or does

it mean that, thanks to the absence of achievement data, no one can *report* which are the best

and which are the worst? Or, to essay a third possibility, does it simply mean that no one *is*

telling which teacher is which? Given Duncan's use of the immediately-visible character of

quality as regards schooling, it seems difficult to imagine that anyone could fail to *know*

fairly immediately and without the use of achievement measures whether a given teacher was

likely to fall into the top or the bottom ten percent of all teachers, if those were the only two

options. If educational quality is immediately perceivable, then knowing outright excellence

from abject failure ought to be an immediate experience.[4] As to the second possibility

pertaining to the reporting of teacher quality, it would behoove me to postpone that

discussion for a moment.

---

[4] This is unrelated to the problems of "all teachers" being "rated the same," which Duncan also cites.
If all teachers garner the same ratings, it does not necessarily mean that no one *can* distinguish high
quality from low quality teachers; it absolutely *does* mean, however, that too few people *are* drawing
those distinctions, whatever the basis.

The third possibility must come first. Duncan's colloquial statement that "there's something wrong with that picture," which refers to the fact that "no one in California can tell you which teacher is in which category" expresses an obvious dissatisfaction with the basic element of the third possibility: *that* no one *is telling* which teacher is in which category. But Duncan raises this problem in the context of calling for the elimination of laws which keep student achievement data separate from teacher evaluation, which would imply that access to student achievement data would have some obvious bearing on teacher evaluation, over and above whatever strikes an observer immediately. *That* no one *is telling* which teacher is in which category therefore seems like a consequence of *not knowing* which teacher belongs to which category. In speaking about teacher evaluation generally, Duncan points to the problem that one would solve by means of access to achievement data as exemplified by the fact that "99 percent of teachers are all rated the same" (Duncan, 2009c). The apparent failure of existing evaluations systems to distinguish among 99 percent of teachers calls for the "robust data systems" that Duncan propounds. Offered as a solution to this rating problem, such data could indicate *either* that the problem is one of inadequate knowledge, which data would help to clear up, *or* a simple lack of courage to tell the (immediately) obvious truth on evaluation forms, which data would obviate by rendering the process nominally impersonal or objective.

When taken with Duncan's comments above on teacher evaluation, the role that immediate, "the-moment-I-walk-in" judgments play with respect to evaluations that include the input of achievement data grows less and less clear. What *remains* indisputably clear is that no one in California *is telling* excellent teachers from disastrous ones. Whether this is a

problem of honesty in reporting or whether this is a problem of knowing (or both) remains murky.

The second possibility – that in which "no one can *report*" on teacher quality with any conviction – now requires some further discussion. Duncan himself has no difficulty (in the abstract) in reporting on the immediate visibility of successful schools. He lists criteria with absolute confidence. He hazards a claim about accessing educational quality, and does so with a measure of certainty in pointing to these schools, these criteria. But again, the immediately observable facts of good schooling appear irrespective to any student achievement measures. We have already seen that, if "classroom observations" are to play any real role in teacher evaluations, they cannot simply reiterate in every instance that which achievement measures reveal. In cases where the data that Duncan cites as evidence of school turnaround – "five point jumps" in the percentage of students meeting standards and dramatic increases in attendance – were not accompanied by orderly or clean environments, or by "positive and welcoming" staffs, or by any teacher collaboration, would one feel comfortable reporting that school as belonging in the successful-turnaround category? In the opposite case, in which cleanliness and orderliness, "open and positive" staff members, teacher collaboration and the like finds no (or little) reflection in achievement measures, would Duncan stand by his confident assessment of the school in question? What means exist to navigate these contradictory impressions? If two differing sets of criteria serve as a foundation for *knowing* and *reporting*, and if they conflict with one another, how is one to resolve the dilemma?

I raise such questions at this point to indicate that in cases like the ones enumerated above, Duncan reaffirms the relatively ordinary sense of educational quality that remains

accessible without recourse to achievement data, which is similar to the view he broaches in the standards and assessments speech under the aegis of "classroom observations." It is in no way outlandish to claim, as Duncan repeatedly does, that good teaching cannot be reduced to visibility according to any single measure. One task remaining before Duncan at this point in the discourse on educational quality will be to establish the appropriate relation among the ostensibly "multiple measures" which will render teacher quality totally transparent, and thus reportable in universal fashion. Despite the way my questions in the foregoing paragraphs drew a distinction between *knowing* quality and *reporting* on it, the combined use of achievement and observational data is clearly aimed at and capable of addressing both issues at once, since as Duncan (and common sense) would imagine it, secure and confident reportage ought to (*ought to*) follow cleanly from a secure and confident form of knowing.

Rather than anticipating the problems in ascertaining teacher quality that Duncan will have yet to tackle, let us return to the assumptions pertaining to educational quality that appear in this particular speech. Duncan has already alluded to a kind of educational quality that is visible according to achievement and attendance data, which is to say, abstractly; and also to a kind of educational quality that is immediately apparent to the situated observer, even one as remotely positioned with respect to the individual school as a district CEO or a Secretary of Education.   But another salient assumption about the nature of school quality – particularly as regards the kinds of criteria he has cited in his immediate ascertainments of quality – surfaces in this speech as well, and this assumption lies in the metaphor of "taking what works to scale."

Duncan uses the particular phrases "take to scale" and "scale up" quite a bit in his thinking about quality schooling. It appears in the 2012 interview on *The Daily Show*

(Stewart, 2012), in which he says that while there "are pockets of excellence" in the American school system, the task is to "take them to scale. . . . It's happening. It's just not happening at scale." In his speech on turning around troubled schools, he praises charter school organizations, to whom he delivers the speech, for "getting to scale" (Duncan, 2009e). In the same speech, he says the following of the great schools he has visited: "The hard part is to replicate those conditions everywhere, and you need to challenge yourselves and challenge each other to turn one success into a hundred and a hundred into 200." In remarks to a group of Oregon educators in 2011, Duncan once again praises the good work they do as specifically worthy of replication, saying, "The CLASS Project is a tremendous example of the successful work that should be taken to scale, because students benefit when teachers work together to share best practices and learn from one another" (Craig, 2011).

Duncan has no monopoly on this metaphor, either: it pervades the reform-movement milieu from which his policy innovations spring. In the speech on turning around the lowest-performing schools, for example, he praises certain charter networks and alternative certification routes, among them Uncommon Schools and Teach For America. The former organization was co-founded by Doug Lemov, whose recent book on teaching technique has been oft-heralded (Lemov, 2010); and the latter organization uses Lemov's methodology in developing new recruits in a very compressed span of time. Says Lemov of the primary issue in teacher training: "The really good question is, can you get people to improve really fast and at scale?" (Green, 2010). The metaphor has likewise made the rounds in education reform discourse quite generally over the past two decades (Elmore, 1996; Goggin, O'M Bowman, Lester, & O'Toole, 1990), and receives extensive airtime in Ford Foundation-financed studies on educational innovations (Glennan, Bodilly, Galegher, & Kerr, 2004).

Other examples are simply legion and do not require citation here. Rather, two things stand out immediately in the uses of this scale-up or take-to-scale trope.

The first thing to stand out is the scaling metaphor's native land, so to speak, the regions of language from which it comes to education policy discourse. To "scale up" or to "take" a given process "to scale" comes to education reform from two contexts: a business one and a network or IT context. In the business context, scaling refers to the process of adding volume to an existing system as a means of or accruing income geometrically while incurring costs arithmetically; in the IT or network context, scaling has to do with generalizing a particular program's or node's function in a limited context to a larger domain. In the business terms with which Duncan's academic background is conversant, the cost of adding volume to an existing system is lower than the cost of building a new system of production from the ground up, but the price of the widget remains the same. The result is an increase in profit, as production volume has doubled, let's say, while production cost has risen by far less. Scaling up is desirable precisely inasmuch as it increases profit by raising production volume exponentially without any commensurate rise in production cost.

In the education reform movement that Duncan champions and codifies in Race to the Top, the notion of scalability retains its basic conceptual shape as it slides over from the economics and business context. Instead of "adding *volume* to an existing system," scalability in education reform most often refers to integrating "successful" methodological innovations into existing methods of conducting the American school. In this context, scaling proposes to locate instances of particular educational success and transfer the structures of practice and administration responsible for that success to other parts of the system, thus substantially lifting the central tendency metrics of the system as a whole

without having to spend inordinate amounts of energy or money on reinventing the wheel in each particular and unique setting. When Duncan says to the charter school assembly, "The hard part is to replicate those [successful] conditions everywhere, and you need to challenge yourselves and challenge each other to turn one success into a hundred and a hundred into 200" (Duncan, 2009e), this is what he has in mind: implementing the innovative conditions that caused success in these particular instances much more generally, so that "one success" turns into "a hundred and hundred into 200." The "replication" of particular examples of success warrants highlighting inasmuch as it points to the means of scaling educational excellence from particular examples to system-wide demonstration. As the title of the Elmore article, with its focus on "good educational practices," indicates – a fact further echoed in Duncan's mention of "best practices" in his remarks to the Oregon teachers highlights (Craig, 2011) – the idea of scalability comes with a built-in sense of the means of its own attainment, which will bear further exploration and critique in Part 2. For the time being, it will suffice to draw out the picture of good teaching that Duncan's comments paint.

In order for good teaching, in particular, to be amenable to something like scaling, certain things must be true of teaching, and it is in this sense that the metaphor of scaling reveals a particular notion of teaching's ontology. In the first place, teaching must be conceived at a fundamental level as a more or less integrated – but ultimately atomizable – set of behaviors. When Duncan wishes to take innovations "to scale," either in the realm of teaching practice or school administration, the working assumption is that a successful innovation in one context will have a similar effect in another, and the basis of that assumption lies in the conception of teaching and learning as *fundamentally* a collection of isolated objects (behaviors or techniques) linked causally to a specific outcome. This is the

only sense in which "structures of practice or administration" can be considered "responsible for" the visible success in any particular context. Such causally-responsible structures of teaching and administration comprise the contents of a given educational innovation, and thus must be taken as isolated or isolable behaviors or techniques. Unless the *contents* of "successful" education can be transferred elsewhere in the "system," there is no sense in talking about scaling.

In the second place, in order for the above assumption itself to function as it does, educational *quality* must be wholly separate or separable from its context of actualization or demonstration and contained within the structures responsible for success. In other words, since the successful innovations themselves are constituted in certain structures of administration or the (behavioral) practices of teachers, their "successfulness" is conceived as an internal feature of these structures of practice or administration. Insofar as scaling successful educational endeavors involves spreading success throughout the system by means of inculcating particular "good" or "best" practices, that which makes the practices "good" or "best" must be internal to the structures themselves, more or less irrespective of contextual factors. After all, scaling theoretically involves spreading the *success* of the innovative techniques to the far reaches of the educational system by means of propagating the *behavioral* or *structural contents* of that innovation. To refocus the emphasis in the final sentence of the preceding paragraph in order to draw out this point, unless the contents of *"successful"* education can be transferred elsewhere in the "system," there is no sense in talking about scaling.

The frequency of the scaling metaphor's appearance reinforces its conceptual potency as it proliferates in policy and public rhetoric. The upshot of the conceptual analysis of

scaling in education reveals a version of the educative process in which (1) the practice of

teaching is comprised of isolated behaviors that (2) are causally related to "successful"

outcomes at both the level of the individual student and the level of the school, and whose (3)

fundamental quality is a predicate of the behaviors themselves as they manifest in (4) any

classroom whatsoever without much respect to context, if any.[5]

The sort of good teaching amenable to scaling, as it appears in Duncan's rhetoric and

in the discourse of education reform at large, is precisely the sort depicted in the subtitle of

Doug Lemov's book: *Teach Like a Champion: The 49 Techniques that Put Students on the*

Path to College. Referencing by coincidence Duncan's "college and career readiness"

criterion in that very subtitle, Lemov's view of teaching reveals itself identical to Duncan's

view: teaching is divisible into 49 (or *n*) "techniques" which, taken together, have an implied

causal relation to the end goal of college-preparedness; and this relation *itself* determines the

quality that inheres in the techniques. Teaching well therefore consists in implementing the

proper techniques in classroom settings, specific behaviors whose "properness" has been

defined by the behaviors' correlation with achievement score growth, as Elizabeth Green

explicitly notes (Green, 2010). Exhibiting the proper isolated behavior in the proper

circumstances, which constitutes "good teaching" in Duncan's rhetorical groundwork for

Race to the Top, is therefore a form of *rule-governed practice*, a term that will re-emerge in

---

[5] When I say "without much respect to context," I mean to indicate that while certain factors like the grade level of the students certainly factor into consideration – one does not use "proven reading techniques" designed for fourth graders on high school students – certain contextual factors often predictively associated with achievement outcomes, as, for particular example, as it pops up in Duncan's speech, socio-economic status, are believed to be irrelevant.

Chapter 2, along with a longer exploration of Lemov's project. Having introduced it at this point, I leave it once more for the time being.

By way of transitioning at last to Duncan's final speech in this preparatory cycle, delivered to the National Education Association on July 2nd, 2009, I wish to make one final observation about Duncan's view of good teaching in light of the foregoing discussion. The Secretary has mentioned many times, as I have already highlighted, that test scores provide an insufficient means of evaluating teacher quality, such that "multiple measures," including "classroom observations," are required as well. He has also, as we have again seen, made reference to a kind of immediately perceptible version of educational quality, such that the "goodness" of a good school is visible "the moment [he] enter[s] a school." Given the above discussion, and particularly the way that teaching as a practice has been revealed to consist in isolated behaviors whose relative quality is determined through established correlations with student outcomes on achievement measures, it is important to sound a note of caution – or outright skepticism – as I transition to the last speech.

At the opening of this section, I referred to the existing, burgeoning concept of teacher quality as visible in two distinct sites: achievement score data and "classroom observation." I also said that, while the emerging concept of teacher quality remained vague as to *what* (criterially) might be visible at each of these sites – according to *what* we would know the level of teacher quality – I noted that in the case of achievement data, presumably a certain measure of growth would indicate good teaching. Duncan stresses the intrinsic distinction between the two loci of visible teacher quality, test scores and classroom practice. The picture of teaching *practice* that springs from the notion of scaling that so predominates

in the reform movement, however, risks merely translating achievement score data into the sorts of isolated behaviors that would then signify "good teaching" to teacher evaluators.

If the "classroom observations," which are offered as a *complement* to student achievement measures, ostensibly intended to round out the assessment of teacher quality, do no more than seek out and check off the desirable behaviors whose desirability is a function of achievement score growth, then such observations offer no complement whatever. Such observations would merely proxy for the achievement score data whose inadequacy to the task of evaluating teachers has been roundly proclaimed by all sides of the discussion.

Nevertheless, Duncan himself also references the immediately visible quality that seems to require no such behavioral prescription, which means that in the concept of teacher quality undergirding the Race to the Top, there remains – rhetorically, at least – something like a version of "good teaching" that is knowable without any recourse to data whatsoever. But if it is indeed the case, as Duncan himself avers, that some kind of educational quality is immediately apparent, classroom observations of the sort I have just described will systematically preclude the possibility of recognizing this quality as such.

<u>1.5 Improving the Teaching Corps: Rewarding and
Weeding</u>

Bearing in mind Duncan's repeated insistences that, in the first place, teacher evaluations must take "student achievement" into account; and that, in the second place, such evaluations cannot *simply* take the result of "a multiple choice test" as the measure of teacher quality, I turn now to Duncan's speech to the National Education Association, in which Duncan specifically wishes to place his "focus… on the quality of the education workforce— teachers, principals, and education support professionals" (Duncan, 2009a). My own focus

lies in arriving at an ever-more-nuanced sense of what Duncan takes *teaching*, conceptually, to be (to consist in). I have already explored his remarks on data, on standards and assessments, and on taking innovative educational solutions to scale. These explorations have revealed certain unaccounted-for flaws in the notion of "college and career readiness" that functions as a criterion of success against which the educative process ought to be evaluable. The flaws pertain to the apparent conflation of preparation for college with preparation for career, as well as to the proxy measures that have (so far) assumed the role of gauging this preparation: namely, the very student achievement measures that the criterion of "college and career readiness" was more or less explicitly designed to supplant.

Additionally, and certainly relatedly, however, the exploration of Duncan's rhetoric has exposed certain uninterrogated assumptions underlying both his discourse on data and on the possibility of "scaling-up" and transferring educational success among a variety of contexts. These assumptions commit him to a particular view of teaching, as the previous section has revealed. This particular view of teaching constitutes the practice of teaching as the implementation of specific behavioral techniques proven – through their correlation with achievement score gains – to "work" no matter the other variables at play in a given classroom, including but not limited to the variable of student socio-economic status to which Duncan relentlessly points. The definitional link between "what works" and the achievement measures that reveal this "working" risks "reducing," in Duncan's words, the recognition of good teaching to the analysis of test scores, without the which good teaching seems somehow unknowable, or at any rate, unknown to the public at large.

At the same time, however, there yet remains the recurrent motif running throughout Duncan's speeches in which the Secretary praises the complexity of the process of teaching

and alludes to the immediate visibility of educational quality, both of which hints seem to augment the otherwise ruthless reduction of teaching to the production of achievement score gains. It would appear, as we turn to the final speech, that the contrast between the everyday availability of "good teaching" and its technical definition represents a major point of tension in Duncan's own rhetoric.

While acknowledging the limitations of raw numbers to reveal "the whole truth," Duncan also claims that "[data] doesn't lie" (Duncan, 2009b), and this apparently ambivalent attitude toward the usefulness of the measures upon which so many of his reforms are made to hinge continues to dog his remarks throughout the summer of 2009 and beyond. The concept of teaching that has so far emerged, particularly in the speech on turning around troubled schools, has been one clearly defined in terms of the behaviors producing test score gains. Insofar as teaching, in accordance with this definition, is considered *only* in light of student achievement measures, the data certainly does not lie.[6] Insofar as Duncan continues to acknowledge that teaching cannot *only* be evaluated according to achievement measures, such data certainly does not "tell the whole truth." The centrality of "robust data" to Duncan's reform plans, evident also in the fact that it furnished the subject of the very first speech in the cycle, continues to outpace whatever concept of teaching that Duncan imagines would complement it, such that the "whole truth" that ought to be told, ideally and specifically in a realm regarding teacher "evaluation, compensation, and tenure decisions" (Duncan, 2009a), continues to await its telling.

---

[6] It absolutely cannot. It is a tautology.

Duncan's final speech leading up to the introduction of Race to the Top features a more fully-developed picture of teaching than has heretofore appeared, inasmuch as it reveals the role of the teaching profession in his approach to improving the national quality of education. Exploring the tensions and symmetries between his "technical" view of teaching – that which emerges according to test score gains – and what I have been calling his "ordinary" or "everyday" conception of teaching – which he often signals with phrases urging the audience to recall an important teacher in their own lives or other such rhetorical devices – will occupy the center of my investigation now.

In Duncan's final speech, he reiterates, nearly verbatim, some of his favorite chestnuts from his earlier addresses. In particular, he emphasizes the role of data that he envisions for the classroom teacher, which reinforces the technical concepts of teaching and learning to which I already alluded:

> Now, let's talk about data. I understand that word can make people nervous, but I see data first and foremost as a barometer. It tells us what is happening. Used properly, it can help teachers better understand the needs of their students. Too often, teachers don't have good data to inform instruction and help raise student achievement (Duncan, 2009a).

Repeating one particular refrain regarding the appropriate use of data, Duncan underscores its ability to "inform instruction," to "help teachers better understand" student needs, and to tell "us what is happening." In this instance, "what is happening" appears in connection with student progress specifically, and the assumption seems to be either (1) that teachers cannot tell in the absence of data "what is happening" with their students, or (2) the "us" to whom the data reports extends beyond the teacher. The vital salience of point (2) will return in the final Chapter of this project.

Duncan's immediate shift following the quoted passage above indicates that the latter assumption – point (2) – is the more likely one. Says Duncan, "Data can also help identify and support teachers who are struggling. And it can help evaluate them. The problem is that some states prohibit linking student achievement and teacher effectiveness" (Duncan, 2009a). In this instance – which we have heard before, particularly the concern with "firewalls" preventing achievement data from being harnessed for evaluation purposes – the data has little to do with informing instruction, at least if one assumes that the teacher is primarily responsible for designing and conducting that instruction. In a quick maneuver, data becomes an administrative tool for identifying "struggling" teachers, offering to "help evaluate them."

None of this is new to Duncan's vision regarding educational quality. What bears reiteration, mainly because Duncan reiterates these points, is that the *same* data that ostensibly serves *teachers* as a "barometer," helping to guide their instruction and increase their understanding of student needs, also and once again, without undergoing any transformation at all, serves *administrators* in an evaluation-identification capacity, where "struggling" teachers are distinguishable according to such metrics. The conflation of teacher and administrator purposes once more indicates a profound propensity to understand the whole of *learning* – the telos of the educative process – more or less entirely in terms of student achievement data. In order for the data to "inform instruction" or help teachers grasp student needs, it must be *revealing* of those needs; it must reveal directly the aspects of learning that teachers are expected to address.   In order for the data to serve *administrative* or *evaluative* purposes, it must likewise reveal the degree to which teachers are in fact addressing the aspects of learning for which they are responsible. The "learning" at work

here, both at the level of student need and teacher expectation, is thus wholly constituted by math and reading scores – that is the only sense in which such data can be put to the service of teacher and administrator purposes simultaneously.

Duncan evinces, here and elsewhere, the certain hyperawareness of this propensity that I have noted before, which is not necessarily to imply any defensiveness on his part, and he indicates his consciousness of the issue in his frequent statements with regards to the "nuanced" and "complex" nature of teaching, as we have already seen. The question remains, however, as to the degree to which this awareness translates into an actual conceptual resilience of ordinary-sense teaching and learning in the face of the propensity just noted.

On the one hand, such conceptual resilience in the face of the temptation to view both teaching and learning wholly in terms of achievement data appears undercut at every turn. During explicit discussions of data's role in educational endeavors, Duncan repeatedly verbalizes his keen awareness of achievement data's fallibility – the need to use it carefully and alongside other forms of measurement. But this appropriately cautious approach to data simply goes out the window when his back is turned, as it were.

For example, earlier in this very speech, referring to administrative encouragement for professional development, Duncan says, "I am a big believer in this program, but let's also be honest: school systems pay teachers billions of dollars more each year for earning PD [professional development] credentials that do very little to improve the quality of teaching" (Duncan, 2009a). Duncan appears to refer to factual knowledge here, something established as true or accurate; and not only that, but also something somewhat obvious, as his "let's be honest" appeal indicates: additional credentials – teacher certification and advanced degrees

– "do very little to improve the quality of teaching." Here, in discussing state and local investment in teacher credentialing and professional development, and not the appropriate application of "robust data systems" in educative processes, Duncan appears to let his guard down with respect to the appropriately circumscribed use of student achievement data, and in the following way:

When claiming that professional development or additional credentials have little effect on "the quality of teaching," he invokes one or both of two lines of research that trace the effects, respectively, of teachers holding graduate degrees and teacher certification status on student outcomes *as measured by growth in student test scores*. A typical finding in the latter instance tells the story like so: "Finally, consistent with previous work, there is little or no evidence that a master's degree raises the quality of teaching. All estimates are small (or negative) and statistically insignificant" (Rivkin, Hanushek, & Kain, 2005), and it is a point echoed in myriad other places both prior and subsequent to Duncan's speech (Chingos & Peterson, 2011). Meanwhile, in terms of credentialing or certification, an oft-cited study[7] concludes, "Our results suggest that the emphasis on certification status may be misplaced. We find little difference in the average academic achievement impacts of certified, uncertified and alternatively certified teachers" (Kane, Rockoff, & Staiger, 2008). When these authors speak of "academic achievement," they explicitly point to math and reading achievement scores.

---

[7] A google scholar search (on March 7, 2012) reveals that this study has been cited in 282 published books and papers since its own publication, making it most likely the one that Duncan has in mind in June of 2009.

When Duncan claims, then, that professional development seems to "do very little to improve the quality of teaching," the teaching he has in mind is the form of teaching defined *only* in terms of the fallible data he is elsewhere careful to bracket with caveats about the need for "multiple measures" of quality. No such multiple measures appear in the accepted truth about teacher credentialing that he claims in this example. It will be notable that when other proxies for the broad concept of learning are employed – equally incomplete proxies, one must add – different conclusions result. A recent study that considers the impact of teacher credentials on student GPA, for example, which takes data from Swedish compulsory schools, finds that certification does indeed have a significant effect on "student learning" (Andersson, Johansson, & Waldenström, 2011). Whether or not credentialing has an impact on "teacher quality" seems to depend – as common sense would predict – on what is taken to constitute or reveal "student achievement" as itself an indirect indicator of "teacher quality."

The quotation marks in the above paragraph are crucial: they serve to indicate that the meaning of "student achievement" and "teacher quality" are *not* the everyday or ordinary uses of the terms, but are rather *technically* defined, visible according to an *explicit and limited* subset of all criteria by which such things *might* be visible. Such scare quotes are notably absent in Duncan's assertion about the effects of professional development on the quality of teaching, which only aids the impression – against what Duncan explicitly insists again and again, at least when speaking explicitly about the uses of data – that, for him and his policies, teaching and learning really does come down to measurable growth in student achievement scores in one form or another.

In other words, despite Duncan's attempts to maintain the integrity of an ordinary concept of good teaching, and to police its distinction from the technical definition thereof,

the technical definition seems to usurp the place of the ordinary sense of the term when Duncan's back is proverbially turned. When Duncan calls for the inclusion of achievement data in teacher evaluations, citing also the caveat that teaching and learning can never be reduced to such a measure, he explicitly wishes to use a technical sense of "teacher quality" to achieve a fuller view of the *ordinary* concept of good teaching; instead, the technical sense comes to function interchangeably with the ordinary sense, and thus threatens to replace it entirely, and by wholly invisible means, with nary a ripple.

In the Introduction, I asserted that a conceptual investigation of good teaching as it appears in the contemporary education discourse would most certainly need to examine both the official rhetoric of the Department of Education and the rhetoric circulating in the popular media, as well, such as the education reporting of *The New York Times*, for instance. In discussing the "technical" sense of "teacher quality" in the paragraphs above, I have cited a number of studies in order to reveal the ways that the technical sense of "teacher quality" emerges in scholarly literature and is taken up into official rhetoric. I wish to note the prevalence of economics-oriented journals among those citations, which is surely more than coincidentally related to the prominent place of the "scaling up" trope I analyzed earlier. I would also like to point out that two of the studies I have so far cited (I will cite more later on) are authored by some combination of Thomas J. Kane of Harvard University and Douglas O. Staiger of Dartmouth. These two scholars specialize in developing value-added methodologies of determining teacher quality – methods of determining quality in its *technical* sense, as revealed in achievement score growth – and applying those methods to a variety of questions (such as the effects of credentialing and professional development, for example).

In order for a technical definition of teacher quality completely to usurp the place of the everyday concept of good teaching to which Duncan continues to allude, not only would such conceptual slippage have to be visible in the official rhetoric – as we have seen that it is – but it would also have to appear in the popular discourse, as well. Questions pertaining to teacher quality would have to be addressed in everyday or ordinary arenas on technical terms, and without any indication of the distinction that Duncan avowedly maintains with the ordinary concept of good teaching.

The popular discourse indeed reveals a considerable amount of this conceptual slippage. Simply looking at the archives of the *New York Times* alone, without examining any other major outlet, Douglas Staiger and Thomas Kane been invoked or sought out for comment in response to questions about teacher quality, compensation, and evaluation in seven articles over the past four years (Dillon, 2010; Kristof, 2007, 2009; Leonhardt, 2010; Medina, 2008, 2010; Otterman, 2010), several of which articles appeared on the front page of the paper. Their high profile in this regard exceeds even that of the leading giant in the field of teacher education and preparation (and, at one time, candidate for the very job Duncan now holds in the Obama administration), Linda Darling-Hammond. She apparently only gets to write letters to the *Times,* or else is merely dismissed as representing an "establishment view," whatever that might mean (Brooks, 2008). Some of the articles in which Staiger and Kane appear also cite Chetty, whose work Duncan and Obama have each invoked in their most recent official discourse, as has been examined in the Introduction. By serving as sources for reporters writing for the public on issues of teacher quality, these scholars enter into the discussion of "teacher quality" with no allusion to any conceptual distinction whatsoever. The feedback loop between official and public discourse echoes and reinforces

and propagates a highly technical sense of teacher quality that radiates outward over the general public in precisely the top-down fashion that Duncan wishes and claims to avoid.

1.6 Ordinary Teacher Quality and the Evaluation
Systems of the Future

So much for the technical concept of "teacher quality" as it appears in Duncan's cycle of speeches. The way in which teaching has been both explicitly and implicitly conceived in terms of student achievement growth in the domains of math and reading has been amply demonstrated by this point, and the tension between this precisely defined sense of teacher quality and some more ephemeral concept to which Duncan often appeals has likewise been made evident. In his fourth and final speech, Duncan takes some steps to clarify the ephemeral nature of this alternative, ordinary concept of good teaching that he evidently has in mind as a goal for the profession. At stake, once more, a theme that has been echoing since the opening paragraphs of this project, is whether or not this everyday concept can withstand or resist its conflation with the technical concept.

I noted earlier that, on the one hand, such conceptual resilience in the face of the temptation to view both teaching and learning wholly in terms of achievement data appears undercut at every turn. Now I wish to say that, on the other hand, despite the persistent undercutting of this everyday conception of teaching through implicit or explicit references to achievement data, Duncan continues to insist upon the importance of the ordinary sense of good teaching in his rhetoric. Despite the robust data systems he prizes, Duncan continues to recognize a simpler, immediately available sense of good teaching that seems, if not orthogonal to demonstration in terms of achievement data, then certainly independent thereof.

In his speech to the NEA, Duncan raises this ordinary conceptual terrain through narrating relevant memories of his own childhood, in which his mother "began an after-school, inner-city tutoring program." He highlights his "community's chaos" and the experience of "early, violent deaths" due to the violence of the neighborhood, but he notes that:

> from the group of friends I grew up studying with and playing ball with, from one street corner at 46th and Greenwood, emerged literally a brain surgeon, a Hollywood movie star, one of my top administrators at the Chicago Public Schools, and one of IBM's international corporate leaders (Duncan, 2009a).

That he and his childhood friends could attain such high-profile careers despite the socio-economic conditions in which they came up militates, first of all, against any objection that one must consider socio-economic conditions when evaluating teachers. Duncan attributes this outlying success to the influence of a meaningful adult involvement in his life:

> How did this happen? Because these children, despite tremendous poverty, despite staggering neighborhood violence, despite challenges at home, had my mother and others in their lives who gave them real opportunities, real support and guidance over the years, and had the highest expectations for them.

Duncan's rhetoric thus far accomplishes two things with respect to his explicit agenda: in the first place, it offers a preliminary depiction of an ordinary sense of good teaching, and in the second place it emphasizes the potential impact of such teaching in terms of "college and career." The distinction between the sense of "good teaching" that appears in the narrative and the highly technical sense alluded to above appears in the numerous places. In the first and most obvious indication of difference, the concept of teacher quality visible in the above narrative only becomes knowable – *tellable* – in extreme hindsight, after Duncan and his childhood friends have already achieved career success. Would such success have been predictable by virtue of math and reading scores (and is becoming a "Hollywood movie

star" contingent thereupon in the first place)? The assumption that underlies Duncan's use of this particular anecdote as a means of *indicating the need* for increased data use in teacher evaluation protocols seems to be that it would. In the second place, though, Duncan alludes to the way the teacher figure in this narrative offered "support and guidance *over the years,*" a fact that distinguishes the concept of "teaching" involved here from that accessible in terms of the technical sense already explicated. Duncan must once again assume that the technical sense of teaching implicitly includes the kind of "teacher quality" that explicitly exceeds the scope of measures upon which his technical definition rests, since what he points to – the over-the-years involvement of great teachers – can in no way be inferred from students' score growth from one year to the next.

It is, nonetheless, a common feature of elementary, middle, and high school experience alike – to say nothing of college and graduate school – for a teacher to function in teacherly fashion in both an administratively official capacity (i.e. one in which the student appears on the class list assigned to the teacher) and, in ensuing years, an unofficial capacity, offering longer-term "support and guidance" of precisely the kind Duncan depicts in the instance of his mother. Duncan is quite right to recognize that teacher quality pertains to this domain as well. This very domain of the concept of "good teaching," however, exceeds the technical definition. While some correlation may exist between the two definitions, it certainly cannot be assumed.

Duncan knows *now*, in the present – and can report – that his mother's after-school program in the past, her presence and the presence of others like her in the lives of his friends, made a tremendous difference, which resulted in the prestigious careers that he names. Are we to imagine that the "teacher quality" of Duncan's mother was either

unknowable or *unreportable* during the period in which Duncan and his friends were growing up together?   She ran an "after-school" tutoring program which generated (or was responsible for) no achievement data whatsoever – even under the reforms Duncan proposes, *her* quality would not be visible, and also, more nefariously, the classroom teachers on whose rolls Duncan and his cohort appeared would have received the credit for their gains, unearned. It seems absurd even to suggest that her quality would have been somehow invisible to anyone locally familiar. Rather, Duncan appears to display an instance in which a teacher's quality would have *required* no data, would have been *immediately visible*, much as when Duncan knows a successful school "the moment [he] enter[s] the school." He introduces, in short, an overt reference to the ordinary or everyday concept of teacher quality, one which is emphatically not a subset of the technical variety.

Duncan in fact emphasizes the ordinariness of this version of good teaching, and also specifies some of the content of the concept, as he reiterates the importance of such teachers:

> All of us remember an educator or coach who changed our life. It stays with us forever. It sustains us, guides us, and inspires us. They're the ones who commit those everyday acts of kindness and love and never ask for anything in return. They counsel troubled teens, take phone calls at night, and reach into their pockets for lunch money for children who are too ashamed to ask (Duncan, 2009a).

By inviting the audience to communally and individually "remember and educator or coach who changed our life," Duncan underlines an experience *so ordinary* that he can assume that his entire audience shares it. He fleshes out the content that constitutes the teacher quality he depicts in pointing to "those everyday acts of kindness and love," such as "counsel[ing] troubled teens, tak[ing] phone calls at night," and so on. None of this content shows up in lesson plans, to be sure, or, quite generally, in classroom observations. The concept of teacher quality on display here clearly and self-evidently differs from the

technical definition – the very content *indicative* of each as such fails to evince even the slightest overlap. Nothing at all warrants the assumption that the sort of teachers Duncan depicts in the concrete example of his mother and in the reverie he conjures in the above quotation would *also and necessarily* be the sort of teachers to consistently lift student test scores.

In a speech on teacher education programs given in October of the same year (Duncan, 2009d), the Secretary once again invokes the ordinary concept of good teaching, and he offers new conceptual content as well. But on this occasion, he immediately runs together the ordinary sense with the technical once more, as he did earlier when his back was proverbially turned.

> Now there is a reason why so many of us remember a favorite teacher forever. A great teacher can literally change the course of a student's life. They light a lifelong curiosity, a desire to participate in democracy, and instill a thirst for knowledge. It's no surprise that studies repeatedly document that the single biggest influence on student academic growth is the quality of the teacher standing in front of the classroom—not socioeconomic status, not family background, but the quality of the teacher at the head of the class (Duncan, 2009d).

The sense of urgency that runs throughout Duncan's speeches has its basis in his recognition of the importance of teachers. In an ordinary sense, he names the reasons and the ways in which great teachers *are* important: "They light a lifelong curiosity, a desire to participate in democracy, and instill a thirst for knowledge." But in the very next sentence, the teacher importance to which he refers has nothing to do with "light[ing] a lifelong curiosity" or any of the other salutary effects that he attributes to great teachers in the previous sentence. He says, "studies repeatedly document that the single biggest influence on student academic growth is the quality of the teacher standing in front of the classroom." Suddenly, Duncan speaks of "student academic growth" rather than "curiosity-lighting" or

"knowledge-thirst." When he claims in the latter sentence that teacher quality represents "the single biggest influence on student academic growth," he entirely departs from the realm in which quality *appears in* any sort of "lighting" or "instilling" or "inspiring." The teacher quality he refers to in the latter sentence is revealed in the amount of variance in student math or reading scores accounted for by variance in "teacher quality," however that variable is (technically) defined. So the two concepts of teacher quality are precisely as (dis)similar as inspiring a lifelong love of learning in a student on the one hand, and accounting for variance in a math or reading score from one given year to the next, on the other.

In a subtle and perhaps unintentional move, he thus abruptly conflates the great teacher "so many of us remember" with a teacher who "influences student academic growth." He, himself, has been adamant that these are different constructs, and with good reason. But his own example above indicates the ease with which the dramatic distinction between these concepts simply vanishes in the course of discussing teacher quality.

In pointing to or evoking the life-changing capacity of teachers, Duncan appeals to another criterion of success than that which appears merely in achievement data, something more akin to the face value of "college and career readiness," which is also intended to peg teaching and learning to real-world outcomes. He raises the above examples in order to highlight the *concrete* importance of good teachers to the future lives of their students. In referencing the "everyday acts of kindness and love," he indicates that the *quality* of the acts are bound up genitively with the contextualized acts themselves, rendering this quality immediately apparent to the situated observer. No mediating, abstract knowledge-producing protocol seems required to understand the acts in question as acts *of kindness*, for example. The real-world outcomes that serve as the ultimate aim of teaching in his rhetoric remain tied

to the immediately visible quality of the teachers whose stories he tells. The immanence and commonality of the kind of teacher quality he has in mind in moments like these distinguishes the everyday concept of "good teaching" from the technical variety outlined earlier.

The distinction between the ordinary concept of good teaching and the technical sort studied by economists such as Kane and Staiger remains important to this discussion precisely because *Duncan himself* insists upon it. I may appear at pains to note and substantiate the profound conceptual gulf between the two notions of teaching, but the driving impetus comes from Duncan's own words on the matter that appear in this very speech: the "complex, nuanced work of teaching can never be reduced to a simple multiple choice test" (Duncan, 2009a). Duncan's use of the term *reduction* in the foregoing sentence emphasizes that the "complex, nuanced" work of teaching *exceeds* the scope of existing achievement measures. If it is indeed the case that good teaching cannot be "reduced" to achievement measures, then neither can it be simply *extrapolated from* achievement data. Duncan avers, in essence, that good teaching is more than lifting achievement scores. This *more than* manifests itself in the both the example of his mother – an extracurricular tutorial figure over several years in the life of neighborhood children – and in his appeal to the audience to recall the transformative, life-changing teachers in their own life.

The question remaining before Duncan and our investigation pertains, yet again, to the ability to resist the gravity of the technical conception of teaching in the search to say something meaningful about the ordinary concept of good teaching. Further consideration of Duncan's speech to the National Education Association on the subject of improving the teaching profession ought to shed light on the relative possibility thereof.

The brunt of his fourth speech – following his laudatory depictions of great teaching – has to do with creating the conditions for an influx of great teachers into classrooms all over America. Teachers requiring replacement will leave the classroom for a variety of reasons, but two in particular urge his attention: the mass retirement of baby boomers, to which Duncan speaks in a later address (Duncan, 2009d), and attrition or termination. Improving the quality of the teaching corps generally thus requires two specific efforts, one geared toward ensuring that the quality of new teachers entering the profession be superior, aggregately, to those leaving it (the subject of the October speech just cited), and one geared toward giving LEAs in particular greater freedom to make decisions with regard to personnel – in terms of both termination and compensation – to the effect of removing poor teachers, retaining good ones, and recruiting promising ones. That comprises the main thrust of the speech currently under discussion.

The bulk of this fourth speech aims in the thorny direction of teacher unions and collective bargaining agreements, which would require reform in their own right in order to free LEAs to make the type of personnel decisions Duncan views as necessary to education reform as a whole. The political events of the winter and spring of 2011, in which the governments of Wisconsin, Ohio, Florida, and a handful of other states introduced legislation that would curtail public employees' rights to bargain collectively in the first place, ignited protests and counter-protests across the country. During this upheaval, teachers unions, and particularly the tenure protections they commonly negotiate with states and districts, came to represent for television pundits the disastrous mediocrity brought about by unions in general. One Fox News commentator put his complaint with teacher tenure succinctly: "How hard would *you* work if you couldn't get fired?" (Stewart, 2011). Duncan announces near the

opening of his speech to the NEA that he "came here today to challenge you to think differently about the role of unions in public education because, when thousands of schools are chronically failing and millions of children are dropping out each year, we all must think differently" (Duncan, 2009a). The implication, obviously, is either that the "role of unions," as it has heretofore existed, has something to do with *causing* the currently unsatisfactory situation, or else that the "role of unions" as it presently stands, will somehow *prevent* reformers from remedying the current state of affairs, or both. The question that Duncan directly addresses, then, as to the appropriate "role of unions in public education" lies mostly beyond the scope of this conceptual investigation. The concept of "teacher quality," however, lies at the heart of the particular *way* Duncan wishes to alter the role of unions, and that is the element to which I mostly wish to attend. After all, if LEAs will have more freedom to hire and fire on the basis of quality rather than any other standard, the determination of quality that comprises that basis is of pivotal importance

As he turns his attention to his view of the teaching profession's future as a whole in his July speech to the NEA, Duncan's anecdotal praise for the kinds of great teaching modeled by his mother's example, as well as the sort of teacher quality exemplified in "instilling a love of learning" and "taking phone calls at night" stands in vivid contrast to a problem at which he has hinted before – a perceived inability to distinguish good from bad teaching, as evidenced by the fact that "99 percent of teachers are all rated the same" (Duncan, 2009c). To reiterate a point once more, the ineffectiveness of current systems for evaluating teachers, whose ineffectiveness is indicated in its apparent failure to differentiate among teachers at all in a world where such outstanding teaching as that to which he alludes is a common feature of school life, offends Duncan and thus calls for a great measure of

correction. The anecdotes he employs in order to underscore the *need* for such improved modes of knowing teacher quality prove telling, precisely inasmuch as they are examples of the ordinary, everyday concept of good teaching that has already been presented. In other words, the rhetorical structure of his speech implies that of central importance to his effort lies the project of knowing great teaching *in its ordinary sense.* The ability to identify and reward teachers like his mother, and the need to recruit and train more in her mold, drives the reform measures he endorses.

In considering the content of this speech, then, one needs to seek out the concept of teacher quality where it emerges and to assess its shape relative to this everyday sense that motivates the approach Duncan takes. In the opening several minutes, Duncan lays out ordinary examples of great teachers, which serve to justify the need to reward them, train more like them, and remove those unlike them. The question at hand has to do with whether *these* teachers in their ordinary sense remain Duncan's focus throughout the speech.

In the second half of the speech, he lays the responsibility for the current failings of the public schools at the feet of the profession itself, in the form of the teacher unions:

> And I'm telling you as well that, when inflexible seniority and rigid tenure rules that we designed put adults ahead of children, then we are not only putting kids at risk, we're also putting the entire education system at risk. We're inviting the attack of parents and the public, and that is not good for any of us.
> I believe that teacher unions are at a crossroads. These policies were created over the past century to protect the rights of teachers, but they have produced an industrial factory model of education that treats all teachers like interchangeable widgets (Duncan, 2009a).

The reference to seniority and tenure specifically pertain to union-negotiated contracts or collective bargaining agreements (CBAs), the most common models of which, in P-12 teaching, include provisions that reward seniority both monetarily (with step-wise

salary increases throughout a teacher's career) and in terms of tenure protection (after a

certain number of years). A 2005 report on the effect of existing CBAs on urban districts'

hiring policies, conducted by the New Teacher Project (whom Duncan has invoked as

praiseworthy earlier) sums up the situation – in words that Duncan fairly straightforwardly

echoes in his 2009 speech:

> At the core of these transfer and excess rules is a single principle: every incumbent teacher is guaranteed a job in a school, frequently in seniority order, even if no school wants to hire him or her…. by adopting factory model protections, these efforts codified an important set of assumptions that structures the work and culture of urban schools to this day. These early contracts determined that job protections and teacher movements would occur according to seniority and required that all teachers be treated as if they were interchangeable in every other respect (Levin, Mulhern, & Schunck, 2005).

The "factory model," in which teachers are "interchangeable widgets," reflects

Duncan's dissatisfaction with the failures of contemporary evaluation methodologies to

distinguish among teachers in terms of quality at all. Thanks to the union-negotiated CBAs,

nothing pertaining to teacher *ability*, to teacher *quality*, separates one teacher from another;

rather, seniority remains the only distinguishing feature.

Interestingly, the implication, when Duncan says that this "industrial factory model of

education… *treats* all teachers like interchangeable widgets" (emphasis added), is precisely

that teachers are obviously *not* interchangeable widgets. I raise this issue not to dispute the

point, but rather to refer back to an earlier part of the discussion in the interest of previewing

an upcoming section: his denigration of the "factory model" seems on the surface to run

directly *counter* to his adoration of the "scaling up" trope that he uses so frequently. If

factory models are insufficient primarily because they fail to make important distinctions in

rendering certain things – teachers, in the current case – interchangeable, it is difficult to see

the immediate appropriateness of the "scaling up" metaphor in reform efforts, inasmuch as it likewise levels potentially decisive distinctions in assuming that "what works" (what defines teacher quality) in one instance will necessarily "work" in another. *Quality teachers*, in Duncan's locution, are *not* interchangeable; *teacher quality*, through the process of behavioral scaling, on the other hand, somehow *is*.

But to return to the point of this section – drawing out the conceptual structure to which Duncan appeals – one must add the feature of relative permanence and stability to the concept of teacher quality. Earlier – in the discussion of the scaling-up trope – I pointed out that in order to be amenable to scaling, teaching must *consist in* what Elizabeth Green calls "a series of bite-sized moves" (Green, 2010) or isolated "techniques" (Lemov, 2010). In order for *quality teachers* to be non-interchangeable, however, such that the *teachers* themselves are being distinguished from one another in an evaluation process, the *having* of these particular techniques or moves needs to manifest itself in a holistic fashion, all at once. A further assumption necessary to this particular view would entail that this *having* of the ability to teach well, once *had*, remains relatively constant. Otherwise one cannot talk about a *good teacher* with any sort of futural sense necessary to the kinds of calculations one would make in contemplating whether to reward or terminate someone.

The phrasing of the foregoing section makes the phenomenon sound perhaps more unfamiliar than it in fact is. It is an ordinary feature of everyday experience in any number of domains, be it sports or knitting or music: novices who enter a profession or any other skill domain get better at navigating that domain skillfully over a given period of time. At a certain point, they cease to improve at the same rate but do not necessarily noticeably decline. We then say that they have "reached their potential." The potential, and the

eventually-attained skillfulness, are *theirs*, such that it makes sense to talk about *them* being good at a given thing, *their* quality in the skill domain. A person is a good knitter, a good tennis player, a good violinist. According to Duncan's assumptions, as well, there are likewise *good teachers*. I point this out primarily to indicate a point of potential conflict or tension with the *technical* sense of teaching to which Duncan has a habit of alluding: the transition from the conceiving of teaching as the exercising of isolated techniques to a view in which teaching is an internalized, self-consistent, and holistic skill such that a person can be or possess *it*, rather than wielding or demonstrating *them* (isolated behaviors), remains to be accounted for. The picture wherein good teaching consists of a set of techniques that correlate with the lifting of test scores (the technical view) does not self-evidently match up with a picture of good teaching as a holistic, self-consistent and relatively constant-over-time quality (the ordinary, everyday view).

The assumption, then, at work in this section of Duncan's speech certainly accords with the everyday concept of good teaching that Duncan continues to invoke. Under that concept, there *really are* such things as good teachers, whose quality, once established or developed, remains relatively consistent. Thus, under a system like the one Duncan proposes, one that offers to reward excellent teachers, a principal would be justified in increasing a given teacher's salary for an upcoming year based on the principal's evaluation of a previous year's performance. The very notion of rewarding good teachers requires that the quality of the teacher remain relatively constant into the future, which fits the everyday concept of good teaching in a way that does not immediately accord with the technical sense. Whether or not this self-contained consistency of teacher quality appears in the technical sense of good teaching remains to be seen.

Duncan's speech aimed at the teachers' unions, then, addresses two pressing needs: the need to reward good teachers for their excellent quality, which assumes that this quality is internal to the teacher (the ordinary sense), and the need to remove poor teachers, which makes the same assumption and is rooted in the same ordinary concept. Therefore, assessing whether the dismantling of the seniority rules in CBAs, which prevents salary and personnel decisions from considering teacher *quality*, and eradicating tenure protection, which on his view has the same effect regarding the flexibility to excise poor teachers in particular, will have the desired effect remains entirely dependent on the specific concept of teacher quality operational in the policy itself, and the means of ascertaining or measuring it. We have just seen fairly clearly that the concept of teacher quality toward which Duncan rhetorically gestures in the speech so far is an *ordinary* one. But he has yet to speak to the possibility of measuring or ascertaining it, and in making the concept operational, there is certainly room for the kinds of conceptual slippage that we have already seen.

In looking more closely at the paths Duncan treads in this speech, I will end up including quotations that have already appeared in earlier sections. This is not an exercise in mere redundancy or repetitiveness, but is rather undertaken in the hope that by returning to Duncan's words as quoted earlier in light of the fleshing-out of the concepts of good teaching that continue to develop and evolve, his quotations will resonate in new and more multifariously revealing ways.

Of the two perceived reform needs pertaining to improving the teacher corps that we have already elucidated, Duncan speaks first to the need to reward great teachers: "Many schools give nothing at all to the teachers who go the extra mile and make all the difference in students' lives. Excellence matters and we should honor it—fairly, transparently, and on

terms teachers can embrace" (Duncan, 2009a). In the interest of demonstrating the fair, transparent and cooperative nature of such endeavor, he points to the program he developed in Chicago with the aid of "24 of the best"[8] teachers in the system: "It was based on classroom observation, whole school performance, and individual classroom performance, measured in part by growth in student learning" (Duncan, 2009a). I have already noted, with regard to this locution, that "classroom performance, measured in part by growth in student learning" straightforwardly refers to student achievement gains *coupled with* "classroom observation." Only a few minutes later in the same speech, he makes the other salient statement to which I have already referred: "I understand that tests are far from perfect and that it is unfair to reduce the complex, nuanced work of teaching to a simple multiple choice exam. Test scores alone should never drive evaluation, compensation, or tenure decisions" (Duncan, 2009a). I raise these twin issues again mainly to highlight once more the weight that such observation is meant to bear conceptually, and in circling around to it here, I wish to carry this idea somewhat further.

As technically defined – in its technical and limited sense – good teaching appears in the student achievement growth to which Duncan refers here. But Duncan also notes that this technical sense is not enough: "it is unfair to reduce the complex, nuanced work of teaching to a simple multiple choice exam." He clearly alludes to the insufficiency of the *means* of ascertaining the technical sense of good teaching to reveal something durable about the everyday or ordinary sense of good teaching, and in doing so he zeroes in on the *ordinary sense*. Rewarding and terminating on the basis of the *ordinary* sense of good

---

[8] How he determined the quality of these teacher, or by what criteria they stood out as the "best," is not addressed.

teaching remains the centerpiece and goal of his strategy. In noting that achievement score growth alone reveals merely the technical sense, such that the exemplary model of a functional evaluation procedure to which he points employs both achievement data (in determining both "whole school performance" as well as, partially, "classroom performance") *and* "classroom observation," the observation piece clearly and obviously plays the role of *rounding out* the evaluation such that the *ordinary sense* of good teaching would emerge into view above and beyond the technical sense. Classroom observations' value therefore lies in bridging the conceptual gap between the explicitly inadequate technical sense revealed in test scores and the everyday concept of good teaching that remains Duncan's ideal. The role of observations, then, is positively central to the ability of a given evaluation protocol to access and address what it is that we really want from such a system: the ordinary sense of teacher quality.

The purpose of this investigation remains to draw out the concept(s) of teacher quality in the rhetoric of the Obama administration's educational policy and to explore the degree to which the conceptual integrity remains intact as the rhetoric moves toward operationalization. As regards the centrality of classroom observation to bridging the conceptual gap between a technical and an ordinary sense of good teaching, where the latter is what Duncan explicitly wishes to function as the criterion on the basis of which personnel and salary decisions can be made fairly and transparently, the relative integrity of such observations – their ability to do their conceptual work – will appear in two places, only one of which I will address at present: academic studies on the effectiveness of classroom observations, and actual state policies enacted with the specific aim of winning Race to the Top funding. For the present, I investigate classroom observations as addressed in the

studies conducted by the scholars on which Duncan so often relies, including Thomas Kane

and his colleagues.

In order for classroom observations to *bridge* the conceptual gap between the

technical and the ordinary senses of good teaching, which is the task they are explicitly

assigned, they must obviously remain free to detect good teaching over and above the

technical definition. Nothing in the scholarly literature indicates that this is the case. In fact,

insofar as classroom observation has been assessed for its value, the value has been

determined in terms of observations' ability *to detect the technical sense of good teaching*

alone. Kane and others conclude *happily* for the sake of classroom observations that, when

"well-executed," observations appear highly correlated with what achievement results reveal:

"We find that evaluations based on well-executed classroom observations do identify

effective teachers and teaching practices. Teachers' scores on the classroom observation

components of Cincinnati's evaluation system *reliably predict the achievement gains made*

by their students in both math and reading" (Kane, Taylor, Tyler, & Wooten, 2011a

(emphasis added)). Where the role that Duncan's rhetoric assigns to classroom observations

has explicitly to do with augmenting a view of teaching revealed in math and reading

achievement scores, the studies that Kane conducts measure – and celebrate – precisely

observations' ability to *conform* to that which math and reading growth reveal. Kane and his

colleagues are hardly alone in this tendency.

Glaringly, despite the fact that some scholars voice "skepticism" around the value-

added metrics by which Kane and company assess teacher effectiveness, this criterion of

quality continues to function as the standard against which classroom observations are

judged. I quote a recent study's discussion of the matter at length to expose the shape of the

thinking that undergirds it:

> We have chosen to define educational effectiveness narrowly as the value a teacher adds to gains in student learning as measured by standardized test scores. We share some of the skepticism surrounding value-added scores, yet we acknowledge that they reflect something that policy makers and much of the public truly value, and we believe that at least in aggregate they measure what they claim to measure. Nonetheless, we also believe it is important to use other methods for measuring student learning. Regardless of how student learning is defined, in the present climate of Race to the Top and other accountability initiatives, the search for effective teachers has become widespread, urgent, and high stakes, and it is our hope that an observational measure that reliably identifies the best teachers will help administrators find and support them, and in so doing advance the cause of education in a meaningful way (Strong, Gargani, & Hacifazlioğlu, 2011)

Strong and his colleagues reveal many of the conceptual problems running

throughout the discourse. Despite their "skepticism," Strong, et al. assert that these value-

added metrics "reflect something that policy makers and much of the public truly value."

This statement is accurate, insofar as the public and policy makers value a technical sense of

"student learning" that can be attributed to "good teaching." But Duncan has already

indicated that, while valuable in terms of "monitoring" or knowing "what's happening," such

metrics are not by themselves sufficient to the concept of good teaching that he and his

policies ostensibly pursue. Duncan puts forth the notion of classroom observations precisely

in order to *add* to the view of teaching revealed in value-added metrics, not merely to

replicate it. Additionally, though, Strong, et al. make a few other interesting statements

bearing mention, notably that "in aggregate," value-added metrics "measure what they claim

to measure." They offer this statement as a means of defending their choice to define teacher

effectiveness "narrowly," in value-added terms; and the "hope" expressed in their study is

that "an observational measure that reliably identifies the best teachers will help

administrators find and support them."

They certainly cannot be wrong in asserting that value-added methods "measure what they claim to measure." Strong, et al., state precisely what this construct consists of in the preceding sentence: "the value a teacher adds to gains in student learning as measured by standardized test scores." When the researchers then refer to "an observational measure that reliably identifies the best teachers," they clearly conflate the highly technical sense of good teaching that they detailed a moment earlier with an ordinary sense of what "the best teachers" means, or, what amounts to the same thing, they assume that these two concepts are identical. Value-added metrics, to which these studies of classroom observations are pegged, do measure exactly what they claim to measure. The problem, from a conceptual standpoint, is that what they claim to measure is explicitly, per Duncan's many examples, not what we talk about when we talk about "the best teachers."

Given the profound differences between the technical and ordinary concepts of teacher quality that have so far been exposed in the discourse, the dismissive tone with which Strong, et al. circumvent the issue is downright galling: "Regardless of how student learning is defined, in the present climate of Race to the Top and other accountability initiatives, the search for effective teachers has become widespread, urgent, and high stakes." They are quite right to gesture to the climatological features – "widespread, urgent, and high stakes" – that characterize the pursuit of good teachers, but to refer to the definition of "student learning" in a dependent clause, as something that can be in any way disregarded or put aside for the moment, utterly misunderstands the role this definition plays in that which they wish to discover. The definition of student learning that contributes to the concept of good teaching is by no means incidental. It is absolutely central.

Strong, et al., for all their concern and skepticism regarding value-added metrics,

have no problem declaring classroom observers "inaccurate" in terms of an ability to identify

"the best teachers." Inaccuracy here is clearly expressed in terms of deviation from measures

of effectiveness ascertained by value-added methods:

> In every case, judges achieved relatively high levels of agreement but were absolutely inaccurate, leading us to question whether educators can identify effective teachers when they see them. This in turn has motivated us to undertake development of an observational measure that can predict teacher effectiveness (Strong, et al., 2011).

While I am struck by the potential significance of the "high levels of agreement"

among the judges, I will leave that discussion for a later Chapter. And while I am somewhat

befuddled as to what sort of "skepticism surrounding value-added" methodologies enables

their conclusion that it must be the *judges* who misidentify effective teachers, I leave that too

for the time being. For now, I merely suggest that judges' "agreement," coupled with their

"inaccuracy" (deviation from achievement score data), might indicate *precisely* the presence

of the conceptual bridge Duncan's policies require. These researchers, though, are after no

such bridge. While in their study, Strong and his colleagues here draw a different conclusion

from Kane's group, ultimately questioning "whether educators can identify effective teachers

when they see them," in both studies, the relative success or failure of classroom

observations is revealed in the observations' ability to *agree with* the kind of teacher

effectiveness demonstrated in "the value a teacher adds to gains in student learning as

measured by standardized test scores."

The emphasis on classroom observations' conformity with value-added methods of

determining teacher quality would straightforwardly preclude the possibility of observations

functioning as one among "multiple measures" of assessing teacher quality, designed in toto

to capture the ordinary sense of good teaching. So long as observations' relative validity rests upon agreement with that which is revealed in the achievement data, there is no sense in speaking of a multiplicity of measures at all.[9] The concept of teacher quality that would result from the combination of value-added metrics and classroom observations so conceived would remain firmly tethered to the technical definition thereof. For all the rhetorical feints in other directions, it would be test scores all the way down.

To return, now, to Duncan's speech to the NEA, we have addressed the problems that appear to have developed with the operationalizing of the classroom observations Duncan espouses, at least in terms of the way observations are studied in academic circles. We will circle around to the issue of observations again in discussing states' policy-making activities in pursuit of Race to the Top funding, as the practice of observing teachers that appears therein speaks both to the relative ability of such practices to bridge the gap between the technical and the ordinary sense of good teaching and also to the ontology of teaching practice itself as conceived in political discourse.

In the meantime, Duncan's speech to the NEA also addresses the second perceived need that compels his proposed reforms: the need to address tenure practices in order to free administrators to remove and/or support underperforming teachers. As Duncan says, citing the need for reform, "When great teachers are unrecognized and unrewarded, when struggling teachers are unsupported, and when failing teachers are unaddressed the teaching

---

[9] Though I emphasize classroom observations, this is but one among the "multiple measures" with which Duncan proposes to augment raw test-score analysis; they all fall prey, though, to the same basic flaw: that their worth is determined by agreement with achievement data. In Amanda Ripley's recent article in the *Atlantic*, she notes that Ronald Ferguson's student surveys were remarkably effective as teacher quality measures. "The responses did indeed help predict which classes would have the most test-score improvement at the end of the year" (Ripley, 2012).

profession is damaged" (Duncan, 2009a). He speaks directly to the historical function of tenure regulations, and, once more invoking the upshot of the New Teacher Project's work – to which he also refers explicitly elsewhere in this speech:

> We created tenure rules to make sure that a struggling teacher gets a fair opportunity to improve, and that's a good goal. But when an ineffective teacher gets a chance to improve and doesn't—and when the tenure system keeps that teacher in the classroom anyway—then the system is protecting jobs rather than children (Duncan, 2009a).

It is worth noting in an aside that a teacher's getting "a fair opportunity to improve" was not, in fact, the reason that tenure rules came about, a fact acknowledged in the New Teacher Project study to which I referred earlier (Levin, et al., 2005), but that is not of central concern here. Of more direct relevance remains the issue that the "tenure system" has the potential to keep "ineffective teachers" in the classroom despite a failure to improve.

These two quotations seem to combine in an interesting fashion the two concepts of teacher quality that I have been tracing here, the ordinary sense and the technical. In the first place, once again, "great teachers" are conceived as having a sort of stable, internally-consistent kind of quality, something that, once developed, appears relatively unshakable. In the second place, this quality is described in terms of its development over time, such that a given teacher receives "a fair opportunity to improve," after which opportunity, the teacher is therefore considered beyond the possibility of meaningful improvement. These constructions both imply that not all presently ineffective teachers are of low quality. Rather, a teacher's latent quality comes to reveal itself through a process of skill development, which is presented as occurring over some (non-specific) amount of time. Following this temporal window for improvement, a teacher's true quality will be manifest. The teacher will then either be of high quality, and deserving of reward, or of low quality, and thus deserving of

replacement (the third category of response, "support," obviously pertains to the window of improvement).

Of particular import here is a notion which must be addressed more fully through considering Duncan's remarks at Columbia's Teachers College in October of 2009, and which refers back to an earlier observation about the point of tension between considering teaching *techniques* (that constitute "good teaching") transferrable while considering *teacher quality* somehow an innate trait rendering teachers themselves non-"interchangeable." Namely, at the bottom of Duncan's rhetoric on developing or supporting currently-ineffective-but-perhaps-remediable teachers lies a certain recourse to teacher quality as something inherent, such that habitually ineffective teachers can be declared to *lack* it. This notion refers to the ordinary sense of teacher quality, as I have said before: good teachers are good teachers and bad teachers are bad teachers, and while allowing for a period to move from beginners to full-fledged quality-display, at some temporal point the true quality of a teacher will appear, for better or worse. Teacher quality might therefore be accurately said to be *latent* – there, but waiting to emerge with the right guidance, training, and practice. The stability of teacher quality after an early period of growth is thus built into the ordinary sense of good teaching. This very stability proves necessary to the concept in that it functions as the justification for removing or rewarding teachers. If an ineffective teacher, as Duncan says, "gets a chance to improve and doesn't," the teacher is considered *stably* or *constantly* or *perpetually* ineffective. When one *knows* something about a teacher's quality, then, one knows something *lasting*, something that *endures*.

I noted above that these two quotations cited earlier reveal an interesting combination of the technical and the ordinary concepts of teacher quality. The previous two paragraphs

have traced the outline of the ordinary concept of good teaching as it appears in the aforementioned quotations. But I must also highlight Duncan's recourse to the terminology of "effective" and "ineffective" teaching in characterizing teacher quality. As one curriculum specialist has noted, "effectiveness is a value-neutral term" (P. S. Hlebowitsh, a gazillion different occasions), which draws attention to the fact that "effectiveness" itself is insufficient to evaluation insofar as it necessarily refers beyond itself to a non-neutral value. "Effective teaching" requires that one have a particular concept of "teaching" in mind, such that "effective" comes to have value derivatively. After all, as Hlebowitsh is fond of pointing out, one "can be an effective axe-murderer," in which instance effectiveness is by no means a positive quality. Effectiveness rather derives its qualitative value from the concept that it modifies, and thus the term "effective teaching" says nothing in itself about the *quality of teaching* under discussion without some shared sense of *quality teaching*, in which the quality is internal to the concept itself. What Duncan appeals to in this case would therefore *appear* to be the ordinary concept of good teaching, precisely because the notion of "teaching" that would render "effective teaching" a positive term is necessarily an *implicitly shared* and thus non-technical, non-explicit notion.

Instead, however, Duncan's particular references to an "effective" or "ineffective" teacher has to do explicitly with a given teacher's ability to *cause* an *effect*, namely student learning. Immediately after the quotation raising the example of retaining a teacher who "gets a chance to improve and doesn't," Duncan segues in the direction of a discussion of accountability in public education: "I told the charter schools they need to police themselves or their progress will be stalled. I told the school boards that if they can't improve student achievement, they have a moral obligation to consider mayoral control" (Duncan, 2009a).

The reference to "improving student achievement" is particularly telling. As he moves in the next sentence from charter schools and school boards to the teaching profession itself – "And I'm telling you as well that, when inflexible seniority and rigid tenure rules that we designed put adults ahead of children, then we are not only putting kids at risk, we're also putting the entire education system at risk" (Duncan, 2009a) – it becomes apparent that "effective teaching" does not reach for the ordinary or everyday sense of *teaching* for its value-ground, but rather toward a technical sense of "student learning." The immediately subsequent topic of this speech is devoted to the need for student achievement data in teacher practice and evaluation; I have already quoted the relevant paragraphs several times and do not need to reproduce them again at this point.

Suffice it to say that the "effective teaching" referenced in Duncan's speech dives below or through the ordinary sense of "teaching" to its technical definition, in which teaching appears *merely* as the *cause* of "student learning," which itself warrants scare quotes in that Duncan alludes to student learning merely in terms of "improv[ed] student achievement," visible, as his segue to a discussion of data implies, in the growth of math and reading scores. Surely student learning results from good teaching; but one must reiterate that "good teaching" is not exhausted in causing the kind of "student learning" visible in terms of "improving student achievement" to which Duncan alludes here. Nowhere in this concept appears the lighting of curiosity or the thirst for knowledge or the taking of after-hours phone calls. These ordinary aspects of great teaching simply fall away from the concept of "effective teaching" that Duncan plies here.

In order to accentuate or establish the way Duncan's discussion conflates fundamentally incompatible concepts of teacher quality, relying upon both an ordinary and a

technical sense at once, the seemingly common-sense notion that districts or schools ought to

be able to expunge ineffective teachers after a certain temporal point will perfectly suffice.

Effectiveness, as we have seen, is revealed in terms of a teacher's measurable effect (through

value-added methodologies) on "student learning," where student learning is defined as

"improving student achievement." In other words, teacher effectiveness is defined as a

function of lifting student achievement scores on math and reading assessment, which

accesses the technical sense of good teaching. At the same time, the justification for

expunging chronically ineffective teachers has to do with the assumption that the *quality* of a

teacher is, after a certain point in the developmental process, fairly stable and consistent,

such that the *chronic* nature of a teacher's ineffectiveness reveals the teacher's "poor

quality," which is thus constructed as irremediable. A school or district expunges a teacher

not only because a teacher *has not been* effective to date, but also because such

ineffectiveness is considered therefore to be a *trait*, something durable, and this aspect or

assumption draws upon the ordinary or everyday sense of good teaching.

However, the results of a number of studies, as summarized in an Economic Policy

Institute briefing paper, reveal these two assumptions to be fundamentally in contradiction

with one another. The passage deserves quotation in full:

> VAM [value-added methodology] estimates have proven to be unstable across
> statistical models, years, and classes that teachers teach. One study found that across
> five large urban districts, among teachers who were ranked in the top 20% of
> effectiveness in the first year, fewer than a third were in that top group the next year,
> and another third moved all the way down to the bottom 40%. Another found that
> teachers' effectiveness ratings in one year could only predict from 4% to 16% of the
> variation in such ratings in the following year. Thus, a teacher who appears to be very
> ineffective in one year might have a dramatically different result the following year.
> The same dramatic fluctuations were found for teachers ranked at the bottom in the
> first year of analysis. This runs counter to most people's notions that the true quality
> of a teacher is likely to change very little over time and raises questions about

whether what is measured is largely a "teacher effect" or the effect of a wide variety of other factors (Baker et al., 2010).

The conceptual contradiction is revealed in the statement that the above evidence of instability in measures of teacher effectiveness "runs counter to most people's notions that the true quality of a teacher is likely to change very little over time." Faced with these results, one must either conclude that value-added measures of teacher effectiveness – teacher quality according to the technical concept – reveal something *accurate* but *fundamentally unstable,* such that terminating or rewarding a teacher (encouraging the teacher either to stay or to go) on such a basis makes no sense; or else that, as the above researchers conclude, and in fact as Duncan himself reiterates again and again, the technical concept of teacher quality cannot proxy satisfactorily for the ordinary sense thereof. A further possibility, of course, would be that this technical concept is accurate and that variation across the years will be distributed around a "true" quality value. But as a minimally-adequate sample size of years (n=30) to draw conclusions about a teacher's value would span the majority of a teacher's career, it is difficult to see how these measurements improve upon our ordinary experiences with teacher quality.

It is simply less likely that "most people's notions that the true quality of a teacher is likely to change very little over time" are misguided than that value-added methodologies do not adequately speak to "the true teacher quality" that we all, including Duncan, aim to speak of. The problems with relying on student achievement data for access to teacher quality appears in both these researchers' work and also in Duncan's speeches, which indicates a certain inchoate wariness around the possibility of conceptual conflation or corruption, at least in terms of the ability to distinguish among teachers on the basis of quality. However,

in Duncan's rhetoric, at least, this wariness fails to prevent the very conflation that undermines his well-intentioned project: to reward the good teachers, support the struggling, and replace the poor.

Not to be forgotten, though, remains the fact that Duncan's rhetoric has also highlighted the ordinary aspects of teacher quality that have the characteristic of being immediately perceptible or apparent, such that the potential for apprehending good teaching in its ordinary sense remains a genuine possibility. The role of classroom observations, on-the-ground evaluation measures, provides the structural possibility of augmenting the inadequate sense of teacher quality revealed in student achievement growth alone. We have already seen, in academic studies, the destructive tendency to understand the relative success of classroom observations in direct proportion to their ability to confirm the findings of achievement-data analysis. Classroom observations and other local means of witnessing great teaching in its immediately-accessible aspect retain, for all that, their utility as potentially fruitful additions to the otherwise insufficient reduction of teacher quality to student achievement measures.

The conditions under which such observations could play this role in ascertaining the ordinary concept of teacher quality have already been enumerated, at least in terms of minimal necessity. The primary requirement involved remains the relative ability to extend beyond the view of teacher quality afforded by achievement data. That academic studies tend to evaluate classroom observations with respect to value-added judgments of teacher quality does not augur for the likelihood that state policies enacted with an eye to winning Race to the Top funding will give sufficient attention to this distinction, but an exploration of the policies of such states certainly remains warranted.

<u>1.7 States' Policies Regarding Data and Classroom</u>
<u>Observations</u>

The National Council on Teacher Quality (NCTQ), a "non-partisan research and policy organization," which focuses on the project of ensuring that "every child has an effective teacher," provides an ideal view into the steps taken by individual states to address Duncan's reforms in the years between 2009 and 2011. While legitimately non-partisan as regards a Democrat-Republican axis, the NCTQ represents full-throated support for Duncan's reforms. Its advisory board includes a veritable who's-who of proponents that Duncan himself, and the popular media, cite relentlessly and hold up as models of the reform movement. The advisory board includes, for example, once-touted and now-former heads of large school districts who became famous for taking major steps toward precisely the kinds of reforms that Race to the Top encourages: Michelle Rhee and Joel Klein, the former appearing alongside Duncan at a very recent panel on education reform (Winerip, 2012) and the latter the subject of a *New Yorker* expose into the challenges of tackling teacher tenure rules (Brill, 2009). The advisory board also includes one prominent curriculum theorist in E. D. Hirsch, Jr., most famous for his advocacy of a core-knowledge curriculum and his authorship of the *Dictionary of Cultural Literacy*. It likewise includes the founders of several prominent charter school organizations and alternative routes to the teaching profession, among them KIPP schools and Teach For America, examples Duncan regularly holds up as innovations deserving of replication; and experts in measurement and, once more, economics (NCTQ, 2010).

The NCTQ, then, not only provides factual insight into the particular policy developments resulting from the implementation of Race to the Top, but also rhetorical

support for such policies as well. The particular outlook on education reform that NCTQ reveals mirrors Duncan's own approach right down to the studies that each cites, the examples of "dysfunction" that crop up in the discussion, and so on. For this reason, NCTQ's view on states' reform "progress" provides a good proxy for Duncan's view, as well.

I turn to an examination of states' policies for the specific purpose, once more, of tracing the role that achievement data and classroom observations, among other measures, play in teacher evaluations with respect both to each other and to the conceptual terrain of teacher quality in general. The task of teacher evaluations generally, as Duncan's rhetoric has constructed it, is to identify good teachers and distinguish them from poor ones. The concept of "good teaching" to which Duncan points anecdotally is an ordinary, everyday one; the concept of "good teaching" accessible by means of recourse to achievement data is what I have been calling a technical sense. Whether or not existing and emerging policies exhibit a concern for the relative ability of the technical sense to shed light on or speak to the ordinary sense comprises the subject of my investigation at this point. The technical sense, as we have seen, has value insofar as it proxies for the ordinary; the danger that continues to stalk Duncan's rhetoric lies in the proxy's usurpation of the ordinary sense, its shift from a partial indicator of what we really want to the thing itself.

The NCTQ report, which summarizes the ongoing state of legislative reform across the states in accordance with the broad goals of Race to the Top, certainly echoes the primary concerns that Duncan has laid out, notably in (1) the problem that calls for a correction, (2) the means of correcting that problem, and (3) the ambivalent or self-contradictory language in the elucidation of how the means of correction will be applied to the current problem,

which speaks in particular to the insufficiently interrogated conceptual problem inherent in

the laser-like focus on the technical concept of teaching.

The NCTQ report opens with a statement of the problem, which reiterates Duncan's

position in the four speeches during the summer of 2009, albeit in much more succinct

fashion:

> The move to rethink how to evaluate a teacher's performance and explicitly
> tie assessments of teacher performance to student achievement marks an important
> shift in thinking about teacher quality. The demand for "highly qualified" teachers is
> slowly but surely being replaced by a call for highly effective teachers
> The change is significant because policymaking around improving teacher
> quality to date has focused almost exclusively on a teacher's qualifications – teacher
> credentials, majors, degrees and licensing. Those criteria would be all well and good
> if they were associated with positive gains in student learning. Unfortunately, by and
> large, they are not. (NCTQ, 2011).

In contrasting "highly effective" to "highly qualified" teachers, the report echoes

Duncan's skepticism of the value of teacher credentials such as advanced degrees and

certification. The skepticism originates, as we have seen, in the fact that such variables have

not been shown to have a significant effect on student achievement growth, which remains

the criterion of "effective" teaching. In support of its claim that qualifications bear little or

no association to "positive gains in *student learning*" (emphasis added), the report cites the

Kane, et al. study that I have already spoken to (Kane, et al., 2011a). The linking of "student

learning" to "achievement score gains" indicates once more that the concept of good teaching

(and student learning) at work in this report is by and large a technical one.

The failures of traditional evaluations – identical to those Duncan cites – function as

justification for the turn to including student achievement data in teacher evaluations: "Their

[The New Teacher Project's] study across a set of twelve districts in four states found that

less than one percent of teachers received unsatisfactory evaluation ratings" (NCTQ, 2011).

As in Duncan's example, the need for more robust systems of evaluation becomes visible in the fact that "99 percent of teachers are all rated the same" (Duncan, 2009c). I have already raised an issue with regard to the presumed shape of this problem: it remains unclear whether the *fact* of nondiscrimination according to quality in teacher education indicates a failure to *know* or a failure to *tell*, but once more I wish to defer particularly the second possibility until a later Chapter. In any case, the use of achievement data in the solving of this problem – an objective means of identifying and reporting – promises to solve both at once. However, as we have already seen, *what* is knowable according to the value-added metrics alone does not conceptually match with the teaching we wish to know something *about*.

In discussing the teacher evaluation practices and policies across several states, the report singles out a few for special praise, linking them explicitly to Race to the Top policies:

> Several RTT winners are clearly at the forefront of efforts to develop and implement performance-based teacher evaluations. Delaware, Florida, Rhode Island, Tennessee, and D.C. Public Schools, for example, all require annual evaluations of all teachers and require that annual evaluations include objective evidence of student learning – not as an option, but as the *preponderant criterion* for assessing teacher effectiveness (NCTQ, 2011).

The report characterizes these teacher evaluations as "performance-based," where teacher performance is visible in terms of test scores. "Evidence of student learning" functions as an (emphasized-in-the-original) "preponderant criterion." The centrality of growth on achievement scores to the concept of "student learning" that represents the effect of "good teaching" cannot possibly be overstated here. Tennessee, New York, and D.C. have all adopted weighted formulas for rating teachers, which establish aprioristically how much of a teacher's quality is reflected in achievement data. Tennessee's system places particularly significant weight on these measures: "A total of 50 percent of a teacher's annual

evaluation must be based on student achievement data, of which 35 percent must rely on student growth data from the Tennessee Value-Added Assessment System (TVAAS)" (NCTQ, 2011).

Teacher quality, on this view, is once more taken as a causal factor whose causal value becomes visible in the teacher's effect on "student learning," represented by achievement measures. Because of the narrow definitions of teaching and learning here depicted, such a view represents merely the technical view of good teaching, when what is needful is access to the ordinary, everyday view. The NCTQ report itself seems to recognize the fallibility of achievement data to a certain extent, and as in Duncan's case, holds up observations as a potential corrective.

First of all, the report includes a subsection called, "States Shouldn't Lose Sight of Classroom Observations." Having already signaled and discussed the necessary function of such observations with regard to bridging the gap between a technical sense of good teaching and the ordinary one that evaluations in fact pursue, an examination of the role observations play in the view of reformers and state policies will be in order.

The report binds up the importance of stringent and rigorous classroom observation with the failure of the older evaluation systems: "The criticism of many current evaluation systems is not just their failure to take student learning into account, but their failure to include high-quality classroom observations" (NCTQ, 2011). High-quality classroom observations once again appear as an *additional* element to be attended to, over and above "student learning," however that is defined. In other words, observations are understood to reveal something *other* than what analysis of "student learning" will reveal. The *distinction*

between student achievement measures and classroom observations emerges even more

strongly in the same paragraph:

> While there is a great deal of attention focused on linking value-added and student growth results to teacher evaluation, it is equally important to gather evidence observing behavior – what teachers do and what students are learning in the classroom – during classroom observation (NCTQ, 2011).

Student growth measures are explicitly taken to reveal something different from what

classroom observations can reveal. What "evidence observing behavior" ought one to gather

in the classroom setting? What counts as "high-quality" classroom observation?

The language of *behavior* becomes particularly important at this point. The above

passage already notes that the aim of such observation is to "gather evidence observing

behavior." A further clarification of this suggestion reveals other commitments in this area:

> A strong observation rubric should focus almost exclusively on teacher practices and student behaviors that can be observed in the classroom. While other criteria are not without merit, they may call for too much subjectivity and guesswork on the part of the evaluator (NCTQ, 2011).

What arises for evaluation is described here in terms of "practices" and "behaviors" –

and those of the sort that do not "call for too much subjectivity and guesswork." In other

words, evaluators will not be looking at any sort of holistic "teaching" or "learning"; instead,

they will be using a "rubric" (interesting that this is simply assumed) that will identify

specific and isolated behaviors for which evaluators to seek and check off. Evaluators look

for teachers and students to demonstrate – in a way that does not require too much subjective

interpretation on the evaluator's part – *certain* behaviors or practices that the evaluator can

witness; the demonstration of these behaviors will reveal an aspect of "teacher quality" that

has been explicitly conceived as differing from achievement data. *That* good teaching can

appear in the form of "behaviors" implies that the *presence* of these behaviors also *makes*

present good teaching itself. Teaching, then, really is "a series of bite-sized moves." But

this is not an ordinary sense of good teaching, in which good teaching is durable and stable

and self-consistent and holistic; this is the atomized view. If states implement classroom

observations by means of developing "rubrics" that aprioristically posit the objective

"behaviors" for which evaluators will non-subjectively look, some connection between these

isolated behaviors and the sense of "good teaching" in which they are implicated must

obviously exist. What makes a certain behavior desirable in a teacher? What qualifies it as

something for an observer to seek out?

By this point, the root *source* of these behaviors that evaluators are tasked with

recognizing will be wholly unsurprising: the value of a given isolated behavior is determined

by its correlation with lifting achievement scores. Says the report:

> New research is encouraging on this front. Well-designed and executed
> classroom observations can be effective at identifying the effectiveness of teachers,
> particularly teachers at the top and bottom ends of the distribution. Recent research
> also finds that good evaluations impact teacher effectiveness– that is, evaluations
> don't have to be just summative report cards. If done well, they can indeed be
> formative tools that drive teacher improvement (NCTQ, 2011).

In substantiating the claim that "well-designed and executed classroom observations

can be effective at identifying the effectiveness of teachers," as well as the claim that such

"evaluations impact teacher effectiveness," the report cites the Kane study I mentioned

earlier, which asks about the visibility of the isolated behaviors associated with test score

growth (Kane, Taylor, Tyler, & Wooten, 2011b). While the NCTQ report explicitly

articulates the need to "gather evidence observing behavior" with the aim of *supplementing*

achievement data in the evaluation of teachers, the very behaviors for which observations

would look are defined *according to* the achievement growth in question. Classroom

observations, therefore, *cannot* function as a conceptual bridge between a technical and an ordinary concept, since the observations themselves are restricted to the hunt for indicators of the technical sense of teacher quality. While cloaking themselves in the appearance of adding something to achievement data metrics, classroom observations in states winning Race to the Top funding, such as Tennessee (Anderson, 2012), remain utterly sealed within the realm of achievement data, the technical view. In a bait-and-switch maneuver, classroom observations function to further strengthen the self-sufficiency of the technical sense of good teaching.

Such self-sufficiency is also revealed in the burgeoning primacy of "effective" teaching in the rhetoric of the discussion here, a phenomenon that emerges even in Duncan's rhetoric from 2009. Effective teaching, as I have already discussed, refers only to a technical sense of good teaching, one in which teaching appears merely in terms of its (presumed) causal relation to its salient effect, "student learning," as revealed in test scores. The ordinary sense of good teaching – the kind of teacher quality *immediately accessible* in a way that Duncan recognizes in his various speeches – has simply disappeared from the discussion. The completeness of this disappearance becomes even more clear in considering a Brookings Institution report (co-authored by Staiger) that the NCTQ report cites on the relation of value-added metrics to other means of evaluating teachers, and also in the NCTQ report's discourse on the precision of measures.

The Brookings Institution report (Glazerman et al., 2011) attempts to answer criticisms of value-added measures as they are applied to teacher evaluations. It declares, among other statements, that

We have previously issued a report that describes some of the imperfections in value-added measures while documenting that: a) they provide one of the best presently available signals of the future ability of teachers to raise student test scores; b) the technical issues surrounding the use of value-added measures arise in one form or another with respect to any evaluation of complex human behavior; and c) value-added measures are on par with performance measures in other fields in terms of their predictive validity (Glazerman, et al., 2011).

Point (a) in the above paragraph simply clamors for attention: the value of this metric lies in the fact that it is one of the "best presently available signals of the future ability of teachers to raise student test scores." So, no other method of predicting a teacher's ability to lift test scores can compare to the teacher's present ability to lift test scores. Two problems arise, one minor and one significant. In the first (minor) place, the fact that there is no better predictor available is both totally obvious and wholly unresponsive to the kinds of criticisms about the validity of value-added measures that I have quoted earlier: simply because there is none *better* in this regard does not render such existing methods of prediction conceptually *valid* for use in teacher evaluations in the first place. In the second and substantially more important place, the *function* of measuring teacher effects under the technical sense of teacher quality has heretofore had ostensibly to do with providing some sort of insight into the *ordinary* sense (of which the technical sense is an admitted reduction). But this locution, wherein the defense of the appropriateness of value-added metrics appears in terms of their predictive power *of themselves*, eliminates the ordinary sense from the discussion altogether. We know that value-added metrics are useful to teacher evaluations because they remain the best predictor of "the future ability of teachers to raise student test scores." The technical view has completely occluded any other concept of teacher quality that might arise.

The use of the measure itself as the measuring stick against which the measure's validity is judged cuts the ordinary sense of teaching out of the picture entirely. Duncan has

already said that the "complex, nuanced work of teaching" defies reduction to a test score;

the measure of value-added methodologies' claim rests in its ability to speak to *that* complex

and nuanced sense of teaching, not in its ability to predict teachers' success according to its

own measure. When teacher quality is only taken to consist in the raising of student test

scores, it is entirely unsurprising that a teacher's established ability to raise test scores is

considered a good measure of his or her teacher quality. But it is precisely at this juncture

that the notion of *construct validity* would like to raise an objection.

I wish to pull out and bookmark point (b) from the above report as well, that the

"technical issues" in value-added methodologies appear in "any evaluation of complex

human behavior." I agree that the sorts of problematic issues to which these scholars refer

also crop up in gauging or evaluating any other skillful endeavor, but I hold that these are

hardly "technical issues," amenable to technical solutions. They are conceptual issues at

bottom, resolvable or dissoluble only by conceptual means. But I leave that for Chapter 2.

The Brookings Institution report, and the language in which the problem is conceived,

indicates that the scholars at work on developing teacher evaluations have very little

consciousness for the conceptual problem they face. The technical definition of teacher

quality to which their studies speak is emphatically and explicitly not the sense of teacher

quality to which Duncan's anecdotes and audience exhortations appeal. The gap between the

two remains unaccounted-for, and worse: the very existence of the conceptual chasm goes

unacknowledged in such studies. In discussing the problem of reliability, the Brookings

Institution demonstrates this fact perfectly:

> The degree to which a performance measure in one period predicts
> performance in the future will depend on both the degree to which the *measure* is

related to true performance and the extent to which *true performance* is stable from one period to the next (Glazerman, et al., 2011).

The authors, in fairness, do not consider *validity* at this point, and reliability is not at all the same thing. But they do note that the relation of a measure to "true performance" shares responsibility with the stability of "true performance." "True performance," here, refers merely to the ability to raise test scores, because *that is what these measures claim to predict*. But in the ensuing discussion of "true performance's" stability, the report conflates "effectiveness" with the *ordinary sense* of good teaching, and without any awareness that it does so:

> This assumption [that true performance is relatively stable] is buttressed, in the case of value-added measures, by the fact that value-added measures from one period predict student achievement in future periods. It is also buttressed by anecdotal evidence that some teachers are simply more effective than other teachers and, as a result, parents work to get their children into these teachers' classrooms (Glazerman, et al., 2011).

The "anecdotal evidence" cited here, in which parents desire certain teachers for their children, is certainly a common feature of school life. But it is an unwarranted assumption that parents are using any sort of "effectiveness" criterion, in the technical sense that these scholars measure, to determine which teachers are the best. After all, value-added rankings are only publicly available in a few places nationwide, and this sort of parental lobbying has been going on since time immemorial. The researchers in this case seem to equate the "true performance" on measures of teacher *effectiveness* with "true performance" in the fully-formed sense of wholesale teacher quality. "Effectiveness" has been partitioned off from the ordinary sense of good teaching in Duncan's rhetoric and in the NCTQ report as well, but that partition appears infinitely permeable in the discourse. The many conflations lead one to suspect that those using the "effectiveness" terminology allow themselves to forget the

circumscribed role of achievement data that they, themselves, have called for. In their minds, in other words, it would appear that an "effective teacher" and a "great teacher" are the same thing, despite their explicit avowals to the contrary.

This type of conflation runs throughout the Brookings Institution piece. "Teacher performance" has a technical meaning here, but it sounds so *ordinary:* "If we are to judge the quality of teacher evaluation systems relative to each other, we have to have a common measure or a set of common measures across those systems that are sensitive to true differences in teacher performance" (Glazerman, et al., 2011). By "teacher performance," here, the authors mean "sensitive to true differences in teacher performance *in terms of raising test scores*." But they do not *say* that, and to the educated layperson, at whom the report is aimed, "teacher performance" means the performance of the teacher, period, holistically. The conflation of the ability to raise test scores with the full-blown concept of good teaching, toward which teacher evaluation systems ostensibly aim, serves to indicate the extent of the usurpation of the ordinary sense of teacher quality that technical notions of "effective teaching" have perpetrated.

The disappearance of the ordinary concept of good teaching that we have seen attend the operationalization of "teacher quality" in psychometric discussions and policy papers becomes glaringly apparent in the NCTQ's justification for using admittedly imperfect value-added methodologies in determination of teacher quality. In 2009, Duncan says, as I have quoted much earlier, "I absolutely respect the concerns of teachers that test scores should never be used solely to determine salaries" and "I also appreciate that growth models as they exist today are far less than perfect." He dismisses the suggestion, however, that "since standardized tests are not perfect, eliminate testing until they are," calling such a notion

"absolutely ridiculous" (Duncan, 2009b). In this speech, the imperfection of growth models appears to speak to problems in accurately measuring student learning; that "standardized tests are not perfect" refers to the tests' imperfect ability to measure something about *students*. Because it is an inadequate measure of student learning, Duncan implies, "test scores should never be used solely to determine [teacher] salaries."

In the 2011 NCTQ report, however, the authors reiterate the rhetorical structure of Duncan's argument almost point by point, but in a profound shift, the focus on the imperfections of "growth models" now speak directly to "teacher effectiveness" rather than to student learning. In other words, the ability of "standardized tests" to ascertain precisely what is important in "student learning" is suddenly and entirely taken for granted. At issue in 2011 is merely the *predictive power* of such measures relative to "teacher quality," wherein it is merely assumed that student achievement in reading and math *alone* determine "teacher quality." This assumption – by now so basic that it apparently requires no explicit articulation – starkly reinforces the complete annihilation of any trace of the ordinary concept of good teaching in the debate.

In the NCTQ report, the authors echo nearly verbatim Duncan's insistence that imperfect measures ought not to prevent us from using student achievement data at all, they repeat Duncan's citation pertaining to the fact that "99 percent of teachers are all rated the same," and then they say,

> Student growth and value-added methodologies are still emerging. However, examining student achievement as a metric for assessing teacher effectiveness, even if measurement is imperfect, represents a big step forward. Indeed, we set the whole enterprise up for failure if we attach unrealistic expectations to the exact precision of every measure – and doom ourselves to the alternative of doing very little to measure and examine teacher performance. At the same time, the reality of measurement and

limits to teacher control over student outcomes do argue for measured caution in developing teacher effectiveness policies (NCTQ, 2011).

"Measured caution" has also emerged in the discourse before, but the evaluation systems in fact under development evince none of the appropriate caution. The authors of this report argue, in a sense reminiscent of Duncan, that the *only* alternative to basing teacher evaluations on students' standardized tests is to forego evaluation entirely. This form of argumentation recalls once more the sense in which teacher evaluation combines problems of *knowing* and *reporting*; it is unclear whether, in the absence of student achievement data, teacher quality would be unknowable or simply unverifiable, nontransferable.

When Duncan sounded a note of caution in his 2009 rhetoric about measures' imperfection, it occurred in a context of questioning whether scores in math and reading were perfect representations of *student learning*; here, that achievement data authentically capture student learning is merely assumed. Now the question appears to be whether teacher *effectiveness*, itself the technical definition, can be perfectly hitched onto this narrowed sense of "student learning." If so, it is implied, "teacher effectiveness policies" would require none of the "measured caution" the report calls for, because policy would then be operating from a position of certainty in terms of meaningful teacher quality, despite any objections from the oft-acknowledged fallibility of one-shot assessments of student achievements. The validity of the assessments themselves in expressing *everything* relevant to "teacher quality" is beyond question. The "teacher quality" under discussion, then, is the technical view. The ordinary sense is no longer even on the table. The calls for caution in treating the metrics as indicators of the concepts we would use them to measure do not seem to translate into policy or practice.

Delaware's teacher evaluation rules perhaps demonstrate this fact better than any other argument. The NCTQ report summarizes beautifully:

Delaware's DPAS II system formula is organized around five key evaluation components or topics: 1) planning and preparation, 2) classroom environment, 3) instruction, 4) professional responsibilities and 5) student improvement. Teachers are rated highly effective, effective, needs improvement or ineffective in each individual area. A summative rating depends on the number of effective or not effective ratings teachers receive in each of the individual component areas. All other components aside, if a teacher does not meet or exceed student growth requirements in Delaware, the teacher cannot be rated any higher than needs improvement overall, regardless of ratings in the other four components.

Despite the fact that student achievement data reflects only math and reading performance (and only that in a limited sense itself), the test score is king. It is curious, also, that the state's policies plan for the contingency in which a teacher could fail to meet growth goals while earning high marks in every other category, since the teacher behaviors according to which a teacher would be assessed in "classroom environment" and "instruction" are directly and intentionally pegged to "student improvement." Even, it is implied, in cases where student growth is subpar *despite* a teacher's *demonstrating the behaviors correlated with student achievement score growth*, the test results would win: that teacher "needs improvement."

It's like Derrida once said: there is nothing outside the test.[10]

1.8 Summary and Segue:

The methodological assumptions undergirding this project bear mentioning once more as I move to summarize the results of the investigation thus far and seek to extend it in the direction of the upcoming sections. The most basic assumption I make pertains to the relevance of Duncan's speeches on his proposed reforms, the studies and lines of research

---

[10] That might not be an *exact* quotation.

that he cites, and the translations of this research and rhetoric into policy. At the very

bottom, I assume that such speeches, studies, and policies pertaining to teacher quality reveal

the conceptual shape of what we take teaching *to be*.

In this respect, my investigation has uncovered two distinct and eminently

distinguishable concepts of "teacher quality" at work in policy speeches, research, and

policies themselves. On the one hand lies what I have been calling the "ordinary" conception

of good teaching, so named in accordance with its particular content and its contexts of

evocation. Content-wise, the "ordinary" concept of good teaching is associated with

references to such educative actions and results as a "thirst for knowledge," the "inspiring" of

students, "everyday acts of kindness and love," the "counsel[ing] of teens," the "tak[ing of]

phone calls at night," the "lighting of a lifelong curiosity," and the ability to "literally change

the course of a student's life." Prevalent features of the ordinary concept include the types of

long-lasting and real-world impacts of the sort Duncan mentions above, and to which his

"college and career ready" criterion is intended to speak; a certain stability over time, such

that teachers can themselves be good or bad; a persisting and influential relationship with

students, as in the case of his mother's guidance "over the years"; and a sort of immediate

accessibility or visibility of the kind Duncan avers he can spot "the minute I enter a school."

The ordinary sense of good teaching, in short, is the one we have in mind when Duncan

encourages us to think of a "favorite teacher" that we remember "forever."

The ordinary concept of good teaching tends to emerge in the context of policy

speeches, and is, for the most part, notably absent from academic studies and state policies.

The absence from the latter two contexts is fairly unsurprising, since what makes the

ordinary concept ordinary in the first place has to do with its independence from rigid

definitions of the sort that academic studies and state legislation tend to require. The ordinary concept, however, is not therefore necessarily insufficient to the purposes of studies and policies; rather, as evidenced by Duncan's effortless conjuring of the ordinary concept through appeals to the teachers that "all of us remember" (Duncan, 2009a) or that "so many of us remember" (Duncan, 2009d), as well as his platitudinous statements about the "complex, nuanced work of teaching" (Duncan, 2009a) the ordinary concept simply exceeds or goes beyond the possibility of definition , and is explicitly the one that occasions the urgency underwriting the Race to the Top Program.

On the other hand, my investigation has also unearthed a concept of "teacher quality" that positively requires and rests upon explicit definition, a concept that I have termed, for this very reason, the "technical concept" of teacher quality. The technical concept of good teaching, as I indicated above, differs from the ordinary primarily insofar as it rests upon explicit characteristics or features laid out aprioristically. The technical concept of good teaching is in every instance defined according to its relation to "student learning," which itself is a technical term, distinct from the sorts of "thirst for knowledge" or "light a lifelong curiosity" criteria that serve to characterize the perceived effect of the ordinary concept of good teaching on students. "Student learning" in its own technical sense is revealed in students' score growth on math and reading assessments. Good teaching in its technical sense is then revealed according to its effectiveness in lifting test scores. For this reason, the technical concept of good teaching also goes by the name of "effective teaching" in the literature.

Due to its inherently close – indeed, tautological – relation to student achievement data, the technical sense of good teaching evinces certain stark differences with the ordinary

concept. These differences include what one might call a mediate nature of teacher quality, where it requires theoretical or algorithmic means of becoming visible; the associations with highly specific teacher behaviors, as evidenced in observation protocols under development at the state level and in academic attempts at linking classroom observations to indicators of "effective teaching"; and also, importantly, the present-tense availability of actionable evidence of teacher quality. This last contrast refers to the notable fact that citations of ordinary-concept "great teachers" of the type explored in this study tend to refer backward from present adulthood, implying that while an ordinary sense of great teaching may be immediately accessible or knowable, it is somehow not relatable or transferable in plausible narrative form until the students involved have become adults (Duncan, 2009a; Kristof, 2012a).

Despite this delineation of the ordinary and technical concepts of teacher quality in terms of their differences from one another, perhaps the most important result to emerge from the investigation so far involves the degree of conceptual interpenetration and usurpation that have emerged in the speeches, studies, and policies explored above. In particular, I have formatively outlined the ways in which the technical concept borrows liberally from – to the point of entirely supplanting, in some cases – the ordinary concept.

In the broadest and most obvious sense, where the self-evident need for greater attention to teacher quality derives its urgency and importance from the ordinary concept of teacher quality, as Duncan's speeches repeatedly demonstrate, the methods involved in ascertaining teacher quality and addressing matters pertaining thereto all refer to the technical concept at bottom. A need with its basis in the ordinary concept is supposed to be adequately answered by means of the technical concept. But other borrowings proliferate as well, as we

have seen: the enduring nature of teacher quality, for instance, is a feature of the ordinary concept; such stability of teacher quality, as critics of value-added methodologies have pointed out (Baker, et al., 2010), is not at all in evidence in analyses of teaching according to the technical sense. Yet when policies and speeches seek to identify and reward good teachers, whose goodness is a stable and enduring trait, they propose to do so by technical means. This conflation emerges once again in the dissonance between Duncan's deploring of the "factory model" of education while discussing innovations in terms of "scaling up," in which quality is supposed to be both intrinsic to a teacher (ordinary sense) and yet also to consist in specific and isolated behaviors (technical sense). Along similar lines, classroom observations are repeatedly invoked in the interest of accessing the ordinary concept of good teaching to which achievement data cannot speak, but in academic studies and state policies alike, these observations are in fact addressed to the technical sense of good teaching, even to the absolute exclusion of the ordinary.

The conflation of the technical and ordinary senses of teacher quality demonstrate strongly that the ordinary sense of good teaching remains the goal toward which the many attempts at knowing represented in the speeches, studies, and policies strive. The occasions that I have highlighted, in which not only Duncan, but also scholars and researchers, simply forget the distinction between the two senses of teacher quality at work in the discourse evince a certain tendency in this regard, of which forgetting the distinction is but the first part.

As we have seen, since the ordinary concept of good teaching serves as the starting point of the discussion, being both the problem requiring address and the rhetorical indicator of teachers' importance, it represents as well the end goal of the reform policies outlined

here. The proposed reforms aim to address the ordinary concept of good teaching.

Throughout Duncan's 2009 speeches, he carefully circumscribes the role of the technical

sense of good teaching, reiterating the irreducibility of teaching to "a multiple choice test," as

he touts pilot programs that involved multiple measures of evaluation, including, importantly,

classroom observations. The academic studies to which he refers in those speeches, and to

which state policy is later geared, however, use the technical sense of good teaching in order

to answer their questions about the effectiveness of classroom observations. In the

Brookings Institution report, the scholars discuss the technical sense as though it were itself

the ordinary sense. Even in reiterating Duncan's cautious tone, the 2011 NCTQ report has

clearly adopted the technical sense of good teaching not as a means to an end, but as the end

itself, a conclusion painfully evident also in Delaware's teacher evaluation policy. The

technical sense of teacher quality, at length, gathers to itself the inherent stability and

justificatory urgency of the ordinary concept, and then simply steps into its place as the alpha

and the omega of the evaluation of teacher quality.

Nevertheless, as we have seen, for all the conceptual interpenetration and usurpation,

the ordinary sense of teacher quality really is distinct from the technical sense, and really has

functioned as the teleological goal toward which reform initiatives look(ed) and around

which the American public has rallied (to the extent that it has). The ordinary sense of great

teaching certainly has an effect on students – life-changing impacts are repeatedly attributed

to it – and certainly part of that effect would show up in terms of student learning, even the

"student learning" represented in the narrow terms of year-to-year math and reading

achievement score growth. But at the outset of Race to the Top's proposals, it was

rhetorically understood that the technical sense of teacher quality required actual

supplementation in order to bridge the gap between its own sense and the ordinary sense

which represented the actual goal, the actual source of urgency. Classroom observations, as

we have likewise seen, were to bear most of this burden.

The most basic assumption underlying my investigation remains that the conceptual

shape of good teaching at work in the reforms touted by the Obama administration and others

will emerge in the rhetoric, academic studies, and policies pertaining to teacher quality. In

tracking the way in which the technical sense of good teaching has come to stand in for the

ordinary, a very specific (and technical) notion of what good teaching is has come into view,

and it is this "good teaching" to which evaluators are instructed to attend by means of

classroom observation rubrics. Simply put, this view of good teaching considers teaching to

be a series of isolated behaviors whose relative quality is determined according to established

correlation with achievement score growth. Such observations also attend to student

engagement behaviors, which are likewise correlated with score growth and demonstrate

whether a teacher is having an "appropriate impact" (NCTQ, 2011). Because score growth

remains the bottom-line determinant of what an observer ought to (is able to) observe, when

the originally-imagined role of classroom observation pertained specifically to a domain of

what teaching is that exceeded the technical sense, it is clear that the types of observations

currently being implemented across the country cannot fulfill that role. Instead, what "good

teaching" is taken to be is straightforwardly restricted to this behavioristic view.

But, in the spirit of that most basic assumption undergirding this investigation, it

remains the case that classroom observations were originally tasked with speaking to a wider

domain, over and above the technical sense. If classroom observations once seemed

appropriate to this role, and if they were searching for the ordinary sense of teacher quality, it

would stand to reason that a conception of teaching that restricts itself to the piecemeal demonstration of particular behaviors will not suffice. Another account of "good teaching" is needful – and perhaps it may also address the reasons underlying Glazerman, et al.'s assertion that "technical issues" always crop up in "evaluation of complex human behavior."

It is now time to ask the questions that will signal the transition to Chapters 2 and 3: if classroom observations were assigned a bridging role between the technical sense and the ordinary sense of good teaching, what aspects of good teaching do we imagine inherently and necessarily to escape evaluation in terms of achievement data, such that observations could function as a conceptual bridge in the first place? If it is the case that the ordinary sense of good teaching exceeds definition of the kind that the technical sense of good teaching requires, how might classroom observations come to seize upon this excessive dimension at all? The ordinary sense of good teaching has been described by Duncan and others in terms of a certain immediate availability; well, what do we take to be immediately available?

Further, we have seen that the ordinary sense of good teaching requires no explicit definitions of the sort that the technical sense needs, and the ordinary sense may in fact even resist attempts at definition; would an ordinary sense of teacher quality be meaningfully reportable on a yearly basis in the absence of such definitions? Can one do something well without first explicitly defining what "well" means? What are the relations among knowing, doing, and reporting?

CHAPTER 2: VERSIONS OF TEACHING WELL

2.1 Teaching as a Function of Learning

We have seen in Chapter 1 that the technical sense of good teaching derives from a highly technical sense of learning, namely one representable in the form of student achievement data. The features of this sense of learning that indicate its "technical" nature go beyond the simple fact that the domains of student skill that the relevant assessments measure are limited to reading and math, to the exclusion of history, art, geography, and the like. As other scholars, like Rothstein, have pointed out, these are important *subject-matter* exclusions. As such, they represent a certain kind of restriction of the domain of learning to which teaching as an endeavor addresses itself: a restriction of *what* ought to be taught and learned. But beyond a domain restriction in terms of content, the paramount importance of student achievement data in answering questions about learning and teaching (such that achievement data can and ought to be the "predominant criterion" for teacher evaluation) evinces a *conceptual* restriction of learning, and therefore of teaching, as well. In other words, the reliance upon student achievement data limits *what counts* as the learning of anything in the first place in ways that we have already seen, namely that the types of learning characterized by developing a student's "lifelong curiosity" and "thirst for knowledge" are either nowhere in evidence or assumed to simply go along with rising scores on reading and math assessments. The "learning" to which teaching is imagined to speak is restricted in terms of both its *content* and its *form.*

The conceptual restriction of learning leads directly to a certain restricted view of the concept of good teaching, which I have also called the technical sense thereof. In a speech at Columbia Teachers College in the autumn of 2009, Arne Duncan makes explicit the technical

view of good teaching. Touting the virtues of exemplary teacher-training programs, Duncan says, "And these programs have a shared vision of what constitutes good teaching and best practices—including a single-minded focus on improving student learning and using data to inform instruction" (Duncan, 2009d). By combining the notion of "using data to inform instruction" with these programs' "single-minded focus on improving student learning," Duncan leaves little doubt as to the type of learning he has in mind: the technical variety, limited to that which is revealed in achievement score data. But here, he makes an additional claim connected to teaching: these programs, characterized by the monomaniacal pursuit of "student improvement" by means of "using data to inform instruction," "share a vision *of what constitutes good teaching and best practices*." The vision of teaching for which Duncan advocates here is thus directly related to the technical sense of learning that I traced in Chapter 1.

Such an understanding of what constitutes good teaching is dangerous in two ways. First, as we have seen, the sort of learning connected to this vision of good teaching diverges from the kind of learning that serves to justify the need for the Race to the Top program in the first place, and it is by no means clear that the technical sense of good teaching will entail an ordinary sense of learning. Secondly, this technical sense of good teaching, as I will demonstrate, relies on deeply-flawed assumptions about the nature of skillful practical performance in general, and teaching in particular. The assumptions that undergird Duncan's view – as well as the view of many scholars in the academic field of teacher-education – bear a striking resemblance to assumptions made in early (and failed) attempts at Artificial Intelligence (AI). Put briefly, these views assume that knowledge of how to do something consists in the having of certain rules or maxims by means of which to navigate a world of

bare facts in order to achieve a particular end. The "consists in" in the previous sentence ought to be understood as implying both that *what happens* during good teaching *is* the application of transcendent and articulable- in-principle rules to a situation comprised of articulable-in-principle facts, and also, and therefore, that the facts and rules themselves *are* a transferable form of good teaching. According to this view, to study good teachers and to develop novices in the profession toward competence is to extract the rules and the salient facts from particular situations and then to give those to the neophyte, in one form or another.

I indicated above that I consider this view dangerous. The danger lies primarily in the fact that it is not obviously wrong, especially inasmuch as the "learning" connected to this view of teaching is itself distorted. The distorted sense of learning to which teaching is accountable in Race to the Top makes the inadequacies of this view of teaching particularly difficult to see. But additionally, even a more robust view of what constitutes good teaching allows for the place of conceiving skillful activity in terms of facts and rules. Thus, when Duncan lauds colleges of education that attend to a view of "good teaching and best practices" by focusing on what the data reveals, he is not incorrect, strictly speaking. New teachers, like new tennis players, must start off by consciously and deliberately following certain maxims or rules.

Whether or not the practice of excellent veteran teachers *also* consists in rule-following behaviors on the same model is what I wish to draw into question. If the "best practices" view of teacher quality only characterizes actual teacher practice up to a certain degree of competence, but not beyond that, then it follows that seeking to evaluate the quality of teachers by means of seeking out maxims and rules operational in their practice will not suffice. Similarly, holding up as a final goal the actualization of certain maxims and rules

and attempting to develop teachers with that goal in mind will run up against a certain limit well shy of the kind of excellence that motivates and justifies the Race to the Top policies in the first place.

To recall once more the Mrs. Grady test from the Introduction, that which made her example and her impact exemplary was specifically *not* amenable to representation in facts and rules; it was not codifiable or transportable, except in terms so broad as to provide no practical, situational direction ("put the needs of the student first," or some such). The Introduction made clear the ways in which Mrs. Grady's excellence would not show up in student achievement data; but neither would it show up in terms of "best practices," and only partly because "best practices" are defined in connection with achievement score growth. If we take Mrs. Grady's actions as a paradigm case of teacher excellence, then such excellence is not limited to that which best practices reveal. Her excellence was typified by *situational* responsiveness, by doing what one would *not* generally do (allow and enable a student to steal) so as to achieve a particular result for that particular child in those particular circumstances.

The discrepancy between the ordinary and the technical senses of good teaching thus entails another consequence: since we *want* good teaching in the ordinary sense but orient our accountability systems and teacher-training programs to good teaching in the technical sense, we foreclose on the possibility of any deliberate attempt to move from a technical sense of good teaching to an ordinary one – we train and measure *against*, as it were, situated responsiveness. Not only do teachers lack *incentive* to deviate one jot from the technical sense of good teaching, but new teachers, administrators and state agencies alike share a kind of systematic, official blindness to the very *existence* of excellence in the ordinary sense. If

Duncan can attribute remarkable skill and impact to his own mother, his policies include no mechanisms for recognizing or rewarding that excellence in the present tense. Her excellence, which is immediately visible to himself and to the audience of his retrospective narrative, would not *count* as excellent practice, officially.

In the subsections that follow, I wish to flesh out the way in which Duncan's technical sense of good teaching relates to student achievement data, which will amount to revisiting some of the salient points from Chapter 1 with an eye toward their consequences for a view of teaching practice. I will then explore the ways in which this view of "what constitutes good teaching" relates to the now-defunct project of symbolic artificial intelligence with a particular view toward highlighting the similarities in the assumptions about skillful behavior on which each project rests. With the assumptions laid out, I will offer a critique of those assumptions in the hope of making explicit their limitations in terms of accounting for the evaluation or production of skillful behavior. Lastly, under the aegis of this section, I would like to offer an alternative view of what constitutes skillful behavior, which ought to have particular resonance for teacher-training programs. In the third and final Chapter of this project as a whole, I will discuss the ramifications of this view for evaluation and accountability in the field of education generally.

<div align="center">

2.2 Teaching, Data, and Scaling: A Theory of Practice
At Work
</div>

I have already spent some time discussing Arne Duncan's vision for the use of data in guiding teacher decision-making, as well as driving reform in teacher-training programs, and I have made preliminary remarks on what the above reveals in terms of Duncan's ontology of teaching. In returning to those remarks now, I wish to make two arguments. In the first

place, Duncan's valorization of student achievement data as the final arbiter and definitional touchstone of "learning" tempts policy in a misguided, behavioristic direction. A behavioristic view would hold that "best practices" – here defined as isolated and extracted behavioral elements on display in a teacher's classroom – comprise good teaching such that the focus of teacher development ought to attend to the inculcation of these skills, where skills are simply these isolated behaviors in action. Absent from such a view, crucially, are considerations of *how* these ought to be combined into a total practice, or *which* situations call for specific techniques. Such a view presumably takes those aspects to be already implicit in the act of teaching itself, in which case the specific techniques are simply the *best version* of what already happens. I think it worth mentioning that I doubt that Duncan would directly espouse any form of raw behaviorism in the form of an argument that teacher training ought to consist of installing fixed responses to broadly-defined situations in pre-service teachers; but I do think, as the example of Delaware's teacher evaluation protocols at the end of Chapter 1 made clear, that Race to the Top policies lead in that direction.

The second argument, however, addresses one of the ways in which such raw behaviorism is finessed, and this argument will lead on to my larger critique. The second argument I wish to make has to do with an ongoing project to distinguish among different varieties of teacher knowledge. In the second model, conceived and developed by Shulman, Hill, and Ball, primarily (H. Hill & Ball, 2009; H. C. Hill, Rowan, & Ball, 2005; L. Shulman, 1986, 1987; L. S. Shulman, 2007), teacher knowledge runs deeper, so to speak, than any broad rules for applying specific behaviors in broadly-defined contexts, and it explicitly takes up questions around *how* teachers know what to do in particular situations. I will argue, however, that as conceived by Shulman in particular and later implemented by Hill and

others, the notion of teacher knowledge remains inadequately constructed. What Ball and

Hill, following Shulman, have meticulously developed is, on my view, a much finer – and

therefore superior – form of rationalist rule-governed behavior: call it a mentalistic view. By

speaking, at last, of rule-following in general, I hope to underscore two very different views

of what "following a rule" in the performance of learned, skilled behavior amounts to, and to

lay out the ways in which Duncan's view of teaching and the preferable view of Shulman,

Hill, and Ball share salient and problematic assumptions in this regard.

2.2.1 What Data Reveals About Teaching

To recapitulate some points from Chapter 1 in a new light, I will return to some of

what Duncan has to say about the relation of data to teaching and expand upon a discussion

of the literature of scaling. Duncan views the utility of data for the practice of teaching in the

following terms:

> [Teachers] need to know how well their students are performing. They want to
> know exactly what they need to do to teach and how to teach. It makes their job easier
> and ultimately much more rewarding. They aren't guessing or talking in generalities
> anymore. They feel as if they're starting to crack the code (Duncan, 2009b).

In Chapter 1, I indicated that the use of a "code" trope both implies a transmission

model of education, which comes with well-known and much-discussed limitations, and also

constructs the practice of teaching as somehow representable in terms of that data, as though

"what they need to do to teach" and "how to teach" were directly tethered to achievement

data. The nature of this connection, expressed in terms of the cracking of the code and

opposed to "guessing," has to do with a certain propositional *knowledge*. Duncan further

alludes to the role that knowledge plays in linking data that explicitly reveals student

progress in terms of achievement measures and teaching as a practical exercise in claiming

that the use of data answers a *dual* need: "we need to monitor progress. We need to know what is and is not working and why" (Duncan, 2009b). Student achievement data not only serves to "monitor progress" but also indicates "what is and is not working."

The first thing to discuss is that the "working" of teaching, thus expressed in terms of student growth on achievement measures, is conceptually problematic in its own right, as such "working" is expressed in terms of a merely technical sense of learning. However, in Chapter 2, I wish to highlight the fact that Duncan conceives of teaching as a "what" – or a collection of "whats" – that either "is" or "is not working." Since achievement data reveals the "working" of teaching, Duncan takes it to simultaneously reveal the "what" of teaching that either does or does not work. The implication in suggesting this connection is that differences in teacher value-added between two teachers will be attributable to differences between the teaching *practices* of those two teachers, where such practice is thought in terms of "what." Thomas Kane, whose role in the development of value-added measures I mentioned in Chapter 1, is quite explicit on this point. Quoted in an interview with the *New York Times'* Joe Nocera, Kane claims that seeking out and studying the teachers with the highest impacts on student achievement data is geared toward "identifying the practices associated with student achievement" (Nocera, 2012). This attempt at the identification of practices is likewise visible in the studies supporting Race to the Top's teacher-evaluation policies that I discussed in Chapter 1, particularly in the way scholars set out to calibrate the role of classroom observations in terms of specific behaviors (Glazerman, et al., 2011; Kane, et al., 2011b).

"Practices" therefore ought to be understood here as isolated techniques or the maxims that guide them. Examining the practices of each teacher (or groups of teachers) will

reveal differences in habits and behaviors, or, one level further down, differences in mental

states (beliefs and desires) that give rise to such differential practices. Thus, when teachers

want to "know what to do to teach and how to teach," achievement score data and its

correlated behaviors can point to an answer by revealing the "what" of the matter – the

behaviors or the mental states underlying them – in which knowing "*how* to teach" is taken to

consist.

According to such a view, in which skillful practice can be dissected into its

component parts, whether imagined as expressed behaviors or as the mentally-located rules

and maxims underlying the appearance of a holistic practice, the development of good

teachers both from newcomers to the profession and from underperforming teachers already

in the profession involves giving the teachers in question something quite tangible: rules to

apply in feature-defined situations or broadly-sketched behaviors to exhibit in any situation

whatever.

The project of teacher development as a whole, as it appears in Duncan's rhetoric and

Kane's many studies, sets itself the task of (a) identifying the teachers associated with "high

growth students," (b) videotaping or otherwise observing the practices of those teachers, and

(c) distilling the practices of those teachers into their component parts. The ultimate aim of

such an undertaking is, as I quoted Duncan saying in Chapter 1, "to replicate those

[successful] conditions everywhere, and you need to challenge yourselves and challenge each

other to turn one success into a hundred and a hundred into 200" (Duncan, 2009e). By

defining successful teaching in terms of the ability to influence achievement scores, and then

by distilling the practices of teachers associated with exceptional score gains into the smaller

parts – behaviors or rules – that comprise them, Duncan and the reformers hope to render

successful teacher practice into a form that can be "replicated" and thus applied in any classroom whatsoever. The assumption underlying such an endeavor is that such replication of teacher practice so distilled will also as a consequence replicate the achievement score gains according to which success is defined.

As I noted in the introduction to this section, there are several assumptions at work in this line of reasoning, and I wish to take issue with them in turn. Since I have already dealt in Chapter 1 with the conceptual problem of defining teacher quality in terms of influencing achievement score gains, I will set aside for the most part the assumption about the commensurability of the ordinary and technical goals at which Duncan's rhetoric and his policies aim for the time being, but will return to it at the end of this Chapter and also in Chapter 3. Leaving that issue to the side, there remain problematic assumptions undergirding the view of teaching as a skillful endeavor that Duncan and Kane espouse. First among these assumptions is the possibility that skillful practice is comprised of isolated techniques or "bite-sized moves" in the first place. I will call this the behavioristic assumption. Second among these assumptions is the notion that while teacher practices might not be defined according recognizably identical *behaviors*, they might rather be defined according to the recognizably identical *rules or maxims* that guide situationally-appropriate behaviors or responses. I will call this the mentalistic assumption.

I highlight here, and then leave for the time being, the bottom-line assumption of which both the behaviorist and the mentalist assumptions are derivative: the notion that skillful practice in a given domain *must be* a product of isolable and identifiable component parts, whether behaviors or mental states. Without such an assumption, attempts to "replicate" skillful practice by extracting particular techniques, behaviors, or rules from

successful teachers for the purpose of transferring them to new or underperforming teachers would be incoherent. In suggesting that neither an approach according to the behavioristic assumption nor one according to the mentalistic assumption can satisfactorily do the job of replicating excellent practice, I also draw into question this larger, overriding assumption about the composite nature of skillful practice in general.

2.2.2 The Behavioristic Assumption in Practice

One way for which a scholar or policy-maker to approach the question of replicating high-quality teaching under the conditions enumerated above would be the discomposing of the recognizably (or data-substantiated, more accurately) high-quality practice into the recurring behaviors within that practice, the behaviors out of which, one might say, the practice in question is made. Doug Lemov, as I briefly mentioned in Chapter 1, has undertaken just such an approach to the problem of developing teacher quality. As one review of his work succinctly notes, "The techniques described in Lemov's study convey the replicable behaviours, mannerisms, classroom management techniques, time management skills, lesson delivery styles, and preparation methods of teachers whose students excel on state benchmark assessments" (Hollabaugh, 2011). The emphasis in the above review lies on the replicability of the specific behaviors in question. As another scholar notes, this view of skillful teaching makes no particular distinction among the various kinds of classrooms one finds in the world: "Mr. Lemov, Mr. Pianta, and others take the position that many of the core elements of effective teaching transcend the subjects taught" (Keller, 2010). Bess Keller distinguishes here – as do I – between Lemov's focus on behaviors or "elements of effective teaching that transcend" the specific subject-matter involved and other approaches that emphasize techniques particular to certain subject-matter. This distinction serves to

highlight the broad assumption underwriting Lemov's approach, namely that the "core elements of effective teaching" are taken to be behavioristic in nature (Hollanbaugh even terms it a variety of "Skinner-style behaviourism"), and thus are, as Elizabeth Green says, meant "to be adaptable to anyone" (Green, 2010).

The specificity of different demands in different subjects reveals one facet of the insufficiency of such broadly-defined behaviorism, but, as Keller notes, Lemov's approach already takes such an objection into account insofar as he assumes a certain "transcendence" of these subject-matter specifics.   But the particular demands of a specific subject-matter hardly provide the only standpoint of objection in this regard, mainly because subject-matter is hardly the only axis according to which one could further specify the teaching environment. One has at one's disposal, for just one example, the age of the students in question. It is difficult to see the outstanding value of any particular behavioral technique that would be equally applicable to students who still have a scheduled naptimes during the school day and also to students who have after-school jobs. Certainly there are such teacher behaviors that one should exhibit in each case, but at the level of abstraction necessary to such transcendence of the students' developmental stages, it is particularly hard to see how the presence or absence of such a behavior can distinguish between excellence and minimum acceptability.[11] In other words, if the behavioral techniques that "put students on the path to college," are to be seen as "core elements" of exceptional teaching, the degree of breadth such transcendence assumes calls into question the exceptional claims of the teaching it is meant to replicate.

---

[11] Of course, distinguishing between unacceptable performance and minimal competence is not out of the question in such an analysis, and this is a point to which I will return later.

A better way to put it: the teaching it wishes to replicate may in fact be exceptional, but the broad transcendence of the universally-adaptable behaviors cannot plausibly claim to replicate the exceptional quality of that teaching. The very generality of such techniques makes them basic to competent performance, not sufficient (nor perhaps, as I will argue later, even necessary) to excellent performance. These observations raise questions about the relative possibility of taking apart excellent practice and expecting its reassembly in another situation to also manifest such excellence, and as such, these observations imply an incompleteness in the view that skillful practice is atomizable along behavioral lines in the first place.

But the form of the behavioral techniques extracted from skillful teachers and offered up to the broader public by Lemov also requires commentary, and in this space further objections to the atomizable assumption about skillful practice arise. Education reporter Elizabeth Green offers Lemov's approach to classroom management as an example of a contrast to the curriculum of the traditional college of education's teacher-training program: "While some education schools offer courses in classroom management, they often address only abstract ideas, like the importance of writing up systems of rules, rather than the rules themselves" (Green, 2010). Lemov's techniques, Green implies, address the issue at hand in a more concrete, less abstract, form. Green points specifically to Lemov's "What To Do" technique, which certainly sounds concrete. A closer examination reveals, however, two important features of this type of behavioral approach. In the first place, no claim to concreteness can legitimately be made on its behalf (nor, I imagine, would Lemov himself make such a claim, since the notion of providing "the rules themselves" for any and all classrooms is hardly his endeavor). In the second place, the behavioral techniques

themselves appear in the form of maxims and rules, along with examples and suggestions. The implication is that, while Lemov himself has spent countless hours observing high quality teachers at work, the fruits of that observation can be so rendered and made useful simply by codifying the observed behaviors. A teacher would simply have to read the description of the technique and apply it, and voila, excellent practice results.

Because Green focuses on the "What To Do" technique, I will likewise attend to Lemov's discussion of direction-giving. My exploration will demonstrate that such techniques represent no alternative view of teacher training to that offered by the education schools Green denigrates, at least along the lines according to which Green derides them. Lemov's view of teaching teachers to teach is equally abstract; or, put another way, schools of education and Lemov's taxonomy are equally concrete in what they offer their students. But more importantly, I wish to underscore the forms in which Lemov offers behavioral solutions to classroom management and question their utility.

At the heart of the "What To Do" technique is an attempt to solve problems of unfocused, undesired, or contrary student behavior. Lemov notes quite correctly that such undesirable behavior is not always the result of some sort of oppositional attitude on the part of the student but lies rather in simple misunderstanding, the result of insufficient clarity on the part of the direction-giver. This is not to imply that all oppositional behavior stems from such misunderstanding, but Lemov avers that "a larger portion [of noncompliance] than many teachers ever suppose" can be attributed to clarity problems in direction-giving. The solution, according to Lemov, is more clarity:

> When you tell a student to pay attention, ask yourself if she knows how to pay attention. Has anyone ever taught her? Does she know your specific expectations for paying attention (having her eyes on the speaker, say)? Has anyone ever helped her

learn to avoid and control distractions and distractedness? The command "pay attention" provides no useful guidance because it fails to teach (Lemov, 2010, p. 178).

It is important, I think, to point out the excellence in Lemov's analysis as well as to probe the insufficiencies, and to that end, Lemov's first question is particularly well-taken. The teaching profession remains, as Arne Duncan notes in a variety of speeches, insufficiently diverse in many ways; it remains principally the domain of white, middle-class women. Insofar as the students in a given classroom differ widely from the teacher and from each other in terms of race, socioeconomic status, and a general style of upbringing (among unreckonable further qualities), a given teacher cannot simply assume that his or her students all share the sort of lifeway (one might say) in which the directive to "pay attention" has meaning for the teacher. "Pay attention" could well imply different reactions on the parts of different students, and to the extent that these reactions diverge from the teacher's expectations for attention-paying, all such divergent reactions will be treated as non-compliance with the direction. In responding to all such non-compliance as misbehavior, possibly the result of an oppositional disposition or disrespect, for example, the teacher risks alienating students, with potentially significant and negative educational effects.

But problems remain in Lemov's account. In the first and most simple of places, as I promised earlier, the technique for direction-giving that Lemov proposes remains equally abstract to the discussions within the "schools of education" of Green's hypothetical example. Green notes that while such schools of education promote the importance of displaying rules in the classroom, they remain silent as to the content of such rules. Lemov, too, proposes a means of giving instructions clearly to encourage maximal student compliance. But he does not propose what those instructions should be. Lemov absolutely

cannot do so, in fact: his techniques are designed to be "adaptable to anyone." Specific

directions, like rules in the classroom, would depend on the teacher, the students, the school

culture, the subject-matter, and uncountable other factors as well. To suggest specific

directions that ought to be given – which is different from providing examples of direction-

giving – would foreclose on the possibility of general application, of the techniques being

"adaptable to anyone." Such generality is the very definition of abstraction.

In moving to the second point, another congratulatory note is appropriate for Lemov.

He identifies a problem facing many teachers, particularly young teachers: the inability to

know oppositionally-motivated misbehavior from simple misunderstanding or confusion.

His solution seems commonsensical, as well: remove the possibility of misunderstanding or

confusion by clarifying expectations to a degree further than what feels ordinarily warranted.

Under such conditions, more students will experience the feeling of success in meeting a

teacher's expectations, and the risk of conflating the uncomprehending and the uncooperative

students disappears.

On the other hand, the criticism involved in the second point concerns the

behaviorism inherent in both the ways in which the proposed technique is to be implemented

in the classroom and the way in which Lemov expects students to follow directions. These

two, I argue, are intertwined. Lemov describes the "What to Do" technique of direction

giving like so: "To be effective, directions should be specific, concrete, sequential, and

observable" (Lemov, 2010, p. 179). He helpfully continues to use the "pay attention"

situation as he explains these four sub-features of effective direction-giving technique. I will

focus on the first two of these four in order to draw out the shortcomings and insufficiencies

of Lemov's view. Lemov says,

Effective directions are specific…. Instead of advising a student to pay attention, for example, I might advise him to put his pencil on his desk or keep his eyes on me. This provides useful guidance that he can take action on and pay attention to doing. (Lemov, 2010, p. 179).

The teacher in this hypothetical situation wants the student to pay attention. The teacher, in point of fact, says to the student, "Put down your pencil and keep your eyes on me." The student performs these behaviors. But is the student paying attention?

The answer to the question depends upon whether or not the isolated behaviors of putting down one's pencil, on the one hand, and keeping one's eyes on the teacher, on the other hand, constitute "paying attention." Certainly, in an ordinary sense, when one pays attention to something, one's eyes are on it. When one pays attention to the closing seconds of an NBA playoff game, one's eyes are proverbially glued to the screen. But not every instance of one's eyes alighting on a particular thing or feature indicate paying attention to it. I can recall many a statistics lecture in which my eyes were dutifully on the professor, but I was drifting around in my head. In no sense was I paying attention.

The point in raising the latter example is that only "pay attention" means "pay attention," and only paying attention counts as paying attention. (It is unclear to me from Lemov's example whether the hypothetical teacher also said "pay attention" in addition to listing the observable, specific behaviors that the teacher wishes to see, which would have the possibility of connecting the isolated behaviors to some concept of attention-paying, of teaching students how to pay attention, rather than simply aping the behaviors associated therewith.)

It seems perhaps like an obstinate point to make, as though I have to refuse commonsense understanding in order to make the claim that separating a request for what

Lemov calls a "complex skill," (i.e. one that is "rarely equated with a single specific action") into a request for those single specific actions does not add up, somehow, to the whole of "paying attention." After all, to most observers, a roomful of students with pencils down and eyes up front looks for all the world like a roomful of students paying attention.

But it is worth recalling that it is specifically such a situation that Lemov intends to solve by means of such specificity – one in which a "commonsense understanding" of a "complex skill" like paying attention does not suffice for both parties in the direction-giving parley. In fact, part of the implicit claim of behaviorism in general is that no "commonsense understanding" is required; by specifying particular behaviors, one circumvents the problem entirely. In raising the obstinate point, I address precisely the issue of whether or not successfully requesting that a child demonstrate the component behaviors that one associates with paying attention constitutes asking the child to pay attention. It is my as-yet-unsubstantiated claim that it does not. The teacher receives the specific behavioral responses that he or she associates with paying attention – the putting down of the pencil and the upfront-ness of the eyes – but these behavioral responses, one might say, are not the same as "paying attention." Paying attention involves listening, as well, for example; it involves focusing one's mind on a particular topic, place, or person. It is certainly and obviously true that wandering eyes and scribbling pencils can indicate the absence of attention-paying, but this is not always so, and it is further not the case that stilled pencils and attentive eyes comprise attention-paying. In asking for the performance of a "complex skill" in terms of a series of "single specific action[s]," Lemov's teacher indeed receives the single specific actions, but not necessarily any performance of the complex skill characterized by those actions. If all that the teacher desires are the specific behaviors – if he or she wants only

quiet children – such methods of direction-giving are perfectly sufficient. If he or she wishes the students to perform some complex skill, the behavioristic account seems to leave out something essential.

To make the claim stronger and more general, I would like to say that on display in Lemov's discussion of "What to Do" is the bottom-line assumption undergirding Lemov's entire project: that the "complex skill" of teaching itself is merely a collection of many, many "single specific action[s]," and that encouraging people to perform each of those single, specific actions (grouped together under each of the "49 techniques") constitutes teacher education.

Lemov holds the assumption, which was just underlined, that "complex skills," "rarely equated with a single specific action," can nevertheless be analyzed by means of decomposing them into such single, specific actions as would render them performable by a rank novice. The assumption, framed another way, is that the complexity of complex skills is a matter of mere (if presumably enormous) quantity of behaviors. Given the 49 techniques of Lemov's book, any college graduate, let us say, could perform the behaviors associated with the excellent teacher practice – and with "high growth students" – from which the behaviors have been extracted. The result will be, on Lemov's view, the replication of excellent practice.

But there remains one further point to make about Lemov's behaviorism. Lemov seems to assume that performing combinations of behaviors in a particular sequence constitutes the performing of a given holistic skill, as we have seen. Lemov likewise seems to assume that teachers can and ought to frame directions such that they are comprised of actions that "any student knows how to do" (Lemov, 2010, p. 179). Students, then, are

expected to demonstrate correctly observable behaviors given only verbal input. As I just noted, this is not at all unlike the project of the 49 techniques as a whole as they relate to teacher practice. The 49 techniques, however, remain insufficient in a variety of ways, only a few of which I wish to raise at this point.

To return to the example of the practice of direction-giving, it is true that breaking down something like "pay attention" into specific, concrete, sequential and observable behaviors" will likely eliminate any misunderstandings about how to accomplish the behaviors the teacher will equate with the "complex skill." The implication is that one should always frame directions in this way, and if the only danger in terms of inappropriate direction-giving were noncompliance of one form or another, that would surely be the case. Another way of putting the previous statement would be: the "What to Do" technique works flawlessly in an artificially limited domain, in which noncompliance represents the only negative response to direction-giving.

In actual classrooms, however, there are other negative responses for teachers to consider. For example, while breaking down complex skills into their simple components works excellently in precisely the problem situation in which students do not understand what the exhortation to "pay attention" entails, it makes students who do share the form of life in which paying attention follows as a matter of course from the exhortation to "pay attention" feel condescended to, which has precisely the alienating effect that the technique was meant to preclude.

Perhaps one would respond to this objection by saying that a teacher has to know the level of specificity required by his or her students and tailor instructions accordingly. Two avenues of response open up on this rejoinder. First, the 49 techniques are meant to

transcend subject matters and so on; they are supposed to be themselves the core elements of skillful practice. To claim that their application depends upon further knowledge is to undercut precisely that which is supposed to indicate their value. It also reveals the threat of a regress, in which actionable knowledge of a technique requires further knowledge about something else, which may well lead on to other necessary knowledge, infinitely. I will return to that idea later on in greater depth.

In the second place, however, even assuming that the teacher has such knowledge of the specificity required and that such knowledge really is the end of the road, do not most (all) teachers address a wide variety of students in any single class, such that providing instructions perfectly tailored to one group necessarily occurs at a suboptimal level for another? If teachers address a variety of students and wish to alienate a minimum number through either confusion or condescension, the teachers require a 50th technique, one for simultaneously appealing to several learning styles and modalities. This consideration points to still further additional knowledge required for the successful implementation of teaching techniques so formulated, and thus, in the direction of the same regress.

The above considerations, brief though they are, highlight the fact that even the simple technique of giving directions in a way that students can successfully follow with a minimum of negative response turns out to be much more complicated than it initially appears. Lemov's 49 techniques address only, one might say, the bare behaviors he has correlated with achievement score growth; it is obviously not enough to describe those behaviors, tell new teachers to perform them, and then expect excellence in teaching. Given these bare behaviors that Lemov has identified, practitioners still require rules for their successful implementation, such that the novice avoids, for the purposes of my chosen

example, the error of talking down to some of his or her students in direction-giving without speaking in some mystery code to others.

The breadth and generality of Lemov's techniques is thus the source both of its appeal and of its impracticality. In seeking to be "adaptable to anyone," it requires something more for its own implementation. In discussing many of the techniques, Lemov seems to assume a sort of background understanding of what schools and classrooms are like in general, their organization and their purpose, the typical problems that arise, and so on. But as no teacher operates in classrooms in general, and as each one faces some particular set of circumstances, choosing the right technique for the present moment in the present situation seems to appeal to a faculty of judgment, which lies beyond a baldly behavioristic account.

Teaching, on Lemov's view, consists in the knowing of specific behavioral techniques. Performing these behaviors constitutes high quality teaching practice. I wish to say, in segueing to the next section, that the argument I have to make about skillful practice includes a place for something like Lemov's view. Lemov's behavioristic view, I would add, is self-evidently correct as far as it goes. The bone of contention, here and throughout, has to do with how far such a view of teaching can take one toward the ordinary concept of "great teaching" outlined in Chapter 1.

I have no doubt that ordinary-sense "great teachers" in practice break down their directions into manageable parts; but I strongly suspect that what counts as "breaking down their directions into manageable parts" varies widely in the working definition of "manageable," the pervasiveness of the technique's application, and so on. Ordinary-sense "great teachers," though, manage to inspire the highest-performing students, support the lowest-performing, and alienate a minimal constituency of students. Any success in applying

the technique[12] (as well as how, given these distinctions in practical applications, such techniques can be counted as instances of the same) remains to be explained; Lemov's behaviorism cannot do it alone.

2.2.3 The Mentalistic Assumption in Practice

By an odd coincidence, Green's 2010 *New York Times* article also alludes to the work of Deborah Ball, one of the progenitors of another approach to discussing, identifying, and developing teacher quality, and one that attempts to address the problem of *applying* specific techniques. Ball represents what she and her colleagues have termed a "teacher knowledge" research project, that takes its conceptual framework and basic assumptions from the 1980s work of Lee Shulman. Shulman's approach, which I will discuss in detail, takes a more subtle angle to understanding what enables good teachers to teach well than that offered by Lemov and other such behaviorists. To make the contrast explicit, Lemov's view holds that applying specific techniques will have consistently positive educational effects; good teaching, on this view, consists in the demonstrations of these particular behaviors. On the other hand, the assumptions of Shulman's approach, from which scholars such as Ball and Heather Hill operate, recognize that teacher knowledge of *specific behaviors* can only account for a very narrow range of teacher quality, because such the full spectrum of teacher quality manifests itself, as I hinted above, in the choosing among potential behaviors to demonstrate. In Shulman's eyes, the knowledge upon which such a choice rests, the knowledge which arbitrates the choice, also requires an accounting. As Ball and others aver in a 2005 article, "Shulman and colleagues' work expanded ideas about how knowledge

---

[12] I want to stop short, here, of implying that I accept or endorse the notion that what a good teacher is doing is best characterized as "applying a technique."

might matter to teaching, suggesting that it is not only knowledge of content but also knowledge of how to teach content that influences teachers' effectiveness" (H. C. Hill, et al., 2005). Shulman's project, then, seeks to break down Lemov's techniques – which are intended to transcend every specific instance of their application – into techniques specific to particular content areas. Additionally, however, he takes up the problem of the *judgment* necessary to select among alternative techniques, which marks an improvement of one kind on Lemov's model.

Since Shulman represents the jumping-off point for this particular point of view, I will spend most of this section delving into his work and the way its assumptions construct the practice of teaching, and specifically the relation of knowledge to practice in general. The notion that "knowledge of content" is potentially equivalent to "knowledge of how to teach content" on the level of *knowledge* arouses my suspicion.

In spite of my own apprehension in this regard, it is worth mentioning that in terms of Lemov's goal of discovering what makes excellent teachers excellent, the measures Ball and Hill, among others, have developed for Shulman's "pedagogical content knowledge" – in their specific case, what they call "Mathematical Knowledge for Teaching" or "MKT" – has proven to be one of very few useful teacher-characteristic predictors of student outcomes in terms of achievement on standardized tests of mathematics (H. Hill & Ball, 2009; Kane, et al., 2008). In other words, if Lemov's extraction of specific behaviors from the practices of excellent teachers represents a certain kind of step in the direction of replicating excellent teaching elsewhere, the concept of "teacher knowledge" toward which Shulman and his acolytes labor represents a more refined step in the same direction, insofar as it addresses the problem of selecting among alternative behaviors and insofar as the statistical methodologies

of determining teacher effectiveness validate this approach. Shulman's assumptions about

the relation of knowledge to teaching practice, in sum, have led to advances in terms of

identifying teacher characteristics associated with influencing achievement score growth.

Whether or not these advances hold the potential to advance much further and prove equally

useful to other content areas remains, for me, the primary subject of discussion.

In a 2009 article, Hill and Ball present their vision of MKT as it relates to other forms

of knowledge:

> We also noted that some MKT was more of a blend of mathematics with other
> kinds of knowledge, such as knowledge of students or knowledge of teaching or
> curriculum. These blended forms of content knowledge—*knowledge of content and
> students or knowledge of content and teaching and knowledge of content and
> curriculum*—appeared as finer-grained categories of what Shulman and his
> colleagues termed "pedagogical content knowledge" (H. Hill & Ball, 2009).

In order to make sense of the notion of "blended forms of content knowledge," it will

be necessary to lay out Shulman's full-fledged understanding of the role of knowledge in

teaching. It is my suspicion, which will emerge again at the end of this section, that the

recourse to the idea of "blended forms of content knowledge" in such hybrid terms as

"*knowledge of content and students or knowledge of content and teaching and knowledge of

content and curriculum*" gestures once more in the direction of a regress, the danger of

which will be visible in Shulman's original discussions of knowledge and teaching, to which

I now turn.

In what Hill and Ball regard as his groundbreaking contribution to educational

research, Lee Shulman recognized a "blind spot" in educational discourse: "In reading the

literature of research on education, it is clear that central questions are unasked. The

emphasis is on how teachers manage their classrooms, organize activities, allocate time and

turns, ascribe praise and blame" (L. Shulman, 1986). Shulman says that "what we miss is the

content of the lessons taught, the questions asked, the explanations offered. […] Where do teacher explanations come from? How do teachers decide what to teach, how to represent it, how to question students about it, and how to deal with problems of misunderstanding?" (L. Shulman, 1986).

Shulman's suggestion that researchers attend to the specific content of lessons in order to explore and discover the requisite qualities underlying high-caliber teaching represents a significant improvement over Lemov-style generality. Rather than speaking on a level ostensibly common to all student ages and subject-matter requirements, Shulman acknowledges that being a good chemistry teacher might require different strengths than being a great physical education teacher or art teacher or history teacher.

Shulman opens an oft-cited article, "Knowledge and Teaching" (1987) with a narrative account of the practice of an excellent English teacher, whom he pseudonymously calls "Nancy." Shulman's account notes that "the observer was well impressed with the depth of [her] understanding of that novel and her skill as a pedagogue, as [the observer] documented how Nancy helped a group of California high school juniors grasp the many faces of that masterpiece" (L. Shulman, 1987). While an interview with Nancy leads the interviewer to conclude that she has a theoretical framework underlying her skillful teaching, the observer claims that this framework is not implemented as a rigid set of guidelines: "Thus Nancy's pattern of instruction, her style of teaching, is not uniform or predictable in some simple sense. She flexibly responds to the difficulty and character of the subject matter, the capacities of the students (which can change even over the span of a single course), and her educational purposes" (L. Shulman, 1987). From Nancy's example, Shulman hopes to elicit

the skillfulness of practice. His goal, like Duncan's and Lemov's, is to see "teaching like Nancy's … become typical."

What "typical" means for Shulman appears slightly different from what it means for either Duncan or Lemov. First and foremost, no particular behavior or technique seems to characterize Nancy's practice, and so simply extracting and repeating behaviors will be insufficient to replicating her excellence. Rather, as the description of Nancy's practice has it, instead of a behavioral "pattern of instruction" that one might expect to be "uniform or predictable in some simple sense," one instead finds that Nancy "flexibly responds to the difficulty and character of the subject matter, the capacities of the students (which can change even over the span of a single course), and her educational purposes." The *flexible responsiveness* to her particular situation defies accounting in terms of isolable, repeatable behavioral techniques. Nothing so outwardly stable undergirds Nancy's practice. What requires replication is not *behavioral*, then.

But Shulman's research questions make clear an assumption that *something* stable must ground the excellence of Nancy's practice, and this stability exists in the form of certain kinds of knowledge. Asks Shulman: "What does Nancy believe, understand, and know how to do that permits her to teach as she does?" (L. Shulman, 1987). Shulman's question establishes two assumptions, only the narrower of which I will here address. The broad assumption is that "to teach as she does" is "permitted" or caused by some further factor to such an extent, one might say, that the appropriate response to Nancy's practice lies in seeking its causal conditions, conditions conceived here in terms of "what," to hearken back to the introduction to this section. The more narrow of the two assumptions is that, specifically, the causal conditions are taken to consist in "What … Nancy believe[s],

understand[s], and know[s] how to do," all of which fall, for Shulman, under a rubric of the mental. Nancy's skillful practice, in other words, is the expression of certain inner mental states or conditions, the replication of which will have the effect of replicating her skillful practice. Shulman, too, is absolutely clear about this assumption: "One of the more important tasks for the research community is to work with practitioners to develop codified representations of the practical pedagogical wisdom of teachers" (L. Shulman, 1987).

Shulman's view, then, holds the following assumptions: skillful practice has specific, identifiable causes; these causes consist in what skillful practitioners believe, understand, and know how to do; the task of the research community is to develop codified representations of the same. Such codified representations will thus enable the propagation of the causal conditions of excellent or skillful practice, and thus skillful practice itself.

The last sentence in the above paragraph, it is worth noting, reiterates Lemov's salient assumption: that developing teachers consists primarily in codifying the separable causal conditions connected to existing skilled practice. Where Lemov's view holds that the relevant causal conditions consist merely in the knowledge of particular behaviors, Shulman's view asserts that additional forms of knowledge prove requisite. Shulman focuses to a large extent on the ways teachers decide what and how to teach; in acknowledging the role of decision-making, Shulman moves away from a Lemov-style behaviorism to what I am calling the mentalistic view. In particular, though Ball and Hill focus primarily on the *pedagogical content* knowledge side of what Shulman requires, I wish to highlight his attempt to make sense of the role of judgment in teacher practice. The form of knowledge involved here goes by two names in Shulman's writing: "the wisdom of practice" and "strategic knowledge."

Briefly put, instead of codifying and replicating *behaviors* under the assumption that behaviors comprise the excellence of excellent practice, Shulman recognizes that excellent practice is not necessarily "predictable or uniform in some simple sense" and that, a fortiori, simply treating teacher practice as predictable or uniform behaviors will not have the desired effect of replicating excellence. However, he maintains the assumption that despite the appearance of unpredictability or non-uniformity in practice, such excellent practice is grounded on something that transcends these differences; namely, certain beliefs, understandings, or knowings-how (all conceived as potential responses to questions that start with "what"). The codification and replication of these mental elements – collectively, "the wisdom of practice" – will yield the replication of excellent practice, despite the appearance of behavioral discrepancies.

The very notion of codifying a "wisdom of practice" raises a problematic philosophical history around attempts at conceiving of practical knowledge as a subspecies of theoretical knowledge, of attempting to formulate a theory of practice. This is, in fact, the crux of the matter in Race to the Top policies quite generally, where the ability to articulate, according to abstract and theoretical tools, *how* one knows something skillful is imagined to access necessary and sufficient conditions for the performance of the skillful behavior in question. All knowledge, including, in Shulman's case, knowledge-how, is conceived as a species of knowledge-that or knowledge-what. This is the salience of the form of Shulman's question: "*what does* Nancy… know how to do that permits her to teach as she does?" – that know-how is conceived as the propositional answer to a question in the form of "what." Nancy's skillful teaching is thus conceived according as the result of a chain of causation which bottoms out in abstract, theoretical knowledge amenable to the form of a proposition.

Shulman's account of "strategic knowledge" and the "wisdom of practice" help us to see the problems in conceiving of skillful behavior in this way.

Shulman's definition of the "wisdom of practice" makes clear his mentalistic assumptions about the nature of skillful activity: "Wisdom of practice" consists of "the maxims that guide (or provide reflective rationalizations for) the practices of able teachers" (L. Shulman, 1986). The practices of able teachers – not merely beginners, but *able* teachers – are held, on this view, to be more or less explicitly guided by maxims. Skilled practitioners thus know maxims or rules, and knowing these rules accounts for the skillfulness of the practice. Knowing *how* to teach consists in knowing the "what" of propositional rules. If this is in fact the case, then certainly the rendering explicit – the codifying – of these maxims, so that they can be transferred to new or underperforming teachers, will enable the replication of excellent teaching that Race to the Top and academic scholars alike strive to attain.

However, Shulman's own academic lineage as a student of Joseph Schwab's leads him to look more deeply into the matter. Schwab, after all, makes a clear distinction between his notion of the practical and the "theoretic,"[13] noting that "the practical is concerned with choice and action, where the theoretic is concerned with knowledge" (Schwab, 1969). In order for Shulman to make the a matter of "choice and action" reliant upon some sort of "knowledge," Shulman must account for the way in which knowledge undergirds choice and

---

[13] Obviously, this distinction does not belong to Schwab – one can see Plato's account of Socrates in *The Euthyphro* as drawing this distinction. Socrates asks Euthyphro for the characteristics according to which the religious man can identify piety; Euthyphro responds with examples of piety, but cannot make these examples into definitions; Socrates concludes that no one knows anything (because "knowing" means propositional knowledge). This despite the fact that Euthyphro clearly *can cite examples of* piety.

action. "Strategic knowledge" is Shulman's term for the interaction of the foundational theory and the resultant practice.

The intervention of "strategic knowledge" becomes necessary because, unlike Lemov, Shulman recognizes that in certain situations, the "rules or maxims" that comprise the "wisdom of practice" may come into conflict. Either the rules themselves do not provide adequate guidance, or else the situation seems to bear features that call for the following of *different* or *contradictory* rules. "Strategic knowledge" provides the means by which Shulman proposes to account for the faculty of *judgment* in skillful practice that Lemov's behavioristic model overlooks. On this point, he is explicit: "When strategic understanding is brought to bear in the examination of rules and cases, professional judgment, the hallmark of any learned profession, is called into play. What distinguishes mere craft from profession is the indeterminacy of rules when applied to particular cases" (L. Shulman, 1986). While Shulman conceives of skilled practice in terms of following (propositionally-known) rules or maxims, he also recognizes the necessity of navigating among them, of making the temporal connection between the theoretical and the practical. *Simply* knowing rules and maxims – or Lemov's behavioral techniques – is insufficient to the performance of skillful practice.

But in developing the notion of "strategic knowledge," Shulman reveals it to remain explicitly on the side of the theoretic, which prevents it from playing the bridging or mediating role between the theoretical and the practical for which it is designed. Shulman describes the nature of strategic knowledge by emphasizing its nature as *knowledge*: "The professional holds knowledge not only of how—the capacity for skilled performance—but also of what and why. The teacher is not only a master of procedure but also of content and rationale, and capable of explaining why something is done" (L. Shulman, 1986). The

navigation of situations, in the face of an "indeterminacy of rules," appears, in Shulman's construction, to bear some relation to an ability to *explain* "why something is done." Shulman differentiates between "the capacity for skilled performance," which he associates with know-how, and knowledge of "what and why," but the purpose of drawing this distinction remains unclear. He seems to imply with his sentence structure that the mere capacity for skilled performance is insufficient to our purposes, and that the real goal is the explanation of the "what and why." The value of a post-hoc explanation is, we can infer, directly related to the notion that "maxims … guide … the practices of able teachers." In fact, this direct relation is further cemented by the way in which Shulman, in discussing the concept of the "wisdom of practice," *conflates* the guiding function of maxims with their rationalization-providing function.

Two problems arise for Shulman at this point: the first is the seeming implication that a post-hoc ability to "provide rationalizations for" or otherwise explain "why something is done" necessarily indicates that this rationalization or reason-why provided the a priori guidance that Shulman attributes to "strategic knowledge." The second is that such rationalizations or reasons-why, being themselves propositional in nature, are made to serve as *rules or maxims themselves*.

To address the second point first, Shulman's conception of strategic knowledge as consisting of, most importantly, the ability to explain "why something is done," itself founded upon a certain "knowledge of what and why," exposes him to the prospect of the regress to which I alluded earlier. In acknowledging that maxims or rules conflict in practical situations, such that a faculty of judgment is needful, Shulman indicates that blindly adhering to maxims or rules proves an insufficient means of navigating an emergent

situation. But his solution to the indeterminacy of these situations appears in terms of "knowledge of what and why," or *principles for action* according to which such indeterminacy is to be settled. In other words, Shulman's strategic knowledge, amenable as it is to "explanation" and "knowledge of what and why," is conceived as a codifiable and *higher* order of rules and maxims. But given that these are also rules and maxims, and that in lived situations, rules and maxims come into conflict, how is the practitioner to navigate among the conflicts among these *higher* principles? Presumably, on this model, with further, higher rules.

This is the regress to which attempts to formulate practical know-how in terms of mentalistic or behavioristic principles that can be theoretically known are famously prone. Hill and Ball's hybrid construction of the knowledge thought to underlie excellence in mathematics teaching echoes this original regress: such skillful teaching, they maintain, must be comprised of certain hybrid forms: "*knowledge of content and students or knowledge of content and teaching and knowledge of content and curriculum*" (H. Hill & Ball, 2009). But having separated the domains of knowledge into content, curriculum, teaching, and students, and having even combined those into overlapping interaction terms like (content-and-students) or (content-and-teaching) or (content-and-curriculum), nothing prevents one from further and infinitely dividing those categories into say, knowledge of (algebra-content-and-geometry-content-and-Spanish-as-a-first-language-students).

The raising of the regress argument on my part is meant to accomplish two things. In the first place, on a point I will continue to reiterate, the finer-grained theoretical specifications of a domain *do indeed* move one closer to the goal of replicating skillful practice. In this sense, Shulman's conceptions of "strategic knowledge" and the "wisdom or

practice," developed later by Ball and Hill into the domain of pedagogical content knowledge known as Mathematical Knowledge for Teaching, represent notable improvements on the tremendously abstract behavioral account of the sort that Lemov offers. In the second place, though, the threat of the regress implies that the pursuit of finer-grained specifications is literally infinite, or at least that the endeavor winds up at the level of the individual student in the particular situation, and so attempting to replicate the excellent skillful practice that *in fact exists* by these means alone are bound to fail.

This bound-to-fail ought to be heard in two ways: failure to replicate excellent practice by these means is inevitable in light of the regress threat because (a) bottoming-out in some sort of foundational domain or fact or rule or principle seems implausible by the very nature of an infinite regress, and (b) this impossibility of bottoming-out, in combination with the *actual existence* of excellent practice implies that excellent skillful practice is not comprised of following maxims and rules in the way that either the behavioristic or the mentalistic account assumes. Whatever would be generated according to either a Shulman or a Lemov model would not be, therefore, identical to the kind of practice from which it is derived. To speak of the replication of excellent teacher practice on this assumption, then, ceases to make sense.

On that note, I need to return at last to the first problem I located in Shulman's arguments around "strategic knowledge" and "the wisdom of practice": that an ability to offer post-hoc rationalizations or explanations of why something is done seems to indicate to Shulman that such rationalizations or explanations also provide the aprioristic guiding function he attributes to the "rules and maxims" of practical wisdom. The regress argument provides one form of demonstrating the implausibility of this assumption: the failure to reach

the bottom of attempts to dissect skillful practice into the rules and maxims that apparently guide it, and then into rules for navigating among conflicting rules and maxims, and so on, suggests that perhaps skillful practice does not start with something like the most general or abstract facts and rules, and thus that rules and maxims may not "guide" practice in the way that Shulman imagines.

But one need not reach into the logical realm to locate the implausibility of the assumption that providing a post hoc rationalization for why something is done necessarily implies that this rationalization also played an a priori guiding role in the doing itself. A situation from ordinary experience will suffice. Every day, for example, and this is a true story, I order coffee, no matter where I am. In Switzerland, as I approach the counter, I have to think hard (embarrassingly hard) about the subtle differences among "*Ich het gärn e grosse Kafi*," "*Für mich bitte e grosse Kafi*" and "*E grosse, schwarze Kafi, bitte*." I even have to mouth the phrases to myself to make sure the pronunciation will come out right. When someone asks me afterward why I ordered coffee in such and such a way, I will be able to come right out with a reason (it might not be correct), and that reason will certainly have explicitly guided my coffee-ordering behavior as a maxim or a rule.

In the United States, meanwhile, I can have conversations with friends as I approach the counter, read a newspaper, check my phone, or do any number of other things. When I get to the counter, I simply order a black coffee. If someone asked me on my way out of the café *why* I ordered coffee in just the way that I did, I would first be taken aback. And then, doubtless, I would provide an explanation. But being a native speaker of American English, I would not feel compelled to concede that I had explicitly reasoned out beforehand the best way to order coffee, as I did in Switzerland. The reason that I give to my interlocutor under

duress, as it were, feels like an invention; in no way was I guided explicitly by rules or maxims, nor would I expect my explanation to make it possible for a Martian, say, to order coffee like a native speaker. It is only when I am *asked* for a rationalization that the very notion that I *might have had* a reason to order coffee in *just this way* appears.

In both cases, I can offer a post-hoc rationalization for why I ordered coffee in such-and-such a way. In only one case – in which I am a decided beginner – was this rationalization tied to explicit, aprioristic reasoning. While sometimes, particularly in the case of beginners in a given skill domain, the ability to provide an explanation of why something is done *does* indicate that a certain rule or principle has played a guiding role, this ability to give a post-hoc rationalization does not *always* indicate that such guiding took place. Evidence of aprioristic guiding by maxims or rules is in fact conspicuously absent from empirical studies of skilled performers in various skill domains. (cf, for example, DeMaio, Parkinson, Leshowitz, Crosby, & Thorpe, 1975).

The fact that one follows rules and maxims as a beginner is nevertheless taken as evidence that such a process occurs *throughout* one's skill development, even if unconsciously. But in ordering coffee in Madison, Wisconsin, say, I no more employ particular rules than I think to myself, "the rabbit goes around the tree" as I tie my shoe.[14] Neither my ability to *do* it, nor an ability to offer an explanation of *how* I did it in terms of particular facts and rules, is sufficient to indicate that the terms of the explanation amount to the causal conditions of the performance itself. Such after-the-fact explanations do not

---

[14] There are, of course, thinkers who would certainly contend that this is *exactly* what happens; my claim is that they must appeal to *logical* grounds, since parties on both sides of this issue agree that nothing *conscious* or empirically available is going on which could support their claim.

necessarily support the assumption that *whenever* one does something skillfully, the skillful performance is a behavioral manifestation of interlocking maxims and rules. Shulman's attempt to codify the wisdom of practice with the ultimate aim of replicating the skillfulness of this practice, however, makes just such an assumption. His conception of knowledge-how, like Lemov's, amounts to a knowledge-of particular rules and maxims thought to guide even the most skillful practice.

I suggest that while such codification of rules and maxims – like Lemov's behavioral techniques – does indeed have value in the realm of teacher training, its utility has limits, as well, and a certain cognizance of the limits of this approach is quite necessary to the development of the excellent teachers that Duncan, Lemov, and Shulman all desire. Seen in this light, Lemov and Shulman's assumption that knowledge of behavioristic or mentalistic rules underlies any particular skilled performance does indeed make progress in the direction of excellent teacher practice. But if one entertains the notion that excellent teachers are not themselves, generally speaking, being guided at every moment by rules and maxims, then attending to their practice only with an eye to *extracting* rules and maxims will prove insufficient to replicating their practice. Without considering what happens as we move from explicitly following rules, like a non-native speaker ordering coffee, to some other form of skilled practice, as in adult-level shoe-tying, we risk misplacing the end goal of teacher training in general. If something *different* from explicit rule-following behavior comprises excellence in a given skill domain, then the progress one can make in the direction of excellence by means of identifying, codifying, and transferring rules will eventually run out and some entirely new approach will have to take over.

While conceding that extracting rules and maxims has value, and that implementing them makes progress toward excellence, I would like to assert – a point I will substantiate in the conclusion of Chapter 2 – that this represents what Hubert Dreyfus will call the fallacy of the successful first step.

2.2.4 Another Sense of Scaling in Duncan's Rhetoric

In Chapter 1, I noted the prevalence of the metaphor of scaling in Duncan's discussion of improving the teaching corps and the quality of schools in general, nation-wide. I likewise noted that the scaling metaphor comes from two different contexts of origin, a business or economics context and a network or IT context. In Chapter 1, I focused primarily on the salience of "scaling-up" in the business sense. In returning to the notion of "scaling-up" educational successes in light of the discussion of the scholarly or academic hunt for the *knowledge* necessary to the skillful practice of excellent teachers, I would like to consider attempts at taking reforms to scale in its computing- or network-derived sense.

In the business sense, as I discussed, scaling-up means adding volume to an existing system; in the network or IT sense, scaling-up has a somewhat similar meaning, but its particular valence has primarily to do with *generalizing* the mechanisms by which a given program achieves results in a certain delimited domain to other domains, and expecting the achievement of similar results. Scaling, in this sense, makes two assumptions from a given fact, and at this point these assumptions will sound familiar. From the fact of demonstrable and obvious success in a given place, a scale-up approach assumes, correctly, because we are speaking here of programming, that the success is directly attributable to the coding involved – that is, to the symbolic representations of facts and to the rules programmed for manipulating them. The scale-up approach then further assumes that generalizing from this

success is a matter of *adding* further symbolic representations to account for additional features in an expanded or altogether different domain and necessarily also *adding* rules for navigating among this expanded set of features. Thus, generalizing from a limited success is imagined to follow from grasping the rules and representations active in the limited domain and expanding upon them as needed for transference to another, wider application.

Duncan's comments on scaling in a variety of outlets reveal the similarities between his view of the problem of educational success and the IT or network sense of scaling. In his interview on *The Daily Show,* Duncan demonstrates the similarity most clearly: he acknowledges that while there "are pockets of excellence" in the American school system, the task is to "take them to scale. . . . It's happening. It's just not happening at scale" (Stewart, 2012). The "pockets of excellence" he sees imply to him that the answers to the questions of how to teach *exist* and all that is now required is some sort of means for making those answers generally available or widely practiced. It will be obvious by this point that "how to teach" is assumed to exist in the form of information, in a set of rules and facts. Duncan's enthusiasm in this interview, the energy driven by a sense of frustration, also drives home the impression that he feels that the large step – discovering the essential rules and facts that underlie "how to teach" – has already been taken. Generalizing ought to be the *denouement*, and thus the frustration with the "pockets of excellence" and the fact that it is "not happening at scale."

Speaking in Oregon, as I also already mentioned in Chapter 1, Duncan says, "The CLASS Project is a tremendous example of the successful work that should be taken to scale, because students benefit when teachers work together to share best practices and learn from one another" (Craig, 2011). Here, too, Duncan emphasizes the *ease* of generalization, once

given the existence of local success. Noting that the CLASS Project represents an example

of "successful work that should be taken to scale," Duncan hints that the *isolation* of the

project's success has to do with insufficient information-transfer among either schools,

administrators, or teachers: "[S]tudents benefit when teachers work together to share best

practices." The concept of "best practices," by this point, is freely understood to consist in

specific, isolated structures of behavior applied by rule. "Sharing" best practices would

consist in simply offering the behaviors in question much the way Lemov does, and

accounting for the local conditions in making minor additions and modifications. In the way

that any necessary adaptations or modifications are assumed to be negligible in generalizing

from one particular context to another, such that the "best practices" maintain the spotlight,

this comment of Duncan's likewise bolsters the similarity between the sense of scaling he has

in mind and that derived from an IT or network context.

Duncan's sense that achieving success in a local context represents the largest step

toward achieving widespread application – generalization problems being, on this view,

matters of ineffective communication or entrenched bureaucratic opposition or the like – is

echoed in the literature on scaling-up educational innovations.

One group, attempting to bring innovations in science education to scale and

proposing ways to deal with the vast varieties of contexts in which the innovation will be

applied, notes the general pattern of efforts to take such innovations to scale: "The stages

involved in 'going to scale'—designing an educational innovation, assessing its efficacy,

testing its effectiveness, and, finally, implementing it on a large scale—are under

development as well" (O. Lee & Luykx, 2005). In the stages cited by the authors, three of

the four have to do with creating something that works in a local situation and demonstrating

that it works. "Implementing it on a large scale," on their view, is not one of Lemov's "complex skills," as it does not appear to require any further specification. It is itself, then, taken to be simple and singular.

Another group expresses similar frustrations with the problems of getting to scale. Noting that "pockets of good educational practice can be found almost *anywhere*," the authors remark that "Nevertheless, good educational practice cannot be found *everywhere*. The incidence of usage of available knowledge, and the rate of spread of effective practices, is grievously low" (Healy & DeStefano, 1997). Having spied successful practice in pockets, and relying on the assumption that such successful practice rests upon explicit knowledge or information, the frustration with the slow spread of best practices is natural. The "usage of available knowledge" is once again taken to be a simple matter.

In noting the frustrations with attempts to take education reform to scale, and in thinking of "scaling" in its network or IT valence, it will perhaps be helpful to make a broad analogy between the ways in which skillful teacher practice is assumed to work – i.e., as a matter of atomizable behaviors or techniques whose implementation is governed by rules or maxims – and attempts at generating skillful or intelligent behavior where no such *assumptions* are necessary. The project of Artificial Intelligence needs *assume* no underlying rules-and-facts ontology to computer programs; such programs *absolutely* run on symbolic features linked by rules.

In spending considerable time drawing out this analogy, I hope to suggest that the problems of scaling local successes in both Education and symbolic Artificial Intelligence are related to one another insofar as both take the discovering of applied rules and facts that lead to success in a limited domain to be the difficult part of the task, and the generalization of

those same to some larger domain to be the simple part. Similarly, I wish to assert that

failures of AI to demonstrate any sort of skillful flexibility of the kind Shulman attributes to

"Nancy" likewise threatens the assumption that our own skillful practice *actually* stems from

the sorts of "rules and maxims" that Shulman considers necessary. That the many attempts

to generate intelligent or skillful behavior according to the same assumptions on which the

studies of Shulman, Lemov rest – as well as the whole of the educational reform movement

typified by Race to the Top policies – have met with only the most limited types of success

ought to call those very assumptions into question. The demonstrable failure of symbolic AI

to achieve its projected goals – repeatedly – ought to caution the education reform movement

away from conceiving of the generalization problem as a simple one, and should also

encourage us to revisit our assumptions about what skillful practice is made of. In this new

light, the concept of "best practices," I intend to argue, thus carries us a much smaller

distance toward our goal of developing excellent educational practice than we often assume.

<u>2.3 Best Practices and Artificial Intelligence</u>

The primary assumption that I have identified, but perhaps without sufficient

explicitness, in the above section is that, following Lemov, "complex skill[s]" are in reality

combinations of the "single specific actions" into which their complexity is (finitely)

divisible. In order to achieve the performance of a complex skill from the combination of

smaller behavioral pieces, as again Lemov notes, one requires a form of directions for putting

them together, directions that are "specific, concrete, sequential, and observable." I take the

ontology of a "complex skill" from Lemov's "What to Do" technique, which might appear

something of a misappropriation or misrepresentation of the intent. Two things auger against

such an understanding, however: Lemov's division of the complex practice of excellent

teachers as a whole into 49 separate techniques, on the one hand, and Shulman's academic

pursuit of the "rules and maxims" – the directions – that guide teacher practice, on the other.

The assumption undergirding both projects has to do with the divisibility of complex skills

into simpler component parts ordered by particular rules or maxims. Understanding the

whole of complex teacher practice, on this assumption, requires understanding the right ways

to combine the right techniques.

Kevin Cahill calls this assumption the "disengaged view" of practice; Hubert Dreyfus

terms it the "detached view" and, elsewhere, a form of "cognitivism"; Charles Taylor names

it the "intellectualist view." Cahill's succinct explanation of the "disengaged view" reveals

its identity with the assumptions upon which Shulman, Lemov, and, ultimately, Duncan's

vision for data usage in guiding teacher practice, rest:

> Perhaps the main intellectual requirement that the disengaged picture makes
> on us is that we must envisage for ourselves a way to account for the rationality
> manifest in our various activities that is completely independent of those activities. In
> the particular case of following an arithmetic rule, the disengaged view requires us to
> be in possession of something that would satisfy the requirement that our
> understanding transcend all of our actual responses when we write out a numerical
> series. (Cahill, 2011, p. 116).

Lemov's seeking for the 49 techniques of great teachers, as well as Shulman's search

for the maxims and rules – the knowledge – that comprise the "wisdom of practice," both

assume, exactly as Cahill puts it, that "the rationality manifest in our various activities"

results from something (the knowing *of* specific techniques, the knowing *of* rules and

maxims) that "is completely independent of those activities," that "transcend[s] all of our

actual responses" in lived situations. The value in having a rule to follow or a technique to

apply depends upon its applicability in all possible situations, including the unforeseen, the

unpredictable, and the entirely new. The value of the *specific* techniques that Lemov

extracts, or the *particular* maxims and rules that Shulman locates, depends upon the view that the skillful practices *from which* researchers derive them are in fact rule- or maxim-guided activities, that the "complex skill" of teaching under observation is *in fact* the manifestation of "single specific actions" combined in practice according to some particular (and transcendent) rules. Only if this is the case does a "best practices" approach to teacher development have the possibility of replicating the excellent practice for which purpose it is designed.

The dominance of a "best practices" approach to the problem, mentioned earlier, of propagating the "usage of available knowledge" (Healy & DeStefano, 1997) in many practical fields, including the field of education, is such an obvious fact that it practically defies through sheer quantity the possibility of adequate documentation.[15] "Best practices" represent an outgrowth of a mode of problem-solving pioneered at the RAND corporation:

> More than any particular discovery or invention, however, RAND's calling card became the method of *systems analysis*. The very height of means-ends rationality, systems analysis is the interdisciplinary science of maximizing efficiency and economy given a particular objective and a set of "system parameters" (Medvetz, 2012, p. 71).

In Chapter 1, I have already detailed the *conceptual* problems inherent in applying such "means-ends rationality" to the project of education. Such an approach will work *if given* specific objectives and particular parameters. Speaking here only of the objectives side of it, the *great* teaching immediately visible to Duncan upon his entering a school proves insufficiently particular or else inadequately defined for the purposes of such "systems

---

[15] A search of the University of Iowa Library's database turned up over 8,000 articles in journals ranging across the fields of medicine, law, business, microbiology, engineering, education, and many, many more.

analysis." Quantified measures of teacher quality, however – where "effective teaching" is defined according to the ability to influence achievement scores – does the trick. Distilling the knowledge-that believed to underlie the practice of such teachers is from this point a relatively simple matter. In order to apply such means-ends rationality to the problem of developing great teachers, we must remember, we have to transform the objectives we *want* into a shape that the chosen problem-solving method requires.

Here in Chapter 2, I emphasize that in addition to the considerable problem that "best practices" are derived from an analysis conducted in terms of a distorted goal, there is also no reason to assume – and substantial reason to doubt – that spreading "best practices" in teaching will be coextensive, one might say, with the dissemination of skillful teaching practice. This is not to claim (to reiterate an earlier point) that Shulman and Lemov's work and the "best practices" to which it is related are without value. It is very much to claim, however, that whatever such approaches can do in terms of teacher development dries up well in advance of the greatness that provides Race to the Top's justification. A good test of the limits of the assumptions about skillful behavior typified by Lemov and Shulman, as well as by the "best practices" regime, will be to examine the relative successes and setbacks of the projects of the AI research community, whose machines operate according to just such a features-and-rules model that the aforementioned scholars believe mirrors our own.

2.3.1 The Rationalist Assumption of AI and the Projects
Involved

The critical perspective from which my analogy proceeds will rely to a substantial degree on the work of Hubert L. and Stuart E. Dreyfus, originally undertaken in a consulting capacity for the RAND corporation, as the think tank sought an evaluation of its AI projects

(Selinger & Crease, 2002, p. 249).   Because the brothers – Stuart an engineer and Hubert a philosopher – present for simplicity's sake much of their work from Hubert's point of view (H. L. Dreyfus & Dreyfus, 1986, p. 1), I will likewise do so, referring to "Dreyfus" in the singular even in cases where both are assumed to have done the authoring.

In addition to offering a considerably pessimistic projection for the ability of AI to accomplish its goals (H. Dreyfus, 1967; H. L. Dreyfus, 1992), Dreyfus also translated his critique into a phenomenological model of skillful or expert practice, a discussion of which appears in Chapter 2.4.

At the outset of his critique of AI, I wish to say plainly that in pointing out the shortcomings of the rationalist or intellectualist assumptions undergirding Race to the Top's teacher-quality policies along Dreyfusian lines, and in describing the ways in which Dreyfus's account of skill development proceeds from a different set of assumptions, I hope to claim (1) that ordinary-sense excellent teaching is simply not reproducible according to a best-practices approach, and that (2) Dreyfus's alternative account points us in helpful and fruitful directions.

Dreyfus's Introduction to the MIT edition of *What Computers Still Can't Do* provides the most straightforward account of the assumptions according to which the projects of symbolic Artificial Intelligence proceed. Says Dreyfus:

> GOFAI [Good Old-Fashioned AI] is based on the Cartesian idea that all understanding consists in forming and using appropriate symbolic representations. For Descartes, these representations were complex descriptions built up out of primitive ideas or elements. Kant added the important idea that all concepts are rules for relating such elements, and Frege showed that rules could be formalized so that they could be manipulated without intuition or interpretation. Given the nature of computers as possible formal symbol processors, AI turned this rationalist vision into a research program and took up the search for the primitives and formal rules that captured everyday knowledge (H. L. Dreyfus, 1992, p. xi).

Dreyfus expands a bit later:

> Rationalists such as Descartes and Leibniz thought of the mind as defined by its capacity to form representations of all domains of activity. These representations were taken to be theories of the domains in question, the idea being that representing the fixed, context-free features of a domain and the principles governing their interaction explains the domain's intelligibility. On this view all that we know—even our general know-how for getting around in the world and coping with things and people—must be mirrored in the mind in propositional form (H. L. Dreyfus, 1992, p. xvii).

Or, as he defines the outlook in another place,

> Cognitivism is not simply a theory of cognition but, as the name, cogniti*vism*, suggests, it is the strong view that all mental activity is cognitive – that perception, understanding, learning and action are all to be understood on the model of fact gathering, hypothesis formation, inference making and problem solving. Such a view is taken for granted by current cognitive psychologists, especially those seeking to program computers to behave intelligently (H. L. Dreyfus, 1988, p. 100).

But one need not take Dreyfus's word for either the assumption that AI researchers themselves see either their own project or the human intelligent behavior it ought to surpass on this view. Allen Newell and Herb Simon, the pioneers, assert this assumption themselves:

> The principal body of evidence for the symbol-system hypothesis [of mental functioning] . . . is negative evidence: the absence of specific competing hypotheses as to how intelligent activity might be accomplished whether by man or by machine. (Newell & Simon, 1985, p. 50).

The claims of AI workers from the 1950s onward that, because digital computers represent the rationalistic view *par excellence*, they would soon match and exceed the human capacity for intelligent behavior have gone unfulfilled, even today, even in the age of Google.[16] Examining attempts and failures at achieving intelligence on these assumptions in the area of NL (Natural Language understanding) and its use of "micro-worlds," which, I will

---

[16] Which is hardly to say that there have been no improvements. Google's driverless car, for example, seems like a triumph of AI in an area Dreyfus and those sympathetic to his view predicted would be impossible. But the driverless car still requires human input in certain situations.

contend, has directly to do with Education's scaling problem, helps one to see the limitations of this view.

According to Dreyfus, proceeding from the "symbol-system hypothesis" fails to account for – and thus has so far failed to replicate – certain aspects of what we freely recognize as intelligent behavior. Dreyfus terms this stumbling block the "commonsense knowledge problem":

> There are really at least three problems grouped under this rubric:
> 1. How everyday knowledge must be organized so that one can make inferences from it.
> 2. How skills or know-how can be represented as knowing-that.
> 3. How relevant knowledge can be brought to bear in particular situations (H. L. Dreyfus, 1992, p. xxviii).

In the first place, given the fact that the Dreyfus critique that I offer here and the AI projects involved are now more than 20 years old, let me assert that the NL problem remains unsolved to this day (despite the iPhone 4S's "Siri," the continual refinement of dictation software, and other such notable achievements), as Doug Lenat maintains in a more recent article (Lenat, 2008).

In fact, as Lenat represents what Dreyfus takes to be the lone holdout of the GOFAI approach in the present day, his own account of the failures of NL projects will provide a useful starting point, particularly because it handily gestures toward the way Lenat conceives of the problem in terms of the rationalist assumptions already discussed, and the way in which he notes a certain tendency to define success without apparent reference to "natural language understanding of the sort we need," i.e. natural language understanding of the sort humans demonstrate ordinarily:

> The problem of understanding natural language (NL) is another chicken-and-egg one. On the one hand, our real AI must be able to surf the web, read books, carry

on conversations, tutor and be tutorable, and so on—things that human beings primarily do through natural language. On the other hand, as we saw above (with numerous examples involving resolving polysemy, anaphora, ambiguous prepositional phrase attachment, noun-noun phrases, interpreting metaphor and analogy, recognizing irony and sarcasm, and so on) understanding the next sentence one reads or hears might require almost any piece (out of many many millions) of prior presumed knowledge.

One of the reasons this impediment is on my list is that AI researchers—NL researchers—have watered down the problem to the point where they "succeed" on every project, and yet natural language understanding of the sort we need is not here, nor is it, frankly, just around the corner yet. They talk about parsing an English sentence into logical form, for example, but then turn out to mean transforming it into something that structurally, syntactically fits the definition of "in logical form." As a caricature: add parentheses at the very start and end of the sentence, and declare success (Lenat, 2008, p. 18).

All of the features that, on Dreyfus's view, plague the Artificial Intelligence field in general are on display in Lenat's comment. The first point to observe about Lenat's assessment is the language in which Lenat characterizes the difficulties in NL: the grasping of a sentence's sense from among the multifarious possibilities, given the ways in which the given speech structure might occur in practice, "might require almost any piece (out of many many millions) of prior presumed knowledge." Lenat conceives, then, of the problem of understanding something like the contextualized situation in terms of mentally accessing pieces of prior presumed (propositional) knowledge. But the massive database of facts supposed to represent commonsense knowledge do not seem to result in an ability to flexibly and speedily arrive at something like correct understanding, which flies directly in the face of the way human beings seem to operate. This is a manifestation of Dreyfus's "relevance" problem:

Indeed, AI researchers have long recognized that the more a system knows about a particular state of affairs, the longer it takes to retrieve the relevant information, and this presents a general problem where scaling up is concerned. Conversely, the more a human being knows about a situation or individual, the easier it is to retrieve other relevant information (H. L. Dreyfus, 1992, p. xxi).

The difference between the way a symbol-system and a person respond to "knowing more about a situation or a person" suggests a difference in the ways in which each entity arrives at a determination of the *relevant* aspects of a given state of affairs. Additionally, however, Lenat mentions the temptation and proclivity to define-down "success" in the NL field, a temptation that he laudably resists. Lenat's characterization of the way in which "They talk about parsing an English sentence into logical form, but then turn out to mean transforming it into something that structurally, syntactically fits the definition of 'in logical form'" recalls the distinction I drew in Chapter 1 between an "ordinary concept" of good teaching and a "technical" one. In seeking to "parse" the English sentence into its "logical form," they operate on the assumption that English sentences, *as they occur in ordinary practice*, have and indeed rely upon such a built-in logical form for their intelligibility. On that assumption, they are merely *excavating* the logical form of the sentence. But the success "on every project," coupled with Lenat's assertion that NL remains far off, belies the assumption that such holistic NL success depends upon or can result from the "successful" logical parsing of sentences.

Lenat's charge that breaking down or "parsing" an English sentence into logical form – such that it becomes functional according to the "symbol-system hypothesis" – somehow distorts ordinary language suggests that when we use language with one another in ordinary, native-speaker situations, we do something different than mentally applying logical rules to symbolic representations. The problem, as he and Dreyfus both note, has to do with the task of *scaling up* – of generalizing a particular success in NL to the domain of language as a whole. Dreyfus predicts that such scaling is impossible precisely because of his sense that *in*

general competent speakers of a language do not use language according to this system of rules and symbols.

Lenat, in another place, offers an alternative understanding of the problem, one that continues to drive his work today:

> Naturally, all programs are built on some primitives (predicates, frames, slots, rules, functions, scripts). But if you choose task-specific primitives, you'll win in the short run (building a program for that narrow domain) but lose in the long run (you'll find yourself painted into a corner when you try to scale the program up). (Lenat & Guha, 1990, p. 15)

On Lenat's view, the problems in scaling-up have to do with the use of task-specific primitives (like Shulman's "pedagogical content knowledge," one might note); Lenat's project is to develop symbolic representations of the *general* primitives underlying language use, while Dreyfus takes issue with the very assumption of the existence of such primitives. Another way of putting the distinction is that Lenat holds the requisite commonsense knowledge to be *in fact* unformalized, as yet. Dreyfus holds that the requisite commonsense knowledge is *in principle* non-formalizable. Dreyfus's response to Lenat and Feigenbaum's discussion of codifying human input in order to solve the relevancy project bears quotation in this regard. Lenat and Feigenbaum suggest that they, as researchers, ought to

> wait until our programs are finding many, far-flung analogies, but inefficiently, i.e. only through large searches. Then investigate what additional knowledge *people* bring to bear, to eliminate large parts of the search space in those cases. Codify the knowledge so extracted, and add it to the system (Lenat & Feigenbaum, 1991, p. 221).

Dreyfus simply points out that "The assumption that people *are* storing context-free facts and using meta-rules to cut down the search space is precisely the dubious rationalist assumption in question. . . .

> In the end, Lenat's faith that Cyc will succeed is based neither on arguments nor on actual successes but on the untested traditional assumption that human beings

have a vast library of commonsense knowledge and somehow solve the scaling-up problem by applying further knowledge (H. L. Dreyfus, 1992, p. xxii).

Notably, Dreyfus originally penned this critique of the Cyc project in 1992; in Lenat's 2008 paper he continues to claim progress in assembling this massive database of commonsense knowledge, as he has been doing since the early 1980s, but the resulting programs and database still seem unable to mimic the alacrity or flexibility (what Dreyfus calls an ability to "zero-in" on the relevant aspects of a situation) of a situated human being (H. L. Dreyfus, 1992, p. 30).

The disjuncture between the digital processor's many superiorities to human abilities – most notably, as Dreyfus concedes, "their rigor, reliability and indefatigability," which allows "computers used as logic machines to do extremely well what human beings do only poorly" (H. L. Dreyfus & Dreyfus, 1986, p. 63) – and its inability to achieve human-like general intelligence or skillful understanding in terms of NL (as well as many other domains not considered here) suggests to Dreyfus that what we do when we develop and apply such skill or intelligence ourselves is not exhausted according to the tenets of this rationalist assumption. The problems encountered with scaling-up successes in skill domains such as NL seem to imply that, for us, something other than combining facts and rules is involved as we become generally proficient in various skill domains.

To close this illustration of AI's progress at developing skillful or intelligent behavior according to the rationalist assumption, and its subsequent shipwreck, Dreyfus's analysis of the SHRDLU program's use of micro-worlds from the 1970s will provide a helpful segue back to the ways in which similar assumptions about skillful teaching practice permeate the

policy, academic, and public discourse on teaching and learning in the age of Race to the Top.

SHRDLU, initiated by Terry Winograd, was a 1971 program in NL, which researchers at the time referred to as "a major advance" in that it employed the concept of a micro-world. Micro-worlds, as the name implies, are strictly limited domains, based upon "primitives" of the kind Lenat mentioned earlier. Dreyfus describes the micro-world like so:

> SHRDLU simulates a robot arm which can move a set of variously shaped blocks and allows a person to engage in a dialogue with the computer, asking questions, making statements, issuing commands, about this simple world of movable blocks. The program is an integrated system which makes use of syntax, semantics, and facts about blocks (H. L. Dreyfus, 1992, p. 5).

And it worked, that is, it was able to carry out commands, identify objects, and report progress within this world. AI researchers saw this as a major improvement upon earlier attempts (like the famous ELIZA program) because, using the concept of primitives to derive success in a micro-world, it seemed a comparatively small (and merely technical) step to code the necessary primitives to all micro-worlds, and thus, to the whole. (This coding of all primitives, as a side note, continues to characterize Lenat's work to this day, some 28 years after he began the project; he has yet to reach the whole).[17] Dreyfus's analysis of SHRDLU's success, his description of the assumptions on which its ambitious scaling-up proceeded, and his explanation of the inevitability of its eventual failure will lead us back to thinking about Duncan, Lemov, and Shulman. I present three paragraphs culled from within Dreyfus's longer, original analysis:

---

[17] One might also note, somewhat grimly, that something like this – formalizing the outer horizon – is what Husserl called, with no discernable hyperbole, an "infinite task."

What characterizes the period of the early seventies, and makes SHRDLU seem an advance toward general intelligence, is the very concept of a micro-world—a domain which can be analyzed in isolation. This concept implies that although each area of discourse seems to open out into the rest of human activities its endless ramifications are only apparent and will soon converge on a self-contained set of facts and relations. (H. L. Dreyfus, 1992, p. 10)

It is true that physical theories about the universe can be built up by studying relatively simple and isolated systems and then making the model gradually more complex and integrating it with other domains of phenomena. This is possible because all the phenomena are presumably the result of the lawlike relations of a set of basic elements, what Papert and Minsky call "structural primitives." (H. L. Dreyfus, 1992, p. 12)

In our everyday life we are, indeed, involved in various "sub-worlds" such as the world of the theater, of business, or of mathematics, but each of these is a "mode" of our shared everyday world. That is, sub-worlds are not related like isolable physical systems to larger systems they *compose*; rather they are local elaborations of a whole which they *presuppose*. If micro-worlds *were* sub-worlds one would not have to extend and combine them to reach the everyday world, because the everyday world would have to be included already (H. L. Dreyfus, 1992, p. 13).

Dreyfus's final paragraph also hints in the direction of my next section, but of

particular note with regard to Shulman, Lemov, and Duncan's thinking is Dreyfus's assertion

that "sub-worlds are not related like isolable physical systems they *compose*; rather they are

local elaborations of a whole which they *presuppose*." In other words, although severely

restricting a given domain such that proceeding by means of given facts and rules may afford

a passable simulacrum of skilled behavior, there is no reason to suppose that it will,

ultimately, prove amenable to being "taken to scale," as Duncan has it, by the same facts-

and-rules processing on which the initial success was predicated. It was, after all, in the very

attempt to compose a larger whole from initial and promising successes in AI, by means of

the same symbol-system hypothesis involved in the restricted domain, at which each of the

successive successes in AI failed to generalize. As Dreyfus has it, this failure is attributable

to the fact that what we call intelligent or skillful human behavior in a given domain is *not* a

composite of smaller (unintelligent?) behaviors governed by rules or maxims; rather, it is a kind of *local elaboration* of a *holistic* sense of a given skill or intelligence, one intimately involved with one's sense of what it is to be a person (what Dreyfus – and Heidegger – elsewhere call an "understanding of being").

The latter claim will await its bearing-out. For now, to conclude this section, the experience of the field of AI research, based as it is upon the rationalist assumptions also underpinning Duncan, Lemov, and Shulman's working theories of teaching practice, provides no evidence at all that the teaching practice that Lemov and Shulman study, the "good teaching" that justifies Duncan's Race to the Top, will be replicable according to this rationalist assumption upon which their work and policies proceed. I remain just as circumspect in the force of my claim as Dreyfus, but not everyone foregoes such sanguinity with regard to the rationalist assumption as a whole:

> When AI workers were trying hard in order to produce a machine that can think like human minds, they have in a way been testing—and testing it up to the last point—the rationalist assumption that the workings of the human mind depend on logical rules. Result: No computers actually function like the human mind. Reason: the human mind does not depend on the formal or logical rules ascribed to computers. Thus, symbolic AI research has falsified the rationalist assumption that 'the human mind reaches certainty by functioning formally' by virtue of its failure to create a thinking machine (Kenaw, 2008, p. 227).

<div align="center">

2.3.2 AI's Scaling Problem and the Development of

Good Teachers

</div>

The analogy I have suggested here between the pursuit of Artificial Intelligence along the lines suggested by what Dreyfus calls the "rationalist assumption" and the political and academic view of teaching ability is apt and germane in a few different ways.

In the first place, to take the case of an individual teacher's practice, it explains the predictive success of Ball and Hill's "Mathematical Knowledge for Teachers" measure: 11 percent of the variance in achievement scores (H. C. Hill, et al., 2005), more than any other teacher characteristic, can be attributed to narrowing the teaching skill domain to merely (a) teaching of mathematics (a single subject area) as measured by (b) standardized achievement tests (i.e. according to no "inspiring" or "instilling a thirst for knowledge" criteria). Even though any claim to *causation* between the measures of "MKT" and student achievement results remains merely implicit in Ball and Hill's work, I might suggest that attempting to influence student math scores by means of developing individual teachers' "MKT" could indeed prove fruitful. As we have seen, understanding skillful behavior according to rationalist postulates has led to relatively successful replication under such particular and severe domain constraints. This does not imply, importantly, that "MKT" or any other subject-specific content knowledge is somehow either necessary or sufficient to good teaching in its ordinary sense. Indeed, the fact that "MKT" only accounts for 11 percent of the variance in test data *even* under these ideal conditions points to its limited utility.

In the second place, to speak of problems in scaling-up local success in a given school building to, for example, the school district, or even more generally beyond that, Dreyfus's account of the failure of facts-and-rules approaches to achieve *generally* successful intelligent practice can provide an alternate explanation to Healy and DeStefano's observed discrepancy between the fact that "good educational practice can be found almost *anywhere"* and the persistent, obstinate fact that it "cannot be found *everywhere."* Tellingly, as a reminder, Healy and DeStefano regard the failure to generalize as a communication problem – "usage of available knowledge, and the rate of spread of effective practices, is grievously

low" (Healy & DeStefano, 1997). It is simply assumed that *knowledge* of "effective practices" will work in one place (in one grade, in one subject) as well as in another. Alternatively, AI's failures in the realm of NL draw the assumption about the value both of locally obtained "primitives" to a general application, and also, due to Dreyfus's relevance problem, of generally-applicable primitives in any *given* local situation, into serious question. When Duncan frames the problem as merely one of teachers needing to "share best practices," he, too, trades on an assumption that has demonstrably failed another field.

To focus once more on the notion of developing a particular teacher's practice according to a "best practices" model, whether this is understood according to Lemov's or to Shulman's framework, in light of the foregoing discussion, I wish to point out once again that both Lemov and Shulman, as well as Duncan, rely on an *atomizable* view of skillful practice itself, such that linking pieces of practice (either knowledge or technique) together adds up to the skillful whole. Like the sub-worlds in Dreyfus's earlier analysis, I will argue later that no such holism results from such an attempt, and that the attempt itself, if taken seriously, actively stifles one's ability to *presuppose* the whole.

Lemov revealingly – and accurately – draws an analogy between the teaching world and the art world in order to situate his emphasis on the "foundational skills" he associates with good teaching:

> Great art relies on the mastery and application of foundational skills, learned individually through diligent study. You learn to strike a chisel with a mallet. You refine the skill with time, learning at what angle to strike and how hard to drive the chisel. Years later, when and if your work makes it to a museum, observers will likely talk about what school of thought or theory it represents. They are far less likely to reflect on the degree to which proficiency with the chisel made the vision possible. But although lots of people conjure unique artistic visions, only those with an artisan's skill can make them real. Every artist is an artisan. And while not everyone

who learns to drive a chisel will create a *David*, neither can anyone who fails to learn it do much more than make marks on rocks (Lemov, 2010, p. 1).

The analogy Lemov draws is apt; his inference is not. Lemov makes the argument that foundational skill with a chisel is a necessary but not sufficient condition for Michaelangelo's *David*. He is surely correct, broadly speaking. Also, presumably, a foundational skill with a brush would be similarly necessary (but not sufficient) to the painting of *Water Lilies*. But would one say that this "skill with a brush" is the same as the "skill with a brush" necessary (but not sufficient) to the painting of *Guernica*? What does one do with Jackson Pollock's skill? And are not these all masters and masterpieces?

Lemov's approach takes for granted that all skillful teaching is identical at a certain level – analogous, ironically, to the presumably identical *artistic-ness* of "*unique* artistic visions (emphasis added)" – and then concludes that "mak[ing] them real" requires an "artisan's skill." Were this the case, and were "artistic vision" not in fact unique at all, certainly the *same* skills would be requisite for working in the same medium. In fact, this is precisely what Lemov assumes:

> I had always thought Picasso was a king of abstraction, of a symbolism that made the ability to draw accurately and realistically irrelevant. His sketches, filling the margins of the pages, bore witness to his mastery of fundamentals and a habitual need to refine his skills. Even in the stray moments of his schooling, he was constantly honing the building blocks of his technique (Lemov, 2010, pp. 1-2).

Despite the fact that "the building blocks of his technique" are nowhere in evidence *themselves* in the masterpieces for which Picasso is known – those that "made the ability to draw accurately and realistically irrelevant," – Lemov assumes that his margin-note sketches *remain present* in the masterpieces as necessary conditions, even if, somehow, beneath perceptibility or consciousness. We will return to this assumption in some detail in the

following section. Irrespective of the appearance of the masterpieces themselves, Lemov assumes that all artistic masters possess the same skills.

But the particular skills brought to bear in *particular* masterpieces, such as the ones cited above, remain manifestly different from one another; in fact, the only thing that makes the techniques *skills* in this sense is the "unique artistic vision" entailed in the finished product itself. Only if the "vision" of art were reduced to some sort of explicit definition toward which each artist separately labors would Lemov's vision of the foundational role of isolated and abstract skills hold.

Here, once more, we encounter the value and the danger of the "technical" sense of learning in the direction of which the technical sense of "good teaching" aims. The value of a "technical" sense of teaching and learning lies in its ability to *establish* the identity of excellence such that all skill-development proceeds in just this direction. The danger lies here also: precisely this direction has been officially (and ordinarily) understood as *insufficient*, as Duncan's many insistences about the inability to "reduce the complex, nuanced work of teaching" to precisely this means of its expression confirm and reiterate. While it would seem a simple matter to call for a *better* definition, one that would be sufficient, the worry shared by Dreyfus and others that such definition is in principle non-formalizable will merit discussion in the next section. For the time being, it is enough to point out an analogy with Lenat's view of the NL field's failure to provide the *general* skills it seeks. His claim, recall, is that the field has "watered down the problem to the point where they 'succeed' on every project," which has not solved the larger problem the methods were meant to address: "the natural language understanding that we need is not here yet, nor, frankly, is it just around the corner." Whatever teacher-preparation success is achievable

through the use of Lemov's technique-extraction methodology, I contend, is achievable because the "relevance problem" has been precluded a priori: the explicit definition of "student learning" as the only relevant feature guiding the discovery and application of teacher skill restricts the domain of teaching and learning to the extent necessary for a facts-and-rules approach to succeed in some passable form.

The notion that by having achieved success in a reduced domain, we have taken a considerable step toward achieving success in the larger domain is rendered dubious by the AI experience. A facts-and-rules approach that functions in a limited domain does not necessarily or easily provide the necessary or sufficient conditions for success in the domain as a whole, the ordinary sense of good teaching. In other words, while there is nothing at all wrong with the facts-and-rules approach in its limited application, something goes very wrong indeed when the limited domain is confused with the general whole of ordinary teaching, which, as Chapter 1 has demonstrated, it regularly is.

To close this line of thinking by returning to Lemov's analogy with art, I would like to suggest that his considerations of the "foundational skill" of chiseling, the physical familiarity with the practice, is indeed requisite to the production of the *David*. But working backward from the *David* to extract the chiseling skills involved will only work to reproduce the practice insofar as the *David* absolutely defines the desired outcome. The *same* skills are not required – nor would they be particularly useful – to Rembrandt or Monet, who also produce *great art*. If one so desired, one could pull back and attempt to locate the piecemeal physical skills that transcend these particular media, the skills that Michaelangelo shares with Rembrandt and Monet. But then, perhaps, one factors in ceramic artists, and must account for the physical skills of *that* medium as well, all under the rubric of necessary skills for the

production of great art. At this point, or perhaps earlier, one realizes that whatever physical skills happen to transcend these media, and are in fact shared among all the artists involved, can account for very little indeed of the individual finished masterworks. Insofar as the reproduction of the transcendent elements of excellent teacher practice represents Lemov's goal, there is no reason to suppose that such an approach by itself will be able to reproduce anything resembling the practice of any one of the particular teachers who are taken to apply the rules or principles or techniques.

The recognition of just such a weakness in a "best-practices" approach inspires Shulman's project in teacher development, where no such transcendence of *all* fields of content is sought, the notion being that, to continue Lemov's analogy, the skills necessary to sculpture are different than those necessary to painting, and so, in specifying the distinction, we can focus on the individual skills necessary to each *field*. This is the sense in which Shulman improves upon Lemov's model. By narrowing the domain further – not simply from "ordinary" great teaching to "technical" great teaching, but now also to "technically" great *math* teaching, for example – one improves the functioning of the facts-and-rules approach. But one need not stop there. By narrowing the domain further and further, the facts-and-rules approach will work progressively better. Eventually and in the extreme, one narrows the domain to a single, individual student (or the 30 such individual domains that comprise a classroom). That would certainly provide optimum functioning of the facts-and-rules approach. But then, if one is only following a such a procedure in a *single* instance that is applicable *only* in that instance, in what sense is one following a rule at all rather than coping with one (or 30) exception(s)?

For every narrowing of the domain, which allows for progressively greater success, one reduces the generalizability of one's techniques. Conversely, attempting to find (and inculcate) more and more general facts and rules progressively reduces their *applicability* in any *particular* setting. The freezing of relevance through the defining of good teaching in its technical sense helps with this problem to a certain degree, but it also and simultaneously precludes any serious attempt on the part of colleges of education or other teacher-training programs to develop "ordinary-sense" great teachers. The remainder of Chapter 2 provides an account of why this is so and offers a formative view of skill development that would address the flaws exposed here.

In closing the entire section, it is worth mentioning again – so that it may be heard in light of the foregoing discussion – Duncan's assuredness about the role of data in teacher development:

> In the end, I don't think the ingredients of a good teacher preparation are much of a mystery anymore. Our best programs are coherent, up-to-date, research-based, and provide students with subject mastery. . . . [T]hese programs have a shared vision of what constitutes good teaching and best practices—including a single-minded focus on improving student learning and using data to inform instruction (Duncan, 2009d).

The "ingredients" of a good teacher preparation certainly are not mysterious any longer, but the solving of the mystery, my discussion suggests, has much more to do with "water[ing] down the problem" in order to make it solvable by a specific, rational procedure than the discovery of a revolutionary approach to understanding teaching and learning. That "what constitutes good teaching and best practices" is available in propositional form *at all* is a consequence of the narrowed domain of teaching and learning achieved by defining "teaching" in terms of its effect on "student learning," and "student learning" in terms of

achievement score growth, none of which has anything explicitly to do with "inspiring" students or "instilling a thirst for knowledge," or any of the myriad other skills that comprise the "ordinary" sense of good teaching.

In turning to address the rationalist assumption head-on, I am primarily following a(n) (admittedly unsubstantiated) suspicion that part of what motivates the conceptual slippage between the ordinary and technical senses of good teaching so evident in Chapter 1 on the part of researchers and policy-makers alike is a generally-held sense that some "specific, rational procedure" for solving the "mystery" of great teaching – both in its evaluation and in its development – is the only genuine and acceptable option available, that any proposed solution to the mystery that *did not* assume a facts-and-rules structure to skillful human behavior would be therefore mere superstition (to say nothing of their probable response to a suggestion that no solution is required at all). But I will follow up on this suspicion explicitly in Chapter 3.

In critiquing the rationalist assumption itself, a critique that AI's many failures have demonstrably warranted, I hope to show that transcendent facts and rules, however general or limited the domain, while they can be *sufficient* to the reproduction of practical competence are neither necessary *nor* sufficient to the reproduction or regularity of our shared practices, including and especially the excellent practices of great teachers, and that developing great teachers from novices will require relying on some other means of accounting for this regularity.

2.4 The Rationalist Assumption and Following Rules

Lemov, Shulman, and Duncan all assume that teaching skillfully, teaching well, is a matter of following the correct rules, whether these are purely behavioral or more overtly

mental in nature. Duncan's rhetorical recourse to the growing wealth of student achievement data as the source of "what works," abetted by Kane's characterization of his own pursuits as "identifying the practices associated with student achievement," indicate that what "following a rule" amounts to here is the expression of a certain behavior given the knowledge of certain situations and the correct heuristics. In other words, teachers – skillful and novice alike – essentially conduct their practice by means of information processing: after recognizing a certain situation on the basis of some formalizable knowledge, including and especially the input of student achievement data, the teacher in question will render output in the form of a particular behavior or technique.

When not conceived on a simple mechanical input-output model such as this, or else on a purely behavioral model like Lemov's, in which the techniques' transcendent qualities obviate the relevance of situational particularity as such, the teacher's practice remains a matter of processing information. Shulman, once more, takes the task of a researcher on teacher practice to be the cataloguing the particular *knowledge* on the basis of which teachers are presumed to act. He is explicit about the reflective processing role the mind plays in working from such articulable knowledge and judgment, portraying skillful teachers as "refer[ring] to the full range of practical arguments. . . as they reason about and ultimately make judgments and decisions about situations they confront and actions they must take" (L. S. Shulman, 2007). In all of these cases, action-taking is the direct result of conscious reasoning, with or without the aid of student achievement data, and the process so described constitutes the conception of following rules at work in the educational discourse in general.

The two specific examples of *recognized* excellent teaching that have appeared in my study so far, however, do not seem to support a view wherein excellent teaching proceeds

according to such an information-processing model. The output of Mrs. Grady's decision to allow Olly Neal to steal a book is, after all, presumably absent from any list of "best practices" one could reason one's way toward, given a certain situation. Additionally, attempts to engineer the kind of "flexible responsiveness" of Shulman's "Nancy," which he takes to be a result of the mental processing of transcendent knowledge, by means of machines that *perfectly* process transcendent knowledge have consistently failed to replicate such flexibility. All of this suggests that what skillful practitioners – and teachers among them – do when acting skillfully in a given domain does not in fact proceed according to the rationalist assumption as outlined here and earlier. Skillful human beings do not appear to follow rules the way that "symbol-systems" follow rules, which is to say, by the means of mental information processing.

Nevertheless, a certain regularity appears in the way that we repeat the character of given actions (utter certain phrases, give appropriate gifts, greet friends or acquaintances) in a variety of familiar situations, and I would like to claim say that these iterations of regular behavior in recognizably similar situations amounts to following rules. I wish to suggest that, *pace* the rationalist assumption, "following a rule" is *itself* a practice rather than a primitive, in no sense an abstract engagement of mind and information connected to a telos of action. Inadequate attention to the distinction between these two views of what following a rule amounts to has a great deal to do, as I will demonstrate later, with the consistent conflation of the "technical" and "ordinary" senses of good teaching and more precisely with our seeming inability to reach the "ordinary" in either evaluation or development by means of addressing the "technical." Ultimately, I hope to argue that the "ordinary" is well within our grasp, but that grasping it requires the relinquishing of the rationalist view of rule-

following. In that sense, what I offer is a form of dissolving – rather than solving – the

problems that Race to the Top seeks to address.

2.4.1 The Chicken and the Egg

The AI failures to reproduce intelligent human behavior have done serious work

toward disconfirming the rationalist view of skillful behavior in general – or, more strongly,

as Kenaw claims, these failures have positively falsified the rationalist view – and returning

once again to Lenat's explanation of what he calls the "chicken-and-egg" problem involved

in AI's attempts to replicate skillful responsiveness sheds light on the specific limitations of

the rationalist assumption. To present Lenat's description of the problem once more,

> The problem of understanding natural language (NL) is another chicken-and-egg one. On the one hand, our real AI must be able to surf the web, read books, carry on conversations, tutor and be tutorable, and so on—things that human beings primarily do through natural language. On the other hand, as we saw above (with numerous examples involving resolving polysemy, anaphora, ambiguous prepositional phrase attachment, noun-noun phrases, interpreting metaphor and analogy, recognizing irony and sarcasm, and so on) understanding the next sentence one reads or hears might require almost any piece (out of many many millions) of prior presumed knowledge (Lenat, 2008).

Lenat here points to the problem of engineering a program, running on the rationalist

assumption, to *learn independently*, to add to its own "understanding." The chicken-and-egg

nature of this issue has to do with an inability to know *what* is being learning, in other words

what the particular input words are instances *of*. To *what* are they similar? As long as one

specifies a finite set of potential relevancies beforehand, such machine learning is possible;

but that also means that, unlike humans, the machine in question cannot learn anything

without another person manually (so to speak) delimiting the range of possibly relevant

features to consider.

Dreyfus puts it in similar terms:

In the end the very idea of a holistic information processing model in which the relevance of the facts depends on the context may involve a contradiction. To recognize any context one must have already selected from the indefinite number of possibly discriminable features the possibly relevant ones, but such a selection can be made only after the context has already been recognized as similar to an already analyzed one (H. L. Dreyfus, 1992, p. 54).

That such an eventuality might stem from the requirements of the rationalist

assumption itself, namely, in this case, the requirement for a sort of transcendent structure

(made of knowledge or features or rules) upon which things like situation-specific relevance

and similarity can be made to depend, was the major project of a number of 20th century

philosophers to point out. As Simon Glendinning puts it, in an account of Derrida's

philosophy of language,

According to Derrida, the problem is that, once one embarks on an analysis of language based on the radical distinction between linguistic systems and linguistic events, the conceptual resources at one's disposal cannot account for the possibility of the very phenomenon it is claiming to delimit. This is because the impossible condition of both the historical priority and the logical posteriority of speech ruins any attempt to provide an internally consistent account of the conditions of possibility for the phenomenon of language (Glendinning, 1998, p. 113).

The phenomenon of language, here, as with the phenomenon of learning *via* language

that Lenat discusses, demonstrably defies an accounting in terms of transcendent rules or

primitives on the one hand and particular situations on the other. It is worth pointing out

once again that when Shulman asks after what Nancy "believe[s], understand[s], and know[s]

how to do," he is in search of *precisely* the transcendent structure upon which the particular

practice of her teaching is *assumed* to draw. The chicken-and-egg problem that Lenat

encounters in engineering human-like learning and that Derrida discovers in traditional

accounts of language suggest that what "permit[s] Nancy to teach as she does" will not

appear in a form Shulman would recognize as a potential answer to his questions.

2.4.2 Wittgenstein and Heidegger on the Practice of
Following Rules

AI's experience suggests that the rationalist assumption of the functioning of our

intelligent or skillful behavior cannot reproduce that sort of excellence; Derrida's project,

among others, claims that neither can the structure implied by the rationalist assumption

explain our ability to perform skillfully or intelligently in practical situations. Two important

phenomena, specifically, that accounts following the rationalist assumption can neither

address nor replicate are (a) a emergence of something *new* at all, as in a novel response to a

given set of circumstances, and (b) a shared or recognized sense of the correctness or

appropriateness of this *new* response in particular. Mrs. Grady's allowing of Olly Neal to

steal a book provides an example of both phenomena: that the action itself resists codification

as a best practice due to its *specific* appropriateness – which also implies that it was not

derivable from any such abstractly codified list of acceptable options – neither prevents it

from actually emerging, *nor*, importantly, from Kristof's using it as an exemplar in a *New*

York Times editorial, which use acknowledges some *shared* sense of the action's

appropriateness. What "shared" means here bears little resemblance to what Duncan means

when he calls on teachers to "share best practices." The "sense" to be shared, I will argue, is

not amenable to codification at all.

Alternative means of accounting for and developing skillful behavior predicate the

regularities inherent in our practices not upon *knowledge* of abstract or transcendent rules or

features, but upon what Glendinning calls a "*structural" "relation to others"* that social

practices "in the deed" *are* (Glendinning, 1998, p. 106). Following Descartes, the rationalist

view of understanding opens with an *isolated,* disembodied consciousness, from which

perspective any recognition of filiation or similarity between self and others (as between any two things) must be *produced or deduced* by means of something like the identification of certain features in common.

The philosophers I appeal to in this section hold that this view is precisely backwards; they claim, broadly speaking, that the rationalist model of behavior both functions only to a certain degree and eventually breaks down precisely because such a view *insists* upon the radically isolated nature of a disembodied consciousness as a starting point – something we, at moments, *can approximate* but that we most basically *are not* – rather than an embodied being embedded in a network of social behaviors and practices. That we can also process information from a detached, objective perspective like a computer can (and must) does not imply to these thinkers that such information-processing constitutes the most *essential* fact about us, nor is it necessarily a model of perfection to which we ought to aspire. Wittgenstein's considerations of what happens when we follow a rule illustrates the shortcomings of the rationalist view from an alternative perspective and suggests a different way of seeing ourselves and our regular practical behaviors.

In the *Philosophical Investigations*, Wittgenstein offers a certain view of the problem of teaching someone the meaning of a word, or how to continue a series, both of which are instances of following rules.

> How do I explain the meaning of "regular", "uniform", "same" to anyone? – I'll explain these words to someone who, say, speaks only French by means of the corresponding French words. But if a person has not yet got the *concepts*, I'll teach him to use the words by means of *examples* and *exercises*. – And when I do this, I do not communicate less to him than I know myself. . . . I do it, he does it after me; and I influence him by expressions of agreement, rejection, expectation, encouragement. . . . Imagine witnessing such teaching. None of the words would be explained by means of itself; there would be no logical circle. . . . Teaching which is not meant to apply to

anything but the examples given is different from that which *'points beyond'* them. (Wittgenstein, 2009, p. 89 (§208))

The notion that in teaching someone to use words by means of "examples and exercises," which Wittgenstein suggests communicates no "less to him than I know myself," points to the *use* of words as the bottom-level locus of a given rule. The rationalist view would locate the rule somewhere outside of a given practice; the rule is, on that view, the necessary and transcendent condition that makes any instance of the practice possible. But the fact that the teacher and the learner *alike* in Wittgenstein's example have nothing more to go on than the practice – that in teaching by exercises and examples one communicates the *whole* of one's knowledge of that practice – has the effect of embedding that knowledge inseparably in the activities of the practice itself, which runs directly contrary to the rationalist view.

Wittgenstein's voice of temptation[18] in fact presses him upon this point in some ensuing sections.

> "But do you really explain to the other person what you yourself understand? Don't you leave it to him to *guess* the essential thing? You give him examples – but he has to guess their drift, to guess your intention." – Every explanation which I can give myself I give to him too. – "He guesses what I mean" would amount to: "various interpretations of my explanation come to his mind, and he picks one of them". So in this case he could ask; and I could and would answer him.
> "No matter how you instruct him in continuing the ornamental pattern, how can he *know* how he is to continue it by himself?" – Well, how do *I* know? – If that means, "Have I reasons?", the answer is: my reasons will soon give out. And then I shall act, without reasons. (Wittgenstein, 2009, pp. 90 (§210-211))

---

[18] The *Investigations* is constructed (nontraditionally) as a sort of dialogue among a number of voices – the voice of temptation often offers questions and comments that one might imagine more traditional philosophers to raise, which gives the "voice of correctness" an opportunity to work things out or pose new problems.

Wittgenstein's voice of temptation suggests that there *must* be an "essential thing" linking together the various examples, a "drift" that characterizes the teacher's "intention." The voice of correctness, however, proceeds to give the voice of temptation what it wants: a scenario in which "every explanation I can give myself I give to him too," which forces both voices to recognize that in order for the ensuing picture to make sense, one would require a model of "picking" among "various interpretations" of one's *own* explanation in order to *have* an "intention" at all. That is the point in Wittgenstein's insistence that his own "knowing" how to continue a pattern is no different in character from that of the pupil instructed by means of examples and exercises. Such a suggestion implies that the type of knowing at issue here is not expressible or learnable or in fact *had* in terms of abstract facts and rules. Wittgenstein is fairly explicit on this point in sections 217 to 219.

> "How am I able to follow a rule?'—if this is not a question about causes, then it is about the justification for my acting in this way in complying with the rule. Once I have exhausted the justifications, I have reached bedrock, and my spade is turned. Then I am inclined to say: 'This is simply what I do.' (Remember that we sometimes demand explanations for the sake not of their content, but of their form. Our requirement is an architectural one; the explanation a kind of sham corbel that supports nothing.)
> Whence the idea that the beginning of a series is a visible section of rails laid to infinity? Well, we might imagine rails instead of a rule. And infinitely long rails correspond to the unlimited application of a rule.
> "All the steps are really already taken" means: I no longer have any choice. The rule, once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space. – But if something of this sort really were the case, how would it help me?
> No; my description made sense only if it was understood symbolically. – I should say: *This is how it strikes me.*
> When I follow the rule, I do not choose. I follow the rule *blindly*.
> (Wittgenstein, 2009, pp. 91-92 (§217-219))

Wittgenstein's use of the "rules-as-rails" picture highlights the ease with which the rationalist view can be confused with one of its alternatives. When Wittgenstein says that when he follows a rule, he does not choose – he follows it blindly – the formulation would

work for either the rules-as-rails rationalist view or for his own. In each formulation, after all, no *choice* is involved. But in the first case, where rules are analogous to rails laid to infinity, the person "no longer [*has*] any choice" (emphasis added); in the second case, however, Wittgenstein simply avers that "when I follow the rule, I *do not* choose" (emphasis added). I follow the rule *blindly*.

In both cases, "following a rule" in ordinary circumstances involves not choosing. In only one case, however, the rules-as-rails view, is such not-choosing *logically* necessary. On that view, one does not choose because *no alternative possibility exists*. On the other view, however, no such logical impossibility arises; the fact remains, however, that in order for it to be an instance of "following a rule" it must be, as Wittgenstein calls it elsewhere, "*selbstverständlich,*" a matter of course. The distinction, subtle though it may be, has to do with the difference between no alternative *existing* and no alternative *appearing for consideration.* "I no longer have any choice" is not the same as "I do not choose. I follow the rule *blindly*." The former view, the rationalist one, makes every correct application of a rule *logically necessary*; the latter view agrees with the regularity of the correctness and with the obviousness of the choice, but not with its causal or logical *necessity* as such – not with the requirement of something transcendent that *guarantees* by means of causation or logic the correct following of a rule. One *can* choose a course of action, on this view; that one *does not* is precisely what makes it an instance of following a rule as a matter of course.

Dreyfus also points to the insufficiency of logical or causal necessity to account for the regularity of correctly following rules, particularly in this example, rules for correct *generalization* (for continuing a series of a certain kind, just as Wittgenstein discussed):

All neural network modelers agree that for a network to be intelligent it must be able to generalize; that is, given sufficient examples of inputs associated with one particular output, it should associate further inputs of the same type with that same output. The question arises, however: What counts as the same type? The designer of the network has a specific definition in mind of the type of required for a reasonable generalization, and counts the network successful if it generalizes to other instances of this type. But when the network produces an unexpected association, can one say that it has failed to generalize? One could equally well say that the network has all along been acting on a different definition of the type in question and that the difference has just been revealed.[19] (All the "continue this sequence" questions found on intelligence tests really have more than one possible answer, but most humans share a sense of what is simple and reasonable and therefore acceptable) (H. L. Dreyfus & Dreyfus, 1986, p. xiii).

What "sharing a sense" comes to is not predicated upon a sort of logical necessity, of which any "symbolic expression," as Wittgenstein calls it in §220, remains merely a "mythological description."

Answering the question of "what counts as the same type" depends upon what Wittgenstein calls agreement in "forms of life," which is similar to what Heidegger terms "common being-in-the-world." The rationalist view holds that any particular social or cultural practices – any particular forms of life – are caused, as it were, by *decisions* to commonly do such-and-such; Hobbesian social-contract theory might represent the apotheosis of this view. The Heideggerian-Wittgensteinian position holds, in contrast, that

---

[19] Dreyfus doubtless has in mind here, as he describes elsewhere, a particular military AI project that attempted to have computers develop the ability to identify enemy tanks in forests, given the input of many, many satellite images of forests and tanks and tanks-in-forests. And the program got better and better at picking out tanks in *new* satellite images of forests, images it had never seen before, which suggested to the researchers that the computer was in fact *generalizing* from the earlier images to be able to spot enemy tanks on command. Then, with certain images, the computer seemed to lose the ability entirely, for reasons no one could fathom. There were clearly tanks in the image, and the images were similar to others that the computer had already seen, but the computer suddenly could not identify the tanks. At length, someone noticed that one set of images was collected on a *sunny* day and one was collected on a *cloudy* day; the computer had been extrapolating from *shadow patterns*, which proved insufficient to the continuation of the series. What was missing, on Dreyfus's understanding, was something essentially informative about the nature of tanks, a sense of *relevance* to what was being sought.

any apparent decisions take such being-in-the-world or "forms of life" as their very condition of possibility. Rationality creates the conditions for practices on the former view; forms of life or background practices make up the conditions or do the stage-setting for what counts as rationality, on the other.

To raise a persistent issue once more: Appealing to something like background practices of forms of life as the pre-theoretical or pre-rational conditions necessary for an adequate investigation into "what works" in the field of teaching and learning will seem, to some, insufficient on its face. As John McDowell notes, "If one is wedded to the picture of rules as rails, one will be inclined to think that to reject it is to suggest that, say, in mathematics, anything goes: that we are free to make it up as we go along." (McDowell, 1981, pp. 150-151). The rejection of a view of skillful practice based on logical or causal necessity will imply to some a rejection of all normativity or regularity altogether. In short, such an approach seems unscientific.

Perhaps it does, in fact, deviate from certain epistemological assumptions that science requires. But it is not necessarily therefore contrary to the spirit of science. As Kevin Cahill says of Wittgenstein's project,

> Wittgenstein is clearly bemoaning what he takes to be our modern fascination with the scientific method, a method that accounts for things by subsuming diverse phenomena under general causal laws. The astonishing success and power of this method for the purposes of explaining certain phenomena can then seduce us into seeing everything this way, and we are "surprised" when other phenomena ("disquietudes" [*Beunruhigungen*]) resist being understood in naturalistic terms (Cahill, 2011, p. 154).

"Unscientific" does not signal any retreat into mysticism. On the contrary, in seeking out limits of the rationalist assumption about skillful behavior on the basis of empirical

inadequacies, i.e. *in practice*, the hope is to suggest new avenues for research into the sorts of "other phenomena" that "resist being understood in naturalistic terms."

The importance of the AI analogy to my exploration here lies in its failure to produce or replicate the skillful flexibility that characterizes intelligent behavior in general and paradigmatic cases of excellent teacher practice in particular. If the rationalist view were correct, and skillful or intelligent practice were the logical result of applying transcendent rules to facts defined by transcendent features, the very superiority of computers to humans in precisely the two areas most germane to such a view – information processing and reliability – ought to ensure superior, indeed *perfect*, skillfulness. The AI field, and Doug Lenat in particular, interprets the comparative rigidity of extant AI systems like his own, their *inferior* skillfulness, as stemming from an inadequate number of transcendent facts or rules. It seems at least equally plausible, though, that the areas in which computers remain superior to actual human beings are themselves insufficient to account for skillful practice, that a view of skillful practice based on information-processing and calculating leaves out something essential, and thus the recognition of the shortcomings of the rationalist view, to say nothing of the pursuit of alternative paths, seems more than warranted.

Dreyfus, in fact, in suggesting that AI has reached an impasse that researchers such as Lenat have trouble recognizing, points to one possibility for this essential omission:

> What hides the impasse is the conviction that the commonsense knowledge problem must be solvable, since human beings have obviously solved it. But human beings may not normally use commonsense *knowledge* at all. What commonsense *understanding* amounts to might well be *everyday know-how* (H. L. Dreyfus & Dreyfus, 1986, p. 99).

When Shulman, for example, assumes that Dreyfus's "everyday know-how" must be codifiable, he begs this very question. Everyday know-how may not *be* knowledge in any

codifiable sense. Here lies the importance of Wittgenstein's example of teaching someone

the *concept* of a word by means of examples and exercises, according to which practice he

does "not communicate less than I know myself." As Charles Taylor understands it,

> The reciprocity [between rule and event] is what the intellectualist theory leaves out. In fact, what it shows is that the "rule" lies essentially *in* the practice. The rule is what is animating the practice at any given time, and not some formulation behind it, inscribed in our thoughts or our brains or our genes, or whatever (Taylor, 1993, p. 178).

Dreyfus's "everyday know-how" refers to the ability to get around correctly or

skillfully with the practical world, including social practices. His suspicion that "everyday

know-how" may not be "commonsense *knowledge* at all" points directly to Taylor's sense

that the rule "lies essentially *in* the practice," rather than being "some formulation behind it,

inscribed in our thoughts or our brains or our genes, or whatever."

If it is indeed the case that skillful or intelligent human behavior is a matter of

practice which does not, as it ordinarily appears, depend upon or proceed from articulable or

abstract rules, it remains necessary to account for the *regularity of* our skillful behavior such

that it qualifies as "rule-following" in general.

From a Wittgensteinian perspective, the shorthand for the basis of such regularity lies

in the concept of a "form of life." He says, in §19,

> It is easy to imagine a language consisting only of orders and reports in battle. – Or a language consisting only of questions and expressions for answering Yes and No – and countless other things. – And to imagine a language means to imagine a form of life" (Wittgenstein, 2009).

The connection here between any given language – the words and phrases it consists

in – and a "form of life" emphasizes its reliance on a way of practical getting around in the

world. But this practical form of life is not that of a Cartesian subject in walled-off

introspective isolation: as Cahill says, "A mark of Wittgenstein's later thought, then, is the idea that "authentic" clarity in philosophy is inseparable from our acceptance that the meaning of our words is in a certain sense dependent on what shows up as mattering to us" (Cahill, 2011, p. 141). Contrary to a rationalist assumption that the meaning of our words is fixed and articulable, and only by such means logically applicable in all future situations – something Derrida also reveals to be false – Wittgenstein's basis for the meaning of our words lies in what matters to *us*.

In terms of the "us" implied here, Glendinning puts it perhaps best: "A central feature of Wittgenstein's account of that behavior which one can call 'following a rule' is, therefore, the way in which *a relation to the behavior of others*, and hence a structural *publicness*, is *a priori* implicated by it" (Glendinning, 1998, p. 103). In the notion that a form of life is a shared phenomenon, such that following rules amounts to a "relation to the behavior of others," Glendinning's "structural publicness," implicit in "a relation to the behavior of others," gestures toward the *regularity* implicit in such rule-following activities. The *nature* of the "relation to the behavior of others" may or may not accord with the "symbol-system hypothesis," whereby this relation is characterized by shared access to codified information, but it is certainly public.

Cahill's sense of "*mattering* to us" here, however, once again renders the rationalist assumption about the "symbol-system hypothesis" an unlikely candidate to explain Glendinning's "structural publicness." Cahill's phrase, after all, recalls the *relevance* problem Dreyfus points out in AI, and I will return to that issue shortly. But the use of this phrase also draws in a Heideggerian understanding of the basis upon which *mattering* is possible in the first place: an account of how and why relevance shows up *for us* (as opposed

to computers). One basic Heideggerian concept toward which Cahill's comment points is that of "Dasein" – a reconception of the human in terms of its "being-*there*." Despite the obscure jargon involved, the original Heidegger is worth quoting:

> The totality of [Dasein's] involvements goes back ultimately to a "towards-which" in which there is no further involvement: this "towards-which" is not an entity with the kind of being that belongs to what is ready-to-hand within a world; it is rather an entity whose Being is defined as Being-in-a-world, and to whose state of Being worldhood itself belongs." (Heidegger, 1962, p. 84).

Just as one cannot imagine a language, on Wittgenstein's view, without also imagining a form of life, Heidegger insists that to think of people – of a person – is always and in every case to think of a person *in a world*, which here implies a "totality of involvements" defined, ultimately, by a "towards-which" – something Heidegger elsewhere attributes to Dasein's unavoidable and omnipresent "understanding of Being." Just as Wittgenstein's form of life is ultimately about practical concerns, Heidegger's "towards-which" also derives from a practical sense: As Kisiel has it, "The basic sense of the factic movement of life is caring (*curare*) being out toward something (*Aussein auf etwas*), and that 'something' is the temporally particular world. The dynamics of caring has the character of getting along and coping with (*Umgang*) its world" (Kisiel, 1993, p. 255).

The importance to my discussion of the inseparability of the "Being" of Dasein from its world – like the importance of the inseparability of a language from a form of life – has directly to do with the issue of whether or not such apparent dichotomies as "Beings" and "worlds" or "languages" and "forms of life" are *independently formalizable in principle.* The rationalist assumption holds (a) that they are, and (b) that such logically rule-governed interaction between such formalized representations, whether consciously or unconsciously accessed, accounts for regular nature of our rule-following behavior. However, as Derrida

has demonstrated in the case of language's iterability, such independent formalization *cannot* account for the rule-following involved in linguistic *practice*. Neither has AI successfully been able to demonstrate that formalizing "Being" and "world" independently *can* account for or replicate the flexibility of our own intelligent behavior. Taylor would ascribe such a failure to the fact that the "intellectualist theory leaves out" a conceptual "reciprocity" between these independent terms.

Derrida's project, as Glendinning reads him, has specifically to do with the fact that the former's demonstration that our knowing of language cannot be reduced according to a "symbol-system hypothesis" *presupposes* that we can in fact *use* language unproblematically. That we *know how to use* language is simply assumed in Derrida's deconstruction. Whether or not our knowing-*how* consists in the manipulation of symbols according to a particular system is precisely the issue that he shares with Dreyfus and Wittgenstein, and with that segue, I return briefly to AI and Lenat's chicken-and-egg problem.

Lenat's discussion of machine learning in natural language, that is, AI's ability to use language to learn things about language, breaks down on what he calls a "chicken-and-egg problem," one in which knowledge *growth* implies the existence of *prior* knowledge, a process of learning which therefore relies upon precisely what it is trying to cause. Derrida's discussions of language also expose this characteristic chicken-and-egg difficulty, in which attempting to extrapolate a system from individual linguistic events cannot logically get started due to the fact understanding these events *as* instances *of something* within a language presupposes some conception of the very linguistic system it is supposed to generate. Similarly, working in the other direction, the system as a series of definitions and rules

cannot include all the possible deviations, exceptions, ambiguities, or *new* instances of

language-use that exist in practice, as Lenat also notes.

When one eschews, however, the assumption that language's regularity implies a

systematic ontology, where something like *essential* rules and definitions are amended or

adorned with accidental features and occasional irregularities, the chicken-and-egg problem

disappears, largely because the task turns from *identifying* and later *formalizing* essential

elements to the examination of particular cases. Wittgenstein's concept of the family

resemblance is particularly important here.

The notion of the family resemblance is quite complicated, but utterly central to

Wittgenstein's philosophy. It is also important to my argument, as it undercuts the necessity

and possibility of arriving at any "structural primitives" or any other general formalization in

the arena of practical human *umgehen*. In the *Blue Book*, Wittgenstein introduces the

concept:

> We are inclined to think that there must be something in common to all
> games, say, and that this common property is the justification for applying the general
> term 'game' to the various games; whereas games form a *family* of members of which
> have family likenesses. Some of them have the same nose, others the same eyebrows
> and others again the same way of walking; and these likenesses overlap
> (Wittgenstein, 1958, p. 17).

In the *Investigations*, he notes, once more looking at the concept of games, that "we

see a complicated network of similarities overlapping and criss-crossing: similarities in the

large and in the small." He goes on to speak of learning to use numbers in similar terms:

"And we extend our concept of number, as in spinning a thread we twist fibre upon fibre.

And the strength of the thread resides not in the fact that some one fibre runs through its

whole length, but in the overlapping of fibres" (§66-67). By pointing out that a family

resemblance is often irreducible to any *single* characteristic feature, Wittgenstein raises for

consideration the facts that, in recognizing the kinship between any *two* members of a given

family, an observer *will* likely be able to point to a given feature, say, eyebrows, that the two

members share in common; but also that in locating the similarity between *another* two

members of the same family, one will have to look for different similar features, such as their

shared gait, for example. When we recognize *all* members as members of a single family, it

is therefore *not* on the basis of any single feature shared in common. One thinks especially

of the research program involved in distilling best practices from videotaped teacher practice

when Wittgenstein says, in the above vein, that

> The idea that in order to get clear about the meaning of a general term one had
> to find the common element in all its applications has shackled philosophical
> investigation; for it has not only led to no result, but also made the philosopher
> dismiss as irrelevant the concrete cases, which alone could have helped him
> understand the general term (Wittgenstein, 1958, pp. 19-20).

It is important to recall, as we consider what it means both to identify excellent

teacher practice and to bring others to exhibit excellent practice, that the rationalist

assumption remains a hypothesis. The notion that underneath all changes in appearance, all

variation according to context, lies some one *thing* that, by virtue of its commonality in all

cases, provides the *essence* of excellent teacher practice is merely hypothetical, a fact that

Dreyfus also emphasizes, as when he notes that "Lenat's faith that Cyc will succeed is

based… on the untested traditional assumption that human beings have a vast library of

commonsense knowledge" (H. L. Dreyfus, 1992, p. xxii). The "traditional assumption," as

Dreyfus also notes, is as untested in AI as it is in questions of teacher quality: "What is

unusual and inadmissible is that, in this case, the hypothesis *produces* the evidence by which

it is later confirmed" (H. L. Dreyfus, 1992, p. 174), a feature we have repeatedly seen in the

use of student achievement data to identify best practices. The problems encountered by AI,

and indeed, the inadequacies suggested by the fact that Hill and Ball's "Mathematical Knowledge for Teaching" only accounts for 11 percent of the variance in student achievement scores, suggests that the hypothesis itself either (a) is mistaken or (b) cannot *do* or *provide* what it is often assumed to. No single thing turns out to be, in other words and in either case, *essential* to excellent practice in the way that it is often imagined. This is importantly and obviously not to suggest that there *is* no one thing common to all teaching practices; merely that any one common feature of teacher cannot serve as the *essential* element of excellent teaching, at least in the way essence is traditionally conceived.

The idea of the family resemblance militates against the default acceptance of the rationalist assumption. As Wittgenstein notes, thinking yet again about games,

> What is common to all of them? – Don't say: "They *must* have something in common, or they would not be called 'games'" – but *look and see* whether there is anything common to all. – For if you look at them, you won't see something in common to *all*, but similarities, affinities, and a whole series of them at that. To repeat: don't think, but look! (Wittgenstein, 2009, pp. 36, sec. 66)

In attending to particular cases of good teaching, "which alone could [help] understand the general term," it is not yet clear that our general grasp of a shareable label is predicated upon any one element common to all instances of good teaching. That is the value of the Mrs. Grady test.

To return to the discussion of Derrida and Lenat, the fact that we *use language* with one another on a daily and ordinary basis and yet cannot reconstruct the practical working of language-use on the rationalist model suggests, obviously, that the model is insufficient. The radical distinction between "system" and "symbol," between "language" and "form of life," between "Being" and "world," requires an explicit unifying or mediating component (such as Shulman's higher order of rules, or something like "reason") in order to regulate and drive

the interaction between them, according to which things like mattering, relevance, and purpose would arise. But when we (Derrida, Lenat) stand outside of such a system and examine it, no such mediating component appears. One response (Lenat's) is to assume that such a third term is merely implicit; another response (Wittgenstein's, Heidegger's) is to suggest that it is our *standing outside* of ourselves that generates the appearance of a problem where none in fact exists: being bound to others through forms of life or totalities of involvements accounts for our unproblematic ordinary getting-around.

The necessity of this mediating component, and its persistent absence, results in the regresses I have already identified. Robert Fogelin, after Wittgenstein, calls this situation the "paradox of interpretation." Assuming, with Lenat, that a third term *does* exist, and that it does perform the role of engaging a set of formalized rules with a formalized situation, proves untenable. As Stern has it, locating any formalization that itself requires no interpretation is "both unnecessary and impossible. It [is] unnecessary because we do not need an unmoved mover in order to follow a rule; it [is] impossible because nothing could perform that task" (Stern, 1995:116).   Fogelin augments Stern's claim about the "unmoved mover" as he avers instead that "for Wittgenstein there is no 'paradox' of rule-following. The thought that it is paradoxical is the product of a misconception, namely, the misconception that rule-following is always grounded in (or implicitly contains) acts of interpretation" (Fogelin, 2009:22). Here Fogelin takes seriously Wittgenstein's claim in §201b of the *Investigations:*

> That there is a misunderstanding here is shown by the mere fact that in this chain of reasoning we place one interpretation behind another, as if each one contented us at least for a moment, until we thought of yet another lying behind it. For what we thereby show is that there is a way of grasping a rule which is *not* an

interpretation, but which, from case to case of application, is exhibited in what we call "following a rule" and "going against it" (Wittgenstein, 2009).

Fogelin highlights the distinction between the Lenat view and what Wittgenstein calls "a way of grasping a rule which is *not* an interpretation." In Dreyfus's view, any attempt to represent to ourselves in formalized terms our own skillful activity raises this insoluble paradox. The contrast between the insolubility of the intractable paradox and the ease of, for example, exchanging small talk about the weather, augers for Wittgenstein's claim that, while grasping rules *does* in certain cases involve deliberative interpretations of propositional orders (this is especially clear when one finds oneself confused about what to do next, say, in assembling IKEA furniture), there are also cases in which any involvement of deliberative interpretation must be, as it were, *grafted on* post hoc, cases in which we proceed automatically, not to say naturally. As Wittgenstein himself concludes in §202 – anticipating my next section – "That's why 'following a rule' is a practice. And to *think* one is following a rule is not to follow a rule. And that's why it's not possible to follow a rule 'privately'; otherwise, thinking one was following a rule would be the same thing as following it." The impossibility of following a rule privately, the insistence that following a rule is a *practice*, emphasizes the central role of the sharing of forms of life.[20]

---

[20] Cavell says in a few different places (Cavell, 2010, 2012) that he urges a "key ambiguity" in Wittgenstein's notion of forms of life: "The concept projects simultaneously, as I take it, an irreducibly horizontal ethnological or conventional axis crossed by an irreducibly vertical or biological axis, which is in effect to picture human existence as that life-form that eternally criticizes itself as it were from below and from beyond or incessantly declines to." I find that striking. Stern has said similar things in early work. The irreducibility of those axes is, I think, the most important element of this picture. Just as one could not successfully give a Martian a set of instructions that would enable him to navigate the American social world fluidly – which is simply to say that one's culture is not formalizable – it is clear that human conventions intersect with biological facts about ourselves. As Wittgenstein says, "a mouth *smiles* only in a human face" – we do not need to share a culture with a Tibetan to recognize and acknowledge the human emotion expressed in a physical Tibetan smile.

Dreyfus, for his part, takes up similarly Wittgensteinian themes in discussing the chicken-and-egg problem:

> In the end the very idea of a holistic information processing model in which the relevance of the facts depends on the context may involve a contradiction. To recognize any context one must have already selected from the indefinite number of possibly discriminable features the possibly relevant ones, but such a selection can be made only after the context has already been recognized as similar to an already analyzed one. The holist thus faces a vicious circle: relevance presupposes similarity and similarity presupposes relevance. The only way to avoid this loop is to be always-already-in-a-situation without representing it so that the problem of the priority of context and features does not arise (H. L. Dreyfus, 1992, pp. 54-55).

Heidegger's sense of concern as the "towards-which" that structures our understanding of Being constitutes Dasein's way of being-*in*-the-world such that, in terms of Derrida's explorations of language or Wittgenstein's more general discussions of following rules, "the problem of the priority of contexts and features" – to say nothing of relevance – "does not arise," and hence that we *have* things like skillful use of language and an ability to get along practically in the social world at all. An inability to *give ourselves* a satisfactory explanation according to the rationalist assumption of *how* we deal practically or cope in the everyday world has no effect on our ability *actually* to do so; and attempting to reproduce that skillful coping by means of what a rationalist *outsider's* view requires of us seems very unlikely to have the desired effect, as AI's continuing struggles indicate.   In succinctly detailing the impossibility of framing our knowing-*how* in the form of knowing-*that* that the rationalist assumption requires, Dreyfus invokes both Heidegger's "understanding of Being" and Wittgenstein's spade hitting bedrock:

> To explain our actions and our rules, we must eventually fall back upon our everyday practices and simply say, "this is what one does" or "that's what it is to be a human being." Thus in the last analysis all intelligibility and all intelligent behavior must hark back to our sense of what we *are*, which is, necessarily, on pain of regress, something we can never explicitly *know* (H. L. Dreyfus & Dreyfus, 1986, p. 81).

The very nature of the chicken-and-egg problem involved in what Taylor calls the "reciprocity" of something like the articulable rule and the visible practice, which Cahill also calls the "conceptual interdependence of our understanding" (Cahill, 2011, p. 116), precludes the possibility of spelling out either one completely and separately, since each always depends upon the other. This is the sense in which, per Dreyfus, "we can never explicitly *know*" the background. "Our sense of what we are," upon which our skilled or intelligent everyday behavior ultimately possible, cannot be captured in propositional form (cf. Heidegger, 1971, p. 64). Rather, our intelligent behavior, our ability to follow rules at all, is a matter of a navigating particular situations according to their family resemblances, which are embedded in and arise out of our shared everyday practices.

2.4.3 Structural Publicness and Everyday Practices

The background of shared practices, to which Dreyfus here gestures as a nonformalizable condition for intelligent or skilled behavior – and which Taylor suggests is implicated in our "strong evaluations" in the domain of ethics (Taylor, 1989) – must indeed be shared and pervasive in order to function as the basis of the intelligibility of everyday practices.   Heidegger and Wittgenstein anchor the source of intelligibility of skillful or intelligent behavior in a practical realm. Heidegger, expounding upon Aristotle's *Politics,* explicitly extends the notion of practical "concern" that structures Dasein's "understanding of Being" to a common arena:

> Heidegger . . . explains why "man is by nature a political animal," more so than the other gregarious animals, like bees. For all these merely have "voice" to sound and indicate their pleasure or pain (being well- or ill-disposed to their environment), whereas the power of speech proper to humans is far more discriminating, in making manifest the useful and the inexpedient, the fitting and the improper, the just and the unjust. In short, human beings alone have any real sense of good and evil. "And such a being-with-one-another cultivates a household and city."

Being-with-one-another is therefore speaking with one another in our common concern with being-in-the-world (Kisiel, 1993, pp. 294-295).

The concern with our own being in the world is taken to be common. More than that, the common concern manifests itself the way our language brings to light the criteria for "the useful and the inexpedient, the fitting and the improper, the just and the unjust," all of which is rooted in a sense of practical *umgehen.*

Stanley Cavell, too, understands Wittgenstein as pursuing a similar investigation of something *shared*, something that grounds the intelligibility of things to an *us*, and he explicitly links it to the pre-theoretical and the pre-rational: "The philosophical appeal to what we say, and the search for our criteria on the basis of which we say what we say, are claims to community. . . . The wish and search for community are the wish and search for reason" (Cavell, 1979, p. 20). Reason, on Cavell's view, is (only) coextensive with the practical community in virtue of whose shared forms of life reason has its particular meaning and takes its particular shape.

In bringing Cavell to bear on these considerations, I wish to make two major points with respect to the shortcomings of the rationalist assumption and the suggestion that the intelligent and skillful behavior that we *in fact* both exhibit in practice and distinguish as a matter of course in ordinary life. The first point is a reformulation of a Wittgensteinian point that I have already mentioned: that the functioning of our practical behavior, such that one can be either skilled or unskilled, is a matter of agreement in forms of life, and that this "agreement" amounts to nothing like arriving at a *mental decision*.

> Wittgenstein's appeal to criteria is meant, one might say, exactly to call to consciousness the astonishing fact of the astonishing extent to which we *do* agree in judgment; eliciting criteria goes to show therefore that our judgments *are* public, that is, shared. What makes this astonishing, what partially motivates this philosophizing on the subject, is that the extent of agreement is so intimate and so pervasive; that we

communicate in language as rapidly and completely as we do; and that since we cannot assume that the words we are given have their meaning by nature, we are led to assume that they take it from convention; and yet no current idea of "convention" could seem to do the work that words do – there would have to be, we could say, too many conventions in play, one for each shade of each word in each context. We *cannot* have agreed beforehand to all that would be necessary (Cavell, 1979, p. 31).

Cavell notes the same problem that Lenat sees: that in order to understand language according to a rationalist assumption – a "current idea of 'convention'" – "there would have to be, we could say, too many conventions in play, one for each shade of each word in each context." Lenat continues to add to the 28 years he has already spent attempting to *compile* precisely these conventions; Cavell's suggestion that "we *cannot* have agreed beforehand to all that would be necessary" is aimed at undermining just the picture of agreement that implies such exhaustive and exhausting mental activity. The sort of agreement he has in mind is the "shared"-ness of Dreyfus's shared background practices. Cavell illustrates: "The idea of agreement here is not that of coming to or arriving at an agreement on a given occasion, but of being in agreement throughout, being in harmony, like pitches or tones, or clocks, or weighing scales, or columns of figures" (Cavell, 1979, p. 32).

The upshot of Cavell's exegesis with respect to the rationalist assumption about the causal conditions that ought to produce functional everyday language is that no articulation of such causal conditions – in the form of a transcendent system with articulable rules and definitions – is *possible* in principle.

The second point I wish to make takes us back in the direction of Fogelin, Stern, Dreyfus, AI, and teaching. The notion that the sort of "agreement" upon which such a pervasive and ordinary skill depends might represent an "idea of convention" that seems to resist formalization suggests, once again, that no such rationalist-assumption "idea of convention" is in fact *necessary* to the ordinary functioning of everyday language. The sorts

of "conventions" at work within the familiar world are *not* the formalizable rules upon which intelligent behavior is assumed to rest; and yet we seem to have access to something better, on the basis of which we have the *actual* skillful behavior that, while if not *theoretically* explicable according what we take to be the requisite assumptions, remains *factually* extant.

In speaking of shared practices, particularly when drawing on Wittgenstein, Cavell, and Derrida, I have tended to restrict myself to *language-use* as the salient example. But the notion of shared practices undergirding skillful or intelligent human behavior in ordinary situations includes vastly more domains. Dreyfus is fond of talking – after Nietzsche – of the *style* of a culture, ways of being that are shared among members but so pervasive that one *never* learns them by means of explicit rules, but rather absorbs them, as it were, in the process of coming up in that culture. Learning a second language, for that very reason, is much different from learning a mother tongue. These, too, are matters of "agreement" in the Cavellian sense, but are, as he says, "so intimate," that one finds one has never in fact *agreed* to them. Such factors account for the kinds of *know-how* that AI has struggled so mightily to imitate through formalization.

2.4.4 Relevance and Temporality

An important difference that seems to distinguish actual skillful practice from artificial attempts to reproduce it, and one upon which I have touched only too briefly in previous section, has to do with a second aspect of know-how: the issue of relevance. Lenat's account of the struggles with developing NL simply indicates that language-using humans have a different and non-inferential means of establishing or determining relevance. Heidegger's ideas are particularly helpful in this regard. His notion of *concern* – rooted in the sense of Dasein's being-in-the-world – entails already a futural and ongoing "project" on

the part of Dasein, in light of which certain aspects simply show up as relevant under

ordinary circumstances. The mention of "futural" introduces Heidegger's notion of

"temporality," which is in itself far larger than my own project requires. Dreyfus is succinct

on the way in which Heidegger's temporality accounts for the human ability to respond

skillfully in ordinary situations, without seeming to have to *decide* upon the relevant factors

at all:

> The important point is that we human beings proceed from the past into the future with our past experience always going before us organizing the way the next events show up for us. So we do not need to deal with real-world situations by listing in advance all possibly relevant features plus rules for determining under what circumstances each feature may become actually relevant and when it can safely be ignored, and rules for when these rules are relevant, and so forth (H. L. Dreyfus & Dreyfus, 1986, p. 88).

Pierre Bourdieu notes something strikingly similar, and he links it back to the earlier

discussion of following rules: "The active presence of past experiences… deposited in each

organism in the form of schemes of perception, thought, and action, tend to guarantee the

'correctness' of practices and their constancy over time, more reliably than all formal rules

and explicit norms" (Bourdieu, 1980, p. 54).[21] Bourdieu, importantly, also holds that the

"formal rules and explicit norms" Derrida exposed as inadequate to the factual existence of

our generally correct language remain likewise inadequate to the correctness of our practices

more generally. The role of experience in contributing to the ease of navigation in ordinary

---

[21] I do not want to dwell on this point, but an account of how we "have" our past experience, or how "past experiences" become "deposited" in an organism may be necessary at some point, but it is not necessary to my current work. *That* we "have" past experiences is the crucial part. It is clear that past experience guides present judgment through something like what Dreyfus calls the organizing of the way things show up. It is likewise fairly clear, given the AI work cited here, that however we "have" these experiences, it is not "stored" in the form of feature-based descriptions nor is it searchable according to the meaningless *features* that confront us. That model would, again, as we get older and have more experiences, make practical coping *less* easy, instead of the opposite.

life as well as in a particular skill domain will feature prominently in Dreyfus's account of moving from a rank beginner to an expert in a given skill domain.

<p style="text-align:center">2.4.5 The Shared Background, Situational Relevance,
and Mrs. Grady</p>

I opened this section by claiming once again that Kristof's appeal to Mrs. Grady raises problems for the understanding of skillful behavior implicit in the "best practices" view of teaching to which Duncan and Lemov subscribe, and also for Shulman's knowledge-based "wisdom of practice" account. Specifically, I noted that the rationalist assumption could not account for (a) the emergence of something *new* at all, as in a novel response to a given set of circumstances, and (b) a shared or recognized sense of the correctness or appropriateness of this *new* response, both of which are required to explain the excellence that Mrs. Grady *displayed* and also Kristof's ability to assume the *shared understanding* of such excellence such that Mrs. Grady's example would serve as such. Where the rationalist assumption fails in regard to these standards, the ideas of the thinkers considered in this section fare considerably better.

With respect to (a), the rationalist assumption puts out of play the possibility of something wholly unpredictable, since relevance is a matter of prior determination and since potential actions are themselves always specifiable aprioristically. Working on the assumption that allowing children to steal would not appear on any list of specifiable potential actions, Mrs. Grady's response to witnessing Olly Neal's pilfering a book ought not to have been possible as an expression of excellent practice. In addition, the fact that the *relevant* features of students' rule-breaking, determined from an aprioristic position, especially as regards damage to or theft of school property, admit the *particular* only with

regard to the monetary value of that which is harmed – in consideration of which the punishment involved will be greater or lesser, but in no case absent and in *no* case foregone entirely in the specific interest of *encouraging* the rule-breaking in question – ought to have not only precluded Mrs. Grady's actual response, but should have (and probably would have, had it been discovered) also garnered Mrs. Grady some sort of disciplinary sanction.

Seen from the perspective of an ongoing situation comprised of a "totality of involvements" and structured by the dynamics of *care* or *concern*, however, in which Mrs. Grady's past experiences precede her into the future, Mrs. Grady's otherwise inexplicable[22] choice not only to allow a child to steal, but also to actively encourage repeated theft, makes sense. Mrs. Grady knew Olly Neal from having taught him; he was a kid with an attitude, whose frustrating demeanor made Mrs. Grady cry on occasion. The facts about himself that Olly Neal cites cannot, however, stand in as representations of Mrs. Grady's *experiences* with him in particular and with teaching students in general as a holistic past coming into the present situation.

---

[22] I anticipate that someone will claim that this is not inexplicable according to the rationalist assumption. Such a claim would likely be made upon the relative value of punishing a child for stealing and encouraging a child to read, and if the relevance established a priori valued encouraging a child to read above all else, such a decision would make sense. That's surely true. But above *all* else? There are all kinds of things above which we appear *not* to value reading. If one valued reading above all else, we'd have kids in school 14 hours a day, for one example; but we also value the family life to which the child returns, it seems, even if that child is not learning to read explicitly. And it's by no means clear that we value schools' ability to encourage reading to such an extent that we would, as a matter of policy, suggest that it's okay to violate rules and laws if one suspects that the child is doing so in the interest of reading. How did Mrs. Grady know that Neal stole the book because he wanted to read it? Neal doesn't seem to know that he wanted to read it. He says that at a certain point in his thievery he "caught the bug" and began to love reading, which implies he didn't steal the book *because* he loved to read. All sorts of problems crop up for such attempts to make Mrs. Grady's decision correct based upon prior knowledge and transcendent rules or principles.

The practical *sight* required to recognize the relevance of Neal's stealing a *book* and the response to re-stock that author such that he might steal further books cannot be of the rationalist variety, as we have seen. That type of sight, a concept Heidegger takes from Aristotle,[23] must be rooted in *practical concern*, the "towards-which" that constitutes a concernful understanding of being.[24] Heidegger says,

> Something like this intentionality of producing, which we have characterized, and the type of understanding of being peculiar to it should be seen simply with a vision that has not been dazzled and made squint-eyed by some current theory of knowledge. No matter how logically rigorous concepts may be, if they are blind then they are worthless (Heidegger, 1982, p. 114).

Heidegger's play on Kant above explicitly ties the sort of sight he has in mind to the notion of the intentionality associated with an "understanding of being." Interestingly, it also recalls Wittgenstein's notion of "following the rule *blindly*"; in Wittgenstein's case, however, the blindness inherent in the practice of following a rule *presupposes* the very sort of practical sight native to understandings of being or forms of life as such. The emphasis on seeing above, in contrast with the *particular* sort of blindness pertaining to an abdication to "some current theory of knowledge," requires practical involvement in an ongoing situation, such that the relevant features that skillful practice require simply *show up* in spite of the nonformalizable nature of the "towards-which" upon which that relevance depends, even in spite of the potential *novelty* of the practical action in question. In this way, involved

---

[23] And, in the quoted passage, from Kant.

[24] Heidegger says in a discussion of the phenomenology of perception (Heidegger, 1982, pp. 54-76) that (a) perception is always intentional, and that (b) intentionality, properly conceived, is always prior to subjectivity and objectivity alike and is "*the ontological condition of each and every transcendence*." When he claims earlier in BT that Dasein is the "*transcendens* pure and simple," one understands that *seeing* a particular situation in a particular way is *practical* at heart and incontrovertibly bound up with a sense of "what we are," as Dreyfus says.

practical *concern*, by structuring relevance, enables excellence in *new* forms, something for which the rationalist assumption cannot account. Or as Dreyfus puts it: "If one is a master of a cultural practice, one can sometimes do what has not so far counted as appropriate and have it recognized in retrospect as having been just the right thing to do" (H. L. Dreyfus, 1992, p. xxiv).

On the note of the possibility of having something new "recognized in retrospect as having been just the right thing to do," I take up the issue of (b), Kristof's appeal to such an example of *new* and *singular* practice as illustrative of excellent teaching in general. My consideration here has less to do with Kristof's own choice to use Olly Neal's example itself and much more to do with the conditions upon which he expects the excellence in Mrs. Grady's example to be visible.

In the Introduction I pointed out the implausibility of Mrs. Grady's excellence showing up on the metrics by which Race to the Top proposes to make such excellence visible; in Chapter 2, I have described the ways in which a "best practices" or "teacher knowledge" view would prove inadequate to either the task of seeing Mrs. Grady's excellence or of reproducing it. At this point, then, I raise the *fact* of the visibility of Mrs. Grady's excellence in retrospect, without the need for the metrics that Duncan proposes and without the justificatory support of generalizable behaviors. I argue here that the fact of the visibility of her skillfulness presupposes the sharing of forms of life that Wittgenstein suggests underlie our practical rule-following in general, and in doing so I hint in the direction of Chapter 3, which will take up the question of *communicating* teacher quality for evaluation purposes. But to more directly preview the following section, I wish to stress the dramatic distinction between a *retrospective* account of her practice and the conditions of its

emergence in the present of 1957, without an actual happy ending to ease the existential

uncertainty of her choice.

To the first point, the *fact* of the exemplary role of Mrs. Grady's practice in the

absence of all the metrical and observational criteria proposed, I wish simply to suggest that

its appearance *as* excellence demonstrates something important about "the astonishing fact of

the astonishing extent to which we *do* agree in judgment," as Cavell has it. In the context of

the narrative, we are invited to see Mrs. Grady as a teacher who did not let the fact of Neal's

difficulty prevent her from believing in him. In that, she appears generous of spirit. Nor did

her negative experiences with Neal interfere with her recognizing not only his potential but

also the conditions which made the realization of that potential so difficult. Lastly, the

practical *sight* that afforded her a view of the correct *way* in which to respond to witnessing

him steal a book is on full display in the article, as well. That Kristof's piece can do what it

does – hold up an example of excellent practice in the full expectation that it will serve as it

is meant to serve – indicates and presupposes the very shared-ness of the shared forms of life

in which great teaching shows up as such. His very assumption that the example will have

the intended effect *at all* evidences this fact. The fact of its recognition recalls the

Dreyfusian point about sub-worlds, as opposed to the micro-worlds of AI:

> In our everyday life we are, indeed, involved in various "sub-worlds" such as
> the world of the theater, of business, or of mathematics, but each of these is a "mode"
> of our shared everyday world. That is, sub-worlds are not related like isolable
> physical systems to larger systems they *compose*; rather they are local elaborations of
> a whole which they *presuppose* (H. L. Dreyfus, 1992, p. 14).

Considering Wittgenstein and Cavell's discussions of forms of life as the ground-

level practical involvements upon which basis we have things like reason and resemblance, I

contend that the agreement in forms of life indicate this Dreyfusian "whole" presupposed by

the particular sub-worlds of "the theater, of business, of mathematics." And of teaching. That teaching might be seen as a *local* elaboration of this presupposed *whole* accounts for Kristof's ability to assume his audience's recognition of Mrs. Grady's excellence: the *whole* of the forms of life in which we agree includes an ordinary sense of teaching, on the basis of which the *shared* recognition of the excellence of Mrs. Grady's example is possible in the first place.

It is worth noting, too, that Neal's story is precisely the *ordinary* sense of good teaching to which Duncan so often points: here we have a teacher going beyond the requirements of her own job (for she was not *his* teacher at the time) in the interest of "chang[ing] the course of a student's life" (Duncan, 2009d). The going-beyond involved cannot be accounted for according to any seeking to meet benchmarks or objectives precisely because in an official sense she was not *responsible* for teaching Olly Neal at all. Duncan thus has it right when he says that "Teaching … is not just a job or even an adventure—it's a calling" (Duncan, 2009d). The sense of answering a call is precisely the sense of the "towards which" inherent in Mrs. Grady's "understanding of being" *as* a teacher. Answering a call here also gestures toward the "whole" presupposed by the particular sub-world of teaching.[25] Mrs. Grady's actions can be accounted for, then, in terms of understanding herself *as* a teacher within a larger "totality of involvements" in which "teaching" has its

---

[25] In truth, the "calling" language here on Duncan's part is related to his insistence in another place that teachers ought not to have to take a "vow of poverty" in order to do be teachers. Teachers, both in his speeches and in the popular imagination, really are making religiously-inflected sacrifices (Sarah Palin once said that teachers are doing "God's work"). That is probably the sense in which Duncan is actually using this trope. But teachers are not special with regard to the nature of "calling." Every profession is a "calling" in the sense in which I use it here – every profession presupposes a holistic understanding of being in light of which the profession makes sense at all.

particular resonance. She didn't help Olly Neal learn to read because he was *her* student, her particular and official responsibility; she helped him learn to read, one might say, because *she* was a teacher, and as such took up a responsibility for teaching *in general*.

But to return to the sub-world idea for a moment, her *particular* choices in that *particular* situation would defy such general visibility *if not for* the criterion of "chang[ing] the course of a student's life." In other words, while agreement in forms of life associated with the "whole" enable the *general* public to see Mrs. Grady's example as excellent from a 2012 perspective, and while such a holistic understanding-of-being of teaching would also have been necessary to see it in 1957, such a holistic view would not therefore have been *sufficient* to communicate her quality to a national audience, as no such *life change* had then occurred. Mrs. Grady's practical excellence would have been visible, and only potentially, in virtue of its being a *local* elaboration of this presupposed whole. With no evidence of Neal's eventual judge-ship or the like, *seeing* Mrs. Grady's actions *as instances of* a generosity of spirit, a going-beyond of her official responsibilities, "chang[ing] a student's life," rather than, say, as instances of a cowed teacher too weary to confront a misbehaving student, requires familiarity with the *local* situation. In the present tense of 1957, when Olly Neal was, for all the world, a chronic behavior problem who stole from the school, only a *locally-situated* person would have had the possibility of access to the *sense* Mrs. Grady's actions made. Sharing the local context *and* the presupposed whole would have been necessary in the present.

To make a similar point in a slightly different way, the *sight* that enabled Mrs. Grady to act skillfully in that situation required intimate familiarity with the particular domain, one might say. That Kristof can refer to it in 2012 has to do with our agreement in forms of life

on the basis of which we have an ordinary sense of good teaching, one of the exposed criteria of which has to do with effecting a transformation in students' lives. But one can assume that Mrs. Grady's practice was excellent *before* its effects were visible, or even objectively probable. That Mrs. Grady could act *skillfully* in precisely the way she did back in 1957, yet without that action seeming to provide any extractable rules or generalizable behaviors – *universal* justifiers – points to an understanding of "skillful" or "excellent" in terms of *responsiveness to a local situation* – a responsiveness based on the kind of ability to *see* that Heidegger and Dreyfus explicitly mention, and to which the responsiveness points.

<div align="center">

2.4.6 Aristotelian Sight, Family Resemblances, and
Measurement Error

</div>

Here it will be helpful to return to some promises I made in Chapter 1. In that Chapter, I raised and immediately deferred several issues for consideration, mainly related to the claims made by advocates of value-added measures of teacher quality that apparent insufficiencies of such metrics ought not to be overstated. Such attention to the "technical" flaws in value-added measures, these advocates imply, cover up the significant progress that these metrics represent. In light of the foregoing discussions of AI and practical forms of life as it relates to teaching, it seems fitting to raise those issues once more.

Glazerman, et al., in the Brookings Institution report, assert that their methods of determining teacher quality are not as frail as their critics insist. In defending their work, they offer the three claims below. In Chapter 1, I addressed claim (a); here, I wish to reconsider claim (b).

> We have previously issued a report that describes some of the imperfections in value-added measures while documenting that: a) they provide one of the best presently available signals of the future ability of teachers to raise student test scores; b) the technical issues surrounding the use of value-added measures arise in one form

or another with respect to any evaluation of complex human behavior; and c) value-added measures are on par with performance measures in other fields in terms of their predictive validity (Glazerman, et al., 2011).

The "technical issues" they refer to, I suggest at this point, are hardly technical at all. They are, rather, what Cahill has called the "intractable difficulties" that arise from the fact that "none of our actual proceedings can meet the requirements for correctness that we imagine [the detached picture] imposes on us." In other words, the lack of fit involved in attempts to measure ordinary-sense teacher quality is not, as Glazerman and his colleagues have it, some inherent property of the complexity of "complex human behavior" itself; it is instead *produced* by misunderstanding the nature and source of the complexity involved. Since the ordinary concept of good teaching, as we have seen, has its roots in agreement in forms of life, it is simply not available from a detached or purely objective perspective, the stance that the statistical parsing of student achievement data requires. Once more, such a statement is not to imply that *nothing* is available from such a perspective; that would be akin to a claim that words in a language cannot be defined, as dictionaries regularly do. Instead, it is to claim that *what* shows up according to such a detached view leaves out something *essential* to "knowing" English, for example, or "knowing" good teaching; that in both the case of language in particular and "complex human behavior" in general, the detached view misses the (engaged, practical) whole that the view itself presupposes.

The danger in misunderstanding the lack of fit between value-added measures of teacher quality and the ordinary sense of good teaching is suggested by AI's manifold problems. That such methods of measuring quality can work *at all* is certainly a feat, analogous to SHRDLU's micro-world success in a domain that is highly specific, structured, and finite. But making such methods work *well and in general* might require, as Dreyfus

suggests, an entirely different set of assumptions altogether. It is by no means clear that a modicum of measurement success in a technically-delimited domain indicates that, in the realm of "complex human behavior," one is even on the right track to capturing the whole picture, much less that the distance between present fallibility and future perfection is both relatively small and merely technical. This assumption, no less than in AI, is nevertheless itself all-pervasive. The NCTQ report cited in Chapter 1 makes it plain: "Student growth and value-added methodologies are still emerging. However, examining student achievement as a metric for assessing teacher effectiveness, even if measurement is imperfect, represents a big step forward" (NCTQ, 2011). Such a conclusion is badly undercut by the foregoing considerations of this Chapter. It is certainly a step, but not a step *forward* in any meaningful sense, insofar as the possibility of *arriving* at the goal of knowing teacher quality is precluded by relying on the very methodology that represents the initial step.

I will develop the following argument more fully in Chapter 3, but I wish to preview it here: since the detached view blinds us in some important ways to the heart of what we wish to study or evaluate, we are only able to access that which we wish to study or evaluate by means of an *engaged* or *involved* perspective, which is the source of the Aristotelian "sight" upon which Heidegger and Dreyfus draw.

Earlier I alluded to the intertwined problems of Mrs. Grady *seeing* the proper course of action vis-à-vis Olly Neal and some third party *seeing* Mrs. Grady's excellence in the present-tense of 1957. The sight involved in each instance is afforded by means of both a holistic understanding of *teaching*, which is not at all the same as the having of a technical definition thereof, and an involved familiarity with the local conditions of teaching practice. The combination of Mrs. Grady's visible excellence and that excellence's invisibility

according to any metrics currently under discussion invites one to raise once more Strong's

discussion of human judges' ability to predict teacher quality from Chapter 1:

> In every case, judges achieved relatively high levels of agreement but were
> absolutely inaccurate, leading us to question whether educators can identify effective
> teachers when they see them. This in turn has motivated us to undertake development
> of an observational measure that can predict teacher effectiveness (Strong, et al.,
> 2011)

Strong's claims highlight another danger of relying too heavily upon value-added

methodologies for the disclosure of teacher quality. In addition to the "fallacy of the

successful first step," as Dreyfus calls it, Strong's work also reveals that the sense of taking a

first step can vanish at a certain point, as well. The judges to whom Strong gestures

"achieved high levels of agreement but were absolutely inaccurate." The inaccuracy of their

human judgments stands in contrast to the *uncontested accuracy* of value-added measures of

teacher effectiveness. The technical sense of good teaching, here, is no longer a means of

moving one toward an ordinary sense; it has become the only sense worth talking about. It is

no longer a *first* step; it is the *final* step.

The suggestion I raise now, and the one I will focus upon in Chapter 3, is that the

judges' agreement in Strong's example might point to precisely the conceptual gap that

classroom observations ought to fill between a technical and an ordinary sense of good

teaching. The human observers, Strong's judges, have access to a holistic understanding of

teaching that is systematically denied to the detached perspective required by value-added

methodologies. That human observers fail to predict who will or will not prove most

"effective" in terms of raising test scores is no great argument against human judgment: that

sort of prediction of rigid outcome measures is precisely a domain for which detached

information processing is designed, and in which algorithmic methods are indeed superior.

Were raising test scores actually the *final* step, such a method would also prove sufficient.

However, Duncan's rhetoric, echoed by the NCTQ, has it the other way around: such

measures are supposed to represent a step *toward* what we finally want: the ability to say

something meaningful to the involved "stakeholders" about the wholesale quality of this or

that teacher. In no way are such technical measures intended as the final word, and thus they

cannot justifiably usurp such a position. We have likewise seen, however, substantial reason

to doubt that value-added measures can function in this first-step capacity either; I will

explore in Chapter 3 whether and to what extent something like "judges' agreement" can

serve in the evaluation of teachers. For now, to return to the story of Mrs. Grady from which

this digression departed, I simply assert (with a promise to provide more substantial support

later on) that *judging* the quality of teaching, no less than the skillful practice of teaching

itself, requires the kind of "sight" that enables *flexible responsiveness* to local situations. The

notion that every excellent teacher will visibly exhibit the same behaviors or reasoning at any

given time, such that an observer could check them off on a rubric, is as unsupported as the

notion that *excellent teaching itself* is uniform and predictable in just this way. That, too, is

tautological.

To return my focus, at this point, to the *development* of skillful practice, rather than

its visibility or evaluation, Shulman's example of "Nancy," one might note, also gestures in

this direction of local, grounded responsiveness: "Nancy's pattern of instruction, her style of

teaching, is not uniform or predictable in some simple sense. She *flexibly responds* to the

difficulty and character of the subject matter, the capacities of the students (which can

change even over the span of a single course), and her educational purposes" (L. Shulman,

1987, emphasis added). I have shown that such responsiveness, in contrast to Shulman's

view, cannot be accounted for in terms of the rationalist assumption about underlying forms of "teacher knowledge." Yet the flexible responsiveness involved certainly characterizes the practice of both of the examples at work here. It follows that *developing* such responsiveness will require more than the inculcation of further or better rules and techniques. A view of developing skillful practice in the direction of such flexible responsiveness comprises the next section of this work.

<div style="text-align:center">2.5 A Phenomenological Account of Skill Development</div>

The sort of sight that characterizes and makes possible the flexible responsiveness of excellent, or, in Dreyfus's term, "expert"[26] practitioners is neither reducible to nor comprised of abstract rules or behavior. Such sight also cannot be articulated without regress, as we have seen, in the form of propositional knowledge. That these assertions hold is suggested by the failure of AI to generate the skillful behavior on the basis of attempts to formalize the commonsense "knowledge" that provides for the intelligibility and possibility of excellent practical behavior. In opposition to this view, Heidegger and Wittgenstein – and their heirs and interpreters, Taylor, Cahill, Dreyfus, Glendinning and Cavell among them – suggest that both the common-ness and knowledge characteristics of commonsense "knowledge" amount to shared public practices, agreement in forms of life, and thus, because rooted in practical dealings in a shared and structurally public world, also cannot be represented in propositional form as the rationalist assumption would require.

---

[26] Although Dreyfus uses this term pervasively, it is one of the points upon which his critics harp, and precisely for the reasons that they point out, as well as for the reason that "expertise" has a problematic history in the education literature as well (Welker, 1991), I wish to avoid it as much as possible. My goal is excellent skillful practice, not the creation of "experts" as others imagine Dreyfus to mean.

Nevertheless, the phenomenological description of skill development I offer below –

which is Dreyfus's through and through, and which he presents in a vast number of

publications over the years (H. Dreyfus & Dreyfus, 1992; H. L. Dreyfus, 1988; H. L. Dreyfus

& Dreyfus, 1986, 2005) – suggests that the sort of sight that affords the excellent practice

visible across many different domains is a common feature of life with skills in general.

Developing such sight intentionally, particularly in the realm of teaching and learning,

requires (a) the recognition that the flexible responsiveness that characterizes the teaching

practice of those we hold up as paradigm cases is neither produced nor reproducible

according to rationalist requirements, and (b) that as a consequence there must be a specific

place maintained for the developing teacher to *have* responsibility for the outcome of

situations in which the outcome *matters* to them.

In regards to part (b), it may appear as though I simply describe the practice of

"student teaching," already requisite in nearly every state for certification. But the

experience of student teaching varies widely in quality, and this variance is more dependent

on specific cooperating teachers than upon the teacher-training programs involved. Some

cooperating teachers have been known to vanish from the classroom entirely; others, to

remain to observe the developing teacher; still others, to resist relinquishing control in any

meaningful way, intervening swiftly when the developing teacher appears to struggle. As

such, the mere existence of a requirement for one or more semesters of student teaching does

not amount to a requirement for the sort of emotionally engaged coping with situations for

which one feels directly responsible. The (existential) experience of *caring* about the

successful outcome of a lesson for which *the developing teacher* assumes responsibility is, as

I will make clear, crucial to bridging the gap between a degree of skillfulness associated with beginners and a degree of skillfulness associated with professionals.

I have been highly critical of Arne Duncan and his policies throughout, and so I welcome the opportunity to point out this particular item on which his rhetoric laudably addresses the practical aspects of teacher development. He singles out as an example of a model teacher-education program a small college in Kansas:

> At Emporia State University in Emporia, Kansas, home of the National Teachers Hall of Fame, the Teachers College is the crown jewel of the school. Roughly 80 percent of students are supervised by full-time education faculty instead of adjuncts—and all elementary education professors are in the public schools every day. Senior year is a 100 percent field-based program in Emporia's public schools, where student teachers do everything from assisting with grading to sitting in on parent-teacher conferences (Duncan, 2009d).

When Duncan cites other programs, he tends to emphasize the way they use data to track their graduates. With Emporia State, the criteria he cites have to do with the instructors' practical experience and with the practical experience the school requires of their developing teachers. However, the kind of combination of "responsibility" in a situation that "matters" to a developing teacher remains absent from his description of Emporia State's program: "assisting with grading" and "sitting in on parent-teacher conferences" is not coextensive with being responsible for *assigning* grades or *conducting* parent-teacher conferences. This is not to imply that Emporia's program is in any way unsatisfactory; it is to imply that Duncan's view of where the developmental value of practical experience lies remains insufficient.

The most crucial elements of Dreyfus's description are the concession to conscious rule-following on something like the rationalist assumption as a starting condition, and the *necessity* of moving beyond such behavior at a certain point. The means by which this

moving-beyond occurs concerns emotional involvement with an outcome and the assumption

of responsibility for it. Since his account claims a Wittgensteinian-Heideggerian basis, the

role of past experiences in going before a practitioner and establishing relevance beforehand

constitutes the (nonformalizable) *sight* that examples of skillful practice seem to share in

common. Rule-following on the rationalist assumption, one must note, characterizes the

early stages of skillful development precisely because it offers a *limited approximation* of

what internalized, engaged experience provides to an excellent practitioner. It is therefore

intended as an *insufficient ersatz* form of experience itself. According to the description

Dreyfus offers, one must eventually begin *doing something different* from consciously

following rules in order to improve in the direction of the excellence for which one strives.

In a move that recalls the distinction I drew in Chapter 1 between the "technical" and

the "ordinary" senses of good teaching, Dreyfus is careful to assert that his model of skill

development

> focus[es] upon the most common type of problem area, sometimes called
> "unstructured." Such areas contain a potentially unlimited number of possibly
> relevant facts and features, and the way those elements interrelate and determine
> other events is unclear. Management, nursing, economic forecasting, *teaching*
> [emphasis added], and all social interactions fall into that very large class. Examples
> of 'structured areas' of decision-making, on the other hand, are mathematical
> manipulations, puzzles, and, in the real world, delivery truck routing and petroleum
> blending. Here the goal and what information is relevant is clear, the effects of
> decisions are known, and verifiable solutions can be reasoned out (H. L. Dreyfus &
> Dreyfus, 1986, p. 20).

The appeal to data that we have seen in Duncan's many speeches has generally to do

with the importance of *knowing* quality in teaching; student achievement data seems to

provide this means by proxy. But it has the decidedly less salutary effect of unwarrantedly

transforming teaching into a "structured" problem area, in which "the goal," if not always

"what information is relevant," remains certainly clear. Mrs. Grady's example, once more,

reminds one that teaching is always an "unstructured" domain, where a sense of a holistic situation and the relevance of the various facts involved remain inherently uncertain and undefined. Doing the right thing for Olly Neal *required*, as we have seen, that Mrs. Grady treat the situation in an unstructured fashion; that her view of the situation and her response to it seem to make logical sense in retrospect is the result of the structuring effects of hindsight, not the inherently "structured" nature of the problem area itself. Dreyfus's account of skill development is germane here precisely to the degree that the "ordinary" sense of good teaching is conceptually different from the "technical" sense, and that the "ordinary" sense is the goal of our teacher development efforts. Chapter 1 demonstrates that both of these conditions are the case.

In addition to the problems with knowing teacher quality according to seeking out behaviors or asking after rules, as discussed in Chapter 1, problems crop up for Race to the Top's view of teaching *practice* as well in this light. Since research and policy alike tend to view practical skill on the rationalist model, and since observation systems like those of Delaware and Tennessee actively incentivize adherence to such a model, Race to the Top's policies run the risk of positively stifling the development of precisely the great teachers who serve as its own justifying paradigm cases.

2.5.1 The Stages of Skill Development:

Dreyfus's view of skill development runs directly counter to the rationalist view of what *generalizing* from specifics amounts to: on the rationalist account, generalization proceeds from particular cases to the abstract rules or essential characteristics that these cases all share in common. As Dreyfus says,

> We must be prepared to abandon the traditional view that a beginner starts with specific cases and, as he becomes more proficient, abstracts and interiorizes more and more sophisticated rules. It might turn out that skill acquisition moves in just the opposite direction: from abstract rules to particular cases (H. L. Dreyfus, 1988, p. 102)

What *practical* "generalization" means on the Dreyfusian assumption would register as something like an expansion of the *domain* – the number of particular future and unforeseeable cases – which would be navigable by one's rule-following behavior, but such rule-following, being bound up with "interiorized" past experiences, would amount to the *responsive application* of "ten thousand special cases" (H. L. Dreyfus & Dreyfus, 2005). The "abstract rules," meanwhile, upon which beginners will rely, will come from the practices of recognizably excellent teachers, and as such, Shulman's research project that seeks out "the wisdom of practice" reveals itself to be not only useful but necessary to teacher development. The argument throughout has been, once more, not that Shulman's approach is *useless*, but rather that what one has when one has the codified "wisdom of practice" is not in any sense *sufficient* to its reproduction.

Rather, as one moves through the five stages of Novice, Advanced Beginner, Competence, Proficiency, and Expert, one relies more and more on experience in both establishing relevance and performing actions and less and less on the rules upon which Novices, lacking experience, *must* depend. A consideration of these stages will bear out the importance of the most crucial aspect: emotional involvement in the undertaking and responsibility for its outcome, both of which aspects are discouraged by the rationalist model of rule following and both of which are simply necessary to the Wittgensteinian-Heideggerian view of practical skill.

Dreyfus's stages of skill acquisition concern primarily the areas that skilled humans seem to navigate with greater aplomb than has so far been demonstrated in AI: the grasping of a holistic situation such that relevance stands out without requiring explicit aprioristic specification, and the ability to "flexibly respond" in such situations, to recall Shulman's characterization of Nancy. The five stages are predominantly distinguished from one another according both to the way a person *grasps* the situation and to the way the person acts in response.

2.5.2 Novice

Dreyfus's Novice stage of skill development is characterized by "learning to recognize the objective facts and features relevant to the skill," as well as "rules for determining actions based upon those facts and features" (H. L. Dreyfus & Dreyfus, 1986, p. 21). Such learning to recognize stems from an "instructor decomposing the task environment into features that the student can recognize without the benefit of experience" (H. L. Dreyfus, 1988, p. 102). For his examples, Dreyfus tends to use skill domains like chess and driving. With respect to driving, the Novice stage would include the consideration of the car's speed in absolute terms, for example, and rules based upon it. "Shift into third at 20 mph" provides an example of the way in which the Novice receives instruction – the relevant information is limited to speed, and speed is characterized as an "objective fact." No information about traffic conditions, anticipated stops or turns, hilly terrain, or other such factors that influence more experienced drivers arise for consideration here. In fact, Dreyfus points out elsewhere that following such abstract rules and facts often leads to poor performance, for precisely this reason. Notes Dreyfus, "A car stalls if one shifts too soon on a hill or when the car is heavily loaded" (H. L. Dreyfus & Dreyfus, 2005, p. 782). For the Novice, lacking the experience

necessary to discriminate in any other way, the situation is defined in terms of a narrow range of "context-free" facts, and action is based upon inflexible rules.

<div align="center">2.5.3 Advanced Beginner</div>

With "considerable experience coping with real situations," the Novice begins to perceive situations as similar to past situations where such similarity is not defined according to the similarities of any particular features. This idea recalls Wittgenstein's notion of "family resemblances," and it is one of the things with which AI struggles. Dreyfus says elsewhere that "Everything is similar to everything else in an indefinitely large number of ways. Why should we suppose that any two items should be compared?" (H. L. Dreyfus, 1992, p. xxvi). *Why* we should suppose that any two items should be compared here, invoking Cavell and Wittgenstein, goes back to agreement in forms of life, the basis of conceptuality as such, which defies attempts to reproduce such similarity-recognition *merely* in terms of abstract or formalizable features.

Although the Advanced Beginner develops the ability to recognize what Dreyfus calls "situational elements" – as opposed to "context-free features" – through a holistic, experience-driven discriminability, he or she must still consciously apply the abstract rules that such a situation calls for. For some examples of "situational elements," Dreyfus provides (a) a driver recognizing the way an engine sounds when a shift is required (H. L. Dreyfus & Dreyfus, 1986, p. 23); and (b) a chess player recognizing "aspects of positions [such] as a weakened king's side or a strong pawn structure, despite the lack of precise and situation-free definitions" (H. L. Dreyfus & Dreyfus, 2005, p. 783). Though such situational recognition appears at the Advanced Beginner stage, the performer must still rely on particular rules or maxims to navigate the perceived situation: "Attack a weakened king's

side" (H. L. Dreyfus, 1988, p. 103) or "Shift up when the motor sounds like it's racing and down when it sounds like it's straining" (H. L. Dreyfus & Dreyfus, 2005, p. 783). The recognition that characterizes this stage seems not to rely upon context-free features, but Dreyfus characterizes the type of rule-following that the Advanced Beginner displays as based on a "detached, analytic frame of mind" (H. L. Dreyfus & Dreyfus, 2005, p. 783). The greater experience provides what Polyani somewhat problematically refers to as "the good practical knowledge of the art" (Polyani, 1962) in light of which rules or maxims make sense in the first place, but the gaining of experience alone does not suffice, on Dreyfus's account, for the kind of *flexible responsiveness* that characterizes genuinely excellent getting-around in various domains; it can only account for the increasingly correct application of abstract maxims and rules, as in the rationalist assumption.

2.5.4 Competence

At this middle stage of skill development, the means of even "situational-element" recognition and maxim-application – detached rule-following according to the rationalist assumption – reaches its limit. It will be important to notice two things: in the first place, this is merely the third of five identifiable stages of skill development, and yet the whole conceptual framework of Race to the Top in accordance with which Duncan, Lemov, and Shulman propose to develop teachers and to which Kane, Staiger, Rockoff and others propose to evaluate them holds perfect detached rule-following to represent the pinnacle of excellent practice. In the second place, the very detachment according to which good teaching is imagined to proceed and render itself for evaluation represents the crux of the problem on Dreyfus's view. Dreyfus proposes that the *detachment* of the performer – teacher, we'll say – inhibits the kind of skillfulness genuinely excellent teachers exhibit;

Nancy's flexibility lies not *merely* in seeing the "situational elements" involved with particular acuity, but also in *responding* with similar alacrity. The responsiveness here requires what Dreyfus terms "involvement" (H. L. Dreyfus & Dreyfus, 1986, p. 27; 2005, p. 785).

The sort of involvement that Dreyfus has in mind stems from a particular kind of breakdown in the detached rule-following process. Specifically, Dreyfus notes that as a developing teacher (for example) becomes more and more experienced, "the number of recognizable context-free and situational elements present in a real-world circumstance becomes overwhelming. A sense of what is important is missing" (H. L. Dreyfus & Dreyfus, 1986, p. 23) Dreyfus clearly heralds the "relevance" problem here, to which problem Simon and Newell offered "heuristic rules" as a means of "cut[ting] down the search space" (H. L. Dreyfus, 1992, p. xxii) and thus limiting the number of facts that could appear for consideration. Heuristics, in AI, failed to work, and they resemble very closely Shulman's concept of "strategic knowledge," particularly in their being offered as a solution to an "overwhelming" situation in which too many facts and rules seem to be at play. These heuristics failed in the sense that whatever advantages they offered in terms of the *speed* of the response were not mirrored in the *quality* of the response: by identifying a given situation according to a perspective based on inflexible and detached means, say, probability calculations, heuristics did arrive at a response action more quickly, but in eliminating possible understandings of a situation through the employment of detached and situationally insensitive means, the programs following heuristics found themselves sometimes offering the *correct* response to a *misidentified* situation, or in other words, *failing* to produce the

most appropriate response.[27] Dreyfus's move from "detachment" to "involvement" provides

a means of accounting for the fact that excellent teachers tend to respond both quickly *and*

correctly to a wide range of situations – their responsiveness is *flexible*.

Dreyfus characterizes the competent performer thusly:

> In general, a competent performer with a goal in mind sees a situation as a set of facts. The importance of the facts may depend on the presence of other facts. He has learned that when a situation has a particular constellation of those elements a certain conclusion should be drawn, decision made, or expectation investigated (H. L. Dreyfus & Dreyfus, 1986, p. 24).

Competent performers continue to see situations as sets of facts; but the situational

facts that must arise for consideration are *produced* by the performer's having "a goal in

mind." But very broad and general goals – even those applicable to a "technical" sense of

good teaching such as "raise reading scores," to say nothing of the "transform students'

lives" goal of the "ordinary sense" – provide too little guidance in particular situations. Even

imagining such goals (erroneously) as composites of smaller, simpler goals ("benchmarks")

is of no help here. What is at stake, even on this view, is precisely *which* "benchmark" ought

to allow the relevance to stand out. As Dreyfus puts it:

> Naturally, to avoid mistakes, the competent performer seeks rules and reasoning procedures to decide which plan or perspective to adopt. But such rules are not as easy to come by as are the rules and maxims given beginners in manuals and lectures. Indeed, in any skill domain the performer encounters a vast number of situations differing from each other in subtle ways. There are, in fact, more situations than can be named or precisely defined, so no one can prepare for the learner a list of types of possible situations and what to do or look for in each. Students, therefore, must decide for themselves in each situation what plan or perspective to adopt without being sure that it will turn out to be appropriate (H. L. Dreyfus & Dreyfus, 2005, p. 784).

---

[27] A wonderful, if anecdotal example: a friend of mine recently posted the following status on Facebook: "Dear iPhone autocorrect: I *never* mean 'Holy Shot.'

The experience of deciding remains *rationalist* here, but the *awareness* of the inherent uncertainty produces the involvement necessary to further development, as Dreyfus notes: "While [the performer] both understands and decides in a *detached* manner, he finds himself intensely *involved* in what occurs thereafter" (H. L. Dreyfus & Dreyfus, 1986, p. 26). The conscious decision-making, the awareness of an insufficiency on grounds by which to discriminate, and yet the *need* to make a choice in order to act at all Dreyfus calls the "combination of nonobjectivity and necessity [that] introduces an important new type of relationship between the performer and his environment" (H. L. Dreyfus & Dreyfus, 1986, p. 26).

The involvement, then, amounts to a recognition on the part of the performer that the success or failure of the application of rules and maxims upon which his or her action will now proceed is directly dependent upon *his or her choice of perspective* according to which such maxims or rules are solicited in the first place. Since this choice of perspective cannot *itself* be produced according to abstract features and rules, the quality of the choice belongs ineluctably to the performer rather than the rules. That distinction, as Dreyfus notes, sets the performance of the competence-stage performer apart from earlier stages:

> Prior to this stage, if the rules don't work, the performer, rather than feeling remorse for his or her mistakes, can rationalize that he or she hadn't been given adequate rules. But, since at this stage, the result depends on the learner's choice of perspective, the learner feels responsible for his or her choice (H. L. Dreyfus & Dreyfus, 2005, p. 784).

Such a description recalls, briefly, my much earlier example of ordering coffee in Switzerland. I know several phrases that will have the effect of procuring coffee. But I am also aware that some phrases are more formal than others, more appropriate for linen-tablecloth restaurants than sitting in a friend's kitchen. I find myself, however, standing at a

coffee kiosk in a train station. I am painfully aware that I cannot *sense* the level of politeness appropriate to this situation, and thus whatever phrase I will offer will be offered "without being sure that it will turn out to be appropriate." But precisely my uncertainty *invests* me in the way the barista will respond. The character of her response, in combination with my general sense of the politeness level of the phrase I attempt, will, I assume, tell me something about the *perspective* on train-station coffee kiosks I ought to adopt in order to produce the appropriate locution for ordering coffee. I freely understand the phrase I use to be *mine*, and also recognize the insufficiency of the means at my disposal for determining its correctness; to the degree that not appearing to be some idiot tourist matters to me (it matters to me quite a bit), I am *involved* in the outcome of the exchange in a way I would not be if I were simply reciting something out of a Lonely Planet phrasebook.

The *responsibility* for the choice of perspective and the resulting *involvement* in its consequences have profound effects on the learner's ability to develop further, on Dreyfus's account. Since Dreyfus's understanding anchors itself on Heidegger's "temporality" concept, in which being-in-the-world inherently involves a futural "thrown project," structured by what Bourdieu calls the "active presence of the past," responsibility and involvement play an intensifying role with respect to memory formation.[28] Says Dreyfus: "An outcome that is clearly successful is deeply satisfying and leaves a vivid memory of the plan chosen and of the situation as seen from the perspective of the plan. Disasters, likewise, are not easily forgotten" (H. L. Dreyfus & Dreyfus, 1986, p. 26). The way the *choice of*

---

[28] Again, as this is a phenomenological description, Dreyfus draws only on the reader's acknowledgment of the descriptions for justification., rather than any neuroscientifically "accurate" account of memory.

perspective itself constitutes the past that goes "before us, organizing the way the next events show up for us" (H. L. Dreyfus & Dreyfus, 1986, p. 88) encourages the repetition of successful perspective-choices in similar situations and discourages the repetition of unsuccessful ones, where situations and perspectives alike have the character of holistic, nonrepresentational paradigms.

Such *personal* responsibility and involvement remain absent from the rationalist or detached view of decision-making, even with regard to the choice of perspective, and because of that, Dreyfus claims that further skill development on the rationalist view stops here. In examining now the fourth and fifth stages of skill development, I wish to specifically offer some salient contrasts between Dreyfus's view and the rationalist one, on the specific issue of the role that Duncan ascribes to student achievement data with regards to teaching.

2.5.5 Proficiency

Earlier I claimed that Dreyfus's model seeks to account for skill development beyond which the rationalist view seems to founder: the ability to grasp a holistic situation such that the relevance stands out immediately, and the ability to "flexibly respond" to relevant situational features as a matter of course. In moving to the Proficiency stage from mere competence, Dreyfus holds that the proficiency involved amounts to precisely the ability to grasp a situation and its relevance immediately. Where the rationalist view seems to bog down on this point, as failures of finding heuristics capable of replicating skillful human practice suggest, the immediate grasping of a situation and its relevance at once results from *involved* experience with making existentially uncertain choices of perspectives in situations where the practitioner bears responsibility for the choice. The ability to "spontaneously see

the point and important aspects of the current situation" (H. L. Dreyfus & Dreyfus, 2005, p.

787) I will call – and have called – a particular kind of seeing – say, insight.[29] Kisiel,

repeating Heidegger's interpretation of Artistotle's *Nicomachean Ethnics 6,* has Heidegger

saying:

> The uncovering of principles must be without speech. Here, it is simply a
> matter of "bringing ourselves before the matter itself," traversing the way that leads
> directly to it. Thus, the resolute choice of my concrete situation of action, which takes
> into account the various circumstances entering into the situation, abruptly terminates
> such an accounting and culminates in a simple "oversight" which takes charge of, and
> acts on, the situation in the "blink of an eye" (Augen-blick), in the instant of insight
> (Kisiel, 1993, p. 286).

Dreyfus's descriptions of the instantaneous and "spontaneous" grasping of the

situation are *strikingly* similar:

> Because of the performer's perspective, certain features of the situation will
> stand out as salient and others will recede into the background and be ignored. As
> events modify the salient features, plans, and expectations, and even the relative
> salience of features will gradually change. No detached choice or deliberation occurs.
> It just happens, apparently because the proficient performer has experienced similar
> situations in the past and so associates with present situations plans that worked in the
> past and anticipates outcomes that previously occurred.

However, if Dreyfus's account of the "just-happening" involved in grasping a

situation and its relevance at once directly recalls Heidegger's, the distinction from

Heidegger's account quoted above lies in the *action* portion of the Proficiency stage. Where

Heidegger describes an "oversight that takes charge of, and acts upon, the situation 'in the

blink of an eye,'" Dreyfus reserves that sort of action for the final stage. In Proficiency, on

---

[29] In *Mind Over Machine*, Dreyfus labels what the proficient performer has acquired "intuition,"
which he explicitly equates with "know-how" (H. L. Dreyfus & Dreyfus, 1986, pp. 28-29). He
notably does not include either of those terms in later publications. Know-how, I think, is too easily
confused with propositional knowledge and "intuition" too often associated with both a *mysticism* and
a *mentalism* that Dreyfus is at pains to avoid. I don't know why he didn't go with either "sight" or
"insight" himself, given that term's Aristotelian-Heideggerian legacy.

Dreyfus's view, the *insight* afforded by experience with involved perspective-choosing in situations for whose outcomes performers were responsible *simply* plays the role that heuristics were designed to play with respect to establishing *relevance*. The insight required for relevance merely enables the fluid and correct functioning of detached decision-making with regard to a particular avenue for action: "The proficient performer, while intuitively organizing and understanding his task, will still find himself thinking analytically about what to do" (H. L. Dreyfus & Dreyfus, 1986, p. 29).

To note one important point that will arise once more in Chapter 3, the sort of insight attributed to proficient practitioners has an *immediate* character: the situation as it is, including relevant features, stands out immediately. I take Arne Duncan to be claiming something similar when he talks about immediately knowing the quality of a school: "We know what success looks like. I see it the minute I enter a school" (Duncan, 2009e). Dreyfus's model and the discussions of Wittgenstein and Heidegger in this section both suggest that knowing what success looks like such that its recognition is immediate does not necessarily entail the possibility of spelling out this "looking-like" in any sort of articulable proposition.

2.5.6 Expertise

In the phase of Expertise, Dreyfus adds the action to the insight, as suggested in the Heidegger passage above: "The proficient performer, immersed in the world of skillful activity, sees what needs to be done, but decides how to do it. The expert not only sees what needs to be achieved; thanks to a vast repertoire of situational discriminations, he or she also sees immediately how to achieve the goal" (H. L. Dreyfus & Dreyfus, 2005, p. 787). The sense in which Dreyfus can claim, as he does elsewhere, that such excellent practice can be

called "skillful coping" has to do with the *transparency* of the situation and its relevance.

Drawing upon the practical situations in which most people are most familiar and quite

skilled, Dreyfus says, "We usually don't make conscious deliberations when we walk, talk,

drive, or carry on most social activities. An expert's skill has become so much a part of him

that he need be no more aware of it than he is of his own body" (H. L. Dreyfus & Dreyfus,

1986, p. 30).

In short, at this level of skill, one finally comes upon Nancy's "flexible

responsiveness," where correctly responding to most situations is simply a matter of course,

something that (in general) demands less cognitive energy than the Novice's rule-following.

The Novice's rule-following is that of inflexible logical necessity: once presented with a

given set of rules and conditions, the Novice "no longer [has] any choice," as Wittgenstein

says. Nancy's skillful coping reveals a different kind of following a rule in which one does

"not choose – [she] follow[s] it blindly," and yet for all this, her practice is something that, in

terms of specific behaviors, resists "rigid" predictability in any "simple sense."

In the particular case of teachers, to suggest that at this stage that excellent practice

usually *just happens* is to imply neither that great teachers do not *think* at all, nor even that

they never *deliberate* or consciously reason their way to appropriate actions. It is rather to

suggest that these conscious mental deliberations are *limited* to unusual or particularly

problematic situations. When Dreyfus characterizes the competent performer as

"overwhelmed," the overwhelming nature of competent practice has to do with an *inability* to

do *anything but* reason consciously about both the correct perspective according to which to

grasp the situation and the action that ought to result, given the relevant features exposed by

the chosen situation. That by the expertise phase, most aspects of teaching do not require the

same level of conscious engagement only helps the expert allocate such conscious reasoning to the situations that genuinely require it.

Here once more we encounter the two different senses of following a rule. According to the rationalist view, the rules exist anyway, independent of practices; following a rule amounts to "engaging mental wheels with the already-existing rails," as McDowell puts it. On the other view, the sense of conscious *effort* drops out of McDowell's formulation. To be skillfully engaged in practical dealings is already to have engaged oneself in following rules, and not so deterministically as the "rules laid to infinity" locution would suggest. The Dreyfusian, Wittgensteinian, and Heideggerian considerations in this Chapter have helped to demonstrate that such a rules-as-rails view remains insufficient to account for the regularity or recognizability of our actual rule-following practices. On the rationalist view, the sort of flexible responsiveness of the type that characterizes Dreyfus's final stage of skill development proves impossible – situational relevance and situational similarity each presuppose the other, and thus the transcendent logical conditions that a rule is supposed to *be* cannot establish any traction with regard to *particular* teacher choices or behaviors.

In the Dreyfusian model, no conscious deliberation appears to be involved at all in skillful responsiveness, which to the rationalist (among others), would also seem to exclude the demonstration of "flexible responsiveness" as an instance of following a rule. Such a condemnation, though, assumes precisely the transcendently-anchored picture of rules required by the rationalist view. Following a rule, however, might amount to no more than doing similar things in similar situations, whose similarity is a matter of family resemblance with a basis in agreement in forms of life, and this is not at all to claim that no *thinking* is involved in such rule-following. As Wittgenstein says, "One follows the rule mechanically.

Hence one compares it to a mechanism. 'Mechanical' – that means: without thinking. But *entirely* without thinking? Without *reflecting*" (Wittgenstein, 1983, VIII-61). In distinguishing between reflection in particular and thinking in general, Wittgenstein hints in the direction Dreyfus would recognize as a division between knowledge-that and knowledge-how, where the former both implies and requires "detachment" or "disengagement" and the latter implies and requires the opposite. Doing, on an involved basis in which one's being is implicated, *is also* thinking.

<div align="center">2.6 Arne Duncan and Achievement Data – a Reprise</div>

In light of the previous paragraphs, I ought to return to one of the salutary aspects of using student achievement data that Duncan cites. In the speech directly addressing the role of data with respect to teachers, Duncan says, "They want to know exactly what they need to do to teach and how to teach. *It makes their job easier and ultimately much more rewarding.* They aren't guessing or talking in generalities anymore" (Duncan, 2009b, emphasis added).

In considering Dreyfus's description of skill development, one can see that "easier" teaching takes at least two forms. At the point of competence, developing teachers (and policies meant to address teacher development) arrive at a crossroads: either the "problems" that confront teachers will be taken as "structured" or "unstructured." The examples of Nancy and Mrs. Grady that continue to arise, as well as Duncan's own propensity to reify an ordinary sense of teaching, suggest that the sort of realm that we expect teaching to address is indeed "unstructured."

Coping with unstructured problem areas requires the existential choice of perspective in which the teacher bears responsibility for that choice and for the outcome as well. The way that Duncan proposes to use student achievement data, however, precisely converts the

realm teaching addresses into a "structured" one, along the lines of a puzzle or a mathematical manipulation. The teacher's goal, here, in light of which situations have their relevance, requires no insight and no choice; it is simply determined beforehand. Therefore, the data that Duncan wishes to use *will indeed* make the teachers' jobs easier, but only through transforming the sort of teaching that we ostensibly want into the "technical" version characterized by the "preponderant criterion" of raising test scores. With the implication that every practical situation ought to be grasped in terms of its relation to that criterion, such that relevance never changes, teachers can never undergo the types of conditions that develop the flexibility in either vision or responsiveness necessary to the "ordinary" sense of great teaching. All of this would be unproblematic if we *knew* that raising test scores *sufficed* as the goal at which teaching aims; but as Duncan's own words and the examples he cites indicate, we *know* exactly the opposite.

The fact that the projects of Lemov and Shulman, which share the rationalist assumption of Duncan's Race to the Top, prove necessary to the *development* of ordinary-sense good teachers is too easily confused with a view that the model upon which they rely also remains necessary to the *excellent practices* themselves of skillful and experienced teachers. Returning to Lemov's art analogy for a moment, one can surely see the sense of his attribution to Michaelangelo of "foundational skills": "You learn to strike a chisel with a mallet. You refine the skill with time, learning at what angle to strike and how hard to drive the chisel." His picture of familiarizing oneself with the tools and media of an artistic craft focuses on the piecemeal aspects thereof: angles and forces. If this is to imply that the *David* would have been impossible without this type of attention to foundational growth in Michaelangelo's youth, Lemov's argument presents no problems. If this is to imply,

however, as I understand it, that skillful activity involved in the production of the *David* actually consisted of the *application* of these foundational skills *as* piecemeal actions, then this argument will not serve.

In terms of sufficiency and necessity, with the problems of AI and the considerations of *involvement* and the role of *experience* in mind, I might put the issue like this: "foundational skills," in Lemov's terms, or the sort of "teacher knowledge" that Shulman, Ball, and Hill all seek in their research, both of which amount to the "best practices" approach to educational reform touted by Duncan and many others, remain *necessary* to developing excellent ordinary-sense teachers; such an approach is also *sufficient* to the actual production of relatively *competent* practice. But they are neither necessary *nor* sufficient conditions for the ordinary sense of good teaching in practice, and in fact, in order for teachers to develop their skillful practice beyond mediocrity, such approaches must be eventually set aside.

In this way, teaching and many other skills resemble bike-riding: training wheels are most often necessary to the development of the skill; with training wheels on, even a novice bike-rider can approximate competence insofar as the rider stays aloft, as it were; but bike-riding as a skill is more than remaining upright, and in order to develop the sense of balance required for proficiency, the training wheels must come off. In just this sense, that proficient bike-riders do not use training wheels, the conditions necessary to the *development* of the skill do not remain necessary to its *proficient practice*, though the training wheels *suffice* to produce competent riding no matter the experience level of the rider. (It may be important to note that in no case would we consider anyone riding with training wheels a "competent bike rider" – the very fact of the training wheels constitutes the salient criterion to the contrary.)

An important aspect of this analogy is that the step from never having ridden a bike before to riding with training wheels is *substantially smaller* than the step from riding with training wheels to riding well in their absence. Dreyfus helps one to see, also, that moving from, say, never having held a chisel in one's hands before to a limited ability to chisel does not lead in any *linear* fashion to the production of the *David*. Adapting a criticism from Maya Bar-Hillel, not only is the step from not being able to do anything *at all* to being able to do it a *little bit* so much smaller than the step from being able to do it a *little bit* to being able to do it *well*, but making that second leap is no mere *technical* extension of the first step. What Duncan images being replicated when innovations are "taken to scale" is doing something *well*. But by transporting the innovations in terms of formalized rules and behavioral facts, educators transport only the conditions for doing it *a little bit*.   This does not suggest that scaling-up educational successes is impossible; it does suggest, however, that merely communicating "best practices" will not, in itself, suffice. Dreyfus allows us to see that making the second leap necessarily involves developing an "entirely different relation" to the skillful activity itself: it *requires* engagement or involvement, and cannot be achieved through the detached or disengaged means that suffice for mere competence and are in fact necessary to beginners.

Despite their shortcomings in terms of a full account of teacher practice, as I have noted, Lemov and Shulman's methods both produce actual, measurable results: instilling behaviors, techniques, best practices, and abstract knowledge in newcomers to the domain proves effective to a certain point. The correlation that Hill and Ball demonstrate between their construct of MKT and student math achievement evinces this fact. But the fact that one has traveled some distance *away* from incompetence ought not to be confused with having

made meaningful progress *toward* excellence. Stuart Dreyfus puts it colorfully in quipping that such a confusion is like a man climbing a tree and claiming progress toward the moon: true, perhaps, but profoundly misleading.

It must be noted, too, that the above criticisms only bear upon Duncan's Race to the Top project insofar as the *examples* upon which he draws in order to rally popular support are invariably examples of "ordinary"-sense great teaching. That he chooses such examples suggests that his audience—the public, broadly conceived—craves excellent teachers, and also that Duncan recognizes this desire in his audience. The distinction between "technical" and "ordinary" rises again at this final point because Duncan effectively pledges "ordinary" excellence through "technical" means, and the problems with the rationalist assumption undergirding the "best practices" approach reveal the implausibility of this project.

Were his stated goal *not* the widespread development of excellent teachers, but instead, say, the *aggregate* "improvement" of the teaching corps visible only in terms of central-tendency data, this "technical" approach might suffice. *Despite* his rhetoric about excellent teaching, there is certainly evidence that this is in fact the aim of Race to the Top,[30] and raising measures of central tendency can be achieved, after all, by *increasing* the number and quality of teachers *above* mean "effectiveness" *or* by *decreasing* the number of teachers *below* the mean, either through improvement or weeding-out.

---

[30] Important criteria for success that Duncan cites include, for example, shrinking "achievement gaps" between racial and class defined groups. "Achievement" has real-world valences in terms of eventual salaries, quality of life measures, and so on; but the gaps *and* their closure are primarily revealed and referenced by way of student achievement data. And since the gaps are gaps between *mean scores*, "improvement" according to such a metric means shrinking the distance *between means*. Insofar as teaching is taken to be a cause of learning, the shrinking of the gap would indicate (Duncan would claim) that teachers in general are of higher overall quality than they were previously.

My intent in raising Dreyfus's model of skill development is to demonstrate, if "competence" represents the mean, (a) that these two approaches are crucially *different* from one another and (b) that the methods involved in *decreasing* the number of sub-par teachers, if applied to the teaching profession as a whole without regard to the role of internalized, involved, and non-formalized *experience* in moving from mediocrity to excellence, will have the additional and unintended consequence of *hindering* teacher development above the mean. Kristof's understanding of the Chetty article cited in the Introduction illustrates how easily these two different things are conflated: "This latest study … not only underscores the importance of education but also illuminates how we might improve schools. An essential answer: more good teachers. Or, to put it another way, fewer bad teachers" (Kristof, 2012b). The latter formulation is not "another way" to "put it" – the "it" is different, and pursuing a "fewer bad teachers" agenda via the policies of Race to the Top will in fact *stifle* the "more good teachers" agenda to which it is imagined to be equivalent. According to a *technical* view – visible in measures of central tendency in terms of student achievement data – "more good teachers" and "fewer bad teachers" appear identical. They are not.

Ultimately, Race to the Top's proposals in the realm of teacher training and evaluation will indeed, as Duncan promises, have the effect of making "their jobs easier" – but the policies will accomplish this feat not by engendering the skillfulness required to navigate the unstructured realm of teaching and learning, but instead through eradicating by fiat all unstructured characteristics and thereby also *preventing* the development of the very skillfulness embodied in his many examples. Duncan presents himself in his many speeches as performing the grim but necessary duty of demanding *more* from teachers and their training programs; he is in fact demanding *less*. The pact his rhetoric makes with the public

suggests a "more good teachers" approach; his policies, though, are aimed at a "fewer bad teachers" solution. The statistical identity of these two agendas belies their profound differences. A "more good teachers" effort *includes* the necessity of conscious deliberation in the development of flexible responsiveness; a "fewer bad teachers" approach, by virtue of misunderstanding the relation between deliberate reasoning and skillful practice, *actively stymies* the development of flexible responsiveness at all. In this respect, Dreyfus's caution about the misguided fears of AI, with its attendant rationalist assumptions, resounds powerfully: "Our risk is not the advent of superintelligent computers, but of subintelligent human beings" (H. L. Dreyfus, 1992, p. 280).

CHAPTER 3: COMMUNICATING GOOD TEACHING

"You cannot make an omelet without breaking eggs. No, nor by breaking eggs."

Stanley Cavell

In Chapter 1, I referred to the problem of ascertaining and reporting on the quality of teaching according to test scores alone; I also noted that Duncan shares this sense of achievement data's inadequacy, which prompts him to call for "multiple measures" of quality in teacher evaluations. In Chapter 1, likewise, I showed that in order for "multiple measures" to be meaningful, measures involving classroom observations would need to remain at least somewhat distinct from measures involving achievement data, since even reformers like Duncan hold that more remains necessary. Lastly, in Chapter 1, I demonstrated that despite these concerns, the forms and criteria of classroom observations are *directly* tied to student achievement data. In this section, I wish to reintroduce the claim that classroom observations so conceived cannot accomplish the complementary feats that policy-makers desire. I wish also to raise once more the issue of Duncan's claims to an immediately apparent valence of teacher or school quality, and to suggest in more substantial form with respect to this fact that such claims appeal to *different criteria* of quality than those revealed by achievement data.

In developing this view further, I wish to harken back to the Wittgensteinian and Heideggerian thinking of Chapter 2. Specifically, I suggest that the criteria to which Duncan's immediate-visibility claims appeal remain inherently bound up with the notion of shared practices, or agreements in forms of life, which are *non-formalizable in principle*. As a consequence, it follows that, while such criteria can appear in *descriptions* of teacher practice, both good and bad, and can even be characterized to a certain degree, as in the notions of "flexible responsiveness" or its opposite, "excessive rigidity," for example, such

characteristics function, one might say, as holistic *paradigms* rather than composites of

otherwise atomizable criterial features. The claim here is that "flexible responsiveness" is a

family-resemblance concept rather than a *particular set* of behaviors or mental states

identifiable according to the presence of any one (or more) essential feature(s). Thus, its

recognition or visibility requires the sharing of shared forms of life; it is structurally public,

to use Glendinning's phrase, but not therefore *universal* in any purely objective sense,

whatever that might mean. It is possible, on the view I will develop, to specify the criteria

according to which we call teachers "good" and "bad" (and all the degrees between), but it is

not possible, as state evaluation policies increasingly hope, to remove the faculty of judgment

or the need for interpretation from the practice and performance of teacher evaluation. In

fact, as I hinted in Chapter 2, the notion of "agreements in forms of life," or of the "shared-

ness of shared practice," and the "us" implicated in every instance of structural publicness,

are shot through with *dis*agreement, as well. The absence of logical necessity (or any other

causally-determinative third term) means that such disagreement, reproof, or disavowal is a

constant and standing possibility.

Assuming the impossibility of eradicating the need for interpretation and judgment in

matters of knowing teacher quality, the evaluation of and reporting on the quality of teachers

and of education generally will require a means of limiting the possibility of something like

pure subjective error or willful interpretive blindness or whimsical favoritism. In fact,

Duncan's citation of the New Teacher Project's statistic that 99% of all teachers are rated the

same might be taken as an expression of his fear of just such whimsicality or blindness in the

evaluation process. Although his turn to achievement data provides a means of dealing with

this fear, we have seen the conceptual problems that arise from attempts to apply it in the

realm of education. The shortcomings of using teacher value-added to address the problems arising from subjective interpretation of teacher quality do not mean that *nothing* can be done to address these problems, but it does mean that turning to "objective" means will not suffice. Joseph Schwab's conception of "the practical" in matters of education, and particularly his notion of deliberation, might be turned to the service of minimizing the risk of interpretive error without turning a blind eye to the necessity of interpretation wholesale. Schwab's deliberation holds out also the possibility of managing the risks associated with the standing possibility of disagreement and disavowal.

Having addressed the insufficiencies of focusing upon student achievement data, as conceived according to the freezing of relevance required by the rationalist assumption, in terms of both *knowing* and *doing* the excellent teaching toward which Duncan ostensibly labors, it will follow in self-evident fashion that using student achievement data in order to ground or justify claims about teacher quality will likewise prove insufficient. This Part will revisit and recollect the arguments around the use of student achievement data in teacher evaluations with an eye specifically to the role of justification in such evaluative purposes. However, I should say at the outset that the character of Chapter 3 will diverge from the earlier portions of this project: where the first two Parts were aimed at leveling a conceptual test at the teacher-quality provisions of Race to the Top – will the rhetorical goals and the means of achieving them match up, and if not, why not? – this third and final piece will extend the discussion in the direction of the ramifications of the first two Chapters, offering the outline of a potential alternative. For this reason, it will be more speculative, one might say, and perhaps less critical in nature.

In dealing with issues around reporting on – *telling* – teacher quality, broader questions arise, primarily pertaining to, as I suggested above, the intertwined topics of the *measurement* of human endeavors and the *justification* of evaluative claims. The first two sections of this Part address the multiple kinds of "data" and evaluative claims at work in Duncan's discourse, and they explore in limited fashion the kinds of justification that are both possible and requisite for each. A guiding suspicion of mine as I embark upon this Chapter – and in putting it into words, it seems almost too obvious to bear mentioning – is that different kinds of measurement, even of ostensibly the same thing, are bound up with different human purposes (different futural projects), and that equally entailed in such an idea is that differing forms of justification will be appropriate with regards to these differing forms of measurement and the differing purposes that give rise to them.

This suspicion stands counterpoised to what I see as Race to the Top's view of the connection between data, measurement, and evaluation: that the sort of data, measurement, and evaluation necessary to distinguish for the Secretary of Education the top 10% of California's teachers from the bottom 10% ought also and unproblematically to function at a microcosmic level as well, as in communicating something meaningful to parents about the relative quality of their child's teacher. Such a claim would amount to (and be a species of), I propose, a sort of universalized, formalized view of education as such, in which (a) what the various parties to the educational endeavor desire will always (or should always) be exactly the same, (b) the means for its achievement or effecting will likewise (because causally necessary) be uniform, and (c) the form and content of its expression will further always conform to a single axis or standard.

In Chapter 1, I somewhat clumsily asked of Duncan's statement about California's 300,000 teachers, "*Where,* precisely, is the need for such data that the California example is meant to make evident?" In this Chapter, I hope to better express the impulse behind my earlier objection: that the need to be answered by distinguishing the best 30,000 teachers in California from the worst is simply incoherent from any perspective *except* that of the state or federal government, and that there is no reason to assume that the means of answering a state or federal need will also or equally suffice to answer a school-specific, community, or parental need.

As Chapter 1 in particular has revealed, however, Duncan himself, among many others, often appeals to *differing* and *incommensurable* criteria in the discussion of teacher quality. That two differing concepts of teacher quality pervade the discourse has been established, as has the fact that Race to the Top and state policy has insufficient awareness of that fact. The question confronting us in this final Chapter is: given the differing human purposes (understandings of being) in terms of which data, measurement, and evaluation make their claims upon us, how are we to navigate these differences in order to say something meaningful to someone else (to whom, exactly?) about whether or not this or that teacher is good at teaching?

3.1: Judgments Immediate and Certain: Differing Senses of Measurement

In the Introduction, with broad reference to Race to the Top generally, I asked: "If these means remain insufficient to demonstrate or secure the conceptual recognition of good teaching we wish to reward, *on what basis* is good teaching so immediately visible in Mrs.

Grady's story? If not communicable in terms of specific practices or measurable outcomes,

how else could our ordinary concept of "good teaching" become visible?"

Speaking broadly, this section marks an attempt in the direction of answering

questions around the relative usability of both the *immediate* visibility of teacher quality to

which Duncan occasionally alludes and the version available in terms of student achievement

data; in other words, this section explores the bases upon which such immediate visibility

finds its foundation and their relation to the basis of teacher value-added in particular and

student achievement data in general.

To recollect the kinds – and examples – of discourse that give rise to the suspicion

that multiple concepts and sets of criteria pervade discussions of teacher quality, I here once

more juxtapose Duncan's view of the role of student achievement data in evaluating teachers

with his insistences that educational quality – at least in some instances – requires or is

amenable to other forms of expression.

Attending first to instances in which student achievement data plays no role in his

own claims about educational quality, I return to two different types of statements that

Duncan makes. The first type amounts to a claim of educational quality's immediate

visibility in the present.

> We know what success looks like. I see it the moment I enter a school. It's
> clean, orderly, the staff is positive and welcoming, and the kids and the classroom are
> the focus. I see award-winning school work on the walls. I see discipline and
> enthusiasm in the children. I see parents engaged and teachers collaborating on
> instruction (Duncan, 2009e).

While I have cited this passage earlier in this project, I wish to highlight at this juncture the

fact that Duncan's claim to the immediate visibility of "success" is accompanied by an

appeal to the criteria according to which this success emerges, none of which is the type of

feature captured in the measures of student achievement currently employed in the evaluation of teachers and schools. These criteria include the presence of "award-winning school work on the walls," "discipline and enthusiasm in the children," engaged parents and collaborating teachers. Duncan offers these factors as a *description* of "what success looks like." Notably, Duncan offers this description of "what success looks like" in the present tense: the school to which he refers is succeeding *right now*, in and according to the very characteristics that he highlights in his description.

The second type of statement in which Duncan professes to recognize educational quality without recourse to student achievement data as Race to the Top conceives it has to do with present-tense *adult* success. In other words, the teacher quality these statements recognize lies far in the past, and is expressed as a putative causal condition of present adult achievement. In earlier Chapters, I aligned these statements, as well as his claim pertaining to immediate visibility discussed above, with the "ordinary" concept of good teaching, and I belabored, in particular, the Secretary's tendency to exhort his audience to recall excellent teachers from their own lives. In this space, however, I wish to return to his comments about the people in his neighborhood, and his mother's effect on them, in particular, with the intent of drawing out the way in which the quality of teaching long gone, so to speak, comes to visibility only years later, in terms of the success of the then-children in their adulthood endeavors.

In elucidating the importance of good teaching, Duncan points to the present *adult* success of his own local cohort, and he offers a narrative about the education of this group, an educative endeavor to which he attributes, in causal fashion, the success he describes. In describing this success, he says that,

from the group of friends I grew up studying with and playing ball with, from one street corner at 46th and Greenwood, emerged literally a brain surgeon, a Hollywood movie star, one of my top administrators at the Chicago Public Schools, and one of IBM's international corporate leaders (Duncan, 2009a).

He explains the remarkable concentration of successful adults in terms of the quality of the educational endeavors – defined broadly, here, rather than confined to teachers at particular grade levels – of the adults overseeing their childhood experiences:

How did this happen? Because these children, despite tremendous poverty, despite staggering neighborhood violence, despite challenges at home, had my mother and others in their lives who gave them real opportunities, real support and guidance over the years, and had the highest expectations for them (Duncan, 2009a).

The presence of a certain kind of adult in the lives of these children – those who "gave them real opportunities, real support and guidance over the years, and had the highest expectations for them" – is offered here in answer to a question of *how* the noteworthy success of this group was possible. In this scenario, the claims regarding the quality of the educative experiences in question here receive their justification from the indisputable present-tense fact that Duncan's peer group has achieved remarkable (professional) success. The specific nature of the causal conditions for this success involve a collection of attitudes, behaviors, and actions that Duncan lumps together in terms of providing "real opportunities, real support and guidance over the years," and the holding of the "highest expectations." The criteria that render visible the excellence of the education that Duncan received thus rest in two sites: the current success of his cohort (as well as the statistical unlikeliness of such a concentration of success) and Duncan's own attribution of this success to the sort of educative adult attention lavished upon this cohort in the past. Olly Neal's story bears the

marks of this species of appeal, as well: a person's striking success is attributed in more or less vivid detail to educative experiences in the past.

The two types of criterial appeals discussed above share the salient feature of an expectation of sufficiency in the absence of the kind of student achievement data envisioned in Race to the Top policies. Neither type of appeal requires reflection or support from any other source; they themselves make visible the quality of the teaching with which they deal. Both types of appeal are also, perhaps obviously, grounded in present-tense observable phenomena. In the first case, the phenomena that count as relevant to the revelation of teacher quality are (immediately) visible *in* schools: Duncan focuses on aspects of the building itself (cleanliness, orderliness), aspects of the students (enthusiastic, disciplined), aspects of the teachers and administrators (positive, welcoming, collaborative, kids-first disposition), and aspects of parental relations (engaged). In the second case, the phenomena that count as relevant are, one might say, restricted to the axis of the student. The primary phenomenon in the second instance is the now-successful *former* student. The success[31] of this former student frames the specific description of the teaching to which the former student attributes his or her present level of achievement. Thus, when Duncan recalls his mother and the other adults that shaped the lives of his cohort, his citation of the "real opportunities, real support and guidance," and "highest expectations" that these educators

---

[31] The adult's success is, one might say, indisputable. Any claim about high teacher quality rests on the acceptance of the former student as a paradigm case of adult success; to dispute it is to fatally undermine the argument as a whole. Since becoming a "Hollywood movie star" and rising to the top of the Arkansas legal field represent widely variant pictures of "success," it is safe to say that "adult success" is a family-resemblance concept, one that also requires agreement in forms of life or a kind of shared background. This fact will be important.

provided to their young charges, these become the names of the causes of the adult success

that Duncan's audience witnesses in the empirical fact of this man in a suit and tie

declaiming on American educational policy. An audience, whether it is Kristof himself, or

Kristof's eventual readership, or Duncan's listeners, takes the former student's perspective

on his or her education and biography as authoritative, as worthy of trust.[32]

In examining the two types of criterial appeals that require no student achievement

data as Race to the Top conceives of it, one can already make an important distinction in

terms of the nature of the visibility of the teacher quality under investigation. In the first

instance, the quality of the education in question is taken to be *immediately visible* to a given

observer; as such, the means of making this quality apparent to a wider public (how wide is

this public, exactly?) rely upon a *description* of the observed features – the school building,

the attitudes and behaviors of students, staff, teachers, and administrators. The assumption is

that, through the process of describing what one sees, as Duncan does in his general case, one

captures and shares not only the empirical facts, as it were, but also the quality of the

education itself, the value component. No less visible than (because inherently bound up in)

the "objective" features of an educational endeavor is the quality thereof.

In the second instance, the quality of the education in question is taken to be, one

might say, more *mediately visible*, visible only indirectly, only in light of the former

---

[32] Imagine, for a moment, the kind of person who would *find themselves in a position to disagree* with a given former student about the *causes* of that former student's success. Of 300 million Americans, for example, how many could claim such a position in relation to any other person, say Olly Neal? If one were *not* in that position, as I am not with either Duncan or Neal, on what basis does one accept or reject the former student's claim to knowledge of the causes of his or her success? But one *does* either accept or reject such claims – there isn't a neutral ground, is there? This may also be important.

student's present success, and only in terms of its presumed causal capacities. In this case, the former student's (or, as in the case of the Kristof article, a second-degree advocate's) *autobiographical attribution* of causation for his or her success to some teacher, group of teachers, parent, or school provides the means of making educational quality apparent to a general audience. In describing the character or narrating the events of one's own education, one might *also* be said to make something like teacher quality immediately visible; the "more mediate" label merely reflects the fact that, in subtle contrast with a third-person describer, an embodiment or physically present signifier of success *grounds* the value component of the narration or description of teacher practice in this case. The former student's current success, combined with the autobiographical attribution of its causation to a given facet of the former student's education, renders the quality of that education visible to the public.

Another way of putting the distinction between these types of criterial appeals appears in Cavell's *The Claim of Reason* as the difference between Austinian and Wittgensteinian criteria. [1]Austinian criteria refer to features by which one thing (a goldfinch, in Cavell's discussion) may be distinguished from other things (a cardinal, for example). These criteria will *not help*, Cavell is quick to point out, in *establishing the existence* of the thing in question. In his example, the criteria by which one recognizes a goldfinch will work equally well with regard to the bird singing in the garden and the bird pictured in a painting: criteria of goldfinch-ness, in this sense, prove insufficient establishing the reality of the *real* goldfinch. The criteria Duncan offers in the immediate-visibility case thus lacks an anchoring sense, one might say: even a poor teacher could, for the space of an observation, correctly demonstrate the behavioral criteria of successful teaching: the presence of the *behavioral features* that distinguish good from bad teaching will not be enough to

distinguish *real* good teaching from *fake* good teaching. (It is unclear to me if criteria can do this at all. Cavell does not seem to think so: "There are no criteria for a thing's *being* so over and above the criteria for its being *so*" (Cavell, 1979, p. 51).)

Wittgensteinian criteria, on Cavell's reading, speak to another dimension of the issue, beyond identification based on the recognition of specific features:

> In a Wittgensteinian context, "call" is related to grammatical criteria and generic objects. The criteria do not relate a name to an object, but, we might say, various concepts to the concept of that object. Here the test of your possession of a concept (e.g., of a chair, or a bird; of the meaning of a word; of what it is to know something) would be your ability to use the concept in conjunction with other concepts, your knowledge of which concepts are relevant to the one in question and which are not; your knowledge of how various relevant concepts, used in conjunction with the concepts of different kinds of objects, require different kinds of contexts for their competent employment (Cavell, 1979, p. 73).

Duncan's and Neal's autobiographical attribution of their adult success to the much earlier interventions of teachers highlights not merely another among a list of extant features of good teaching, but rather addresses another axis bound up in the concept of teacher quality. The concept of good teaching, now, is revealed not only in terms of its distinctive *practices or features* but in terms of its specific *effects*. Criteria for good teaching thus extend *beyond* the teacher and acts of teaching themselves, and take an external valence in the (future) characteristics of students. "Teaching," conceptually, is not only supposed to *be* something, but also to *do* something. Adult success, per Cavell above, appears as one of the concepts that Duncan and Neal, among others, relate to the concept of good teaching. It is not obviously therefore of the same criterial kind as Austinian features, but rather serves to bring out or demarcate, one might say, the logical or conceptual space that "good teaching" occupies in our lives.

It is clear, then, that Austinian criteria can be used to challenge the assertion that "Mrs. Smith is a good teacher" by doubting, for example, whether the person making the statement was correctly applying the definitional features necessary to distinguish good teaching from teaching of any other kind. In order to lodge an objection to the same statement on Wittgensteinian grounds, meanwhile, the skeptic would have to draw into question the asserter's right to apply the concept of "teaching" to the activity the asserter has witnessed. The technical sense of good teaching, predicated upon test scores, attempts to answer this latter level of doubt through establishing by fiat the effects that teaching is supposed to entail, according to which the Austinian criteria of good teaching, as opposed to bad teaching (or teaching as opposed to sailing, for example), will also be defined. The point of drawing out the two levels of criteria in Cavell's analysis in juxtaposition with Duncan's words on the matter is to make evident the fact that we *do not* doubt criteria of the Wittgensteinian sort when Duncan describes his mother's teaching quality. The technical sense of good teaching is a distortion, on this view, precisely because it attempts to establish certainty where there is no coherent possibility of *uncertainty.* The kind of knowledge of teaching that Wittgensteinian criteria access is not open to doubt in the way that propositional claims to knowledge – "Mrs. Smith is a good teacher" – are open to doubt, nor for that reason, will the same procedures of warrant or justification apply.

Perhaps this fact – that on a Wittgensteinian level, good teaching, positive outcomes, and adult success are criterially linked – accounts for the prima facie difficulty in raising objections to Duncan's "effective teachers" category. One feels that teachers *ought* to be effective. Even when one finds fault with the notion of "effectiveness" as established in teacher value-added, one hesitates to deny that teachers have *effects* and that these effects

speak to a dimension of teaching in which a teacher's quality will be, at least partially, visible.

The connections between *measures* of effectiveness, as between indications of various types of success, become salient issues at this point. Agreeing that teachers ought to have positive effects settles very little. How one establishes the link between *particular* teachers and *particular* instances of students' adult success, however – what *counts* as success, and how such success might be apportioned among the manifold educative interactions with adults – creates a new problem. Immediately identifying good teaching, as Duncan does in the first example considered in this section, remains, one feels, insufficiently tethered to any indication of that teaching's effects on students, except obliquely. Pointing to present adult success and attributing it to teaching in the past, even to specific interventions, on the other hand, given the number and pervasiveness of educational experiences with teachers, coaches, tutors, counselors and so on, seems to lack the justificatory strength even to imply a suggestive causal connection.

It may be necessary also to highlight the *temporal* insufficiencies of each of these forms of appeal in addressing the felt need to say something meaningful about the quality of teachers currently at work in the classroom, in light of which insufficiencies student achievement data is to make its entrée. The latter case – that of Duncan's citation of his mother and Neal's of Mrs. Grady – addresses the quality of teachers long-since retired, and so comes vastly too late to be of use in assisting with tenure, retention, or compensation decisions in the present. What it lacks in terms of timeliness, however, it makes up for in terms of the felt *authority* of its evaluative capacity: if the goal of education, ultimately, is the development of successful adults, what more powerful evidence could there be than the very

embodiment of successful adulthood pointing to his or her favorite teacher as the driving

force behind his or her success? The opposite is the case for the "moment-I-enter-a-school"

judgment. While capable, in a sense, of sizing up teachers *right now*, it feels somehow too

hasty, too irregular or unregulated, too haphazard, too unconfirmed. Occurring in the middle

of the story, as it were, say at third grade, one cannot know how these children's narratives

end, and the observer's judgment lacks the same claim to certainty, or the same purchase on

our trust. As Cavell says, once more, "There is something [criteria] do not do; it can seem

essential" (Cavell, 1979, p. 83). In the case of observational descriptions of a teacher's

quality, one finds oneself in the uncomfortable position of *merely* taking the observer at his

or her word.

In other words, and by way of expansion, even if a present-tense evaluator feels

confident in declaring this or that teacher "excellent," and even if that evaluator's description

of the teaching garners agreement or acceptance from the relevant parties (from whom?),

such that (some) other people also laud the excellence of the teaching in question, such

claims treat future student outcomes as irrelevant (or else they attend to them, one might say,

implicitly or indirectly, as though accounted for by the excellence of the teaching itself rather

than the other way around). These value claims make no attempt, in fact, to relate explicitly

to future conditions or outcomes at all.

In this sense, such claims cannot cope with any questions that emerge from an

assumption that a child's eventual adulthood (broadly speaking) is produced by and in

accordance with the relative quality of the child's education. Such questions would take the

form of, "But how do you know that it's *working*?" – by which the questioner implies that a

lesson ought to have this or that (visible, measurable, futural) effect, and that the evaluator's

claim appears not to attend to any sort of "working" that the questioner would recognize. In fact, this is precisely the case. The evaluator would point to the (obvious) "working" of the visible quality of the good teaching; but the questioner would see *no relevant criteria* of a given lesson's "working." The request for a modicum of certainty about the futural aspects of good teaching would go unanswered precisely because of something like (but not exactly) *disagreement in criteria*.

The backwards-looking view available in the personal testimony of Duncan and Olly Neal, for all its untimeliness, appears to take care of both of the above concerns: the quality of teaching is immediately visible, one might hazard, in Neal's descriptive narrative of Mrs. Grady's intervention, but its claim to truth or certainty is buttressed by Neal's adult social position, by the career success he attributes to his own education. Despite, however, such warrant behind the claim to certainty upon which Mrs. Grady's quality becomes visible, further consideration reveals that even in Neal's case, problems with dispelling uncertainty remain.

However weak the immediate-visibility case's claim to some sort of certainty or worthiness of trust might appear with respect to its counterpart in the backwards-looking successful-adult judgment, I wish to point out the certainty problems faced by the autobiographical claims, as well, the upshot of which is that one finds himself no less uncomfortably dependent upon another's word in the backwards-looking successful-adult case.

In the first place, a given high school math teacher in an urban or suburban district will teach 120-150 students per year, assuming five sections of 24-30 students with no single-semester electives or enrollment changes. The type of claim that Duncan and Neal

assert is the perspective of a single student in a single year. If a teacher's charge is the development of successful adults, plural, then even the *accurate* autobiographical claim of any single student cannot secure the kind of certainty that one feels tempted to attribute to those of Duncan and Neal. Was Mrs. Grady an excellent teacher *in general*? Neal's own experience with Mrs. Grady is insufficient to address the question. Either the narrative would have to make evident that Mrs. Grady's intervention – the practical sight that allowed her to see Neal in the right light – was somehow characteristic of her, such that similar effects on other students were likely (which the autobiographical nature of the claim precludes), or we find ourselves once again in the discomfiting position of either *assuming* such generalization or rejecting the standing of Neal's claim, neither of which one feels particularly warranted in doing.

Even were such generalization possible, however, Neal's claim about Mrs. Grady's quality remains dependent on a granting of its accuracy. We must hold, to use the other example of this type of claim, that Arne Duncan is *correct* in crediting his mother and other adults with the success he embodies. This amounts to discounting, for example, the interventions of luck, or freaks of fortune, or networking skills, or any other factor, whim, decision, or circumstance that carries a 12-year-old boy from his middle-school years to the office of the Secretary of Education. Surely Duncan has not partitioned out the effects of all possible covariates in order to make his claim. Nothing beyond his word, his own sense of things, compels one to accept *as accurate* the claim that his mother and others like her can be considered responsible for the success he describes.

Lastly, what counts as a "successful adult" is a matter of family resemblance, and thus subject to all manner of definitional slipperiness that I described in Chapter 2. That

becoming a Hollywood actor, a corporate leader, and a major political figure can all count as instances of "success" is not produced by the sharing of any one characteristic. What identifiable, isolable characteristic would link these three positions? One might, for instance, try to claim that all three forms of success share in common the essential characteristic of high income. But one would not call major cocaine smugglers emblematic of the adult success toward which education strives. "Success" itself, as a goal of the educational enterprise, remains itself bound up in agreement in forms of life, shared understandings of being, and so on.

The immediate-visibility case suffers from an intractable inability to dispel uncertainty about the *future outcomes* of the educational endeavor presently at hand because claims to the immediate visibility of teacher quality do not wait upon or predict the future outcomes of the education in question. If it is assumed that excellent teaching in this sense will have visible, positive effects further down the road, the value of good teaching is not taken to depend upon the emergence of those (particular?) effects. Thus, claims to the immediate visibility of teacher quality cannot address skepticism about the future with anything that would satisfy the demands of the interlocutor.

But the backwards-looking autobiographical perspective is also beset by its own limitations in terms of establishing certainty. Not only does one have to accept a single autobiographical narrative as expressing something *general* about a given teacher in the absence of evidence, and not only does one have to accept the causal conclusion of the autobiographer's (self)-analysis, but in order for the claim to have purchase at all, one must also share the sense that the autobiographer does indeed embody something worth educating toward. None of these acceptances is compelled by logical or natural law, any more than the

observer's claims about immediately-visible educational quality can be taken as predictions of future performance with any basis in causal or logical necessity. *Both* claims remain strikingly, revealingly frail.

The confluence of criteria at work here points, I suggest, to the ineluctable complexity of the educative endeavor as a whole, a complexity that Duncan also acknowledges in his various statements about not wishing to "reduce the complex, nuanced work of teaching" to any single measure. Ways in which we *ordinarily* speak of good teaching – in terms of witness, one might say, either as a third-person observer or an autobiographical confessor – seem perfectly adequate to the situations in which they are offered; and yet also *inadequate* when faced with what is essentially a request for something like *totality* with regard to attributing to this or that teacher the value of "good." ("She might have been a good teacher during that day, but is she a good teacher *overall?* She might have been a good teacher for that student, but is she a good teacher *overall*?" Or the converse: "She might be a good teacher overall, but is she a good teacher for *that student?*")

The purpose that evaluative claims are meant to serve seems, if not *determinate* of a claim's form, then certainly important to consider. Duncan's "we know what success looks like" list of criterial features is offered primarily as a statement about good teaching's *prevalence*, its everydayness, and insofar as his audience agrees with him (and the criteria are broad enough that they surely do), the statement *works* in this capacity. However, as it is disconnected from the other criterial areas that bear upon the concept of good teaching, particularly with regard to eventual effects on students, one feels it inadequate to the task of evaluating teaching *as a whole*: it simply omits a portion of the field of criteria that reveals to us what we take the concept of "good teaching" to entail. This is also the case with the

autobiographical account: as Kristof offers the Olly Neal story as a means of dramatizing the importance of good teaching, the first-hand account suffices beautifully. When one wonders whether this unique experience generalizes to Mrs. Grady's effect on other students, however, or whether Olly Neal is correct to discount the influences of, for example, his later college professors or professional mentors, the criteria elicited according to Kristof's narration of the story cannot be summoned as an answer. Both forms of what I earlier called "ordinary-sense" views of teacher quality, despite their sufficiency to the contexts in which they appear, seem to be wanting when one attempts to apply their criteria *generally*, as in the project of defining good teaching.

3.1.1 Achievement Data and Questions of Causality and Generality

Before turning to a consideration of the kinds of appeals made in terms of student achievement data as Race to the Top envisions it, let me make clear the stakes of the discussion. In his many speeches touching upon or directly dealing with the use of data, Duncan treats the availability of longitudinal data as offering something profoundly *new* to the education profession. We have seen, in a limited sense, that both of the criterial appeals discussed above require a certain *acceptance* or *agreement* on the part of those to whom educational quality is reported. In the immediate-visibility case, where only description is required to communicate the quality of the educational endeavor, a very broad level of agreement is simply assumed in terms of what *qualifies* as "good," as well as what *counts as* the kind of features that would reveal quality in the first place. Without belaboring the point too much, it bears mentioning that Duncan's list is radically incomplete. To use Duncan's "I know it when I see it" description of a successful school as an example, what counts as

"teaching," what counts as "collaborating," and what counts as "enthusiasm" are all matters requiring agreement, to offer a few further instances of this fact. Likewise, in the second instance, a person's claim to knowing educational quality rests upon the sharing of the senses of the adult success that grounds such a claim, as well as the hearer's acceptance of the attribution of causation to the former student's education, neither of which is by any means a foregone conclusion. The longer such an autobiographical claim about past teaching excellence continues – the more involved the description – the more terms emerge that require agreement. In Olly Neal's case, one issue that I have broached in the Introduction and elsewhere, though not in these terms, has to do with Mrs. Grady's crucial decision to allow Neal to steal a book. Regardless of whether such an action or non-action on Mrs. Grady's part is taken as an instance of educational excellence or professional malpractice, it is obvious that the action's eventual status requires *assent* in order to function as it does, and assent, it must be said, on a variety of levels: does one accept that there is a relatively direct link between this action and Neal's eventual judgeship, the mark of the "success" of his "outcome"? Does one accept Mrs. Grady's statement 35 years after the fact that she knew Neal had stolen the book and restocked the author with the (implicit) intent of helping Neal catch the "reading bug?"

The felt limitations in each of these kinds of appeals are readily voiced: whatever else they lack (in Duncan and Neal's appeals to their upbringings, for example, the autobiographical claim founded upon present adult success can only tell us about the quality of teachers long since retired, for example), they leave substantial room for doubt, as the explicit reliance upon agreement makes clear. It is neither by logical nor by natural law that one is compelled to accept the claims rendered in these two forms; more than this, the

proliferation of levels of agreement required in each case seems to defy the very possibility thereof. Perhaps Cavell puts it best with his simple truism that "We *cannot* have agreed beforehand to all that would be necessary" (Cavell, 1979, p. 31).

I take Duncan's depictions of the promise that student achievement data offers to the practice of teacher evaluation as responding to the sort of slipperiness to which Cavell alludes. In other words, Duncan's proposals surrounding student achievement data are predicated upon the ability of such data to eliminate the uncertainty of "the uncertain profession," as the very title of one of his speeches suggests (Duncan, 2009d). The first two Parts of this project, which seek to highlight the conceptual problem engendered in the attempt to address an ordinary sense of good teaching through resolutely technical means, might be taken as one sort of objection to data so conceived: whatever certainty such data offer cannot seem to get any purchase on the ordinary terrain that generated the need for certainty in the first place. That represents one kind of failure at the overcoming of uncertainty that the use of student achievement data sets out to accomplish.

In this Chapter, however, I wish to reveal not the *differences* between Duncan's use of an "ordinary sense" of good teaching and its "technical sense," but rather the *similarities* that they share, specifically around the kinds of uncertainty that pervade the uses of each. If the use of data generates a new kind of conceptual uncertainty through putting an "ordinary sense" of teaching out of reach, I wish to argue that it also *fails to solve* on its own terms the kind of uncertainty to which it claims to address itself, the sort implied in Duncan's and Neal's narratives of adult-level success. Its claim to represent something *new*, something that we have *at long last*, falls flat.

In the two previous Chapters, I have raised and critiqued the use of student achievement data in terms of its ability to distinguish without recourse to human judgment good teachers from poor teachers, and also in terms of its ability to account for and reproduce the practical examples of excellent practice that scholars and politicians cite in indicating the need for such measures. If my focus here seems to hew closely to the critical aim of Chapter 1, that fact indicates the conceptual interactions inherent in the activities of *knowing* and *telling* a given thing. What one takes as necessary to the (justifiable) reporting upon the quality of a given teacher, say, is inherently bound up with what one takes as necessary to the knowing of that quality. At the same time, what one takes as necessary to the (justifiable) knowing of teacher quality has to do with an ability to reveal that quality to someone else. In the discussion of data in Chapter 1, I alluded to the problems associated with attempting to disentangle these notions from one another: in reference to Duncan's statement pertaining to California's 300,000 teachers, wherein he claims that "no one in California can tell you which teacher is in which category" and then asserts that "there's something wrong with that picture," I noted that it is unclear whether Duncan's "*no one can tell you*" refers to a problem of knowing the differences between good and bad teachers or a problem of finding oneself capable of reporting upon them (or, as I mention in that Chapter, a simple problem of unwillingness on the part of the parties from whom Duncan expects oversight to draw the necessary distinctions and to in fact put those distinctions into print, as it were).

In Chapter 1, I used the difference between the ordinary and the technical concept of teacher quality to make the claim that the use of student achievement data could not satisfactorily secure knowledge of teaching happening *right now*. Conceptual lack of fit makes that which data seizes upon meaningfully different from the sort of teaching that

policy-makers and the public wish it to reveal. However, whatever the ramifications of that distinction, student achievement data still makes a demonstrable claim to secure *something* involved in life and learning that we as a polity value, even if it is not the ordinary sense of good teaching. Crucial to the sense of securing this something is data's relative ability, quite literally, to predict the future. It is tempting, for this reason, to prefer data-driven measures of teacher quality to something along the lines of on-hand observations or other instruments that rely upon human judgment. The use of student achievement data seems well-positioned to overcome *both* of the problems raised in Duncan's citation to the immediate availability of success, on the one hand, and Neal's post-hoc reflection of Mrs. Grady's educational interventions, on the other.

Put succinctly, just like Duncan's we-know-what-success-looks-like argument, the use of achievement data to generate teacher value-added *also* promises present-tense visibility of teacher quality, albeit including the conceptual slippage I underscored in Chapter 1. But offsetting that slippage, or assuaging one's anxiety about it, is the crucial second promise that data makes: to make justifiable inferences about students' *future* outcomes from a present-tense perspective. That second promise underlines the value of Neal's type of backwards-looking narrative: a measure of certainty about adult success performs a justificatory role with respect to claims about educational quality. But *unlike* Neal's narrative, the use of student achievement data seems to dispatch the worries associated with the singularity of the autobiographical perspective and with the attribution of causation. Even though statisticians and economists would insist that correlation does not imply causation – correctly, of course – the existence of significant correlations between data-deduced measures of teacher quality and certain measures of future student outcomes provide

sufficient grounds to justify claims that teacher value-added tells us something worthwhile and meaningful about the quality of teachers, even if that justification is not taken to rest upon something as solid as causation. Additionally, because teacher value-added is calculated using aggregate data, the claim to a given teacher's quality takes into account the experiences of *all* students enrolled in that teacher's classes, thus disarming the threat of having to take the narrative of a single, albeit successful, adult as representative of wider student experiences.

It is in these forms and on these grounds that the use of student achievement data offers to address the perceived insufficiencies of our ordinary discourse on teacher quality. In contradistinction to our ordinary discourse, student achievement data seems to afford a greater measure of certainty in offering claims about the quality of this or that teacher, a certainty grounded in the data's generality across students and its predictive ability, as well. Where the uncomfortable dependence upon shared understandings of being or agreements in forms of life permeate the offering and acceptance of ordinary-sense claims pertaining to teacher quality, the use of student achievement data promises a more secure, more certain, foundation.

To lay out the matter clearly: the use of teacher value added, grounded in and derived from student achievement data, makes its claim to *superiority* with respect to the ordinary measures discussed above on the grounds that (a) it enables inferences about the relation between current teachers and future outcomes in a way that is denied in ordinary evaluative means and (b) its objectivity and generality avoid the pitfalls or the definitional slipperiness that arises with respect to relying upon something as nebulous, uncertain, and mutable as "agreement in forms of life."

The contention throughout Chapter 3 amounts to the following: claim (b) is straightforwardly false; and claim (a) is, one might say, misleading. It is *accurate* to suggest that the use of data enables one to make inferences about the future in a form previously unavailable to us. It is, however, unclear that this new form is different in *kind* from what has heretofore served for educational evaluation, given the falseness of claim (b) – if claim (b) is indeed false, as we shall see, claim (a) amounts to no more than the laughable suggestion that in prior ages a teacher's probable effect on students' eventual adulthood never factored into ordinary judgments about his or her quality. Moreover, given once more the falsity of claim (b), it is unclear why one should regard the use of student achievement data for judgments of quality as deserving of priority – much less an exclusive claim – in teacher evaluation.

When Duncan and Obama address the need for student achievement data, in fact, they refer to precisely these grounds for preferring them as a basis for claims about teacher quality: the measure of certainty they provide about teaching both with respect to the *present* and to the *future.* However, in considering Duncan's thoughts on the use of data once more, one sees myriad problems in terms of claims (a) and (b) bound up together.

Taking the most starry-eyed of Duncan's statements about data's possibilities reveals both the temptations and the pitfalls associated with turning to putatively objective measures of teacher quality. Two temporal axes of certainty show themselves in the following quotation, and I will deal with them separately. The very fact, however, that Duncan takes the use of "robust data systems" to secure knowledge and quality claims pertaining to both the present and the future speaks to the tempting aspects of data's usage.

We can one day do a better job of understanding what makes great teachers tick, why they succeed, why they stay in the classroom and how others can be like them. Hopefully, we can track good programs to higher test scores to higher graduation rates. Hopefully, one day we can look a child in the eye at the age of eight or nine or 10 and say, "You are on track to be accepted and to succeed in a competitive university and, if you keep working hard, you will absolutely get there" (Duncan, 2009b).

Duncan wishes to *do* things with data, that is, to take present action on the basis of data. This is a point that must be raised loudly and explicitly: evaluating teachers, making claims about their quality, is a matter of *doing* things. For Duncan, the use of student achievement data holds out the kind of promise that would allow him to do the following present-tense things: (1) understand what makes great teachers tick, which is also bound up in understanding "why they succeed," (2) knowing what makes them stay in the classroom and "how others can be like them," and (3) draw a correlative or causal line from "good programs (of teacher education, presumably) to "higher test scores" to "higher graduation rates."

The grasp on certainty – for the first time in human history, per Duncan's tone – that data offers to the educational endeavor then makes possible new kinds of knowledge claims and new forms of reporting on the quality of teaching programs, the quality of teachers, and also the specific components of successful teachers themselves, such that one of the things that we can know and tell is "how others can be more like them." These knowledge claims and any reporting thereof obviously require a kind of grip on reality – the certainty undergirding them is only useful provided that the knowledge or judgment secured speaks to what one might call on-the-ground conditions, and there is substantial reason to doubt that the type of knowledge-claims that data can be used to insure, as it were, do indeed reflect or

respond to the phenomena they mean to address, which speaks against both claims (a) and (b), as they remain at this point, intertwined.

Take, for example, Duncan's attributions to data of the power to "understand what makes teachers tick" as well as "how others can be like them." Data, on this view, reveal something about teachers that a given administrator or program can first isolate as at least correlated with excellent teaching and then transmit or transport to another teacher – in order to allow other teachers "to be more like them." We have already seen in Chapter 2 the names of what such data can reveal: "best practices" and "teacher knowledge." While test scores can make evident something about student learning and, in a limited sense, about teacher quality – an unproblematic concession – the problem presented in this case is that *what* data is taken to reveal is something that stands behind the practice of excellent teachers, something that, to recall Shulman, "permits" a given teacher to teach as he or she does. Chapter 2 demonstrated that it is entirely unclear, and in fact extremely doubtful, that *anything at all* stands behind teacher practice in the way that both the "best practices" and the "teacher knowledge" approach assume. The failures of rationalist-assumption approaches to successfully account for or to reproduce skillful practice, combined with the fact of such skillfulness's pervasiveness, suggests that the flaw lies in the model.

Despite these failures to account for or reproduce skilled human behavior, which, as we have likewise seen, has to do with the same "agreements in forms of life" and "shared understandings of being" that threaten the certainty of ordinary-sense claims to teacher quality, study after study demonstrates correlations between this or that abstract characteristic of a teacher's practice – behavior or knowledge – and growth in student test scores (Kane, et al., 2011b; Lemov, 2010; Nocera, 2012; Rivkin, et al., 2005). This has ontological

implications for "best practices" and "teacher knowledge," and profound epistemological implications for claims to certainty based upon putative knowledge regarding such characteristics.

The "best practices" case provides the best means of seeing the limitations that Duncan's hopes for data overlook. Basing value or knowledge claims upon the information yielded by the use of student achievement data promises a kind of actionable certainty that eludes both the autobiographical narrative of a former student and the immediate sense of a third-party observer. The elusiveness of certainty in the latter two cases, while having also to do with the issue of temporality – an inability to speak to the future and the present, respectively – comes down to a reliance on agreement rather than something "natural" or "objective." Data offers something, then, that will *not* boil down to such insubstantial or contingent agreement – that is the basis of the certainty it promises, the essence of claim (b).

Correlations require two variables. In the case of seeking out "best practices," one variable is always growth in student test scores: the dependent variable. The other is the particular "practice" under investigation, say, to recall the Lemov discussion from Chapter 2, breaking down instructions into "simple" rather than "complex" skills: this is the independent variable. A researcher then compares the test-score growth of the students of teachers who perform those behaviors regularly with those who do not. If the mean difference between the two groups' test scores rise to statistical significance, the extra gains of the former group is then attributed to the intervention, to the instructors' use of the practice itself.

But *just what* does the practice consist in here? Assuming that a given researcher observes real, live teachers ordinarily teaching in many different classrooms, to many different students, how is that researcher to *identify* the practice of "breaking down

instructions into simple rather than complex steps"? The teachers under investigation will

not be giving *precisely* the same instructions (the same words concerning the same tasks) to

their many different students.   According to *what* can a researcher count two different sets

of words, offered in the interest of directing student activity, as *similar enough* to represent

two instances of "breaking down instructions into simple rather than complex skills" in the

way that the notion of a practice as an object would require? Perhaps one acknowledges that

telling students to "design a policy brief" will not provide the students sufficient clarity to

succeed in the task (but perhaps it will; that depends on prior learning, the course content

itself, many factors to which the observer may not be privy at all). But what is to count as

adequately simplifying the task? "Take out your pencils, then design a policy brief" is of no

more help, though one has broken a task down in a sort of rudimentary fashion. At the other

end of the spectrum, one might imagine a teacher instructing students to prepare to write an

essay in these terms: "Use your eyes to locate your backpack; unzip your backpack; find a

pencil; take it out of your backpack; place it on your desk; remove a sheet of paper; recall the

events of your summer vacation; choose one of these memories that you think you would like

to write about; think of a title…" and so on. There are, doubtless, ways of giving directions

that are *too* complex for the students at hand, and there are also ways of giving directions that

are *too* simple.

That which enables and authorizes a researcher or observer to classify two different

manifestations of two different teachers' practices as both examples of "breaking down

instructions into simple rather than complex skills," a singular "practice," is not available in

behavioristic or context-free form. Rather, it is yet another instance of a family-resemblance

concept. Instructions, as the most recent example demonstrates, can be *endlessly* atomized;

and *all* instructions can, in reverse fashion, be revealed as simplifications of some more complex skill. Some teachers give directions at the *right* level for their students and others remain too vague. That right-ness, one might say, is established or perceived as a matter of situational *fit*, rather than by means of comparison to some external or formalizable or empirical standard that transcends classrooms. The researcher will categorize two different teachers in two different classrooms as direction-givers who break instructions down into manageable pieces. But that these two hypothetical teachers can be said to have *done* a given thing, to have demonstrated a particular practice, is irreducible to abstract characteristics. There is no bright definitional line circumscribing any "practice" of direction-giving.

Where to *begin* decomposing instructions and where to *stop* decomposing cannot be expressed in formalizable terms. Nor can it (accurately) be seen according to explicit features. What enables the researcher to classify two different sets of practical words and behaviors under one category is itself grounded no more (and no less) solidly than in forms of life. If the use of student achievement data seems to tell us something more certain about "what works" than the ordinary discourses, this greater claim to certainty is predicated upon student achievement data's working with learning and teacher practices as clear, distinct, and solid ontological objects; the above discussion, however, reveals that any ontological solidity in the practices that these studies consider is in fact *produced by* a kind of agreement, by the intervention of human judgment. In point of fact, the ontological foundation of best practices as revealed in student achievement data remains no more solid than that of any practice described in third-party or autobiographical claims.

Though it perhaps goes without saying, the above paragraph addresses only the problematic (in terms of certainty-establishment) conditions for a *single* researcher's

recognizing what counts as "breaking down instructions successfully." Should a significant correlation appear between the collection of actions that the researcher counts under this concept and growth in test scores, the breaking down of instructions will then be taken to have earned a spot on a "best practices" list. Thus, *how* to break down instructions *in practice* and *how to recognize* when teachers are performing the practice will have to be *given*. Chapter 2 has dealt with the problems of attempting to recreate *performance* in a given domain based on the rationalist assumption upon which the above example operates. In attempting to *generalize* something *practical* that one researcher has successfully accomplished in *formal* terms, the reliance upon agreement in background practices or forms of life simply multiplies exponentially. In no sense does this process lessen such reliance – if studies correlating a given practice with student achievement have merit, that merit cannot claim the degree of objectivity claimants would like; thus, the straightforward falsity of claim (b).

Chapter 1 addressed the limitations of the concept of "student achievement" visible in terms of test scores, with which teaching practices are correlated, upon which students' futures are rendered predictable, and according to which teacher quality might be judged. The above examples are meant to reveal the shaky ontological status of "teacher practices" that make up the other side of the correlative relationship. The *correlations* that justify knowledge and judgment claims about teacher practice, including those undergirding "best practices" approaches, thus establish a measurable degree of certainty (and uncertainty) and also a secure link between two profoundly *unstable* ontological objects. The ability of data to provide any more certainty about "what makes good teachers tick" and "how others can be like them" thereby loses its claim upon us. The failure of the rationalist assumption as

embodied in artificial intelligence projects implies that teachers *may not have at all* anything

as secure or isolable or formalizable as a set of "practices" standing behind their actual

teaching to which one might attribute their success. The considerations above imply that

what researchers *call* "teacher practices" are not isolable objects of pure empirical

investigation: the commonalities among successful teachers are recognizable as such only on

the basis of *agreement*. While correlations between teacher practices and test growth purport

to offer an objective, stable alternative to the felt inadequacies of something like the recourse

to agreement necessitated in appeals to immediate visibility, these correlations *also* come

down the same frail notion of the shared-ness of shared understandings of being.

Student achievement data, then, cannot deliver on the promise of certainty it holds out

with respect to knowing and judging teacher quality in the present tense. With the foregoing

discussion in mind, the National Council on Teacher Quality's report on state evaluation

procedures bears returning to. The report notes the importance of classroom observation, and

makes particular suggestions as to how these might by implemented:

> A strong observation rubric should focus almost exclusively on teacher
> practices and student behaviors that can be observed in the classroom. While other
> criteria are not without merit, they may call for too much subjectivity and guesswork
> on the part of the evaluator (NCTQ, 2011).

In Chapter 1, once more, I emphasized the fact that classroom observations so

conceived cannot – and, despite implicit claims to the contrary, make no attempt to – address

the ordinary sense of good teaching. The claim here is quite different. In this Chapter, the

claim is that, if "too much subjectivity and guesswork on the part of the evaluator" is a

problem that observation rubrics are meant to overcome, they do not succeed. The very idea

that a teacher practice could be *sufficiently* described so as to make the intervention of

"subjectivity" or "guesswork" (whatever that may mean) on the part of the observer

unnecessary, superfluous, or even minimal proves impossible. Since *what counts* as the

demonstration of this or that practice cannot itself be described in behavioral or abstract

terms, that which allows a given evaluator to check *just this* box when a given teacher

performs a given action or says a given phrase has to do with the evaluator's shared sense of

the classroom environment, the teacher's purposes, and a notion of what "teaching" and

"learning" look like. Even 116-point evaluation rubrics stop short of the fullness of

description that an *objective* observation of "teacher practices and student behaviors" would

require.

In terms of its ability to establish communicable certainty about teacher quality in the

present tense, student achievement data ultimately offers us nothing *superior* to Duncan's

the-moment-I-enter-a-school judgment. Both are matters of agreement in forms of life. It

does offer a different *expression* or *register* of teacher quality, and that itself has value, a

point that will return in a later section, but since Duncan's claims on its behalf take the form

of "cracking the code," as opposed to "guessing" (Duncan, 2009b), this exploration suggests

that no *such* distinction is warranted.

3.1.2 Achievement Data and the Prediction of Adult Outcomes

If data seems to promise the ability to cope with both the uncertainty about the

present and the *future* that remains in the cases of describing teacher practice and in recalling

excellent teachers from the past, we have seen that, whatever its ability to dispel uncertainty

in the present, this ability comes down to the same *agreement* grounds as the view it would

supplant. Its ability to predict future student outcomes, however, would truly be a coup, and it is to the futural promises made on its behalf to which I now turn.

The second half of Duncan's plans for the use of student data – a project that he presumably sees as lying somewhere over the horizon, but made visible by recent advances in the ability to use student achievement data – speak directly to student achievement data's ability to predict future student success. Duncan prophecies that, "one day we can look a child in the eye at the age of eight or nine or 10 and say, 'You are on track to be accepted and to succeed in a competitive university and, if you keep working hard, you will absolutely get there'" (Duncan, 2009b). This statement ought to be heard in light of Duncan's insistences about the need for honesty and responsibility in reporting on educational quality. Earlier in the same speech, he says: "We must tell the truth and we must tell it clearly" and also "When states lower standards, they are lying to children and they are lying to parents. Those standards don't prepare our students for the world of college or the world of work" (Duncan, 2009b). Through (1) the use of student achievement data in *setting* (qualitative) standards, based, given the context of the latter statement, on some alignment with college acceptance and/or graduation rates, and then by (2) *generating* data for particular, current students – by measuring students against those standards and analyzing the particular data that result – we will at last find ourselves in a position to "stop lying" and "tell the truth" with regard to the quality of our children's education. At last, as in Duncan's vision, we will be able to look a third-grader "in the eye" and say *both* that the child is "on track to be accepted and to succeed in a competitive university" and that if he or she "keep[s] working hard," the child "will absolutely get there."

Further context is necessary here: when Duncan speaks of "lying" to our children and the need to "tell the truth," he is once more specifically bemoaning the colloquially-termed "race to the bottom" that I discussed in Chapter 1, wherein NCLB accountability measures unintentionally encouraged states to lower standards against which students were measured for proficiency. Such artificial lowering of standards amounts to "lying" insofar as, in Duncan's words, for example, "[The data] tells us that many adults who do graduate go on to college but need remedial education." What states often labeled "proficient" under NCLB, then, amounts to a lie to the degree that it does not entail the results we attribute to the *concept* of proficiency, including and perhaps especially the preparedness to move seamlessly to the next educational and professional levels without "remediation." The student achievement data now available to educators promises, for Duncan, the ability to banish uncertainty with respect to what proficiency *really means* (or where it *really is* on a given scale), so that, in combination with a measure of "courage" – as he says in the same speech – we can finally "stop lying to children." This subsection seeks to assess the firmness of the ground on which Duncan's faith in the promise of data to cope with uncertainty about the future stands.

Duncan's own considerations of data's utility render such minimization of uncertainty extremely unlikely. Keeping Duncan's futural goal in mind – the ability to tell "a child of 8 or 9 or ten" that he or she is "on track to be accepted and to succeed in a competitive university" and that, given *only* the child's continued demonstration of a work ethic, the child "will absolutely get there" – the means of providing this sort of certainty about the future that Duncan cites prove incapable of securing the standard according to which we might "tell the truth" or "stop lying to children."

Two particular elements frustrate the attempts at securing futural certainty, or even a claim to warrant that is superior to ordinary means, that Duncan imagines data might provide. The first element is the well-documented – and previously discussed in Chapter 1 – notion of *validity* in measurement instruments that generate the data. The second element is a species of the chicken-and-egg problem that objective analyses of language encounter, and that Lenat, Derrida, and Glendinning address in Chapter 2; this second element emerges immediately upon attempting to solve the problems implied in the first element and it stymies finally the endeavor to secure the kind of certainty about the future that Duncan believes necessary to telling the truth.[33]

The first problematic element confronting the use of student achievement data in order to secure qualitative claims with regards to future outcomes has to do with disagreements within the data itself, which indicates a certain validity problem. Student achievement data – which, as Chapter 1 indicates, ultimately comes down to test scores – is produced by assessments that, in the words of psychometrician Daniel Koretz , "are only limited measures of the latent construct of interest, which is some measure of student proficiency. . . . They become meaningful only to the extent that one is justified in generalizing from the score to the latent construct of interest" (D. M. Koretz, 2002, pp. 754-755). The issue of justifying *generalizations* from limited conceptions of a given domain

---

[33] It may be worth pondering what notions of truth and lying are at work in Duncan's discourse on *telling* educational quality. If I say that the Mets are going to win the World Series in 2013 and they do not, have I *lied?* If I say that Roger Federer is a great tennis player but he fails to win a single major for the remainder of his career, have I failed *to tell the truth*? What kinds of *necessary* consequences, specifically, do we expect good teaching to entail?

comprises the second problematic element of using data in the way Duncan wishes to use it. In terms of the first problematic element, that of bare validity, Duncan's own citations of data speak for themselves.

In his speech on the importance of capitulating to student achievement data in order to make truthful or honest claims about a student's learning, Duncan makes the following separate claims, the first pointing to high-stakes state-level exam results and the second pointing to a no-stakes national-level test:

> Now, we know the news isn't all bad, of course. We also know that children of all age groups across the country have improved their performance in reading and that younger students are posting strong gains in math (Duncan, 2009b).

> The results from the long–term NAEP show that we have a lot of work left to do, particularly in raising the achievement of our students at the secondary school level, whose test scores have barely moved over the past three decades (Duncan, 2009b).

These claims come moments apart in the speech itself; each purports to *know* something about "what's happening" in American schools; yet they offer contradictory views on the matter. According to state-level exams, the news isn't "all bad": "children *of all age groups across the country* have improved their performance in reading" (emphasis added). According to the NAEP assessment, meanwhile, the scores of students at the "secondary school level. . . have barely moved over the past three decades." Both statements cannot be true. Different assessments seem to result in different conclusions pertaining to the "latent construct of interest," as Koretz calls it. Since both test instruments purport to measure the same constructs, some other means will be necessary to put this disagreement to rest; appeals to these same measures cannot settle the matter.

In order for student achievement data to perform the futural task that Duncan sets for it – predicting adult outcomes on the basis of student skill levels in given domains – the assessment protocols that generate this data must accurately measure the construct that they claim to measure: proficiency in certain skill domains. Koretz puts it a slightly different way in the same article:

> Thus, if gains are meaningful, they should generalize from the specific test to other indicators of mastery of the domain in question. Because an exhaustive measure of most domains is impractical, one cannot test the degree of generalization from operational tests to the ideal, complete test. One can, however, examine the degree to which gains on a specific test generalize to other tests and to nontest measures of performance in the domain in question. (D. M. Koretz, 2002, p. 757)

Koretz points specifically at two means of grounding justifications of generalizations from test scores. One requires that results on a given test match up with the results of other assessment instruments that purport to measure the same construct. Duncan's statements regarding state-level exams – which he takes as telling the truth about current levels of student achievement above – indicate that correlating the results of different assessments is not sufficient.   The other method Koretz points to involves employing "nontest measures of performance in the domain in question." Duncan is not unaware of either problem. To address the first issue, he is explicit on the problem of test-to-test correlation:

> When we match NAEP (National Assessment of Educational Progress, also known as the Nation's Report Card) scores and state tests, we see the difference. Some states, like Massachusetts, compare very well. Unfortunately, the disparities between most state tests and NAEP results are staggeringly large (Duncan, 2009b)

He attributes the "disparities" to the "race to the bottom" impulse to which I alluded earlier, which amounts to an accusation that states really are "lying to children," and at a fundamental level, about what reading *is*. If states were simply honest about what

"proficiency" means, Duncan implies, then all states would demonstrate the correlation with NAEP that Massachusetts does.

But this is an oversimplification; taking other factors into account will be important. Firstly, however, it is important to acknowledge that differences in the cut scores defining proficiency across states during the NCLB era *do* exist; they *do* contribute significantly (even if not exclusively) to the massive disparity between NAEP results and those of state-level assessments, such that, in 2003, for example, Mississippi ranked third in the nation in the percentage of 4th graders proficient in reading according to its own *state* measures, while according to the NAEP exam the same state ranked 47th; and such stark disparity in the criteria of proficiency *do* amount to differences in the concept of " reading proficiency" itself. Duncan is not incorrect on any of these points, and it is no mischaracterization or exaggeration to label such deliberate malfeasance with respect to criteria a type of "lying."

Koretz, however, points also to the intervening factors of measurement error and score inflation, the latter of which is the result of what is commonly known as "teaching to the test": a phenomenon in which a given test score rises as the result of test-specific coaching without any comparable rise in one's proficiency in the "latent construct of interest." Score inflation has demonstrably pervaded data-driven reporting of on educational quality in the past (Cannell, 1988; Linn, Graue, & Sanders, 1990; Shepard, 1990), and so it is hardly obvious that simply "being honest" about the real nature of proficiency would *suffice* to solve the validity problem that the use of data to make judgments about educational quality presents. Bearing in mind that the NAEP exam entails no sanctions for failure, while the stakes attached to state exams continue to rise, it is perhaps reasonable to suspect that one might just as well attribute some portion of the discrepancies between the data from NAEP

and from state-level exams to the effects of score inflation (not to say outright cheating). In

recalling that the states of Delaware and Tennessee, among others, explicitly require that

student achievement data represent a significant portion of a teacher's evaluation – upon the

basis of which Duncan wishes to empower LEAs to make personnel, tenure, and

compensation decisions – the *incentive* for teachers to engage in the practices of score

inflation has never been higher.

Since the potential for mischief in terms of test scores' ability to accurately reflect the

construct of interest is not fully accounted for by merely acquiescing to Duncan's exhortation

to honesty in *setting* the criteria of proficiency, it is at this point entirely unclear how one

might confirm the validity of student achievement data on a large scale[34] by means of

nontest measures in the domain. Take reading, a skill regularly assessed by means of the test

instruments that generate the data in question. In the present, one might attempt to correlate

reading scores with, for example, English/Language Arts grades. But the number of

confounding variables at work in this relation renders such an attempt futile. Since the

giving of specific grades would depend *not only* upon the student's reading skill but also rate

of homework return, performance on quizzes, and myriad other student habits, behaviors, and

proclivities, such a grade cannot reasonably proxy for reading skill alone. (These stand in

---

[34] This is an important caveat: The problem Duncan seeks to address is operative at the state and federal level, of which he treats the local as a species, I think. It would seem to be relatively easy for a given teacher, at the end of a given year, to assess the validity of the reading data for his or her class based on a great deal of personal experience with the students' lived abilities to read. But when adopting the long-distance gaze of the state or federal official constrained by the absence of personal, daily contact with the data points, as it were, one must seek out other *measurement instruments* that would pertain to the same skill domain. But the federal or state official is hamstrung by the fact that, as far as nontest demonstrations of proficiency in *just that* skill domain go, there simply aren't any. If one *must* ultimately ground a question of validity in nontest measures, people at the state and federal level are simply *in the dark.*

addition to the problems of grading variation among individual teachers, as well as the issue of grade inflation over time.)

Duncan, however, treats the use of student achievement data as though it *were* verifiably or justifiably valid; in fact, its verifiability or justifiability is strong enough that he can couch the debate over whether and to what extent to make use of data in terms of "lying" versus the "telling the truth." Given the limitations of the long-distance view necessitated by the requirement to demonstrate something about teacher and learner success to a state or federal official, how might one establish the validity of student achievement data using nontest factors? How would one go about "being honest" as to the *real* nature of proficiency against which one would measure students, such that one could secure the certainty with regards to future outcomes that Duncan imagines student achievement data to provide? This point raises the second problematic issue to Duncan's proposals for the use of data, one that will recall Lenat's chicken-and-egg problem from Chapter 2.

Koretz highlights the importance of ascertaining the validity of a given measure by comparing the results of that measure with "nontest measures of performance in the domain in question." Duncan's awareness of this dimension of the validity discussion manifests itself in his many citations of "college and career readiness" as a criterion of student educational success; he points often to indicators such as graduation rates and the percentage of college freshmen requiring remediation, as well. Duncan refers throughout his several speeches to various other measures of student outcomes that might anchor the reading scores (for example) of particular students to a notion of "college and career readiness" – a grand criterion, a nontest measure of student performance.

The disagreement between what these nontest measures indicate and what state exams of proficiency report demonstrates to Duncan that states are lying about what proficiency means. It is worth noting – but not belaboring – Koretz's point about comparing measurement outcomes to "nontest measures *in the domain in question*." State exams, at the moment, assess reading, mathematics, and, in high school, science. Why high school graduation rates, for example, should count as a nontest measure of any of those domains in particular (or of all of them together) requires an explanation that is simply not forthcoming from educational reformers. There appears to be an assumption at work here, according to which it is possible to generalize from "limited measures" of specific academic skill domains to the vastly wider domain of educational quality writ large. The measures to which Duncan appeals in order to buttress, even if implicitly, his validity claims with respect to the student achievement data's relation to "proficiency," as the state accountability systems generate it, are insufficient to secure that sort of validity. What "proficiency" *means* – which ought to be heard broadly, as referring to both its implications for signaling something about success in the future and also to its conceptual sense in the present – remains in need of an anchor at this point. Anchoring either of the sides just mentioned, either what proficiency is taken to *consist in* (presently) or to *necessarily entail* (in the future), would perhaps suffice to stabilize the other.

Duncan's emphasis in insisting that we stop "lying to children" rests on the notion that we are (in the present) deceiving them about what their so-called proficiency will allow them to *do* (in the future). The college-and-career readiness criterion speaks loudly here: his rhetorical tactic involves tying proficiency to the *outcomes* it ought (by what power of logical or causal necessity?) to entail. Establishing the truth about proficiency represents the first

step in the establishing the concrete criteria according to which students ought to be measured so as to assess their relative on-track-ness to "be accepted to and succeed in a competitive college," to continue to use Duncan's example. The proficiency construct thus entails the *eventual* ability to gain admission to and succeed in college. One needs to know some *facts* about proficiency levels in relation to college acceptance and success rates, in other words, in order to generate the data necessary to evaluate the actuarial probabilities of a given student's achieving college-level success and higher career earnings at some point in the *future*. Where, then, does one look in order to establish facts about what the future requires?

The answer, perhaps not surprisingly, is data itself – data originally gathered, it goes without saying, in the *past*. Take, for one very rough example, the Chetty research that generated so much discussion in early 2012 and that President Obama cited in his 2012 State of the Union Address. The academic task of this research is to see whether differences in teacher quality as represented in teacher value-added are connected to differences in adult outcomes. Adult outcomes are expressed like so:

> Our data on students' adult outcomes include earnings, college attendance, college quality (measured by the earnings of previous graduates of the same college), neighborhood quality (measured by the percentage of college graduates in their zip code), teenage birth rates for females (measured by claiming a dependent born when the woman was still a teenager), and retirement savings (measured by contributions to 401[k] plans) (Chetty, Friedman, & Rockoff, 2012, p. 60).

The specific student achievement data under investigation consist in a nigh-overwhelming number of data points: "The school-district data include approximately 18 million test scores. Test scores are available for English language arts and math for students in grades 3–8 from the spring of 1989 to 2009" (Chetty, et al., 2012, p. 60).

In the first place, to address Duncan's call to "tell the truth" about proficiency such that the standard might be useful in generating data, the *actual* nontest measures of "proficiency" that Duncan has in mind are both *extremely* broad and empirically available only in the future, not in the present. To address the merely the second part of this objection,[35] Duncan's notion of proficiency has explicitly to do with such future life events as graduating from high school, being accepted to and succeeding in a competitive college, and not requiring remediation beyond high school. In order to render this concept available in the present, one has to tie it to something empirically available *in* the present; say, student test scores generated yearly. But the available means of establishing a connection between student achievement and the outcomes Duncan (as well as Chetty) has in mind involve using achievement data *from the past* and connecting it to *present* outcomes. What will count as a proxy measure of "proficiency" will then be established in accordance with measures of achievement that have *in the past* proven sufficient to secure the outcomes that Duncan takes to be entailed in the concept of proficiency.

Since Duncan's striking example of the benefits of using achievement data involves looking a young child in the eyes and saying something *with certainty* ("you will absolutely get there") about his or her progress toward acceptance to and success in a competitive college, one might offer two points with respect to the insufficiency of correlating past student achievement with present adult success (as indicated partially by college enrollment

---

[35] To address the first point, which turns out to be peripheral to my argument: the ability of proxy indicators for holistically understood adult outcomes to secure the validity of measures of reading and math skillfulness is laughably flawed. There is no world with which I'm familiar in which an appropriate response to "show me how well you can read" might consist in gesturing to the zip code in which you live.

and quality) and taking those as offering secure and scientific warrant for making inferences about the future.

In the first place, the criteria for admittance to college is an unpredictable and moving target. In addition to the most obvious fact that federal policies, particularly in the area of equity and non-discrimination, (a) affect such criteria and (b) have a habit of changing over time, sometimes dramatically, sometimes even to the point of establishing criteria *where there were no criteria before*, colleges and universities themselves experience serious fluctuations in, for example, the number and quality of their applicants, their acceptance rates, and the achievement measures of those that it accepts. In the spring of 2012, the University of Pennsylvania, surely among the "competitive colleges" to which Duncan alludes, announced that it had accepted 12.3% of its applicants for the coming semester (L. Lee, 2012). The same article notes that in 1991, one of the years from which Chetty, et al., take their data in order to construct teacher value-added, the same university admitted 47% of its applicants. In sum, Duncan wishes to tie "proficiency" to the criterion of "college and career readiness." The data from which he derives proficiency, under the rosiest of circumstances, might serve as an adequate ad hoc proxy for "college and career readiness" *now*. But the students and teachers to whom this quantified criterion will be applied as a measuring stick are not concerned with college readiness *now,* in *today's* world; data seems to offer some sort of knowledge about "college and career" readiness, but that concept itself is too unstable to be so cornered.

Another wide-ranging study documents the same increase in selectivity across the entire range of higher education, and, significantly, also depicts a marked rise in college entrance exam scores associated with students accepted to "competitive" colleges (Bound,

Hershbein, & Long, 2009). Further, it remains far from clear that the *value* of acceptance to and success in college – as expressed in the economic data Chetty employs – remains constant over time. A recent study of the effects of the Great Recession on the job prospects of Americans who graduated college between 2006 and 2010 identifies significant differences in pre- and post-recession figures in terms of unemployment and underemployment (Godofsky, Zukin, & Horn, 2011). Considering the "career" half of the college-and-career readiness indicates that the connection between college graduation and other indicators of adult outcomes, such as earnings, fluctuates in its own right, as well.

Both of the above objections boil down to the following: correlations of the kind Chetty and his colleagues rely upon to establish certainty about the value of education, and of teaching in particular, always contain *ceteris paribus* conditions, which means that these correlations are valid, all other things being equal. The studies mentioned above document massive shifts over the course of five years (in one case) and over 20 years (in another) in the criteria for acceptance to college and in the criteria for postgraduate career success. It is manifestly the case, in other words, that to claim that the correlation between a given level of academic achievement, as represented in math and reading scores, and specific future outcomes will is valid, *all other things being equal* is as much as to claim that "these things *might* be so, or they might not; we really don't know." All other things will emphatically *not* be equal. *That* is a statement of certainty. The *ceteris paribus* conditions here absolutely undermine such studies' ability to infer future consequences in light of such correlations.

But, even beyond this objection to the above problem, as Hubert Dreyfus notes, "what 'everything else' and 'equal' mean in any specific situation can never be fully spelled out" (H. L. Dreyfus & Dreyfus, 1986, p. 80) because "these ceteris paribus conditions never

capture, but rather presuppose, our shared background practices" (H. L. Dreyfus, 1991, p. 75). This is related to the ever-shifting criteria upon which colleges and universities admit students, and the ways in which institutions of higher education weigh those criteria against each other. The unpredictable changes in college admission *in general* (which the Bound article implies is linked to the shifting place of a "college education" in the American form of life) and in the labor market for college graduates are merely two of the unreckonable number of contextual features that make up the "everything else" which would need to remain stable in order for the correlations to remain meaningful. It is difficult to see why one ought to assume that achievement levels that have *previously* led to the desired adult outcomes would seamlessly continue to serve to secure those outcomes in the future.

In "telling the truth" with regards to "proficiency," then, which one must do in order to *establish* the criterial measures by which a child's future will be predictable, one looks to present examples of adult outcomes that one imagines the notion of academic "proficiency" to entail and traces them backward to educational outcomes expressible *merely* in terms of student test scores (as opposed to student GPA or any other available fact about student achievement). Correlation, for these purposes, proves perfectly sufficient. To look at these adult outcomes *is also* to look at levels of academic proficiency in the form of student achievement data; that is what the establishing of the correlation between academic success as children and real-life success as adults facilitates. One can thus isolate a given level of adult outcome – say, earnings – and identify the range of reading or math test scores in grades 3-8 (or *growth* rates, or teacher VA; these are all variations on a theme) associated with these outcomes. Some specific number in that range will then become the visible and

present representation of "proficiency" against which current students will be measured.

This is what "telling the truth" or, conversely, "not lying to children" amounts to.

For policy purposes, test scores associated with Duncan's massively broad notion of

"proficiency" must rest on some sort of ground in lived experience. Chetty's study proves

useful here because of the specific adult outcome measures they take to indicate that

experience:

> Earnings, college attendance, college quality (measured by the earnings of previous graduates of the same college), neighborhood quality (measured by the percentage of college graduates in their zip code), teenage birth rates for females (measured by claiming a dependent born when the woman was still a teenager), and retirement savings (measured by contributions to 401[k] plans) (Chetty, et al., 2012).

The researchers expect that *what a successful adult consists in* will be captured in

terms of earnings, college attendance, college quality, neighborhood quality, teenage birth

rates, and saving for retirement. Without objecting to any of these as criteria of successful

American adults, I wish to point out the function of the parenthetical "measured by"

disclaimers in the list of criteria they explored. While it is *also* the case that the list of

indicators itself includes an implicit ellipsis – it is not supposed to be (one hopes) the sum

total of all characteristics of successful adults – the "measured by" caveats indicate that *each*

individual characteristic opens up on its own list of potential "measured by" conditions.

How would the data (and thus the concept) pertaining to "neighborhood quality" change, for

example, if it were calculated according to the number of restaurants within walking distance

(and what is walking distance?) or the average level of job satisfaction, or the average home

price, or any other characteristic? Neighborhood boundaries, it might also be noted, are

defined in the study by zip code boundaries. Perhaps this is, in some metropolitan areas,

something close to a reasonable approximation. But I think many would balk at the

uncritical acceptance of the designation of Northfield, Minnesota, and all of its environs (zip code 55057), which contains thousands of acres of land, two colleges, three area codes, and a population in excess of 20,000 people, as a single "neighborhood."

The truth-telling about proficiency *upon which* standards will be based, and thus according to which students will be assessed and test scores generated, claims to ground itself on *facts* about the conditions for adult success as well as *facts* about the nature of adult success itself: for Duncan's purposes, these are the nontest factors involved. The point of raising the chicken-and-egg specter is that "the conditions of adult success" depend upon both the way we establish *measures* of adult success as well as what we take the (relevant) conditions thereof to be. The generation of data-based criteria for "telling the truth" in education requires the generation of data that *reveals* the truth in the first place. Where are the actual *facts* upon which these criteria should stand, absent other *ceteris paribus* measures of the same? Duncan extolls the role of data analysts in exposing "the research and the facts," in moving beyond "ideological statements and the surface conclusions and find[ing] out what is really happening for our children in our classrooms" (Duncan, 2009b). But it remains far from obvious that these "facts" – on the basis of which we will be able to *know* "what is really happening for our children" – escape a reliance upon radically incomplete approximations and fallible indicators of the concepts that matter. As Wittgenstein notes at one point, "the ground keeps on giving us the illusory image of greater depth, and when we seek to reach this, we keep on finding ourselves on the old level" (Wittgenstein, 1983). Duncan's hopes for data's ability to dive below what he calls "ideology" to the "facts" do not seem to yield the appropriate depth he imagines necessary to his project. Even if Chetty himself were to concede the incompleteness of his "neighborhood quality" measure, for

example, it is entirely unclear how one might go about *completing* it, since the factors that contribute to what we mean when we talk about neighborhood quality, as Dreyfus suggests, cannot themselves "capture, but rather presuppose, our shared background practices."

The chicken-and-egg problem thus recurs: Establishing the appropriate quantitative reflection of real-world success requires converting real-world outcomes into a form amenable to quantitative reflection. But no such real-world outcomes are present in Chetty's study; merely quantitative proxies thereof. For reasons of both the necessary *futurity* of the actual non-test measures that will reveal the quality of teaching (if we concede this straight-line causality, and in which case we would have no need of achievement data anyway), and of the necessity of using proxy measures or signifiers of a certain type of "adult outcomes," the "college and career readiness" criterion simply never touches the ground.

This notion strikes at the heart of the primary claim to superiority that Duncan makes on the behalf of student achievement data. The advantage that the use of student achievement data with regard to teacher quality promises has to do with a superior ability to evade problems of uncertainty and doubt that seem to render the other forms of discourse on good teaching insufficiently trustworthy. As we have seen, while Duncan feels comfortable enumerating the criteria of "what success looks like," such claims to immediate visibility are *only* authorized by virtue of something like common sense or agreements in forms of life, and thus they feel somewhat inadequate when brought to bear in a justificatory capacity upon personnel, tenure, and compensation decisions. The autobiographical and narrative accounts of good teaching made by both Neal and Duncan have greater warrant, it appears: the autobiographers are *themselves* successful adults and their first-hand accounts underscore the connection between the teaching of an adult in their past and their current success. However,

such accounts, pertaining to teachers long since out of the classroom, are of no use at all in the present (unless, perhaps, for the generation of lists of highly-general good-teacher characteristics). Accepting the linkage between the teacher's practice and the adult success also, as we have seen, requires a measure of trust in another's perspective. Finally, were this not sufficient reason to doubt, such acceptance depends upon agreement in the sense of "success" that the narrator embodies.

Duncan's discourse in the lead-up to Race to the Top offers the availability of longitudinal student achievement data as an answer to these myriad problems, a way of eliminating the uncertainty that has heretofore seemed to preclude effective means of evaluating teachers. However, the extensive analysis in this section reveals the use of student achievement data to exhibit the *same* insufficiencies, the *same* flaws, as the other forms of discourse that appear in Duncan's rhetoric.[36] The data supposed to ground the "facts" of the relation between "proficiency" and "successful" adult outcomes fall prey to all three of the problems with the narrative autobiography: the correlations established between past student test scores and present measures of adult "success" are (a) of extremely limited use in predicting *future* correlations, given the wide vagaries of chance or fortune, the ever-shifting ground of social and technological change, and so on (I focused merely upon the implications of changing criteria for college admission and the changing monetary value of a college degree, but obviously possible axes of change simply proliferate exponentially), and (b) dependent upon a given researcher's viewpoint with regards to the components of adult

---

[36] The *forms* of the flaws' appearance and non-appearance, though, are not the same, and perhaps this is an important point to make. The anxiety one feels at the prospect of relying on another person's judgment vanishes (in most cases, or in the case of the general public) when one looks at the Chetty study. One does not *feel* that Chetty's conclusions are suspect in the same way. Why not?

success – to say nothing of the list of "student factors" such as ethnicity, socio-economic status, and others – that he or she finds relevant (or available in the necessarily quantifiable terms) and consequently "accounts for," statistically speaking.

Most importantly, however, like both the immediate visibility claims offered by an independent observer and the autobiographical narrative of a currently-successful adult, accepting a judgment of teacher quality derived from student achievement data *also* requires the kind of haphazard, unstable agreement in the concepts of success and proficiency,[37] for example, to which the use of achievement data is supposed to offer a remedy. Though student achievement data interposes an interim layer between its results and its reliance upon agreement in forms of life, an interim layer which can seem, at first blush, as solid as natural science, this interim layer of measures of adult success and student proficiency itself rests upon nothing more solid than the sharing of shared background senses of these concepts.

3.1.3 Achievement Data and the Problem of Conceptual Change

The panacea visible thanks to the availability of student achievement data will allow for a situation, as Duncan says, wherein "one day we can look a child in the eye at the age of eight or nine or 10 and say, 'You are on track to be accepted and to succeed in a competitive university and, if you keep working hard, you will absolutely get there'" (Duncan, 2009b).

---

[37] I've been wondering throughout this section whether or not I'll ever get around to pointing out explicitly the dramatic gap opened up between what Duncan appears to attribute to "proficiency" – which, since it entails so much in the way of holistic future consequences, cannot properly be said to be limited to *a single* skill domain at all – and the concept of "proficiency" as Koretz and psychometricians generally use the term. Proficiency in the latter sense would simply be proficiency at *reading*, for example, which would hardly seem to entail overall *academic* success: high GPA, participation in extracurricular activities, lucid reasoning, and so on. It is easy to imagine a very good reader who is not a particularly good student. In any case, the point is that Duncan's use of "proficiency" itself is extremely problematic, but I leave that for another researcher.

In an odd coincidence, in 1988, when I was nine years old, my own third-grade teacher made exactly the kind of absolute claim about the future that Duncan imagines in his own scenario. I have a clear memory of this incident. She looked me right in the eye and told me (not in these exact words) that my future success depended, in part, upon improving my penmanship. She in fact said, "In high school, all of your teachers will require your essays to be done in neat cursive. They will take away points for sloppiness." That *is* a direct quote.

By the time I entered high school in 1993, of course, my teachers required that I turn in all of my essays in typewritten form. The school had a computer lab on the second floor, and many of the students had personal computers and printers in their households. By the time I entered college in 1997, some of my professors insisted that I submit papers by email; by the time I entered graduate school in 2003, the university used dedicated educational software that included secure means of submitting files electronically. Over the course of 10 years, the process of writing and its submission moved from a world of dot-matrix printers to one of electronic file transfer. White-Out's use, one presumes, declined precipitously.

Handwriting is hardly the only aspect of the skill and concept of writing to suffer, if one can call it that. It is no longer as necessary as it once was to outline or to generate multiple (distinct) drafts as part of the writing process: erasing typos, reconstructing arguments, shifting entire paragraphs from one place to another no longer require the investment of time or energy that they once did. Over the course of 15 years, the educational world simply ceased to accept handwritten essays as finished writing products; the process of writing itself has changed. In my academic life, from elementary school to the present, I have never once handed in any form of written assignment in cursive. My penmanship remains, at best, questionable.

Applying Duncan's standard of truth-telling, the obvious conclusion to draw from my experience is that my third-grade teacher *lied* to me about what success in high school and college would require. Duncan touts student achievement data as a *remedy* to this problem, as something that would provide teachers (or parents, or administrators – anyone with access to the data) with the knowledge necessary to "be honest" and "stop lying to children." In the above example, however, one confronts the weakness of the available student data relative to its appointed task of predicting the skills, aptitudes, and proficiency levels in each that successful adulthood will require, even in the short term. My third-grade teacher was doubtless relying upon common sense and twenty-plus years of experience in the classroom in order to guide her pupil to attend to some of the conditions upon which scholastic, and thus later-on adulthood, success would depend. She was incorrect: whatever successes and failures I have experienced, academically, have been entirely irrespective of my penmanship.

But it is especially worth noting that she was not incorrect in suggesting that *writing* would prove crucial to academic success. She looked at my third-grade writing and judged it to lack proficiency: whatever the quality of my sentence structure, my writing was at points, apparently, at least difficult to read, thanks to my poor penmanship. Penmanship was *part of* the skill of writing; it was explicitly one of the factors upon which one's work was graded, along with grammar and style.

Thanks to the re-emergence of writing as part of statewide standardized assessment protocols, however, students once again have to generate timed writing pieces by hand. Tennessee's scoring rubric, according to which students' writing proficiency is rated,[38]

---

[38] Tennessee's is a specific iteration of the 6-trait writing rubric, which is common to most state assessments: thus these criteria generalize beyond Tennessee's particular example.

discriminates among levels of proficiency in terms of rhetorical organization, supporting

evidence, sentence variety, "facility in the use of language," and the number of errors in

usage, grammar, and sentence structure. Illegibility does indeed arise on the rubric, but only

as a means of *disqualifying entirely* an essay from evaluation itself; in this capacity, it is

listed as a criterion among "blank or refusal," "off-topic," and "written predominantly in

another language." In order to be disqualified, an essay's illegibility has to render it akin to a

blank sheet of paper or something in *another language*. Penmanship is no longer among the

qualities of writing *to be evaluated*; it is a now a background condition *for the evaluation of*

writing itself. It no longer spans a value spectrum from "extremely poor" to "excellent."

Penmanship either makes the essay's reading possible, or else it does not. It no longer admits

to variation beyond that minimal standard. The word "cursive" does not appear at all.

In drawing out the differences in the requirements for writing visible among what my

third-grade teacher told me in 1988, what my high school required in 1993, what my

undergraduate institution demanded in 1998, what my graduate school made possible in

2003, and what the state of Tennessee takes as demonstrative of writing's proficiency in

2012, one sees something like an evolution of the *concept* of writing itself. This conceptual

fluidity makes itself visible by means of the changing criteria according to which writing's

quality is judged, criteria directly tied to the sense of what writing, as a skill domain, *is*.

Technological developments in word processing and network connectivity, I would suggest,

have led to a separation between the skill of "writing" itself and the skill of "letter- and word-

forming" by means of physically inscribing a mark onto one surface or another. Even in

timed-writing situations that *require* such mark-making skill, that skill is only taken to apply

to the concept of "writing" as a necessary condition, not itself something subject to evaluation.

Student achievement data, as a provider of uniquely justifiable warrant with regard to the future – and therefore as a credible basis on which to evaluate student progress and teacher quality in the present – founders on precisely this sort of *conceptual,* and therefore *criterial,* change. "Reading," "writing," and "math" have claims to something like permanence in academics, as skill domains related to successful adulthood, however one defines the latter. But none of these skills are *conceptually* stable over time, even over the short period of a single student's academic life. Neither is this phenomenon limited to the domain of writing. To continue to use Tennessee as an example, the state's reading standards for grades 6-12 currently include the skills of "previewing and reviewing print *and non-print* texts" (emphasis added) and "understanding visual representations" (*Reading in the Content Area*, 2012), neither of which skill one can imagine necessary apart from a world of prevalent digital media. What we take reading, writing, and even mathematics to *consist in* changes along with our worlds at large.

Student achievement data as represented in test scores, as we have seen in Chapter 1, requires an *explicit definition* of a given skill in order to generate a measure of a given student's proficiency in that skill area. But the generation of an explicit definition requires the freezing of *relevance*, a problem that I discussed in Chapter 2 with regard to the failures of Artificial Intelligence and a "best practices" approach to teacher development. The problem with the freezing of relevance in the area of measuring student achievement has directly to do with the *manifest fact* that relevance even in such seemingly stable domains as reading, writing, and math remains fluid. What was taken as *relevant* to evaluating the skill

of writing in 1989 (the first year of Chetty's longitudinal data) had become in fact *irrelevant* to such evaluation by 2009 (the last year of Chetty's data). Statistical methods of smoothing out differences in tests over time cannot cope with this type of variation; *the domain* has changed.

The requirement for an explicit definition of proficiency in a given skill domain at a particular point in time significantly limits the *predictive* utility of student achievement data: my third-grade teacher may have been wrong in believing that cursive was necessary to my academic success, but in this capacity achievement data could not have offered a corrective. Duncan's notion of an integrated data system that will enable a person to *honestly* tell a third-grader whether or not he or she is on track for success at a competitive university cannot provide the degree of certainty about the future with which he associates the concepts of "truth" and "lying."

I would not feel comfortable claiming that my third-grade teacher lied to me in telling me that I needed to work on my handwriting in order to succeed at the high school level; I would, however, freely say that she was wrong. The distinction might be expressed in terms of her (perceived) ability to predict the future. All of her past experience in the classroom with students probably similar to me led her to the incorrect conclusion about the relevant writing skills to which I should attend. Duncan's claim about the possibilities afforded by this newfound wealth of student achievement data is that future conditions will be at last predictable, next to which standard, of course, one might really *be* in a position to deceive or not. The fact of conceptual change over time, however, belies such utopian claims. Using student achievement data (or, really, any objective measure) fails to solve the problems that crop up when relying upon fallible human judgment.

3.1.4 Conclusion

I noted in an earlier section that I wished to focus upon the *similarities* shared between the forms of data-driven teacher evaluation that Duncan's Race to the Top initiatives propose and the ones that such methods are intended to supplant. This section seeks to highlight the degree to which – to hearken back to the beginning of the section – *what* one takes as indicative of educational achievement and good teaching has ineluctably to do with Wittgensteinian and Austinian criteria, as Cavell calls them, with both the *use* of the concept in everyday life (the concept's relation to other concepts) and with the protean features by which individual states propose to *identify* these concepts. The implying of a causal relation between particular teacher quality (either in terms of particular behaviors or according to contribution to score growth in a particular domain) and particular future student outcomes (in any of the terms of the Chetty article, for example) falls apart on the encounter with temporal criterial change.

While one might make the argument that the future will not be *so* radically different from the past as to dampen the case for taking *former* correlations to be roughly indicative of *future* ones, three objections arise to this line of thought. In the first place, the future as a whole does not have to shift a great deal in order to radically destabilize all sorts of concepts[39] – one does not, for example, have to imagine that the concept of adult outcomes will in the future be *unrelated* to income in order to concede that the particular skills

---

[39] The sneakily hilarious remake of *21 Jump Street* provides an unexpected example. Only five years removed from high school, the former prom king returns as an undercover cop to discover (correctly, in real life) that everything he knows about what it takes to "succeed" in terms of coolness has changed. The place of violence in male relationships, the role of political knowledge in social situations… all of it leads Channing Tatum's character to feel suddenly that he cannot find his feet in this school, the place that for all the world is *just* like the school he ruled five years prior. That these scenes function as comedic at all speaks to the recognizability of the phenomena depicted.

necessary to achieve high incomes will change, a shift that will certainly reorder (but unpredictably) the relevance and place of *each* skill, disposition, and talent in the practical cosmos. Secondly, and relatedly, even conceding the unlikeliness of a radically different world, the use of former correlations as stand-ins for the future straightforwardly discounts the possibility of *anything* new arising – this, I take it, is similar to the Chapter 2 argument against adopting a "best practices" approach to teacher development. Strictly defining skill domains and concepts in the present (or based on the past) unwarrantedly restricts recognition of actual quality in the future as the conditions for the demonstration of such quality invariably arise.

In the third and final place, one perhaps *ought* to concede the merit of using student achievement data in measures of teacher quality based on precisely the claim that former correlations will likely be *roughly* similar to future ones. But having conceded this fact, it is difficult to see the warrant for Duncan's *prioritizing* of these measures to the exclusion of all others, his assumption that with the availability of student achievement data we have solved some great mystery and stand on the threshold of knowledge. If a ragged basis for rough estimation (albeit a highly technical and mathematically demanding basis) is, in the end, what the use of student achievement data offers, then it is welcome, it would seem, to join the extant chorus of evaluation protocols, as this is what they offer, too. But achievement data seems to merit no *special* place therein.

Duncan suggests, and scholars echo him in this, that the use of student achievement data puts us in the position to say something relatively certain about teacher quality, something objective that perhaps approaches truth, and that, in this respect, it is a *new* development – *progress*, one might say. The use of student achievement data offers a

general view of a teacher's impact on students, which autobiographical narratives cannot; and it holds out the possibility of predicting the future, which third-person observations dare not. But its generality *cannot* speak to the experience of any singular student, which other methods *can* do; its notion of what a teacher's "impact" might consist in remains much more restricted than a notion available to existing methods; and its ability to see the future remains frustratingly – merely – *human* in its scope. That it offers something that other methodologies cannot means that it ought to speak up; that its limitations are real means that it perhaps ought to raise its hand like everyone else.

<u>3.2: The Non-Formalizable and Appeals to Criteria in</u>
<u>Measurement</u>

To return to a point that I have made earlier, but one that bears reiteration in transitioning to the next section, I wish to say once more and loudly (through cupped hands, even): the point of the above is emphatically *restricted to* a serious critique of the *assumptions of overall superiority* in terms of justifying evaluative claims that Duncan (among others) makes on the behalf of student achievement data. It would be easy to misunderstand the extensive delineation of the *weaknesses* of student achievement data with respect to evaluative purposes as amounting (or leading) to a claim either (a) that it is impossible to justify through empirical normative means any value claims about a teacher's practice at all, or (b) that student achievement data in particular ought to be banished from the playing field entirely. I am in fact making neither claim. The argument I am urging has to do with recognizing and responding to the limitations inherent in the generation, analysis, and use of student achievement data *so that* it can play an appropriate role with respect to other forms of evaluating teacher performance.

Duncan's tendency to overstate the benefits and possibilities associated with student achievement data has at least partially to do, I suspect, with a restricted view of measurement itself, particularly as it relates to education. Duncan's view of measurement seems limited to the sense of comparing a given object of investigation (say, a tree or a pig or a 40-yard dash or a student's reading ability) to some *absolute* and external measure (say, a yardstick or the kilogram or the output of a stopwatch or the cut-score for proficiency). The sense of requiring a *universality* of this measure – its "reality," if you will, beyond a merely human scope – invites commentary from the world of measurement philosophy in general, the field of metrology. Robert Crease's wonderful book on the history of measurement, for example, describes the millennia-long search (nearly completed, except for measures of mass) for unchanging, naturally-occurring physical constants that one might use to standardize all physical measurement – a system of *absolute* measures – and it concludes with worthwhile thoughts Duncan's implicit view of measurement.

Crease differentiates between the type of measurement described above, the sort I take Duncan to have exclusively in mind, and another sense of measuring. Duncan, and all those who seek out the absolute standards according to which objects of exploration can be measured, are involved in a project that Crease calls "ontic measurement, after the word applied by philosophers to real independently existing objects or properties," which, as he says, "establishes that one property is greater than or less than another, or it assigns a number to how much of a given property something possesses" (Crease, 2011). Duncan's quest, echoed in Chetty's research, to *ground* the criterion of "college and career readiness" in some concept of "proficiency" requires that "proficiency" *be* such a "real independently existing. . . property," a property existing independently, that is, of something like human forms of life,

from which they must be somehow separable or isolable. Over the course of this study, we have been observing failed attempts to *establish* or *effect* that separation.

But Crease discusses a second type of measurement, as well: "ontological measurement." His description thereof ought to be given space:

> This is the kind of measurement that Plato said is guided by a standard of the "fitting" or the "right." This kind of measuring is less an act than an experience; the experience that things that we've done, or we ourselves, are less than they could or should be. We cannot carry out this kind of measuring by following rules, and it does not lend itself to quantification. . . . . Ontological measurement connects us with something trans-human, something *in* which we participate, not something *over* which we command (Crease, 2011).

In tethering the notion of ontological measurement explicitly to the region of the trans-human, which he nicely describes as "something *in* which we participate, not something *over* which we command," Crease gestures to the region that Duncan's attempts at bringing student achievement data to bear seem unable to penetrate. Race to the Top's teacher evaluation protocols require an *explicit* standard, explicit definitions of "student teaching" and "teacher quality" against which to measure. The ways in which we have seen the process of explicit definition distort by necessity the conceptual terrain to which it wishes to speak reveal the serious limitations of ontic measurement in terms of assessing those endeavors in which, as Crease also notes, "our being is implicated."

It is nonetheless obvious that one *can* generate ontic measures of teacher quality and student achievement to some extent: constructing standards, norming test instruments, making statistical inferences, and calculating teacher value-added are all evidence of the impulse to and possibility of using ontic measurement in the domain of education. Even if one perceives the insufficiencies of such ontic measurement, and even if such a perception might express a wish to *deny* the place or appropriateness of ontic measurement, one would

still want to recognize the real *need* of which the recourse to student achievement data is an expression. In the same way, given what is surely felt to be an insufficient basis for *generality* in simply comparing a given teacher to some paradigmatic "example," as ontological measurement would involve, one would also do well to recognize the real *need* to which ontological measurement speaks, a need wholly unaddressed by means of ontic measurement.

Perhaps it would be well to consider that the introduction of this second sense of measurement forces us, in the particular case of education, to *decide whether and in what circumstances* teaching and learning are matters best suited to ontic, as opposed to ontological, measurement and vice versa. After all, it seems equally obvious – or, this obviousness is something I am seeking to bring out – that teacher quality and student achievement, no matter how precise or complex the statistical manipulations brought to bear, is *also* always a matter at once conducive and inescapably tethered to ontological measurement, which, as Crease says, *defies* attempts at quantification and other forms of abstraction or external representation.

At the heart of the issue expressed in Duncan's laudatory view of using student achievement data in the evaluation of teachers lies, once more, differing senses of where the generality in *standards* lies or originates. For Race to the Top policies quite generally, as we have seen repeatedly over the course of this study, such generalizability, such standardization, is securable *only* by means of establishing or constructing (not to say discovering) a purportedly ontic property or concept (if one may so call it) that stands over and outside of the educative endeavor itself, against which matters requiring evaluation can be compared. The lack of something *ontic* seems to Duncan to invite, if not necessitate, the

sort of malfeasance expressed in the New Teacher Project's 99% statistic that has arisen

throughout the study. Addressing the need to get serious about evaluation requires a

scientific basis of certainty in order to combat the lack of seriousness evident in that statistic.

Crease, however, speaks of ontological measurement as though it, too, were

standardizable on grounds simply other than the sort of universal constant to which Duncan

aspires.   Crease continues in his description of ontological measurement:

> Is this only "metaphorical" measuring? It is comparison against a standard. Placed alongside the fitting or the right example, our actions—and even our *selves*—do not have enough being; there is more to be. . . . Ontological measuring involves no specific property, in a literal-minded respect, for it involves nothing quantitative. Calculate all we please, we will never produce this kind of measurement. No method can lead us to it. Ontological measurement connects us with something trans-human, something *in* which we participate, not something *over* which we command (Crease, 2011).

It is not at all obvious that teaching and learning are the kinds of things describable,

ultimately, in terms of the properties of light or energy; and yet the ontological standard to

which we compare ourselves is nevertheless *standardized* according to, as he calls it,

something "trans-human," something "*in* which we participate, not something *over* which we

command." While positing the real existence of a standard, he grounds the *constancy* of this

standard in the "trans-human," something all-pervasive and frustratingly non-formalizable.

Crease seems to acknowledge a normativity or standard that is externally given, and

yet does not appear to be *located* anywhere in particular, specifically anywhere separable

from the activity with which it is bound up. I contend that his "something trans-human"

points in the same direction as Cavell's and Wittgenstein's "agreement in forms of life." In

contrast, after all, to the "absolute physical constants" sought by ontic measurement,

ontological measurement, as Crease suggests above, involves placing "our actions—and even

our *selves"* "alongside the fitting or the right *example*" (emphasis added). Measuring according to a holistic example (of beauty, justice, masculinity or what have you) invites a contrast with the Chetty study cited earlier.

Chetty and his colleagues, as I have explored, take a number of factors to be indicative of adult outcomes. As I once more noted earlier, it is manifestly the case that the list of factors is incomplete with regard to characterizing what counts as adult success and that each indicator requires a "measured by" caveat, implying the incompleteness of each and every individual indicator itself. "Adult success" seems to defy *full* spelling-out in terms of abstract characteristics. In even *offering* the suggestion, however, that there is more to the concept, it might be noted that I have in mind *other* possible indicators than the criteria Chetty cites. This is as much as to say that I have *examples* of successful adults in mind – outcomes worth aspiring and educating toward – that either are not exhaustively described according to Chetty's factors or would be misidentified by those factors. Insofar as the list of factors that Chetty compiles does not *capture* what and who these examples *are*, the success that they *embody*, I am explicitly involved in ontological measurement as we speak, weighing Chetty's criteria against my own, against the successful adults that I know.[40]

---

[40] Part of what motivated my rebellion against the "neighborhood quality" indicator in the earlier section, I now realize, was a disparity between an example of a successful adult I had in mind and the neighborhood quality criterion. An economics professor that I know, a good man and wise, lives in a little house outside of the rural college at which he teaches. Not only would the measure of college-grad percentage not favor him, but the breadth of the zip code boundary encapsulates so much land, so many people and households and farms, that it in no way could be called a neighborhood. And it struck me that Chetty's measures *assumed* that "successful adults" would live in densely populated urban areas, where zip codes *would* be a (more) reasonable approximation of neighborhoods. In the conflict between Chetty's implicit claim to have come up with an (ad hoc) ontic measure of adult success and the example of my econ-prof friend, it was *Chetty's criteria* and not my example that felt suddenly, arrogantly, infuriatingly arbitrary. Chetty's measure assumes that successful adults live in cities that have neighborhoods that align more or less with zip codes; my successful-adult friend does

The examples from which Chetty presumably *takes* or *extracts* his criteria for success in the first place seem, given the measures, to be economically-motivated urban dwellers. An example I have in mind is better characterized as a risk-averse rural professor. Despite these differences in characteristics, I take it that I would be able to recognize Chetty's examples as successful and vice versa. But on what basis? Citing a contrast between *examples* and *ontic measures*, Crease in effect notes that the recognition of the paradigmatic status of examples is *irreducible* to ontic properties: it is a matter of *family resemblance*, available (or not) according to the sharing of shared forms of life. The notion of *sharing* invokes Crease's appeal to the "trans-human," and the invocation of *examples* recalls Wittgenstein's discussion of teaching someone else the meaning of words: "But if a person hasn't yet got the *concepts*, I'll teach him to use the words by means of *examples* and by *exercises*. –And when I do this, I do not communicate less to him than I know myself" (Wittgenstein, 2009 sec. 208). Wittgenstein and Crease share a view in which the holistic example, rather than an example as a composite of isolable features, provides the best (if it would be too much to say the *only*) ontological measuring stick.

---

not live in a city. I expect that Chetty the person (not the data analyst) would recognize him *as* successful, and I expect no less that I would recognize as successful those that Chetty the person takes to be successful. That his measure of "adult success" would fail to identify my friend, and that they would also stand at least a strong chance of misidentifying as "successful" some people that I would never hold up as examples worth aspiring to, indicates that the measure (indeed, all such measures) is imperfect, that it is subject to *disagreement*. But precisely the supposed value of ontic measurement is its ability to establish reasonable certainty and thus put an end to disagreement. Since it does no such thing, and since the *basis* of the possibility of disagreeing about the measure is *itself* a comparison between the measure of success and the paradigm cases one has in mind, it is unclear why we ought to prefer the use of the intermediate measure to direct comparison of this or that case with some paradigm. (It is not *entirely* unclear – one sees easily why one needs such measures to make broad claims about broad groups of teachers; what is unclear is where the measure gets its authority with regard to *specific* teachers, *specific* students.)

Wittgenstein's contention, however, that in pointing to examples of the quality in question, he "does not communicate less" than he knows himself casts a sort of negative light on the possibility of communicating something similar by means of ontic methods. Where ontic measures of "adult success," for example, might succeed at getting at *some* aspects of successful adulthood, despite remaining and demonstrable incompleteness, Wittgenstein suggests that through ontological measures we might have it all (or at least no *less* than we ordinarily have). I raise this issue now in order to bookmark it for later development. If pointing to examples were, in fact, simply the superior method of defining a "good teacher," and if that really were the core issue at stake, the need for ontic measures, for a universal standard, would not have the grip that it does. For whom – for which person or class of person – will the pointing out of examples not serve?

The way that Crease distinguishes the *in*-which from the *over*-which recalls McDowell's comments on the rules-as-rails view that I quoted in Chapter 2: "If one is wedded to the picture of rules as rails, one will be inclined to think that to reject it is to suggest that, say, in mathematics, anything goes: that we are free to make it up as we go along" (McDowell, 1981, pp. 150-151). The *rails* picture, in which the rule would cause, as it were, some particular response or behavior in a particular circumstance by absolute necessity, invites comparison with the historical project toward an absolute system of measurement that Crease cites, in that the drive for the *absolute* system speaks (also, among other things) to the notion that something *is needed* to stand behind all instances of measurement, something beyond *this* person and even *this* meter-stick, something without the which measurement as such cannot be said to take place.

The discomfort with rejecting the rules-as-rails view that McDowell cites is precisely associated with transforming rule-following practices into what Crease calls "something *over* which we command," as though all uses of the meter prior to its tethering to a natural constant (which is to say, the entire history of the meter, excepting the past 30 years), *must* have entailed the coherent possibility that every single person on every single occasion might have simply "felt free to make it up" in a given instance. Turning away from a rationalist view in which the standard for comparison *must* be ontic in nature seems to amount to an assertion that "we are free to make it up as we go along." The fear, then, that the rationalist assumption expresses and to which it attempts to respond is that, if a measure or an explanation is not *inhuman* or naturalistically objective, then it must be *merely* human, whimsically or corruptibly or fallibly *subjective.* Without denying historical facts of mischief in measurement for devious purposes (for which and from which we also have set phrases, like "putting one's thumb on the scale"), it seems equally the case that no such *feeling* free from the trans-human constant of the meter has ever pervaded human society such that radical subjectivity with regards to the evaluation of length has ever reigned. Our *having* of a normative standard does not seem, therefore, to depend on our ability to explain *how* we have such a thing. Nor, it might be said, does something like the converse apply: *having* an explanation for a given standard does not *ipso facto* make such a standard the *correct* one from case to case.

Crease's point in referencing ontological measurement is that this rationalist discomfort – wherein to lack an *inhuman* measure is to lack the capacity to *make meaning* from the measure – is unwarranted: what Crease will refer to as the "ontic measurelessness" of ontological measurement does not amount to measurelessness *as such*, or require denying

the existence of a meaningful standard. To acknowledge the axis of the *in*-which that ontological measurement registers, the trans-human normativity, is also to acknowledge and respond to the phenomenological fact that changes in shared background practices or ways of being in the world, in light of which ostensibly stable skill domains *are* what they *are,* bring about criterial changes in those skill domains themselves, rendering ontically-aspirational measures derived or constructed at any particular point in historical time *obsolete* at a certain point, in addition to their inherent flaw of being, by necessity, radically *incomplete*.

Ontological measurement cannot claim the "absolute" status of physical constants; but then, reading, writing, and math – to say nothing of "teaching" and "learning" – are not objects of natural science, related to one another according to the natural laws of physics. It is not obvious that their measurement *ought* to be expressible in terms, ultimately, of the properties of light, for example; nor is it obvious that attempting to construct ontic measures on the basis of which to ground judgments of educational quality requires the sort of "courage" that Duncan attributes to the endeavor. Such an attempt may, in fact, express something better characterized as *fear* in the face of the necessity of adapting to changing *ontological* conditions.[41] In either case, however, Duncan's mention of *courage* in connection with the need for objective measurement and data analysis marks, with its eyebrow-raising incongruity, *something* that will require exploration in an ensuing section.

---

[41] The point is not yet to *say* whether the impulse to the ontic represents bravery or cowardice. Duncan uses this courage trope both in his speech on data and in his speech to the NEA. He criticizes the NEA for being hostile to change, claiming that adapting to circumstances requires this sort of courage to deal honestly with the topic at hand. I simply suggest that "hostility to change" and a failure to "deal honestly with the topic at hand" are also the criteria according to which one could label Duncan's data-driven reform a version of cowardice.

The re-emergence of discourses on rule-following makes visible a salient similarity between prevailing views on educational measurement and teacher practice in the reform movement to which Duncan belongs. Chapter 2 dedicated itself to highlighting the shortcomings of the rationalist assumption with regards to explaining and reproducing excellent teaching: the rationalist assumption, recall, holds that skillful practice (as a form of intelligent behavior) is both reducible to and explicable in terms of external or abstract representations – in fact, the rationalist view holds that the *having* of these abstract representation of a skill domain and the rules for navigating it *account* for such skillful practice. Duncan's view of measurement, wherein he claims a privileged position for the use of student achievement data, also requires something external or abstract in order to justify value claims. Just as the rationalist-assumption requirements for teacher practice proved unable to account for or reproduce skillful behavior in Chapter 2, the ability of external or abstract standards and criteria to successfully dispel doubt with respect to the value claims about students and teachers it purports to justify has been rendered dubious at best in this Chapter. The arguments against each of these abstract representations of skilled practice and criteria of success boil down to the same form: while it is possible to *derive* serviceable versions of those abstractions[42] (to codify teacher knowledge and best practices, to generate proficiency cut scores based on lived practice and experience), any attempts to reverse the relationship and *generate* skilled practice or *define* skill domains and outcome criteria by means of those abstractions, once codified, fail to respond adequately to ontological-level

---

[42] To *produce* something ontic based on the ontological, which is different from *discovering* something like a "best practice."

changes, evolutions in criteria, alterations in the *whole* of a given shared form of life that each domain and outcome presuppose.

The ontic/ontological distinction that Crease brings into the discussion of measurement includes one further, important point – it is one that I have made earlier in this Chapter; indeed, perhaps throughout – but it deserves reiteration in these philosophical terms. The ontic – what *is* – rests upon (or, in Dreyfusian style, appears in light of and is inseparable from) the ontological – or *how* what is is. Earlier in this Chapter, I described the breaking-down of directions in terms that Crease recalls:

> [Breaking-down directions] is yet another instance of a family-resemblance concept. Instructions, as the most recent example demonstrates, can be *endlessly* atomized; and *all* instructions can, in reverse fashion, be revealed as simplifications of some more complex skill. Some teachers give directions at the *right* level for their students and others remain too vague. That right-ness, one might say, is established or perceived as a matter of situational *fit*, rather than by means of comparison to some external or formalizable or empirical standard that transcends all classrooms.

How an observer is *able to tell* whether a given teacher either has or has not successfully broken down instructions cannot be explained by recourse to the ontic, as the concept of "breaking down instructions" cannot be expressed in exhaustive, abstract, context-free terms; rather, *that* the observer recognizes that a teacher either has or has not successfully decomposed instructions *at all* testifies to a shared ontological sense, agreements in *what counts* as successfully decomposing instructions in just *this* situation.

Note that this argument specifically does not disqualify the *use* of something like an observation rubric in the evaluation of teachers; it rather militates against the notion that such rubrics derive their authority from anything *more than* or *outside of* shared understandings of being. The argument suggests that the rubric's correct *application* in a given classroom relies at bottom upon the same agreement in forms of life as any other form of observational

evaluation. Its grip on a claim to *objectivity* such that it offers some sort of contrast with the *kind* of measurement or evaluation previously available to researchers and administrators is weaker than it appears.

In returning to Crease's thoughts on the nature of ontological measurement, the author notes that,

> While in ontic measurement we compare some object with another object exterior to it, in ontological measurement we compare ourselves, or something we have produced, with something in which our being is implicated, to which it is related—such as some concept of the good, the just, or the beautiful. Ontological measurement is ontically measureless. (Crease, 2011).

This section of Chapter 3 has detailed the ways in which even the ostensibly ontic forms of measuring teachers available by means of the statistical parsing of student achievement data fail in practice to rise to the security of the ontic level upon which Duncan and others rest arguments about these methods' *unique* access to what he calls "the facts." Thus, the various predictive and justificatory claims that Duncan makes on the behalf of these methods and this data do not *stand in contrast* to the means of evaluating teachers that have been available to the profession previously. In part, Duncan's calls for multiple measures of educational success might be understood as an acquiescence to this fact.

This, the most important conclusion to draw from these considerations, takes us back to the problems addressed in Chapter 1. A major aspect of the endeavor in Chapter 1 has to do with uncovering the way that, contrary to Duncan's many citations of the need for "multiple measures" of teacher quality and student achievement, all of the measures that wind up in public policy are either derived from or responsive to student achievement data. The multiple measures of educational quality are better characterized as various iterations of

the same limited version of quality, one defined purely in terms of growth on achievement measures.

If student achievement data were similar to the speed of light in terms of representing a natural constant – if "student achievement" itself were somehow definable in terms of a natural constant – the utter *reliance* on the accuracy of this student achievement data would pose no problems. Chapters 2 and 3 have drawn out both the ways in which the skill domains addressed resist description in such naturalistic terms and the ways in which they depend on agreements in forms of life and shared understandings of being. It is, then, not the *existence* of student achievement data, the *giving and taking* of standardized tests, that represents the major objection to the teacher evaluation protocols encouraged by Race to the Top. Rather, it is the epistemological and axiological *status* arrogated by policy makers to such achievement data that calls for an objection. That *only* evaluations of teacher quality reliant ultimately, if not directly, upon student achievement data will *count* as valid, despite the manifest validity problems such data seem unable to extinguish, appears on the face of it an exercise in folly.

3.2.1 The Call for the Use of Achievement Data and Its Implications

The indications so far available in this Chapter suggest that Duncan's proposals for the use of student achievement data to anchor or justify claims pertaining to teacher quality – as well as student progress toward discernible, if general and incomplete, adult outcomes – invest undue faith in the predictive and reflective powers of ontic measurement in the domain of education generally. Yet it remains the case that Duncan's turn to this sort of measurement and the data it produces to justify such claims about student outcomes and

teacher quality emerge in response to specific needs. In particular, as I proposed briefly in Chapter 1, the interventions of such data respond to one or more of three felt inadequacies surrounding questions around teacher quality: an anxiety that no one *knows* which teachers demonstrate real quality and which are abject failures; a fear that no one has sufficient grounds to *report* which teachers perform at a high level and which do not; or simply a suspicion that no one *is telling* the good teachers from the poor ones.

Thus far, I have sought to highlight that the use of student achievement data in establishing *knowledge* of good teaching in general cannot meaningfully improve upon existing, even informal, means immediately available to students, colleagues, administrators, and even government officials. The indeterminacy of the eventual world for which one prepares students restricts the validity claims of achievement data on the one hand, while on the other, the restrictions of the domain addressed by an ordinary sense of good teaching to the merely technical realm of improving test scores in highly-specific, artificially-formalized academic skill domains diminishes the strength of the student achievement data's grip on the construct it seeks to measure. In these ways, the use of student achievement data cannot adequately establish the sort of *knowledge* of good teaching necessary to address a kind of epistemological anxiety in the superior manner that policy-makers imagine that they require.

Similarly, with respect to the ability to *justify* claims about teacher quality (and student learning), achievement data faces the same impasses enumerated above and thus cannot secure the basis presumed necessary to an *ability to report with accuracy* on teacher quality. Where the criteria to which Duncan and Olly Neal appeal in perceiving educational quality either immediately or with authoritative hindsight seem unable to latch onto either the type of relevance to the present day or a claim to legitimate generalizability, claims grounded

in student achievement data fare no better in addressing these inadequacies, although they

fail, interestingly, in different ways. Temporality poses problems for achievement data not

with respect to (but neither wholly distinctly from) the present's difference from the past, but

rather with respect to the indeterminacy of the future. That same future indeterminacy and

the fact of conceptual and criterial change over time prevents couching teacher quality in

terms of adult outcomes: we have too little idea *what specifically* will come together in

comprising a "successful adult" and therefore what skills will be necessary to attain that

which makes up this concept. Even allowing that reading, writing, and math will *almost

certainly* factor into adult success proves inadequate to the purpose: that which these basic

skills require is also susceptible to change along with the worlds in which they will have their

purchase.[43]

Despite, then, the fact that student achievement data claims to access the "facts" that

will enable "honesty" in reporting – or others offer claims on its behalf – it remains unclear

that what data-based reporting can offer meaningfully improves upon either of the other

forms of reporting that Duncan and Neal acknowledge in their narratives and descriptions.

Meanwhile, given the incapability of data's generality to secure a grasp of either the

---

[43] It will *certainly* be worth noting that the basic claim that I've made here about the relation of the present to the future and the inability to predict the future is not at *all* new with respect to education, although these ideas have primarily been bandied about with respect to curriculum rather than measurement or evaluation (as though those things ought to be separate). Bobbitt's efforts at formulating curriculum based on empirical job analysis – by looking exhaustively to catalogue the specific skills and attitudes necessary to success in the workplace *at present* – turned out (unsurprisingly) not to provide adequate grounds for educating in the direction of the future (P. S. Hlebowitsh, 2005). A similar argument is lodged against Shulman's teacher-knowledge project as well, whose title ("Existing Practice Is Not the Template") is simultaneously right (in that existing practice cannot serve adequately to speak to the future) and wrong (in that existing practice provides the *best* examples on which to build). More detailed study of that paper may be warranted at a later point (Evans, 2007).

conceptual or the futural aspects of good teaching, the turn to generality also loses its hold on the particular and the specific altogether. But beyond *knowing* good teaching and *justifying* claims pertaining to it, there remains the third possibility: that no one *is telling* good teaching from bad.

Here Duncan's call to "courage" reemerges for exploration. The third possibility on offer here, that no one *is telling* the good teachers from the bad teachers – even with the non-technical means available – seems at first glance the least likely of the candidates to account for the anxiety that data addresses. But Duncan's words on the matter, which I reproduce below, suggest that the urgency of Race to the Top may require *simply* getting serious about the need to *report* on teacher quality, irrespective of the particular manifestations, in epistemological or justificatory terms, of the existing lack of seriousness. The unspoken and therefore unsupported suggestion of *equivalence* between using student achievement data as the justificatory and epistemological foundation for claims about teacher quality and something like a general "getting serious" about evaluation requires that we explore the notions of "getting serious" and "courage" at work here. Says Duncan,

> There's so much opportunity for growth and progress in this area. We have the money and we have the technology. The biggest barrier, the only remaining barrier in my mind is whether we have the courage. It takes courage to expose our weaknesses with a truly transparent data system. It takes courage to admit our flaws and take steps to address them (Duncan, 2009b).

While Duncan associates the opportunity at hand with the availability of (new) money and (new) technology, which will enable a "truly transparent data system," the point of emphasis in these closing remarks pertain to the courage "to admit our flaws and take steps to address them." I wondered aloud earlier whether the turn to the use of student achievement data indicated courage or cowardice, but in either case, it might be taken to represent a

response to a type of situation that calls for *telling*, a situation in which one can either demonstrate the courage to do so or the cowardice to decline. Whatever else it may represent, the call for the creation of a "transparent data system" certainly amounts to a resolution to tell *something*, to provide a vocal response to the sort of evaluative vacuum signified by Duncan's repeated citation of the statistic that 99% of all teachers are rated the same. The bald-faced absurdity of *that* fact requires a response. *That* fact provides the warrant in light of which data-based reporting is offered in an ameliorative capacity.

It is, however, certainly worthwhile to note that, while the 99% statistic seems to offend all parties involved, it remains far from clear that *telling* teacher quality by recourse to student achievement data will speak with *equal adequacy* to the many parties whose offense provides the warrant in the first place, given the failures of achievement data to establish certainty with regard to teacher quality *any more solidly* (on any more solid basis) than the means, for example, of the-moment-I-walk-in perceptive observation.

I noted earlier that, where Wittgenstein seems to assert that pointing to examples provides a way of communicating *no less* than the pointer knows him- or herself, Duncan's recourse to data implies that there is nonetheless something that using examples cannot do, which might be as much as to say that there will invariably remain people untouched by the particular example in question – even Kristof's *depiction* of Mrs. Grady's quality cannot take the place of a student's or colleague's or a parent's everyday *familiarity* with Mrs. Grady's whole practice. As far as it goes, such an objection to the (universal) applicability of particular examples on these grounds presents no problems. Pointing to an example of justice or beauty or good teaching requires something like personal contact with the

paradigm, which is in principle non-universal and thus, in an important sense, incommunicable to some subset of people.

But Duncan's discussions of achievement data also seem to admit of something that the *data* cannot do, or at any rate cannot do alone; this awareness underlies the calls for "multiple measures" of student achievement and teacher quality. We have already seen the ways in which the multiplicity of measures that states' evaluation protocols adopt reifies instead the sovereign place of achievement data as "the preponderant criterion" upon which all other measures are grounded. This fact may suggest that something is amiss with Duncan's notion of multiple measures, namely, that he takes the relevant axis of multiplicity to involve the multiple *expressions* of a *singular* measure.

Confronting us, and confronting Duncan, then, is a sort of inadequacy in the idea of multiplicity at work in or implied by his proposals. Duncan's call for multiple measures responds to an obvious fact: that, in his own words, it would be foolish to "reduce the complex, nuanced work of teaching to a simple multiple choice exam." The declaration that teaching is "complex, nuanced work" – what such a statement might mean – remains under-interrogated throughout the discourse on teacher quality in general. While the notion of measuring teacher quality according to multiple perspectives still seems to permit a view in which teacher quality (and teaching itself) is fundamentally singular – multiplicity of measures aside – such a view is more difficult to maintain when one considers the circumstances surrounding the *telling* aspect of teaching and learning, the reporting on educational quality.

When Duncan speaks of what data can accomplish in terms of revealing facts about teaching, his assumptions about the *nature* of teaching (its availability or hiddenness, the

salient or relevant features of its practice, its purposes and effects) may be meaningfully

different from those of, for example, a parent. If this suggestion is accurate, teaching *itself* is,

in a meaningful (that is, non-platitudinous) sense, multiple. The observable nexus of teacher

and student behaviors and responses taken to indicate this or that about the quality of both

learning and teaching will, if evincing a certain degree of overlap, be non-identical according

to the perspectives of a state- or national-level official and the parent of a particular child. It

will then be worthwhile, I suggest, to consider that the relative utility of student achievement

data depends *profoundly* on *who* asks for the telling of teacher quality and under *what*

circumstances. This suggestion further acknowledges the possibility that data *is*, for all of

the limitations highlighted throughout this work, *among* the most appropriate means of

evaluating teachers in *some* cases. This suggestion also, however, simultaneously leaves open

the possibility that in some cases and for some purposes student achievement data cannot

meaningfully address the needs of those demanding the report and thus will represent a

woefully inadequate and inappropriate response. Cavell cites the projects of Austin and

Wittgenstein in gesturing to this dimension of things:

> My stress on the time, or time and place, of depiction is meant to capture what
> Austin means in tirelessly demanding the context (he would often call this the story)
> of an utterance and what Wittgenstein means by repeatedly asking to whom an
> utterance is made. When Wittgenstein asks, "How is telling done?" he is in effect
> asking how it is that saying something, speaking, is done; how it is that someone is in
> a position to be told something. This turns out to be a good question (Cavell, 2010, p.
> 60).

3.2.2 On Multiple Measures and the Singularity of Teaching

If "this turns out to be a good question," it is a question persistently avoided in the

reformers' thinking about measuring and reporting on teacher quality. More accurately, if

the question is sometimes posed, it is simultaneously and in nearly every case begged.

Duncan's view of the charter school debate proves enlightening in terms of the assumptions under which he operates with respect to teaching and data, assumptions that belie any robustness of the view of multiple measures to which he sometimes alludes. At bottom, despite the rhetoric of multiple measures, Duncan's view of teaching's ontology is avowedly singular:

> I recently spoke to education writers about the search for truth in education. I challenged them to go beyond the ideological statements and the surface conclusions and find out what is really happening for our children in our classrooms.
>
> It's kind of like the debate around charter schools. Advocates say they outperform traditional schools. Opponents say they don't. The plain facts show that some charter schools do, and some of them don't. But rather than acknowledge the obvious, we devolve into an ideology debate and somehow forget that this is about children and learning. If something helps children, let's do it.
>
> That's where all of you come in with the research and the facts. Education reform is not about sweeping mandates or grand gestures. It's about systematically examining and learning and building on what we're doing right and scrapping what hasn't worked for our children (Duncan, 2009b)

The above is a reiteration of a basic outlook that we have seen Duncan offer before: he imagines that "systematic examination" through the use of achievement data will yield the truth about "what is really happening for our children in our classrooms" lying beneath "surface conclusions" – it yields what he calls the "facts" about what "helps children." This view, in which data generates something uniquely *true*, leads Duncan to treat disagreement as a matter of "ideology," in which, judging from the tone and context of the statements above, one's political leanings or personal feelings with regard to charter schools *prevent* them from "acknowledg[ing] the obvious": "some charter schools do [outperform traditional schools], and some of them don't." *If something helps children*, he says, *let's do it*.

Duncan's pejorative labeling of disagreement as "ideological" in nature recalls a comment that Paul Standish levies against the uncritical use of "achievement data" by reformers of Duncan's ilk:

> Ideas of effectiveness or of improvement or of "what works" are predicated on assumptions the contestation of which is ruled out of court on the grounds either that these are so obvious as to not be worth discussing by people of good will or that they are matters of personal point of view (Standish, 2001, p. 571).

Duncan explicitly frames his complaint in terms of the "so obvious" that Standish mentions, which necessitates a view of such "contestation" in which those contesting the claim must either lack "good will" or express *merely* a "personal point of view." It will also go without saying (but seems nonetheless to call for saying) that Duncan's contrast of "surface conclusions" with "facts" reifies the rationalist assumption that something permanent and entity-like stands behind superficial appearances. Stopping at a surface conclusion is therefore not quite the same as being *wrong;* it simply evidences that one does not *have* the actual facts, that one has not reached *bottom*. Thus, for Duncan, the argument over the merit of charter schools is an argument between parties whose *question-asking behavior* stands in need of transformation. That it is possible to ask heretofore unasked questions is what allows Duncan to call it an example of "ideology" instead of simply being *wrong*. Neither party to an ideological debate can be *wrong*, properly, because neither party has put itself in the position, through probing the appropriate depths, to be *right*. But this view of the matter is badly confused.

Duncan cites "facts" twice in the excerpt quoted above. The first instance refers to the "plain facts" that demonstrate that some charter schools seem to be achieving remarkable success while others are not. These are also the plain facts, "the obvious," that he takes

ideologically-motivated antagonists to ignore in their wars of words. The second instance, though, while also referring to facts, seems to refer to facts of an entirely different sort. As an answer to the intractability of the "ideological" debate originating from a disagreement over a denial of "plain facts," he says to his data-analyst audience, "That's where you come in with the research and *the facts*" (emphasis added), which apparently stand in some sort of contrast to the "plain facts" that reveal that some charter schools outperform certain traditional schools and some do not.

In order to *settle the matter* by means of employing "the research and the facts," Duncan turns to his audience and the student data innovations they represent. But the *plain* facts have already demonstrated that "some [charter schools] do [promise real improvements], and some don't." It is (a) wholly unclear that the debate Duncan references is ideological in a pejorative sense (i.e. about feelings *as opposed to* facts) since, given the evidence on both sides, the "plain facts" seem unable to settle it at all, and also (b) equally unclear that drawing on *deeper facts* will prove any more capable of settling the matter. Though it is not clear exactly *what thing* Duncan has in mind when he suggests that "if *something* helps children, let's do *it*," what remains patently obvious is that the *something* is singular and that the children *it* helps are identical to each other with respect to *its* helping.

The *plain* facts that Duncan references above show that some charter schools live up to their revolutionary promise while others fail to outperform traditional schools. On the basis of contradictory evidence, an "ideological" debate begins. In settling the debate over the value of charter schools, Duncan turns to "the research and the facts." He must imagine that "systematic examination" of the matter will help in one of the two following ways: (1) it will reveal that the studies employed by one side or the other are inaccurate, or else (2) it will

identify some lower-level characteristic of "successful" charter schools absent from "unsuccessful" charter schools. Outcome (1) seems highly unlikely, since the methodology associated with the deeper-level version of "research and the facts" is of the same *kind* as those currently being employed in the "ideological" debate. In other words, because the sort of research that would generate the deeper "facts" is based on the statistical parsing of student achievement data, and because those facts already contradict each other, what Duncan calls "the research and the facts" cannot mediate between them. Outcome (2) thus seems more likely, though two problems emerge, if this is indeed the case.

In the first place, outcome (2) accomplishes nothing in terms of *settling the charter school debate*: the more fundamental characteristic it would identify, after all, would be displayed by the successful charter schools but not by the unsuccessful, which means that it is not a feature *of charter schools as such*. The debate, then, would cease to hinge on *charter schools*, since charter-school status is no longer the relevant factor in the "working" of "what works." But if settling the charter school debate is accomplished by means of changing the subject, as it were, it is unclear how "the research and the facts" of the *deep* variety that Duncan refers to in saying to his audience, "That's where you come in" are supposed to *improve upon* the existing situation. The *plain facts*, after all, *already show* that some charter schools are successful and others are not, which clearly implies that charter school status cannot be the crucial factor at work. This conclusion, in fact, is what Duncan calls something "obvious" that people need to "acknowledge." If advanced data – more fundamental facts about what works – do nothing *more* to settle the debate than what is already obvious, it cannot be the case that settling *just this* debate generates the warrant for data's use, which brings us to the second problem with outcome (2).

The second possible consequence of outcome (2), since settling the charter school debate is out of the question, involves turning the focus of the discussion to a deeper-level characteristic of the successful charter schools, a characteristic absent from the unsuccessful ones (and from all traditional public schools as well?). The problematic assumption upon which this outcome turns is that if it is not *this* characteristic that will *help children*, then it must be *some other* singular characteristic or amalgam of characteristics, which constructs a situation in which *all* children are formalized as a unity on one side of an equation while possibly relevant characteristics of schools are formalized on the other, directly in line with the dubious, dubious rationalist assumption taken apart in Chapter 2. In addition, the foregoing discussion of the path by which attempts to identify best practices and take them to scale vanishes into an infinite regress suggests that the *data* that Duncan calls upon to answer the question of *what helps children* such that the answer would be *generally* applicable will prove unable to sufficiently address the problem as he conceives it.

Duncan sees something standing in *contrast* to "surface conclusions," which he calls "facts" amenable to proper "research." But the kind of research he has in mind is already implicated in the superficiality of the surface conclusion: such research will never be able to reach the bottom that he imagines. In fact, it is not too much to suggest that, to Duncan and those working on his assumptions, no one has *ever yet* reached bottom; thus the quest. It is perhaps further not too much to suggest, therefore, that Duncan has no idea what sort of "fact" it *is* that he wishes *for*, as no facts heretofore available have ever proven capable of the feats he imagines them to accomplish: allowing him to lay out the *actual* and specific student skills and teacher behaviors that will lead to successful adulthood. The unlikeliness of arriving at some singular ground upon which such general broadcasting of specific skills and

behaviors has been well-established by this point. But the imperturbable assumption that there *is* such a bottom is precisely the begging of the question that short-circuits or precludes robust considerations of the meaning of "multiple measures" in the evaluation of educational quality.

Duncan's conception of the problem, then, represents *one* of the warrants for reporting on teacher quality and student progress. The obviousness of "plain facts" cannot serve to adequately address a question of what *really and in all cases* helps children, and so Duncan wishes to bring more sophisticated analysis of student achievement data to bear, which will access the *less plain* (but no less factual) facts that *will* be able to tackle that question. The suggestion I make above is that whatever the analysis of student achievement data yields will share the same ontological level with what Duncan calls the "surface conclusions," rather than providing the type of external or deeper or higher perspective from which clarity on the matter would be possible. Wittgenstein, it might be noted, perceives something similar in attempts to explain what Duncan calls "the obvious" to ourselves: "The ground keeps on giving us the illusory image of greater depth, and when we seek to reach this, we keep on finding ourselves on the old level" (Wittgenstein, 1983).

Even if Duncan attempts to label them some different sort of fact, it is also unclear how even these *mysterious* facts would address his question. The assumption that founds Duncan's quest is that there *is something* entity-like that *helps children*, where such helping is both *general to all children* and *revealed in student achievement data*. A questioning of this assumption, to Duncan, must feel like the effects of "ideology," in that it ceases exploration at the level of "surface conclusions." But the possibility remains that the threat of the regress brought about by the reification of teaching and learning as formalizable

objects, *immune to* and *separable from* shared forms of life, indicates that any particular

characteristic correlated with student growth would *not generalize* to students as a whole and

would also *not be generalizable* to teaching practice or schooling itself. Another way of

phrasing this possibility is that there *is no single thing* that helps *all* children equally.

The failures of "surface facts" to settle debates such as the one over the value of

charter schools provides one type of warrant for the research project Duncan lays out to the

analysts, the quest for deeper facts. The second type of warrant is best expressed by a parent.

While Duncan's suggestion of bringing "the research and the facts" to bear on the matter

implies both that the 99% statistic is the *result of* inadequate knowledge and also (but this,

notice, is something different) that *adequate knowledge* somehow *entails* reportage.[44]

Duncan's view thus turns on a claim that adequate knowledge has been heretofore

unavailable. Carol Penskar, a parent responding to an article on teacher accountability

systems in a *New York Times* letter to the editor, begs to differ:

> Teachers' unions, according to your article, argue that "no one has figured out
> an accurate way to determine" who the best teachers are.
> This is ridiculous. My son and daughter each attended 13 years of public
> school. The parents always knew who the best teachers were. If the parents can figure
> it out, certainly the education bureaucracy could do the same if it really wanted to?
> (Penskar, 2010).

Where Duncan seems to imagine having lacked adequate knowledge for the *telling* of

teacher quality, Penskar profoundly disagrees. Parents (plural), Penskar claims, "always

---

44 It will, perhaps, strike one that in major school districts like Los Angeles and New York City, public "reporting" on teacher quality according to Duncan's strictures *has consisted in* the publication in major newspapers of teacher effectiveness ratings in terms of value-added. What counts as *knowing* teacher quality is also the *form* of its telling. This is in contrast to, for example, reporting on the quality of a movie. The form of the report of a movie's *quality* is not an academic or statistical *analysis*; it's a *review*.

knew who the best teachers were." The warrant that calls for the revelation of *what really works* now finds expression by way of the *utter obviousness* of teacher quality even to nontechnical observers such as parents. Penskar's claim is an appeal to the common-sense, immediate visibility of teacher quality to those who are familiar with the particular teachers in question. In this way, Penskar is scandalized by roughly the same phenomenon – and in the same way – as Duncan: the plain fact that "the parents always knew who the best teachers were," which mirrors Duncan's claim to know a school's quality "the moment I enter a building," has not found expression in a *general* algorithm or method.

Penskar's notion that "certainly the education bureaucracy could" figure out how to measure teacher quality seems to follow from the premise that "the parents can figure it out"; the assumption is that the parental understanding of teacher quality must be, on some general level, the same as the kind available to policy-makers. But this is precisely the problem that gave rise to the breakdown of a best-practices view of teacher development in Chapter 2. Penskar's assumption thus takes the same form as one that Dreyfus addresses in his phenomenology of everyday expertise: that common-sense understanding and rationalist-assumption theoretical knowledge share the same structure. Says Dreyfus, almost, in this light, directly *to* Penskar and Duncan:

> What hides the impasse is the conviction that the common-sense knowledge problem must be solvable because human beings have obviously solved it. But human beings may not normally use common sense *knowledge* at all. What common sense *understanding* amounts to might well be *everyday know-how*. (H. L. Dreyfus & Dreyfus, 1986, p. 99).

What Penskar and Duncan both express in their different ways amounts to an insistence that the means of knowing and reporting on teacher quality available to a *parent* have the same basis and ought to be equally available to a state- or national-level *politician*

and, importantly, *vice versa.* The assumption is that the knowledge and justificatory grounds had by *each* is had by way of theoretical representation, and thus it ought to be solvable in a *general* capacity, as a *general* problem. Dreyfus's reemergence reminds us of the distinction between the nonformalizable background on the one hand – on the basis of which, Dreyfus would say, parents have the commonsense understanding in virtue of which to evaluate their teachers – and the rationalist requirement on the other, that both ends and means, teachers and student learning, be independently formalizable, on the basis of which formalizations, "student learning" and "teacher quality" would be both explicable and knowable *generally*.

Dreyfus's contention is that human beings have not, in fact, *solved* the common-sense knowledge problem precisely because human beings do not *have* a common-sense knowledge problem. The *problem* involved in formalizing things like "successful adults" and the criteria of "writing" only crops up when one seeks a way of talking about something like educational quality from the perspective of a radical outsider, far away from the lived experience thereof.

Arne Duncan wishes to know the relative quality of 3.7 million teachers at work in schools around the country. Carol Penskar has knowledge of her children's teachers that she is in a position to communicate. Why ought Duncan to take Penskar's word, as opposed to measures of teacher value-added derived from student achievement data? Penskar's experience with her children's teachers is, especially through elementary, and to a lesser extent, middle school, ongoing, personal, and intimate. Penskar finds herself in a position to know and comment upon the aspects of teaching that simply go beyond reading and math improvement; Penskar's position allows her access to the ordinary sense of good teaching, but her view is restricted to only a few of the 3.7 million teachers to whom Duncan must

attend. On the other hand, why ought Penskar to take Duncan's measures as legitimate?

Duncan's student achievement data has a claim to a much wider experience of teaching and

learning, next to which Penskar's sample size, so to speak, is woefully inadequate. Duncan's

data speaks to a sort of teacher quality that transcends jurisdictional boundaries, that has a

claim on *absolute* normativity, one might say – Duncan can communicate to Penskar how her

child's teacher stacks up against *all* other third grade teachers, not simply the paltry few that

Penskar has known or heard about.

The obviousness of the quality of her own children's teachers seems to Penskar to

indicate that the type of particular and ongoing experience on the basis of which such

knowledge *is* obvious ought to make it sufficiently obvious for everyone (and certainly most

parents would agree with regard to their *own children's particular* teachers). But one has

then merely a constellation of particular instances without any claim on generality at all. The

obviousness of the quality of teachers as it appears according to student achievement data, on

the other hand, has large numbers and correlations with adult outcomes behind it. Teacher

value-added has the power of the general on its side. But studies have pointed out that in

attempting to speak to particular instances, to the cases of particular teachers, such methods

do not conform to any sort of ordinary-sense perception of teacher quality, which is what I

take the findings of Strong, et al. to demonstrate: "In every case, judges achieved relatively

high levels of agreement but were absolutely inaccurate, leading us to question whether

educators can identify effective teachers when they see them" (Strong, et al., 2011). The

"judges" in this example agreed with each other but disagreed with the measures, leading

Strong and his colleagues to conclude that the judges (not the measures) were "inaccurate."

The stalemate above, in which Strong – who, recall, is "skeptical" of teacher value-added – has to choose which of the two bases of judgment he will accept as the standard of accuracy to which the other will be compared (or contrasted), highlights the fact that this appears, to recall Dreyfus's phenomenology of skill development once more, something like an ungrounded choice of perspective. The standard of accuracy itself is in question. Strong's conclusion could just as well have been that measures of teacher value-added were self-consistent but completely inaccurate, as they diverged from the judges' agreement with respect to the quality of the teachers they evaluated. We are stuck, it seems, in a position where neither value judgments about teacher quality generated from a highly general perspective nor those generated from a highly particular perspective have adequate claim to universal acceptance, something that will suffice to both the purposes of a Secretary of Education who will never meet most of the teachers whose quality he needs to know and also to the purposes of a given third-grader's parent who is intensely concerned with the quality of one particular teacher.

3.2.3 Summary and Segue

While the purpose of drawing out, over the space of hundreds of pages, the extent to which the use of student achievement data cannot live up its own claims to represent a universally superior form of knowledge or adequate justification for value judgments in all situations has had structurally to do with anticipating the concluding section of this piece, the deconstruction of the assumptions on which student achievement data rest also includes a positive ontological statement about the nature of education that I wish to make explicit, one that speaks to the need for courage that Duncan cites.

One sometimes hears people utter the truism that "teaching is not rocket science," which is ordinarily offered as an implication that teaching is in some senses simpler than rocket science. But my analysis shows that the *ordinariness* of teaching, the fact that it is pervasive, public, bound up in being human, in fact renders it "complex, nuanced work," as Duncan calls it. In light of the fact of its ordinariness, one might say, it is *different* from something like rocket science. That each may be considered "complex" or "nuanced" ought not to imply that such complexity is of an identical nature, amenable to identical forms of representation.

Rocket science, after all, requires highly specialized knowledge in highly specific fields, such as calculus, engineering, and advanced physics. It is a highly restricted field, as well, employing exponentially fewer people than the teaching profession (there are more teachers in the United States than waiters, a startling fact). Rocket science is something for the very few; it is something very few know anything about; its practices, its norms, and its projects are remote from the lives and concerns of most people.

Meanwhile, to paraphrase John Dewey, we are born to be teachers in a way that we are not born to be rocket scientists. Teaching and learning pervades all of our lives, in ways both formal and informal: our pasts as children, our interactions with our peers, our experiences as mentors and mentees. What teachers do is, in a certain sense, something that we all do. The world in which education exists could not be more common to something like the experience of being a person.

I take Penskar's objection that it is "ridiculous" to imagine that we do not know what good teaching is to speak to the pervasiveness of teaching in our lives. I take Duncan's frustration with the 99% statistic to speak to something similar.

Rocket science, one might note, for all of its mythical difficulty, for all of its specialized knowledge, for the rarified space it occupies in the popular imagination, comes down to puzzle-solving. It is about (immensely challenging) calculations of forces on bodies in motion, about taking into account chemical properties of materials, about highly sensitive measurements of space, time, mass, and energy. The above is not, of course, to minimize the amount of knowledge required to fly humans to the moon, nor the difficulty of the calculations and measurements involved. But it is very much to highlight the fact that rocket science is a realm of structured problems and ontic measurement.

Duncan's policy approach treats teaching as though it were rocket science. He thinks that he thereby does the profession a favor, in contrast, perhaps, to Penskar's view in which the common-sense accessibility of teacher quality speaks to something just *anyone* could do: "Our larger goal," Duncan says in introducing a new program designed to increase the quality and appeal of the teaching profession as measured by "student outcomes," "is to make teaching not only America's most important profession but America's most respected profession" (Duncan, 2012b). The fact that standardized tests are not yet perfect, which he acknowledges, he treats as a technical problem on which experts are working, and for which they will someday find a (singular) solution. That scaling-up educational success continues to move sluggishly calls (simply) for teachers to "collaborate" more, to do a better job of "sharing best practices."

But teaching is different from rocket science in important ways. Education, as this project repeatedly demonstrates, resists the ontic measurement associated with structured problem areas. It is instead an unstructured domain, responsive not only to a politically-contested past, but also to a radically indeterminate future. What teaching and learning

should take up and consist in are matters of endless debate and negotiation, debate official and unofficial, local and national, political and personal. Teaching and learning come down on no firmer ground than shared understandings of being, agreements in forms of life. Whatever standards one establishes, and whatever criteria one associates with them, both of these are always susceptible to change: the domain of teaching and learning is therefore always also a matter of ontological measurement in precisely the way that rocket science is not. Rocket science's exclusivity, its prohibitive entrance requirements (which educational reformers also yearn to mimic), are associated with its mythical status. In light of these facts, as Duncan likes to say, "let's be honest": Teaching is not rocket science.

But perhaps it is just this fact that calls for *courage*.

Duncan proposes the use of student achievement data as a means of getting beneath the "surface conclusions" of "ideological debates." That, he says to the group of analysts, "is where you come in with the research and the facts." But this is a poor construction of the problem, as it seems to deny that we have the requisite knowledge of "the plain facts," a suggestion that Penskar calls, rightly, obviously, "ridiculous."   The paradoxical position here is that we seem to have *enough* knowledge and yet seem unable to *do* something important, something that we usually take adequate knowledge to allow us to do, namely, to *say* (broadly, generally, publicly) with some measure of certainty whether this or that teacher is good at what he or she does. That the "plain facts" have not sufficed to preclude disagreement about these judgments is taken to necessitate more or better (or deeper) facts, as evidence that we have yet to reach the bottom. But Wittgenstein sees such matters differently.

Wittgenstein frames the misconception this way: "It is, rather, essential to our investigation that we do not seek to learn anything *new* by it. We want to *understand* something that is already in plain view. For *this* is what we seem in some sense not to understand" (Wittgenstein, 2009, p. 48 (sec. 89)). Wittgenstein's position agrees with Penskar that we *do* have the facts we need; it also agrees with Duncan in that the facts that we have seem unable to *do* something essential that we expect them to do. *Understanding* something already in plain view, contra both Penskar and Duncan, might lead to a *dis*solving of the problem rather than any *solution* to it. Understanding teaching and learning as a matter, ultimately, of ontological measurement and as an unstructured domain leads to a view in which the *solving* of a problem by means of *providing* an answer (what good teaching *is*, what student achievement *looks like* in every case) becomes somewhat beside the point, an unattainable aim rather than a necessary precondition.

Penskar asserts that the "education bureaucracy" could figure out who good teachers are "if they really wanted to." Duncan says that real teacher accountability requires "courage." It is not obvious, however, if it really is the case that education either certainly *is* like rocket science (as Duncan seems to think) or certainly *is* simply a matter of everyday obviousness (as Penskar suggests) why any particular *gumption* or *heart* would be required. Joseph Schwab's thinking about teaching and learning helps not only to see why courage might be necessary, but it also offers a means of dissolving the intractable problems that seem to accrue to attempts at knowing and evaluating teaching. In short, Schwab enumerates the ways in which the process of teaching and learning, while certainly more ordinary than high-level astrophysics, nevertheless requires fundamentally different kinds of knowing and doing, and therefore also different preparations for and forms of telling. Seeing teaching and

learning as "complex, nuanced work," and yet very differently complex from the paradigm case of complexity that rocket science embodies, is what calls for courage.

3.3: Joseph Schwab's Deliberation and Teacher Accountability

It may seem counterintuitive to offer Joseph Schwab's ideas as an ameliorative strain of thought to the present discursive field consisting of parents and policy-makers, educators and economists, think thanks and theoreticians. Schwab, in the first place, contributed his work in the field of curriculum, which, while it certainly *ought* to include concerns around evaluation and has, dating at least to Ralph Tyler, a long history of doing so, has become ever more limited in that regard since Schwab's writing. Additionally, the curriculum field does not so clearly overlap with, for example, the exploration or practice of teacher development, and has little to say in the field of psychometrics or educational measurement, except insofar as these are all inherently bound up together. Further, Schwab's writings are more than a generation old by this point: his seminal essays – the Practical Cycle, as they are known – were published between 1969 and 1983. His former graduate students and intellectual heirs include giants in the field of education such as Lee Shulman and Elliot Eisner, which is as much as to say that his intellectual line expresses contradictory impulses with regard to the discussion at hand.[45]

---

[45] There is a time and place, I think, for exploring the ways in which Shulman and Eisner can both an claim inheritance from Schwab; perhaps they do not perceive themselves to be pursuing opposing intellectual projects, as I perceive them to be. I considered discussing Eisner at length. I suspect I will have cause to mention him, but I will put off serious engagement with him until another time. I fear, frankly, that while my engagement with Shulman was necessary in order to bring out salient problems with the assumptions on which his project relies, I would treat Eisner too much with an eye to my own ends, picking and choosing the places in which he sounds closest to Schwab, smoothing over his many departures and innovations. I have no desire to do that.

Nevertheless, Schwab seems to speak directly to the impasse that I have traced throughout this work. That might express the degree to which the focus of the anxieties and uncertainties pervading educational discourse have moved from concern over matters of curriculum conceived as *content mastery*, wherein arguments over *what* to teach predominate, to matters of *evaluation* founded upon *proficiency,* in which questions over *what counts* as adequacy when it comes to skill-development, and how one might arrive at a position to know or judge such a thing, hold sway. If one recognizes the salience of Schwab's thinking at this point, as offering a means of dissolving the problem confronted by the fact of competing claims of differing standpoints for telling and hearing (as one might awkwardly call it), that fact too will speak to the notion that the *problem* Schwab identifies and addresses in 1969 has simply been transposed from the aegis of curriculum into the realm of teacher development and evaluation. The problem he identifies, I argue, remains the problem that Duncan, Penskar, and an army of economists attempt to *solve*.

Schwab's thinking demands *courage* inasmuch as it requires us to adopt a view of ourselves in which we are neither individually "free to make it up as we go" nor collectively (or individually) privy to some underlying infrastructure of all knowledge. Both of these alternatives, it is not too much to say, amount to hubris: the first a dream of omnipotence; the second of omniscience. Letting go of these dreams does not entail turning away from potential glory or anything of the sort; it does, however, require turning to one another as legitimate sources of understanding. If, ultimately, ontological measurement accesses something trans-human; if our concepts come down to agreement in forms of life; if the quality of skillful practice stands out on a background of everyday know-how; then one will do best to seek such lofty things as knowing, doing, and evaluating – at least in the realm of

education, in the realm of the human – in conversation with other people, in direct contact, one might say, with the trans-human.

3.3.1 The Theoretic and the Practical

In order to see the ways in which Schwab speaks directly to the matters of concern to this project, out of which his proposed (but, to my knowledge, only rarely enacted) practice of *deliberation* emerges, an introduction to Schwab's own terms and concerns will be necessary.

In 1968, Schwab opened the keynote address at the annual meeting of the American Educational Research Association by declaring the curriculum field "moribund." On Schwab's view, the field's current morbidity, the story of its arrival in this place, and the potential way out all hinged upon the distinction between his conceptions of the "theoretical" and the "practical." At the opening of this address, he lays out his terms:

> The curriculum field has reached this unhappy state by inveterate and unexamined reliance on theory in an area where theory is partly inappropriate in the first place and where the theories extant, even where appropriate, are inadequate to the tasks which the curriculum field sets them (Schwab, 1969, p. 1).

Schwab goes on to highlight several specific ways in which looking to theory to solve problems in education field fails the curriculum field. In discussing, for example, the shortcomings of a certain revival of Herbartian theories proposing both a theory of mind and a theory of knowledge that merge into a seamless whole, Schwab comments:

> A theory of mind and knowledge thus solves by one mighty coup the problem of what to teach, when, and how; and what is fatally theoretic here is not the presence of a theory of mind and a theory of knowledge, though their presence is part of the story, but the dispatch, the sweeping appearance of success, the vast simplicity which grounds this purported solution to the problem of curriculum (Schwab, 1969, p. 7).

He offers the above example in contradistinction to several other efforts at deriving educative methods from theoretical grounds, most efforts centered around creating objectives

according to, for instance, that which is taken as necessary to "live in the modern world" (Schwab, 1969, p. 7). That one, in particular, sounds familiar.

The insufficiencies on display for Schwab have primarily to do, as he says above, with the "dispatch, the sweeping appearance of success" of *each* theoretical viewpoint. As totalizing views, each constitutes or composes a self-enclosed universe, and thus proves incommensurate to any of the concerns of any other theoretical view. Schwab points out the incompleteness of each, and then proposes, more or less directly in line with the thinkers referenced in Chapter 2, that the dividing or atomizing of something that used to be called a life-stream – lived practical experience in time – will lead to a sort of understanding that is *necessarily* incomplete. The purpose of quoting Schwab at length here is to draw out the similarities with Dreyfus, in particular, though one also sees the reflection of Glendinning's notion of "phenomenon-splitting" at work. Unlike the others, however, Schwab speaks *directly* to the milieu of educational research.

> It is clear, I submit, that a defensible curriculum or plan of curriculum must be one which somehow takes account of all these subsubjects which pertain to man. It cannot take only one and ignore the others; it cannot even take account of many of them and ignore one. Not only is each of them a constituent and a condition for decent human existence but each interpenetrates the others. That is, the character of human personalities is a determiner of human society and the behavior of human groups. Conversely, the conditions of group behavior and the character of societies determine in some large part the personalities which their members develop, the way their minds work, and what they can learn and use by way of knowledge and competence. These various "things" (individuals, societies, cultures, patterns of enquiry, "structures" of knowledge or of enquiries, apperceptive masses, problem solving), though discriminable as separate subjects of differing modes of enquiry, are nevertheless parts or affectors of one another, or coactors. (Their very separation for purposes of enquiry is what marks the outcomes of such enquiries as "theoretic" and consequently incomplete.) In practice, they constitute one complex, organic agency. Hence, a focus on only one not only ignores the others but vitiates the quality and completeness with which the selected one is viewed (Schwab, 1969, p. 9).

Schwab points in particular to the ways in which the individual and society mutually influence one another; neither can be accounted for *purely in terms of* the other. Roughly speaking, this revelation mirrors the relation that Derrida traces between linguistic events and linguistic systems. Schwab's goal, in fact, aligns closely with the one Glendinning reads in Derrida: to reveal that the traditional ways in which we *study* or *observe* the practical phenomena "in which our being is implicated," as Crease would say, remain insufficient to the *doing* of anything with regard to those phenomena. This is to say that the knowledge we take studying or observing to yield does not suffice (nor is, beyond a certain point, necessary) to the knowledge involved in *doing*. The Dreyfusian critique of AI implies that what we commonly take as *necessary* to intelligence – namely a separate realm of something like pure knowledge on which action is predicated – proves unable to *produce* intelligent behavior: something essential is left out. Derrida's project shows us that what we take as necessary to the use of language – namely knowledge of definitions and grammatical rules – cannot satisfactorily account for language use in practice: something essential is left out.

Wittgenstein's call to "*understand* something that is already in plain view" rather than attempt to "learn anything *new*" by proper investigation draws one's attention to the type of thing that remains both *essential* to understanding practical phenomena and constantly overlooked in traditional forms of accounting for such phenomena. One way of reading Wittgenstein's distinction between *understanding* something in plain view and learning something *new* has to do, as I will draw out more explicitly in a later section, with what one takes to be necessary for reporting or telling. For now it is enough to suggest that Schwab's proposal of what he will later call the "arts of the practical" represents a novel approach to

these matters in the realm of education: where Wittgenstein seems to offer a means of

releasing oneself from the clutches of flawed questions,[46] however temporarily, and Dreyfus

offers a phenomenological description of skill development with broadly painted suggestions

as to its applicability to this or that skill domain, Schwab remains focused upon the problems

facing the field of education directly.

My claim regarding Schwab's contribution thus divides into two areas of emphasis:

the first is, along Wittgensteinian and Dreyfusian lines, a means of dissolving the problem –

a means of showing that the formulations of the problems that the field in question takes up

renders the problems insoluble as such – while his second area of emphasis has to do with

sketching a means of addressing the problems that the field has heretofore formulated poorly.

In dissolving the unsuccessful or insoluble formulations of the problems that curriculum

faces – what should education *do*? How ought we to settle disagreements about what and

how to teach children? – he retains his sense of the *need* out of which these questions arise.

But his reformulation of the problem is offered in a practical light, which is merely to say

that it neither waits upon nor is answerable to the realm of the universal. Further exploration

of Schwab's own words will bring this out more clearly.

Dreyfus's critique of the structure of the problem according to early AI researchers

returns at this point to highlight Schwab's contentions about the "different subject matters"

addressed by different theories. Recalling Dreyfus's comments on the construction of micro-

worlds from the SHRDLU project, one might see the analogy between the notion of micro-

---

[46] It would be fair, I think, to claim that one cannot finally say *what* it is that Wittgenstein is offering.
The literature on *all sides* of this question are voluminous and well-reasoned. I recognize this fact,
and offer my statement both in good faith and in the full knowledge that while some would provide
vocal support for this reading, others would respond with full-throated objections.

worlds as self-contained atomizations of some sort of larger whole and what Schwab calls the

"subsubjects" of the human that "the theoretic" addresses. That which limits the utility of

micro-worlds in arriving at actual artificial intelligence bears marked similarities to Schwab's

view of the same problem with regard to the use of theory in guiding educational practice.

Given the competing standpoints or "commitments" that each theoretical view assumes,

Schwab notes that,

> There is no foreseeable hope of a unified theory in the immediate or middle future, nor of a metatheory which will tell us how to put those subsubjects together or order them in a fixed hierarchy of importance to the problems of curriculum (Schwab, 1969, p. 10).

Dreyfus more specifically locates the dead-end that the "theoretic" faces in the

rationalist assumption about the very relation of theory and practice, namely, that a

representationalist view always and by necessity leaves out the *whole* of which it is

ineluctably and mutually a part: "That is, sub-worlds are not related like isolable physical

systems to larger systems they *compose*; rather they are local elaborations of a whole which

they *presuppose*" (H. L. Dreyfus, 1992, p. 14). The inseparability of the part and the whole

is also what Charles Taylor has in mind in asserting that "this reciprocity is what the

intellectualist theory leaves out" (Taylor, 1993).

Schwab's own view of the shortcomings of the theoretic as applied to the teaching

profession recalls the thinkers of Chapter 2 in other uncanny ways, as well. Mirroring a

Dreyfusian view of skillful coping in which a facts-and-rules approach will eventually

overwhelm a given practitioner, Schwab notes of teaching:

> Moments of choice of what to do, how to do it, with whom and at what pace, arise hundreds of times a school day, and arise differently every day and with every group of students. No command or instruction can be so formulated as to control that

kind of artistic judgment and behavior, with its demand for frequent, instant choices of ways to meet an ever varying situation (Schwab, 1983, p. 245).

Eisner likewise sums up Schwab's view of the insufficiencies of the theoretic, but he expresses Schwab's objections in the specific terms of education's differences from a "cybernetic system," which once more invites comparison with Dreyfus's project. Says Eisner:

> In contrast to the "certain" regularity of a positivistic view of planning and teaching, Schwab's work reminded us of what we knew but seemed at the time to have forgotten: that planning of school programs is neither a modified version of an input-throughput-output model nor is it a cybernetic system that will eventually right itself when it goes off course or over-corrects (Eisner, 1984, p. 202).

Schwab's view of the shortcomings of the prevailing notion of education's relation to theory, a view still in evidence – indeed, even to the exclusion of all others – in the policies of Race to the Top as Duncan proposes them, thus serves to underscore the fact that the rationalist approach to education has a long-standing and unsuccessful history of its own, a history that the educational field shares with scholarly approaches to many human endeavors, as Dreyfus notes specifically with regards to AI; and as Taylor, Cahill, McDowell and others note with respect to rule-following practices in general.[47]

Where Duncan purports to offer something *new* to the educational field, then, in the form of more and better data, more rigorous assessments and standards, and more advanced statistical techniques, one might note that the earlier failures in thus imposing "theoretic" solutions to the problems of education had little to do with the quality of data or any means

---

[47] The list of shared fields is *hardly* limited to these examples. For a *much wider* perspective on the rationalist lineage in the American intellectual tradition, including especially the shortcomings of "systems analysis" as developed by RAND during the Second World War, see the following substantially diverse historical monographs: (Levy, 2012; Martini, 2012; Medvetz, 2012).

of interpreting it. Rather, as Schwab in particular notes, the absence of any "metatheory" for

combining the vast "subsubjects" embodied in the educational endeavor renders the

rationalist assumption itself, upon which all such data-driven policies rest, fatally incomplete.

Not only does such an assumption leave something out of a *particular* analysis – in which

case it might be solved by new and better data, improved statistical techniques and so on –

but its incompleteness is *necessary* and immeasurable, manifest in *every* analysis.[48] The use

of methods relying upon the rationalist assumption precludes a priori any hope of speaking to

the broad, one might say ordinary, concept of educational quality. In revealing and labeling

"theoretic" the rationalist formulation of the "disquieting and disconcerting conditions" that

confront the educational endeavor – a formulation that he, directly citing Dewey, attributes to

inadequate "problemation" (Schwab, 1983, p. 257) – Schwab offers a *dissolution* of the

problem so formulated. On his view, a productive approach to the "disquieting and

disconcerting conditions" will found itself on what he calls "the practical," a description and

analysis of which follows here.

The standard to which Schwab holds his "practical" approach is exceedingly high,

and one ought to be reminded of that which the theoretic cannot do, and what, therefore, the

practical must take up. In Schwab's own words:

> It is clear, I submit, that a defensible curriculum or plan of curriculum must be
> one which somehow takes account of all these subsubjects which pertain to man. It
> cannot take only one and ignore the others; it cannot even take account of many of
> them and ignore one. Not only is each of them a constituent and a condition for
> decent human existence but each interpenetrates the others (Schwab, 1969, p. 9)

---

[48] There is an awesome point to be made about this – the "immeasurable" claim is literal. In stats,
for a technical reason I cannot now recall, the error term in cases like these is literally inestimable,
which would be a great point to make if I could recall my intermediate statistics knowledge.

In exploring Schwab's notion of the practical, its ability to respond to the felt need to "somehow [take] account of all these subsubjects which pertain to man" will, I think, provide an important test.

In contrast with this early sketch of theory, then, Schwab elucidates his sense of the practical. The fundamental distinction between the practical and the theoretic as he conceives of these terms has not only to do with what I have laid out above – roughly, the inability of the theoretic, by virtue of its generality, to account for or to result in appropriate, flexible responses to actual on-the-ground conditions in any particular instance, along similar lines or argumentation utilized by Dreyfus and others – but also to do with maintaining a focus on the *need* or *call* to address particular situations. In other words, it is not sufficient, in the realm of education, to point out the shortcomings of the "theoretic"; there remain "disquieting or disconcerting conditions" that require responses. Schwab's conception of the "practical" seeks a mode of posing and addressing such disquietudes. He is quick to note, in the first place, that his notion of the "practical" ought not to be confused with a retreat from thinking as such:

> By the "practical" I do not mean the curbstone practicality of the mediocre administrator and the man on the street, for whom the practical means the easily achieved, familiar goals which can be reached by familiar means. I refer, rather, to a complex discipline, relatively unfamiliar to the academic and differing radically from the disciplines of the theoretic. It is the discipline concerned with choice and action, in contrast with the theoretic, which is concerned with knowledge (Schwab, 1969, pp. 1-2).

If the "practical" requires a type of rigor that differentiates it from the "curbstone practicality of the mediocre administrator and the man on the street," and if it requires that its rigor will not be expressed in terms of the construction of theory, which, as Duncan's notions of the rigor embodied by student achievement data helps to suggest, continues to function as

a meaningful criterion of the very concept of rigor, it will perhaps not be overstating the matter to say that Schwab's version of the practical represents something new and potentially radical with respect to the questions under discussion.

Most radically, Schwab proposes to begin the task of "problemation" with the *particular* rather than the general, as is habitually the case with the "theoretic." If the generality inherent in "theoretic" responses to problems renders their application questionable in any specific circumstance, then Schwab suggests that one need not construct each disquieting phenomenon in the classroom as a particular instance of a *general* feature or class. Schwab is explicit on this matter: "What we know is not facts *sub specie aeterni*, but facts as parts of a practical problem and as means to its solution" (Schwab, 1959, p. 150). The "choice and action" he has in mind, then, are called for with respect to a "concern," which is itself always grounded in particular conditions encountered by particular people. Putting it another way, Schwab avers that "The construction of needed diversities entails attention by planners of curriculum to the *local* (Schwab, 1983, p. 242). Against a view in which responses or prescriptions or practices are conceived as applicable generally, and are offered as solutions to, one might say, central tendencies extracted from massive compilations of specific conditions, Schwab's "practical" seeks to forego the step back from *this* situation, *this* particular instance. We have seen, throughout the discussion immediately above, as well as that of Chapter 2, the ways in which generally- or theoretically-constructed responses lose purchase on particular situations; in amelioration of this issue, Schwab offers the practical as a means of remaining close to the "local," the "disquieting or disconcerting conditions" that require response in the first place.

Taking the local as a starting point and offering the "practical" in opposition to the "theoretic" might tempt one to read Schwab as banishing theory from educational problemation and problem-solving altogether. But this is hardly the case. Schwab's discussion of the practical and the theoretic is best conceived as pertaining to *when* and *how* a disquieting situation draws upon the theoretic. Schwab's mistrust of the theoretic is directed precisely at theory's "sweeping appearance of success," which is to say, the sweeping appearance of success visible according to *each* incommensurable theoretical viewpoint. This fact suggests to him, not inappropriately, that each such viewpoint omits relevant factors immediately visible in the disquieting local conditions. Each theoretical viewpoint, in other words, serves to *construct* the problem as well as to solve it. Schwab's call to begin with the particular, the local, or the concrete is aimed at defusing this particularly nefarious aspect of reliance upon the theoretic, rather than turning away from theoretical knowledge wholesale.

The correct application of theoretical knowledge, in Schwab, comes by way of what he calls "the arts of the eclectic," which pertain explicitly to overcoming the limitations inherent in employing any single theoretical viewpoint through drawing upon many of them instead. In an article detailing the processes of the practical arts, Schwab puts it this way:

> First, the particularities of each practical problem can be sought in the practical situation itself, the search guided by resources much richer than any one theory can afford. Second, in each instance of application of a borrowed theory to a practical situation, incongruities can be adjusted by mutual accommodation. Third, restricted subject and limited treatment so characteristic of behavioral theories can be transcended by using more than one such theory (Schwab, 1971, p. 495).

Taking a particular situation as calling for the application of theoretical knowledge – a disquietude calling for a response – the role of theory is conceived in terms of offering an actual and singular response, which ought to be understood as different from (and not

necessarily coextensive with) answering a question or providing a self-consistent explanation.[49] In suggesting that the search for the problem is to be "guided by resources much richer than any one theory can afford," Schwab seems to allude to something which, in the test I mentioned above, allows one to "somehow take account of all the subsubjects that pertain to man." The linkage between them points in the direction of agreement in forms of life, shared background practices, the trans-human. It is important to express the sense of this direction now, since Schwab's language can be misconstrued as seeking firmer (formalizable) ground, as perhaps Shulman understands him.[50]

Such a response, as Schwab notes, insofar as it requires "mutual accommodation" and "using more than one such theory" will mean establishing some way of selecting among such theories, processes of enacting the "mutual accommodation" he mentions. This requirement of a method of selecting among such theories seems to beckon in the direction of the same regress that threatens accounts of skillful practice based on the rationalist assumption. In fact, one might take Shulman's "strategic knowledge," which does indeed fall into this regress, as a metaphysical account of that which enables a particular practitioner to successfully navigate such indeterminacy. Schwab himself avoids such prescription,

---

[49] In Wittgenstein's *Nachlass* 219, 81: "That is the fatal thing about the scientific way of thinking (which today possesses the whole world), that it considers every disquietude to be a question and is surprised not to be able to answer it."

[50] It may also be important at this early point to militate against an overzealous criticism of Schwab's project on the basis of the obvious utopian impulse behind the urge to "take account of all the subsubjects that pertain to man." Without delving into detail so early, I wish merely to suggest that a consideration of what it may mean to "take account" of all subsubjects may reveal that such a view does not entail the possibility of *having* a sort of panoptic view of all these subsubjects together. This panoptic perspective, after all, would be awfully similar in ontological status to the "metatheory" that Schwab puts out of play early on.

however, which enables him to avoid the regress entailed in providing a theoretical means of deciding among necessarily incomplete theories, much to his credit. He is explicit on this point:

> The methods by which these ends might be achieved have, however, a complication of their own. Although they can be described and exemplified, they cannot be reduced to generally applicable rules. Rather, in each instance of their application, they must be modified and adjusted to the case in hand. Because of this complication, I call them *arts* (Schwab, 1971, p. 495).

The "eclectic," as Schwab terms the process of selecting portions or aspects of various theoretical positions in order to respond to the disquietude at hand, is not *itself* theoretic in nature, and that fact marks an important break with the various means of addressing educational problems according to the rationalist assumption. By invoking the notion of an "art," Schwab refers to a region of knowledge or experience that is inveterately tied to *practice,* as well as what Crease might call the "trans-human," or agreements in forms of life. This follows also from Schwab's note that the invocation of "art" is directly motivated by the fact that such "methods. . . can be described and exemplified, [but] they cannot be reduced to generally applicable rules," which more or less directly maps onto Dreyfus's description of proficient and expert performers' reliance upon irreducible, holistic, situational paradigms, as well as onto Wittgenstein's insistence that in teaching someone the meaning of a word, he does so by means of "examples and exercises," and in doing so "does not communicate less than [he] knows [himself]." In describing the "arts of the eclectic," in other words, Schwab joins the thinkers of Chapter 2 in proposing alternatives to a theoretically-grounded view of practical, concrete situations.

The "arts of the eclectic" hold out the promise of another tremendous improvement upon any reliance on the "theoretic" as conceived according to the rationalist assumption

upon which Duncan's Race to the Top policies rely. Earlier in this Chapter, I explored the

temporal limitations that beset the studies that serve to justify the use of teacher value-added,

to take that aspect of teacher evaluation as an example. Specifically, I pointed out that

because such studies are necessarily based upon *past* academic competencies and, at best,

present adult outcomes, they have a particularly weak claim to futural conditions. Another

way of expressing this weakness is to assert that the use of such studies to lay out the

particular teacher behaviors associated with "successful adult outcomes" for classroom

observers freezes the relevance of all the skills involved in teaching, learning, and being a

successful adult, which a fortiori renders this use of data a particularly ham-fisted way of

dealing with questions of educational quality. Schwab's notion of the practical arts,

however, in refusing to acquiesce to a *single* theoretical viewpoint and in drawing instead

upon a *variety* of such theories – as warranted by a particular disquietude – leaves open,

explicitly, the possibility of accounting for such changing relevance in the act of combining

and selecting among theoretical viewpoints:

> What remains as a viable alternative is the unsystematic, uneasy, pragmatic, and uncertain unions and connections which can be effected in an eclectic. And I must add, anticipating our discussion of the practical, that *changing* connections and *differing* orderings at different times of these separate theories, will characterize a sound eclectic (Schwab, 1969, p. 10).

As the concrete, local situation plays the role of calling for the intervention of

theory in the first place, it stands to reason that the ability to respond to shifting

relevance embodied in the "*changing* connections and *differing* orderings at different

times of these separate theories" would present a contrast with the kind of rigidity in

this regard necessitated by the attempt to formulate a universal and abstract means of

addressing educational disquietudes.

3.3.2 Aristotelian Sight, the Theoretic, and Emotional Involvement

One further salient point marks an area of overlap between a Heideggerian-Wittgensteinian perspective on the relation between "knowing" and "doing" in Chapter 2 and Schwab's discussion here. I noted above that Schwab avoids the regress that I read Shulman as falling into; it is likely relevant to this distinction between Schwab's avoidance and Shulman's succumbing that Shulman imagines "strategic knowledge" as something *had* by an *individual* teacher, while Schwab never explicitly restricts the exercise of the eclectic to a single practitioner (nor does he proscribe it). Ultimately, as Schwab also notes, however, a single teacher will be the one to take a particular action in a particular classroom, to respond to a concrete and local disquietude, and this presents a certain problem in terms of the human limitations that Schwab would like to recognize. In fact, he points out, as it were in sidelong fashion, one of the supreme difficulties in practicing the "arts of the eclectic": "The problem is to *see* [the various factors]—to take note that each is there and to honor it as possibly relevant to our concerns. This is difficult because we normally see only what we are instructed to look for and we are instructed by theory" (Schwab, 1971, p. 496).

In addition to the way that theory seems to guide or "instruct" our sight, such that only certain factors or aspects appear for consideration at all, an additional danger emerges in thinking of the loneliness of ordinary classroom practice: when a given teacher is alone in the classroom and facing a "disconcerting" situation, there is no (or little) possibility of alternative perspectives or aspects arising at all, such that an *individual* exercise of the eclectic would be imaginable. In recalling the discussion of Aristotelian sight from Chapter 2, I noted there that a "detached view" blinds a practitioner to the practical heart of the

matter, that attempting to see all aspects of a given situation at once precludes the possibility of responding to the *meaningful* center of the disquietude. Schwab's reference to what is "difficult" in seeing, I take it, responds to a danger lying at the other extreme: that sight, inveterately, partly because of a reliance upon a given theory, is limited to the perception of a single aspect. In Dreyfus's account of skillful practice, this limitation does not necessarily present a problem[51] – the emotionally involved practitioner, in effect, learns over time to discriminate among similar situations and to adopt the correct perspective, thanks to personal investment in the outcome, which entails a considerable reinforcing effect.

One undiscussed assumption in Dreyfus's account, however, that emerges for consideration here calls our attention to the fact that the "success" of a given perspective-adoption, to use Dreyfus's term, is not necessarily as obvious in an educational setting as it would be, for example, in chess. Schwab says, correctly, that the difficulty of seeing is connected with the fact that we "see only what we are instructed to look for, and we are instructed by theory." Given the unknowable long-term outcomes of one's teaching, it is not necessarily the case that a competent teacher, faced with a disconcerting situation and responding in such a way as to have some sort of effect on the situation, would immediately *know* whether or not his or her intervention had resulted in "success" or "failure," such that the emotional consequence that performs the reinforcing role in Dreyfus's account would come into play at all.

It is in precisely this situation, precisely here, that a competent third-year teacher would be tempted to see his or her practice only in terms of the governing theoretical light of

---

[51] Nor is it at all clear what would *oppose* this limitation – seeing the "whole" at once?

the present age – in this case, one in which success is defined in terms of the demonstration of best practices, whichever justificatory means are called upon to buttress them. It is unclear whether a lone teacher would be able to overcome the threat to "seeing" posed by any given theory on the one hand and the temptation to understand one's own perception as an inadequate reduction of some omniscient, panoramic view on the other. Dreyfus's account of skill development does not provide the conceptual tools to handle this sort of exigency. His account requires that the practitioner have some immediate sense of the success or failure of the choice of perspective, and also, it must be noted, that the practitioner *only* requires his or her *own* immediate sense of the same in order to develop.

In fact, others have noted this weakness in Dreyfus's model. In yet another strange coincidence, none other than Robert Crease has words to say on this matter:

> Oddly, given Dreyfus's philosophical commitments, his account lacks hermeneutical sensitivity. The flaw in his assumption that skilled behavior crystallizes out of contextual sensitivity plus experience without contribution from individual or cultural biography can be traced to a failure to take into account the fact that the embodied subject, even when behaving expertly, brings to the situation what has been historically and culturally transmitted to it, and in a way such that the subject can never grasp cognitively all at once. The individual expert performer, as a consequence, does not have a complete purchase on his or her own expert behavior. Therefore, contra Dreyfus, it will always be possible in principle for an expert performer to learn about one's own performance from another, contextually sensitive person – though, again, this is not because the other has managed to obtain an objective position, but on the contrary, because the other is differently situated (Selinger & Crease, 2002, p. 262).

Selinger and Crease protest that even the expert practitioner, to say nothing of the competent performer, will be able "to learn about one's own performance from another, contextually sensitive person." Perhaps no other field has as much to gain from this insight as teaching quite generally: even the most skillful teacher must respond to the given situation from *within* a given perspective, which means, as Selinger and Crease say, "it will always be

possible in principle" for another to fill out or augment even the expert's view of a given situation.

Schwab's view of the necessity for the eclectic, as we have seen, has to do with the insufficiency of any given theory to adequately address a given disquietude even and especially to the point of prescribing a particular response. But Schwab carefully notes that the practice of the eclectic is an "art" – no abstract method can secure, as it were, the successful selection of theoretical perspectives, nor their proportions relative to one another. If Selinger and Crease note a danger in Dreyfus's view of having to defer to an expert's (singular, limited) view of a given situation, it will be worthwhile to point out that Schwab anticipates the potential issue of relying upon any single perspective, even the expert's. Central to the practice of curriculum development for Schwab – and central, likewise, I will propose, to the successful evaluating of teachers – is the notion of "deliberation," a convening of people whose perspectives on a given thing, say, the relevant aspects of classroom practice, is likely to differ, and convening them in the interest of addressing a need which is – ultimately – local, concrete, and particular.

3.3.3 Deliberation

With Schwab's conception of the theoretic's insufficiencies laid out, then, and a satisfactory relation established between these insufficiencies and those of the rationalist assumption as explicated in Chapter 2, I return now to the test I suggested in introducing the practical/theoretic distinction, the high standard to which the practical ought to be held. Schwab says that an adequate alternative to a theoretic position must "somehow [take] account of all these subsubjects which pertain to man." Deliberation, then, provides an example of one of the "arts of the eclectic," a means of avoiding the theoretic error of

"arriving too late," as Heidegger calls it. The intent of this subsection is to reveal the implications, both salutary and otherwise, of Schwab's deliberation with an eye toward the evaluation of teachers in particular. I oppose this particular intention to any purpose of fully spelling out the intricacies of the deliberative process itself (a project to which Schwab himself devotes a significant portion of three full-length papers), of defending deliberation against all possible objections (which I will nonetheless name), and especially of proposing a detailed policy of what one might call deliberative teacher evaluations. The point, in other words, is to highlight the promising ways in which Schwab's deliberation seems to improve upon the sort of theoretic approach to teacher evaluation one finds in Race to the Top policies, as well as to point out the places in which deliberation leaves unanswered or underdeveloped questions in this regard. I propose, in short, to suggest deliberation, along with what Schwab would call an eclectic approach to the evaluation of teachers, as a possible corrective to the rigidity of Duncan's policies, which may perhaps make up the basis for an alternative approach to policy without formulating a counter-policy as such.

Schwab's deliberation, by way of introduction or description, is both and at once a frighteningly large and unwieldy institutional and intellectual reformation, and also a minor, locally-actionable suggestion. Its intimidating scope manifests itself in the conclusion of Schwab's first article on the Practical, in which he suggests that deliberation requires "the establishment of new journals, and the education of educators so that they can write for them and read them" (Schwab, 1969, p. 20). In addition to journals (and the education of educators to contribute to and profit from them) Schwab sees the need for "similar forums" operating in person, composed

of the teachers, supervisors, and administrators of a school; of the supervisors and administrators of a school system; of representatives of teachers, supervisors, and curriculum makers in subject areas and across subject areas; of the same representatives and specialists in curriculum, psychology, sociology, administration, and the subject-matter fields (Schwab, 1969, p. 21).

Schwab's deliberation proposes, in effect, a massive restructuring of the educational scene as an antidote to what Harold Bloom once called the "intellectual apartheid" represented in a tendency to hyper-specialization in higher education. Since the educative enterprise draws upon, or consist in, all of several specialized elements in combination, Schwab's proposal amounts to the creation of a field (in Bourdieu's sense of the term) upon which for these various specialties to interact. The relative feasibility of generating an *institutional* effort to achieve this sort of space on a national level through the formation of new journals, new organizations, and so on remains beyond the scope of my concern.[52] Rather, the feasibility of drawing and relying upon representatives of what Schwab will call (following Aristotle) the "common places" of the school in order to evaluate its work concerns me here. This aspect, one might say, comprises the locally-actionable aspect of deliberation.

Schwab's notion of the "common places" emerges from his attempt to access the *whole* of which each "subsubject" remains merely a part. In his second article on the practical, Schwab describes the means by which one discovers relations between the concerns at the heart of the enquiries undertaken by otherwise distinct fields. The notion at

---

[52] It may be the case that one *must* imagine an entirely different relation among specialists and stakeholders (which includes, I suppose, everyone) in education in order to speak of the microcosmic level of Schwab's deliberation, but I do not think this is so. What results at the microcosmic level (which is what Schwab spends most of his time discussing) will be incomplete in the same fails-to-take-account-of-absolutely-everything way that approaches such as Duncan's is incomplete; but my aim is not to consider whether such completeness is a real or ideal limit, but rather to suggest that Schwab's deliberation is *more* complete, and especially accesses the *essential* shared background that any rationalist approach leaves out.

which Schwab arrives consists in performing a "reflexive enquiry of enquiry" in order to "press past the 'answers' propounded in the paper to the suppressed questions (problems) to which the answers speak" (Schwab, 1971, pp. 513-514). In this way, the practical *concern* ought to emerge as something – what Schwab would call a "problem," correctly conceived – requiring response. For him, then, the "common places" (the space between the words in this term disappears in subsequent writings) seem to represent a marginalized area of educational research, marginal precisely because tangential to the theoretic concerns of this or that particular field.

The invocation of the Aristotelian term, as well as Schwab's vision of "enquiry" as responding at heart to a practical concern, aligns Schwab with both Heidegger and Wittgenstein, though most explicitly with the former, as the notion of responding to a sort of concern recalls the "toward-which" that grounds the "totality of involvements" in Heidegger's analysis of Being. The way in which Schwab speaks of arriving at such common places, however – "it is constructed by a certain mode of systematic comparison of principles, premises, methods, and selections used by and in each enquiry" (Schwab, 1971, p. 513) – seems to indicate that this alignment is by no means secure. The reliance, for example, upon a "systematic comparison" of a certain and articulable list of theoretical features, a comparison that will "generate a set of factors to be called 'common places' or 'topica' (Schwab, 1971, p. 513), arouses immediate suspicion. The "systematic" nature of the exploration, the finite and aprioristically declared list of issues for exploration, and the notion that "common places" in an Aristotelian sense are articulable as "a set of factors" strikes one as markedly similar in *kind* to the forms of enquiry beyond which Schwab's deliberation seeks to go. Put simply, Schwab's discussion of the common places in his

second article constructs the common places as something (only?) available to a detached or disengaged perspective, as an epistemological object to be "constructed" or "generated" according to a fixed procedure. Schwab proposes to reach the "practical" by theoretic means, it seems.

Schwab's third article in the cycle, however, casts what he now calls (with no space) the "commonplaces" in a different light. Foregoing the systematic, methodical language of the previous article, he opens his "Translation into Curriculum" paper by referring to the need for deliberation as emerging from the fact that "Scholars, as such, are incompetent to translate scholarly materials into curriculum" (Schwab, 1973, p. 501). He proceeds to say that scholars "possess one body of disciplines indispensable to the task. They lack four others, equally indispensable" (Schwab, 1973, p. 501). In the first place, one notes that Schwab does not speak of bodies of *knowledge*, but of bodies of "disciplines," which speaks to his effort to foreground the way these bodies concern *doings*, practical matters. Where in the previous article, commonplaces were conceived as a "set of factors" generated by a given enquiry, here he notes that they are "concerns, values, and operations, which arise from. . . experience" (Schwab, 1973, p. 501). In taking these five areas – the subject-matter, the learner, the "milieus" ("milieux" in the earlier article), the teacher, and the curriculum-maker – to differ on the axes of "concerns, values, and operations," and in insisting that each must learn something from the other's conception of the same, he seeks to move in the direction of the *whole* presupposed, as Dreyfus would say, by each individual common place.

While it might yet appear that his reduction of this whole into five specific areas risks unwarrantedly imposing something of the theoretic upon this picture, Schwab's five "bodies of disciplines" establish names for the sake of convenience rather than posit logical

boundaries or something of the like. A closer look at his "subject matter" commonplace

alone, by far the simplest and least involved, reveals the inherent instability, incongruity, and,

one might say, infinitude that commonplaces offer, conceptually speaking. (I suggest this,

obviously, as a *positive* aspect of Schwab's approach.) But it would also be helpful to note,

before taking up the example of the "subject matter" commonplace itself, that each of these

commonplaces, as Schwab imagines them, differs radically from any mere "area" of enquiry;

each commonplace is conceived as a *person*; it is representable by a member of its

community, but is irreducible to propositional abstraction. This, I take it, is similar to what

Hlebowitsh has in mind when he lauds Schwab for focusing on people rather than ideas (P.

Hlebowitsh, 2012, p. 3).

Schwab's notion of accounting for the "subject matter" as a commonplace, for

example, opens up on its own universe of concerns, values, and operations – and even these

terms seem suddenly inadequate. Schwab says,

> There must be someone familiar with the scholarly materials under treatment
> and with the discipline from which they come. Suppose the materials under
> consideration are historical; then a member of the group must be familiar not only
> with this body of historical material but must also know what it is to be a historian
> (Schwab, 1973, p. 502).

Schwab emphasizes, first of all, that the subject matter commonplace really is a *place*,

occupied by people.[53] This is especially important, as the person to serve as a representative

of his or her particular "discipline" brings two different forms of knowledge to bear: a

familiarity with "this body of historical material" and also "what it is to be a historian." The

---

[53] One really could bring Bourdieu to bear on Schwab, since he really uses the spatial notion and
gained prominence in the American academy after Schwab's time. A quick search seems to indicate
that this topic has not been dealt with.

latter of these two forms of knowledge is precisely the kind that Chapter 2 suggests is non-

formalizable, the sort that the theoretic cannot touch. As Dreyfus says, relying upon

Wittgenstein,

> In explaining our actions we must always sooner or later fall back on our everyday practices and simply say "this is what we do" or "that's what it is to be a human being." Thus in the last analysis all intelligibility and all intelligent behavior must be traced back to our sense of what we are, which is, according to this argument, necessarily, on pain of regress, something we can never explicitly *know* (H. L. Dreyfus, 1992, p. 57).

A representative of the subject matter, Dreyfus implies, *has* a sense of "what it is to

be a historian," but this sense is, "on pain of regress, something we can never explicitly

know." A detached or theoretic approach to dealing with matters of schooling would require

– indeed, they *do* require, in the present tense, and explicitly rely upon – formalizations of

such things as "what it is to be a historian" or "what it is to be a successful adult" in order to

measure the practice of education. In formalizing the non-formalizable, as the first two

Chapters of this project seek to demonstrate, the "sense of what we are" that the historian of

Schwab's example ineluctably carries with him or her into a given room becomes distorted;

or else, worse than this (or perhaps not), its resistance to formalization alternatively leads it

to be left out of such matters entirely. Schwab's call to convene embodied representatives of

the many commonplaces involved in schooling, based precisely upon *differences* among the

"concerns, values and methods" that characterize their respective disciplines, provides a

measure of hope that the resulting deliberation might yield something that speaks to the

"common" in the commonplace, something that accesses, at long last, the ordinary.

For contrast, consider the Chetty research once more, and in particular the way in which a "successful adult" is represented according to a variety of metrics. The list of constituent parts that comprise adult success is indeed long:

> Earnings, college attendance, college quality (measured by the earnings of previous graduates of the same college), neighborhood quality (measured by the percentage of college graduates in their zip code), teenage birth rates for females (measured by claiming a dependent born when the woman was still a teenager), and retirement savings (measured by contributions to 401[k] plans) (Chetty, et al., 2012).

Long, as I indicated earlier, but hardly exhaustive, and not only because more *indicators* could be added (infinitely) to the list, but also because more *measures* of each indicator are theoretically available, as well, to provide a different aspect of each individual factor. Any attempt to translate the sense of "what it is to be" *anything* in a human sense (successful adult, historian, a human being itself) into articulable or propositional form *necessarily* distorts by omission the sense this translation seeks to represent. The inadequacy of Chetty's list is not only visible in its omission of potentially relevant factors, but also in the very fact of that omission's necessity. Being a mathematical model, it *cannot* include anything that resists modeling, such as, for example, a sense of "what it is to be a historian."

Schwab's deliberation, occurring in conversation between people, on the other hand, *can* draw upon precisely this nonformalizable "understanding of being." Deliberation, consisting of a conversation among representatives of the educational commonplaces, manages to avoid the pitfall Chetty falls into, what Dreyfus calls "preanalyzing all situations in terms of a fixed set of possibly relevant primitives," an incomplete, inflexible, and insufficient means of dealing with the practical. Dreyfus, in fact, in the same paragraph, sets in opposition to this pitfall precisely what Deliberation accomplishes: "The only way to avoid this loop is to be always-already-in-a-situation without representing it" (H. L. Dreyfus,

1992, p. 55). Through convening people, rather than abstractions, Schwab's deliberation holds open the possibility of bringing both the knowledge of a "body of historical material" and the knowledge of "what it is to be a historian" to bear upon the addressing of a given disquietude.

The language of Schwab's first two pieces in the practical cycle differs substantially from the language in the later two. The earlier ones contain the language of dividing subjects into subsubjects, a process which has no obvious endpoint; as well as that of prescribing a particular "systematic" means of uncovering the commonplaces, which seems to hint in the direction of timeless universality. The later two pieces speak a more practical language, one might say: he speaks now in terms of "disciplines" rather than subjects or subsubjects; he alludes to "concerns, values, and methods" rather than "a set of factors" that would denote a given commonplace. While it is possible to object to Schwab's method on a variety of grounds – the National Council on Teacher Quality, for example, would likely find too much "subjectivity or guesswork" in such a process, and it must be noted that Schwab's examples of deliberation would not help Duncan tell the best 30,000 of California's teachers from the worst – I think it mistaken to read, as it may be tempting to do, the notion of the commonplaces, in particular, as aprioristically fixed elements in the deliberative process.

In other words, one could reasonably object that the existing list of commonplaces is itself incomplete, that Schwab has missed or ignored something crucial, and that this fact undermines the approach. I take issue with such an objection on two grounds: first, there is no indication in Schwab's work that these ought to be the *only* commonplaces that one ought to consider, though there is also no *need* to create more, since his notion of the "milieus" is so expansive. He considers the representative of this commonplace to have "experience of

the milieus in which the child's learning will take place and in which its fruits will be brought to bear" (Schwab, 1973, p. 503) – in other words, the local (and likely larger) world.   While Schwab notes some of the milieus by name – "the family, the community, the particular groupings of religious, class, ethnic genus" – he also notes that "the relevant milieus are manifold, nesting one within another like Chinese boxes" (Schwab, 1973, p. 503). This certainly does not seem to imply any limit on what may be taken as relevant to the concern at hand. In the second place, Schwab's most elementary commitment is to the idea of attending to the local concern, the problem at hand, the need requiring response. Nothing about Schwab's approach seems to preclude the possibility of eliminating or adding commonplaces, should the need arise. Should a given situation seems to require the voice or experience of some heretofore unpredicted (or as-yet-nonexistent) domain, nothing about Schwab's ideas would seem to impede its inclusion. In this way, Schwab's deliberation is structurally able to bear the unpredictable issues around shifting or evolving *relevance* in a way that mathematical modeling cannot.

I noted earlier that Schwab sets a high bar for his notion of the practical: it must, he says, "somehow [take] account of all these subsubjects which pertain to man." While one must acknowledge that *simply* locking people of different scholarly, professional, and social backgrounds in a room and assigning them a problem to which they all might speak will not necessarily succeed in taking the *whole* into account in every instance, the way in which deliberation by its (embodied) personal and local nature turns away from the requirement to representation renders its results "whole" in the specific and important sense that shared understandings of being are no longer left out by logical necessity. In fact, Schwab's naming and selecting of commonplaces based on their *differences* speaks to a concerted effort

(perhaps quixotic) to reach this "whole." These differences, one could say, and as in fact

Schwab does note, light up divergent aspects of a shared issue, a shared task. The

differences among the commonplaces, and the conversation among them, thus presuppose

the very agreement in forms of life upon which such conversation is founded; this criterial

agreement may be, in fact, the only sort of "whole" of which it makes sense to speak.

3.3.4 Justification and Responsiveness

In bringing a Schwabian perspective to bear upon the evaluation of teacher quality

that remains the primary concern of this work, I would like to raise again the sense of affront

that Duncan and Penskar seem to share with regards to the inability of researchers to

differentiate between good and bad teachers and to recall Wittgenstein's distinction between

"learning anything *new*" by a given investigation and seeking "to *understand* something that

is already in plain view." I wish, in doing so, to circle back at last to the most pressing issues

I articulated in the introduction to Chapter 3:

> In Chapter 1, I somewhat clumsily asked of Duncan's statement about
> California's 300,000 teachers, "*Where,* precisely, is the need for such data that the
> California example is meant to make evident?" In this Chapter, I hope to better
> express the impulse behind my earlier objection: that the need to be answered by
> distinguishing the best 30,000 teachers in California from the worst is simply
> incoherent from any perspective *except* that of the state or federal government, and
> that there is no reason to assume that the means of answering a state or federal need
> will also or equally suffice to answer a school-specific, community, or parental need
> (for example).
> As Chapter 1 in particular has revealed, however, Duncan himself, among
> many others, often appeals to *differing* and *incommensurable* criteria in the discussion
> of teacher quality. That two differing concepts of teacher quality pervade the
> discourse has been established, as has the fact that Race to the Top and state policy
> has insufficient awareness of that fact. The question confronting us in this final
> Chapter is: given the differing human purposes (understandings of being) in terms of
> which data, measurement, and evaluation make their claims upon us, how are we to
> navigate these differences in order to say something meaningful to someone else
> about whether or not this or that teacher is good at teaching?

This Chapter of the project has sought to bring out the ways in which Race to the Top's use of student achievement data amounts to an attempt to subsume the various and meaningfully different projects of and parties to the educational endeavor under a single measure, a single mode of both knowing and telling, which Duncan and others associate with something like *new* knowledge, something akin to truth, upon which we can (at last) act with a measure of certainty. The critical thrust of this Chapter has been to demonstrate (a) that these purposes and parties remain somewhat distinct from one another with respect specifically to the way in which a given child's education is taken to *matter,* a fact to which Schwab responds with his notion of plural commonplaces, and (b) that the commonality shared among these purposes and parties, while *extant*, does not relate to these manifold differences in a way that one might characterize as grounding, erasing, subsuming, or overarching. The region of commonality, in other words, cannot render the differences either merely *superficial* or *apparent,* next to which the common or shared would be able to claim the superior status of the real or the primary. Moreover, as I suggested in Chapter 2, (c) the common or the shared, which comes down in every case to a notion of family resemblance embedded in forms of life or shared understandings of being, is nonformalizable in nature, which makes the means inherent to the generation of Duncan's notion of student achievement data incapable of expressing or measuring this region of commonality. Far from representing the *whole* or the metaphysical *singularity* of the educative endeavor, Duncan's student achievement data reveals itself to be merely one among a variety of measures, ontic and ontological alike, of educational quality. It is, in fact, one of the "milieus" that Schwab explicitly considers in laying out the kinds of questions that one ought to ask in exploring the area of practical concern:

What are the conditions, dominant preoccupations, and cultural climate of the whole polity and its social classes, insofar as these may affect the careers, the probable fate, and ego identity of the children whom we want to teach? A dominant anti-intellectualism, a focus on material acquisition, a high value on conformity to a nationwide pattern and on the cloaking of cultural-religious differences are possible influences (Schwab, 1973, p. 504).

The primary difference between the way Schwab addresses the futural view of a general cultural climate of the "whole polity" in terms of "career" and "material acquisition" and the way Duncan and the reform movement quite generally treat these concerns has precisely to do with the relative possibility of entertaining or accounting for other versions or views of "outcomes," of *what matters* in education.

The general outline of Schwab's deliberation, itself an outgrowth of his conception of the practical and the eclectic, differs from the rationalist approach that Duncan adopts in two important ways that render it recommendable to the general purpose of evaluating teachers and reporting on teacher quality.

In the first place, Race to the Top's use of achievement data establishes a priori the primacy of the general in terms of both educative purposes (which are economic in nature, visible in terms of careers and earnings and savings, as the Chetty research has demonstrated) and in terms of an overall evaluative context (which, for student achievement data, is always massively broad: 300,000 California teachers, the discourse of national economic security, international comparisons). The primacy of the general, thus established, allows the knowing of student learning and teacher quality to take place from a radically wide-angled and distant perspective. Diane Ravitch notes the way in which the use of data is taken to supplant any need for familiarity with local context: "With their sophisticated tools and capacity to do multivariate longitudinal analysis, they did not need to enter the classroom, observe teachers,

or review student work to know which teachers were the best and which were the worst, which were effective and which were ineffective" (Ravitch, 2010, p. 180).

In contrast, while Schwab includes a place for the "whole polity" under the commonplace of the milieus, his notion of deliberation would bring a representative of that milieu – say, Duncan himself, or more likely an academic or think-tank researcher with expertise in the interpretation of student achievement data – into direct conversation with teachers of the particular students in question, in addition to child psychologists, administrators at the local and state level, children themselves, members of the community, and other representatives of further commonplaces to address the need to say something meaningful about the quality of the particular teachers in a particular school. Schwab's view does not, then, stand in direct *opposition* to the teacher-quality policies of Race to the Top. Such opposition would mean that Schwab's deliberation would *only* take into account local conditions and be amenable to a local telling, much in the way that Race to the Top proposes to use broadly general resources – whose relation to the local involves something like the mere statistical approximation of limited and fixed "local conditions" – in order to speak simultaneously and univocally to national and local purposes as though they were in every case identical. Rather, Schwab's deliberation saves a place, explicitly, for the concerns of the general – concerns that find their appropriate response in precisely the use of student achievement data to generate measures of teacher value-added. Part of the deliberative process would involve balancing *this* type of response to the question of teacher evaluation with others, and achieving this balance *through interpersonal conversation*.

The assumption of the reformers holds student achievement data in such high regard that no such discussion is necessary: in the face of disagreement, student achievement data

simply has it right, where local ignorance of one kind or another is responsible for the unnecessary restiveness of disagreement. In an op-ed advocating for better dissemination of achievement data in reference to proposed "parent trigger legislation," Peg Tyre laments with respect to school choice or open enrollment policies that "sometimes the parents don't even know that the 'choice' they are making is a bad one." Tyre's solution is standard fare: "parents need unbiased, accessible information about what solid research tells us works best in schools — even if they don't have a computer at home or if English isn't their first language" (Tyre, 2011). In cases where parental judgment as to the quality of their children's education conflicts with "what solid research tells us works best in schools," the parents are simply wrong.[54] It is merely a manifestation of the parents not "even know[ing] that the choice they are making is a bad one." Disagreement in judgments, for the reformers, indicates ignorance on the part of those who lack the data. Disagreement in judgments, for Schwab, by extreme contrast, expresses the diversity of needs to which education responds; such disagreement *expresses the need* for deliberation itself (though arriving at agreement is

---

[54] This may perhaps seem similar to other examples of parents' relation to the best interests of their children suffusing the national news, particularly around the relative safety of vaccines. Scientific studies of vaccines have shown no link between vaccines and, for instance, autism; yet certain parents insist that such a link exists, that the studies have it wrong, and as a result they choose not to have their kids immunized. I think Tyre is likely using this type of scenario as a model – parents simply have an inadequate faith in, or knowledge of, the scientific findings about "what works" in education. That these cases are meaningfully different ought to be obvious by this point: the vaccine case hinges upon whether or not a *specific* outcome – autism – is entailed. The teacher-quality case, however, admits to dependence upon a multitude of specifiable outcomes, sometimes severally and sometimes all at once. It is simply not possible to *have* the sort of argument about the standing of scientific evidence that one sees in the vaccine case with respect to teacher quality. A parent could *readily concede* that such-and-such behavior is correlated with raising test scores, admit that the teacher in question rarely exhibits that behavior, and yet insist that the teacher is of high quality, pointing to a multitude of other positive educational outcomes entailed by his or her practice. Whether or not a given vaccine is linked to autism in children is a question science can answer. Whether or not this or that teacher is of high quality is something that science can perhaps marshal evidence for, but not something it can *answer*.

not necessarily the goal thereof). In other words, although this is perhaps a different claim,

the *standing possibility* that parents *can* be incorrect in their judgments calls for *conversation*

rather than an attempt at the absolute *foreclosure* of this possibility via "accessible

information," as though parents could *never* be correct in disagreeing with data's verdicts.

 The second area of contrast between Duncan's proposed use of student achievement

data and Schwab's concept of deliberation proves to be the most salient, and it follows

directly from the above. I have said before, and in fact in many places, that it remains

unclear throughout Duncan's many public statements whether the problem he seeks to

address by recourse to the use of achievement data is one of *knowing* or one of *telling*.

Throughout the literature on best practices, as Chapter 2 explored in some detail, one finds

that knowing and telling are simply bound up together – the slow spread of "best practices" is

taken (problematically) as a communication problem, for example. Duncan's statements in

the *Daily Show* interview echo this view, as does Tyre's commentary that "accessible

information" provides a solution to incorrect parental judgment. Once *known*, the *telling* of a

thing is the denouement. Tyre's construction of the matter in terms of "accessible

information," in fact, beautifully renders the notion in appropriately passive form. Telling is

hardly required at all; the publication of the data in a location accessible even to parents who

"don't have a computer at home" or for whom "English isn't [the] first language," in Tyre's

words, will do the trick.

 The centrality of *conversation* to the deliberative enterprise, as Schwab has it,

presents a radical departure from the view behind Race to the Top's policies. On Schwab's

view, the right and responsibility to enter into the *telling* of a thing – a judgment, for example

– is afforded by a certain standing with respect to the educative context in question. The

leveling of a judgment is an expression, in fact, of that standing, that right, that responsibility. But precisely this standing, along with its rights and responsibilities, renders one *beholden* to acknowledge the standings of the other representatives of the further commonplaces. Rendering a judgment in the deliberative process, then, is not an act of finality, nor a pronunciation even of a subjective fact; it is rather the staking of claim in the space of deliberation. Because each commonplace (and subgrouping within each commonplace) relies upon its own somewhat discrete "concerns, values, and operations," each representative's judgment is also always an expression of *just these* concerns and values, which, as an expression of *finite* concerns and values, simultaneously acknowledges the standings of others.

The most basic idea here partially agrees with Duncan's particular fear of subjective error: no one can do things pertaining to education, be it curriculum-making or teacher evaluation, all alone; the responsibility is shared; we have to do it together. But Schwab's stance also profoundly disagrees with Duncan's: neither is it the case that someone might create something (algorithmic, robotic, or otherwise) that would then do the work *for* us; the shared responsibility is inalienable; *we have to do it together*. Neither recourse to the incommunicably private judgment (whatever that might mean) nor recourse to an inescapably absolute or objective judgment will suffice. Evaluating something connected "with the trans-human," as Crease says, requires trans-human intervention.

Returning to Cavell's thinking here helps to emphasize the departure from Duncan's scientistic chauvinism (if I can call it that – I badly want to) that Schwab's deliberation represents. At many points over the course of his career, Cavell has had occasions to muse on the significance of Wittgenstein's contrast between learning new facts and understanding

something in plain view, and his considerations of these matters prove fruitful for the

discussion here.

In his *Little Did I Know,* for example, to return to a passage I cited earlier, Cavell

highlights precisely the issue that Schwab's deliberation takes up, and which Duncan's

student achievement data assumes to be unnecessary:

> My stress on the time, or time and place, of depiction is meant to capture what
> Austin means in tirelessly demanding the context (he would often call this the story)
> of an utterance and what Wittgenstein means by repeatedly asking to whom an
> utterance is made. When Wittgenstein asks, "How is telling done?" he is in effect
> asking how it is that saying something, speaking, is done; how it is that someone is in
> a position to be told something. This turns out to be a good question (Cavell, 2010, p.
> 60)

Duncan's view of the matter seems to assume that we are all, inevitably, by virtue of

our human reason (one might say), in a position to be told something; moreover that the

universality of our position with respect to being told something means that the something

itself will be singular in form and content alike; and that the task, then, becomes simply to

know what there is to be told. The elucidations of the myriad shortcomings and manifest

incompleteness of this view do not require recounting here, though stating the matter clearly

certainly suggests a possible explanation for the emphasis on *knowing* in the reform

literature, and the denigration of *telling*.

But as Cavell says, "how it is that someone is in a position to be told something" does

indeed "turn out to be a good question." Schwab takes it up in somewhat Heideggerian

fashion, by locating one's position relative to such telling and hearing on educational matters

in terms of a particular *concern* with the educative endeavor, a version of the Heideggerian

towards-which that shapes differential positions with respect to education.

Cavell's view, like Schwab's, acknowledges that different people – representatives of different commonplaces – will have differing positions relative to any particular telling. In *The Claim of Reason*, Cavell interprets Wittgenstein's enigmatic "Essence is expressed by grammar" to indicate that "You have to know certain things about an object in order to know anything else about it (about *it*) (Cavell, 1979, p. 77). Grammar, on this reading, refers to a certain kind of region of knowledge that is always had beforehand, it seems, the on-the-basis-of-which (as Dreyfus would say), through which one comes to understand anything new. On this view, the "differing positions" among the various representatives of the commonplaces might amount to different arrangements (*zusammenstellungen*) of something shared, something like forms of life or background practices. The sharing, here, brings members of the commonplaces into "a position to be told something." The differences among them ensure that the "something" will be what Cavell later and colloquially calls "news" to them.

This notion recalls Cavell's sense of rationality (what counts as rational) as inherently bound up in the what is shared or shareable. As he puts it in *The Claim of Reason*,

> The philosophical appeal to what we say, and the search for our criteria on the basis of which we say what we say, are claims to community. And the claim to community is always a search for the basis upon which it can or has been established. I have nothing more to go on than my conviction, my sense that I make sense. It may prove to be the case that I am wrong, that my conviction isolates me, from all others, from myself. That will not be the same as the discovery that I am dogmatic or egomaniacal. The wish and search for community are the wish and search for reason (Cavell, 1979, p. 20).

In reflecting upon *The Claim of Reason* in *Little Did I Know*, Cavell sounds still more Schwabian:

> I still would not discount the explicit idea of those chapters, to the effect that what makes moral exchange rational is not whether it leads to an agreed upon conclusion but whether the exchange is itself pertinent and clarifies incongruent

positions, that is, clarifies the responsibilities for holding respectively these positions (Cavell, 2010, p. 377).

What is shared, or the fact that it is shared, does not ameliorate the differences or "incongruent positions" that exchange or conversation manifests and responds to. It does, however, acknowledge a communal stake in the conversation itself. *Telling*, then, the expression of judgments, always also expresses one's position and one's stake; telling speaks in response to particular needs emergent in a shared concern. The version of *telling* here, as in the reporting upon the quality of a given teacher, is vastly more involved, more fraught, than the version of telling that appears in Duncan's rhetoric and the reform literature generally.

One further aspect of Cavell's interpretation of Wittgenstein requires explication before returning to Duncan's and Penskar's responses to the 99% statistic, but perhaps Wittgenstein expresses the impulse best himself:

> The difficult thing here is not, to dig down to the ground; no, it is to recognize the ground that lies before us as the ground.
> For the ground keeps on giving us the illusory image of greater depth, and when we seek to reach this, we keep on finding ourselves on the old level.
> Our disease is one of wanting to explain (Wittgenstein, 1983, VI, sec 31).

Wittgenstein contrasts a notion of difficulty that consists in getting beneath "the ground" with a notion of difficult in recognizing that "the ground that lies before us" *is* the ground. Cavell tends to interpret this Wittgensteinian idea in the terms of denial or refusal, the denial of what we ordinarily know to be true, the things that we do not, under ordinary circumstances, doubt. Cavell addresses this explicitly in *The Claim of Reason:*

> Any criticism of the classical investigations of knowledge, entered from an attention to what is ordinarily said when, must account for the fact that the traditional philosophers, masters of at least the language they write, have accepted a question as

requiring an answer which under other circumstances, *they* (as well as "the ordinary man") would have rejected as absurd (Cavell, 1979, p. 57).

Cavell sees the sort of denial at work in the picture of "traditional philosophers" accepting "a question as requiring an answer which under other circumstances, *they* (as well as "the ordinary man") would have rejected as absurd" as pointing to a dissatisfaction with knowledge itself, with our attachment to the world of objects and especially (for Cavell) of others. He expresses this intuition more clearly in *Little Did I Know:*

> My contrary sense [to other early interpretations] was that Wittgenstein was articulating what human assurance amounts to, that it does not stop short of conviction in the world, but that we ourselves become restive with this assurance, and we have the power to undermine ourselves in the name of an, as it were, unachievable or rather illusory rationality (Cavell, 2010, p. 376).

This particular interpretation of Wittgenstein suggests a potential (albeit uncomfortably psychological) impulse for Duncan's views on teacher evaluations as expressed in Race to the Top's policies. Duncan's offer of student achievement data as a solution to the problem that "99% of teachers are rated the same" implies that we have heretofore lacked adequate "conviction in the world," as Cavell calls it, and that such conviction might be achieved or restored by way of proper statistical or objective knowledge. As we have repeatedly seen, however, the methods of deriving teacher value-added, or measuring student learning in general, do not succeed in reaching what Wittgenstein calls the "greater depth"; rather, "we keep finding ourselves on the old level."

Cavell would not argue with Duncan, I think, in viewing the 99% statistic as expressing a failure of conviction, as he would call it. I very much suspect that he would disagree with Duncan that such conviction could be restored by means of correctly identifying the metaphysical kernel of "effective teaching" revealed in teacher value-added.

Rather, by way of recalling Penskar, Cavell might see *both* the 99% statistic and

Duncan's chosen mode of response in terms of denial, as denials of "what we have long been

familiar with" (Wittgenstein, 2009, pp. 52, sec 109) and of the communal nature of shared

forms of life in which "what we have long been familiar with" is visible in the first place –

the community *to* which, it might be noted, expressions of knowledge are *owed*. When

Penskar calls the notion that no one can figure out who the best teachers are "ridiculous," she

counters the assertion of the data-analysts (or the "unions," as she has it, for some reason)

with her own conviction in the world: "The parents always knew who the best teachers were.

If the parents can figure it out, certainly the education bureaucracy could do the same if it

really wanted to?" Penskar's rebuke speaks to the fact that *of course* we know who the best

teachers are, and also to a disappointment, to couch it in Cavellian terms, with the fact that

scientific methods seem incapable of returning that knowledge to us. In fact, Cavell reads

precisely this sort of tension as integral to Wittgenstein's work in differentiating

philosophical from scientific problems:

> Philosophical here contrasts, as explicitly in Wittgenstein's *Philosophical Investigations*, with the scientific, on the ground that philosophy does not seek to tell us anything new but rather to understand what human beings cannot on the whole simply not already know. Yet we are shown repeatedly in the *Investigations* that one cannot tell another something unless it is news to that other. (Cavell, 2010, p. 204).

Duncan's response, seen in this light, amounts to a denial insofar as recourse to a

universalizable sense of good teaching represents an instance of "undermin[ing] ourselves in

the name of an, as it were, unachievable or rather illusory rationality." While the 99%

statistic itself, one might say, denies conviction in the world by manifestly failing to *tell* (us)

the good teachers from bad teachers, Duncan's response denies the crucial importance of,

and differences among, each particular *us* who require the telling in the first place.

Cavell presents a paradox in such telling: that in order to be told, something must be "news" to someone else, but that it is simultaneously something that no one could *fail* to know. This paradox expresses the limitations of the teacher evaluation protocols in Race to the Top policies. The teacher "effectiveness" measures will either be inaccurate (thanks to statistical error of one kind or another, to say nothing of conceptual distortion) or, in the best case, they will be beside the point, expressing *merely* what is already obvious to people on the ground.

Schwab's deliberation, on the other hand, offers a means of simultaneously establishing this Cavellian "conviction in the world" by voicing the obvious, and yet managing to *tell* someone something that counts as "news." In regarding the evaluation or (ontological) measuring of teacher quality as a matter subject to deliberation, one might come to see the deliberative process as, rather than arriving at *knowledge* of something heretofore mysterious (whether or not this teacher is good), coming "to *understand* something that is already in plain view" (the particular quality of this particular teacher).

Deliberation, on this view, amounts to providing *news* to someone else – despite the fact that the person in question could not fail to *know* it already – insofar as the discussion that Schwab imagines involves an assembly or organization of the facts according to *differing* perspectives on the matter. Cavell has something like this in mind, I think, when he says:

> I wanted Wittgenstein's speaking of a task of philosophy as "assembling reminders" to participate in some adjacent register of calling attention to something repressed in the present. We need to understand what we know, to take on our education. (Cavell, 2010).

"Something repressed in the present," while redolent of Freud, is hardly the metaphysical structure or realm of ideas imagined in the rationalist assumption; rather, since every way of seeing, every standpoint, "both reveals and conceals," as Bruce Uhrmacher is fond of saying, the repression of something in the present is internal to seeing as such. But the particular repressed something is peculiar to one's standpoint with respect to, say, the particular teacher in question. In conversation with other commonplaces around the purpose of evaluating teachers, one finds oneself in the position to *tell* what is obvious to oneself, which will simultaneously be something that to varying degrees is *less than* obvious to the others present. Schwab's deliberation, in this way, offers a way of "assembling what we have long been familiar with" as a route to "*understanding* what is already in plain view."

CONCLUSION: COURAGE, CONVICTION, EVALUATION

Central to the discussion throughout Chapter 3, surfacing occasionally to demand a response only to find itself deferred or postponed, was the question of *courage* in relation to the project of evaluating teachers. Duncan proposes that all other elements necessary to the endeavor of reporting on teacher quality are in place, and now:

> The biggest barrier, the only remaining barrier in my mind is whether we have the courage. It takes courage to expose our weaknesses with a truly transparent data system. It takes courage to admit our flaws and take steps to address them (Duncan, 2009b)

The courage to which Duncan refers pertains to transparency, the exposure of weakness, and the admission of flaws. The formulation obviously takes for granted that, in the absence of the resources available thanks to the use of student achievement data, the concepts of teaching and learning – as well as qualitative evaluations of the same – have been opaque, weaknesses hidden, flaws concealed. The *repeated* use (most recently in a *New York Times* editorial on Aug. 5, 2012 – two days ago, as of this page's composition) of the New Teacher Project's statistic that 99% of all teachers are rated the same is taken (correctly, understandably) as justifying this view. However, ordinary claims to the immediate visibility of educational quality offered by both Penskar and Duncan belie the suggestion that the kind of knowledge offered by recourse to student achievement data has heretofore been lacking at all. A given person with some sort of relation to schools and schooling, whether a Secretary of Education or a parent, seems to have adequate insight into the quality of the schooling that confronts them, adequate, that is, to report on it with conviction. A given person with some sort of relation to schools and schooling, in fact, cannot be imagined to *lack* the kind of knowledge Duncan proposes to provide.

It is worth considering the idea that I broached earlier: that if the 99% statistic represents one sort of denial of a responsibility to tell good teachers from bad teachers, Duncan's proposal of student achievement data simply amounts to a different form of the same denial. Though bolder, one might admit, than absolutely refusing to differentiate among teachers on the axis of quality at all, the construction of abstract and inflexible definitions of both teaching and learning for the purpose of generating ontic measures of "teacher effectiveness" to be applied universally is merely a different inflection of the refusal of responsibility to *know* and to *tell*. It is a refusal to *know* in that, beyond the (limited standpoint of the) original construction of the measures, no one makes any attempt to accommodate the manifold commonplaces upon which the educative enterprise bears. It is a still more profound failure to *tell* in that *telling* is imagined to occur without any further exertion beyond the knowing. The knowing, in this sense, *is also* the telling.

The multifarious problems, however, with the means of *knowing* something about educational quality by recourse to student achievement data make the notion that Duncan's policies amount to a *refusal* to tell more nefarious yet. The myriad inadequacies in attempting to use abstract means to grasp an understanding – and thereby ground claims – pertaining to the quality of a given teacher in a given context, as I have noted before, render the report that such achievement-data measures can offer *at best* old news, a mere reiteration of the painfully obvious. In other words, thinking back to Cavell and Wittgenstein, such a report cannot amount to telling at all. In all other cases, from worst- to merely suboptimal-, such a report also seems to fail to measure up to the notion of "telling," and in revealing ways.

Where the official rating of a given teacher as expressed via teacher value-added conflicts with the informal ratings of that teacher's colleagues, of that teacher's students, of those students' parents, of local administrators, and so on, where is the recourse to some sort of relief? How *might* that disagreement be settled, *if* it were the case that such disagreement were even taken to require settling (which, as the Tyre position suggests, is not at all a foregone conclusion)? Can the verdict offered by such sanitized and absolute means be appealed? How would such an appeal differ from merely running the numbers again? In cases where measures of teacher value-added conflict with the "plain view" knowledge of those on the ground, one might note that the report will *certainly* come as news to those involved. But to whom does the teacher in question then reply? To whom does he or she defend herself? *Who* would we say has *done* the telling? Those familiar with his or her teaching practice already know her quality; those with whom the teacher might speak on an everyday basis are not those assailing the teacher's quality.

In point of fact, *no one* is impugning the teacher's quality, as evidenced by the fact that there is no one to respond *to*. Once the equations are established and the data entered, the process, from knowing to telling, is entirely impersonal. In an important sense, Duncan's recourse to achievement data cannot count as *telling* anyone anything because *no one* is doing the telling. The measures of teacher value-added represent *no one's* view; it is the famous view from nowhere, which Duncan seems to confuse with the (equally nonexistent) view from everywhere. A version of telling in which such recourse to an abstract methodology is imagined to simply render telling unnecessary suddenly appears as mere dogmatism or obstinacy, as a withdrawal from the body politic, a refusal to engage with others.

Precisely because the "news" of a teacher's quality as expressed in measures of teacher value-added (or according to any other distillation of student achievement data) comes from *no one* – there is no one to whom one might plead or defend oneself, no one whom to refute; no one bears responsibility for the judgment – it is *exceedingly* difficult to see the association of such uses of data with Duncan's appeal to courage. In the first place, thinking of the example above, in which verdict expressed in teacher value-added conflicts with the lived experience of a given teacher's quality, it is particularly onerous to see anything first-person at all about the exposure of "our weaknesses" and "our" flaws. In these conflict situations, teacher value-added would clearly be pointing to *your* weaknesses and *your* flaws. It cannot be first-person plural, as there is *no one* doing the telling. Further, as *no one* must finally answer for the proffered telling, the issuing of the verdict that has generated the conflict in question, it is still more difficult to see how the implementation of such a system would require any particular *bravery*. The use of teacher value-added is, as I indicated earlier, certainly bolder than refusing entirely to distinguish among teachers at all, but this boldness more closely resembles something like *chutzpah* than courage.[55]

A view of telling that Cavell and Schwab share, meanwhile, does require precisely the sort of courage that Duncan sees as necessary, the sort of courage associated with exposure and transparency. This picture of what telling entails implies, first of all, the

---

[55] For policy-makers to rely in conflict situations on the word (as it were) of a statistical analysis of something as complex as teaching and learning, something absolutely resistant to modeling, and to thereby suggest that they, the policy-makers, know the truth of the matter better than any of the many stakeholders familiar with the teacher in question and his or her practice, though? That's something more than chutzpah.

necessary finitude of any given standpoint or perspective with respect to knowing, which is to say, a human limit with regard to knowledge, perhaps especially in human affairs. Race to the Top's policies, even in states like Tennessee and Delaware, as well as in Duncan's rhetoric, in fact express a desire to recognize such human finitude: thence the call for "multiple measures" of educational quality. That these measures recognizably fail to *be* multiple in deed as they are in word – reducible, as they invariably turn out to be, to the lifting of test scores[56] – speaks to a certain discomfort, as I take it, with the groundlessness implied in actual multiplicity. While expressing the requisite appreciation for the need for multiple measures, and thus the finitude of any single perspective, Race to the Top's policies nonetheless fall back upon a view of knowing in which something like *real*, impartial, overarching truth is available, and the accuracy of this assertion is nowhere so visible as in the restricted sense of telling highlighted in Chapter 3.

The ways in which Duncan's advocacy for the use of student achievement data seeks to present these methods of knowing as both different from and superior to the means formerly available to evaluators claims a certain pride of place for the judgments issued according to these metrics. The *working* of the metrics thus arrives *simultaneously* at knowledge and its dissemination, as it were. But as we have seen, these means of judging teachers are no less fallible, nor any more comprehensive, than their predecessors; the use of teacher value-added thus cannot claim the pride of place that would obviate the need for the

---

[56] Some states have begun to survey colleagues, parents, and teachers in order to draw in multiple views of a given teacher's quality, which certainly seems like a step in the right direction. The *questions* asked on the survey, however, and their *derivation*, will prove decisive in this regard. We have already seen the way in which the salutary practice of classroom observation can become merely an ersatz form of accessing that which is already measured according to student achievement data.

difficult work of telling; and for that reason, any corrective to Race to the Top's reforms with respect to teacher quality must include a more substantial and robust view of telling, one that includes the arduous work of balancing multiple and competing perspectives.

Schwab's deliberation provides a model of a forum in which such balancing work might take place. By explicitly starting with the *local* context, with the colleagues, parents, administrators, and members of the community with and within which a given teacher works, Schwab envisions a collection of the parts that imply a (non-additive) effort toward the whole, a convening which in the act itself acknowledges the finitude of each part. The gravity of the term "deliberation" – which is not idle chatter, or prattle, or mere talking – expresses instead, I think, what Cavell has in mind when he speaks of "serious communication" (Cavell, 1996, p. 45). But Schwab's deliberation is also designed, like Duncan's appeal to student achievement data, to result in a particular *saying*, a particular curricular decision; or, as I envision its adaptation, a particular evaluation or measure of a particular teacher's quality. In this sense, like teacher value-added, deliberation provides a *measure* of teacher quality that is, in the end, communicable to a wider public beyond the deliberative context. *Also* like Duncan's favored metrics, however, as Schwab himself notes, this resulting communique will be, in its *own* right, incomplete:

> Curricular purposes, and reasons for them, must be communicated by language, by formulation. Such formulations will inevitably fall short of encompassing the full meanings and real intentions of the parties to the curricular deliberation. . . . These meanings lie in the whole course of the deliberations which created them. The meanings lie as much in what was decided against as in what was decided for. They lie in the reasons for rejection of alternatives as much as in the reasons for preferring those which are preferred. They lie in nuances of expression in the course of the deliberation. These are meanings which are impossible to encompass in a formulation to be read and acted upon by individuals who were not privy to all the deliberation and become related to it only later through a terminal formulation of its chosen purposes and reasons (Schwab, 1973, pp. 505-506).

That Schwab's deliberation renders something tangible, something final, and yet something also incomplete by virtue of what Schwab notes is "impossible to encompass in a formulation to be read and acted upon by individuals who were not privy to all of the deliberation and become related to it only later through a terminal formulation" recommends it as equivalent, in this sense, to the teacher value-added metric that Duncan prizes. Both are incomplete expressions or final formulations of a complex process.

However, two crucial factors distinguish Schwab's deliberation from Race to the Top's protocols with respect to its view of telling. Firstly, and perhaps the less important of the two (a point that I find myself willing to raise, but unwilling to spend much time arguing), Schwab's deliberation is open (or "transparent") about its own limitations, the necessity of losing the "meanings" that arose in deliberation in the process of formulating something final. One might note that Duncan also considers the translation of the "complex, nuanced work of teaching" to "a single test score" to be a "reduction." But this would amount to an *admission* of incompleteness only if the multiplicity of the "multiple measures" by which he would augment this singular measure were genuinely multiple, and they, as we have seen, are not.

Secondly, and of decidedly greater importance, the results of Schwab's deliberative committee are issued *by* the local and deliberative committee, which means that in this case, there is *someone* who in fact *does* the telling to parents, teachers, students, administrators and the community in general. This fact is decisive insofar as it *actually* amounts to an exposure of "our weaknesses," of "our flaws." These weaknesses and flaws belong to an actual first-person plurality. If this seems a minor factor, I argue that the importance of including the

someone in the notion of saying something, particularly something expressing a judgment of value, in fact amounts to a radically different view of what *telling* consists in, and that this, like Schwab's deliberation, qualifies as what Cavell means by "serious communication."

In reflecting lately on his struggle to arrive at what he would eventually call the concept of "acknowledgement," Cavell elucidates the difference between what one might call *mere* communication and "serious communication":

> I had little idea how to clarify my intuition that Austin's idea of the claim of knowledge as "going beyond" the cognitive accomplishments in being or making sure and certain was not to be modeled on the act of promising as going beyond expressing an intention, but was a separate interpretation of excess, say, of my stance toward my accomplishment, my stake in it, expressing authority toward it (Cavell, 2010).

Cavell insists on the distinction between proffering a *claim* of knowledge and the "cognitive accomplishment in being or making sure and certain," a distinction, I argue, that Duncan and the reform movement quite generally fail to see. The bare availability of teacher value-added, once rendered, obviates the need for anyone in particular taking any type of stance toward what the number or metric expresses, or owning up to a stake in it; the claim to authority, while present or implied, is *impersonal*, an *offloading* of responsibility rather than its taking-on. In other words, the metric is simply taken to *have* authority, and thus capable of *bestowing* authority on those who publish effectiveness ratings. No one needs, explicitly, to *give* or *express* authority toward it.

A sort of *telling* in which one expresses authority toward what is told, in which one takes a stance and holds a stake, is precisely the kind of telling that requires *courage*. Such a telling exposes one's weaknesses, one's flaws, one's incompletenesses. Such a telling invites and indeed requires other viewpoints, the actuality of multiple measures. Such a telling

might be said, without any distortion at all, to *invite* a productive response, and this stands in

diametric opposition to the use of effectiveness ratings in teacher evaluations as proposed by

Race to the Top. Schwab's deliberation, the acknowledgement of the need for the *multiple*

commonplaces and the centrality of trans-human *conversation* among them, embodies the

possibilities afforded by such telling. The flaws, weaknesses, and incompleteness of *each*

commonplace are voiced, accounted for, and responded to – along with, one must add, the

unique and productive view of the multifarious practical situation. The eventual

pronouncement of a judgment on teacher quality, the resulting evaluation, would then

express a process of "settling judgments," not by means of an *in*human or purportedly

objective methodology, but rather by *admittedly* fallible and incomplete human beings.

Perhaps strangely, this ought not to be taken as expressing a shortcoming in terms of human

knowledge, but as expressing rather something about the knowable *nature* of teaching and

learning itself, something I take Wittgenstein to allude to in §208 of the *Investigations:* "The

fact that we cannot write down all the digits of π is not a human shortcoming, as

mathematicians sometimes think." By virtue of the admission of finitude that such "settling

judgments" embodies, and by virtue of staking one's personal and human authority in a

judgment, such a process renders the resulting "exposure" legitimately "ours." Such a mode

of judgment neither attempts nor pretends to be, in contrast to the use of teacher value-added,

objectively final, whatever that might mean. Cavell's thoughts on the matter sum it up

nicely:

> The establishing of criteria makes the process of judging more convenient,
> more open, less private or arbitrary. One might say: here establishing criteria allows
> us to *settle* judgments publicly, not exactly by making them certain, but by declaring
> what the points are at issue in various judgments, and then making them *final* (on a
> given occasion). That is a practice worth having; human decisions cannot wait upon

certainty. But it is therefore one which can be abused. In assuming the burden of finality in the absence of certainty, an authority stakes the virtue of its community: if its judgments are not accepted as scrupulously fair, in its criteria and its application of criteria, the community is shown to that extent not to provide a secure human habitation for its members; it fails to take up the slack between the uncertainty of judgment and the finality of decision (Cavell, 1979, p. 31).

The teacher evaluation protocols advocated in Race to the Top express *precisely* the problems to which Cavell offers the "establishing of criteria" as a response, and to which, it must be said, the deliberation that Schwab proposes also ameliorates. The esoteric statistical manipulations, comprehensible only to advanced psychometricians, that generate teacher value-added render the measures themselves effectively "private," the criteria "arbitrary," and the consequent judgments therefore equally private, equally arbitrary.

In fact, Duncan's reliance on student achievement data, the view in which such data represents something radically new and unquestionably valuable to the prospect of evaluating teachers, represents, in this light, something like the *opposite* of courage. As we have seen with the indicators that Chetty's research relies upon, the correlations that justify the use of teacher value-added in the evaluations of teachers rest upon nothing less frail than agreements in criteria. But teacher value-added is not taken as an expression of any "uncertainty of judgment"; quite the contrary. The use of achievement data represents a *shunning* of responsibility for the "burden of finality": its value is taken to lie in its *in*human objectivity, its ability to *provide* certainty on which to offer judgments of teachers. In the way in which all "multiple measures" are pegged to student achievement data, its use indicates precisely a retreat from responsibility to all communities and their members; it is a retreat from the trans-human; no one stakes authority or risks exposure in communicating the results of the calculations. For all that, it still fails to *escape* the trans-human, relying, as we

have endlessly seen, on the same agreement in forms of life as the methods to which it considers itself superior. It is thus something akin to a bumbling burglar caught red-handed, honorable in neither intention nor execution.

Nonetheless, Duncan is not wrong when he insists, first, on the need to get serious about teacher evaluation, and second, on this endeavor's requiring of courage. The 99% statistic indicates that, whatever else is taken to be entailed here, distinguishing among teachers in terms of their quality *as* teachers has not been taken as seriously as the obvious variation among teachers demands. That the means of *knowing* good teaching upon which Duncan's Race to the Top relies prove no less able to escape the vagaries of the trans-human or agreements in forms of life, as demonstrated in both its conceptual limitations in Chapter 1 and its practical limitations in Chapter 2, indicates that it cannot serve either as a singular measure itself, nor (what amounts to the same thing) as the ontic ground of the various "multiple" measures. Rather, the criteria expressed in growth in student test scores ought to be brought into *conversation* with other forms of criteria in a public or shared protocol for settling judgments, something Schwab models with his notion of deliberation.

Recalling Strong, et al.'s findings that, despite pervasive agreements among judges, these human judges all disagreed with the verdicts of teacher value-added, one might see the judges as, contra Strong, expressing different "concerns, values, and operations" rather than being *simply* "inaccurate," as Strong, et al. has it. Seen this way, each party to this disagreement speaks for and as a mere part of a whole, neither rightfully the occupier of any superior place and each in need of the other's augmentation. Both, one might add, have a *view* of the whole of which they are part (this is entailed in *seeing* themselves as a part of a whole), but these views differ; this is what, in fact, *requires* conversation.

Throughout his thinking on Wittgenstein's view of philosophy, Cavell pictures the philosophizing, as it is embodied in "serious communication," perhaps with Luther more or less explicitly in mind, as requiring "conversion," a perpetual "turning-toward" another. Just as Dewey highlighted the etymological relation between "common," "communication," and "community," one might point out at that "conversion" and the type of "conversation" embodied in Schwab's deliberation share close etymological ties. In precisely the way that Schwab's deliberation envisions the type of telling involved as expressing *simultaneously* the fullness of a unique claim to certain "concerns, values, and methods" and also the consciousness of the kind of lack that *calls for* the unique claims of others to different "concerns, values, and methods," Cavell's "serious communication" pictures "telling" as expressing a request for a response as much as what one might call a *position*. Cavell thus notes that "[the philosophical impulse's] signature mode of criticism is one that is simultaneously turned toward oneself and toward one's contribution to the communal" (Cavell, 2012, p. 213). The simultaneous turning toward oneself and one's relation to the communal, I suggest, embodies the very process of Schwab's deliberation.

The crowning problem with the rationalist assumption, upon which the justificatory claims leveled on behalf of student achievement data are founded, has to do with, as I have noted before, the unwarranted freezing of relevance by virtue of ossifying either the whole in order to understand the part or vice versa, a freezing that renders the result both insufficient in the present and in principle closed to the future. The solution, as Dreyfus has it, is to be "always-already-in-a-situation without representing it" (H. L. Dreyfus, 1992, p. 55), which amounts to what Glendinning, quoting Derrida, prescribes: "instead of acceding to the 'craving for generality,' we must learn to 'think *at once* the rule and the event'"(Glendinning,

1998, p. 107). The thinking *at once* the rule and the event is hardly as exotically postmodern

as it may seem: Dewey points out almost the same thing in 1929:

> But it does mean that educationally speaking, the problem of attendant radiations, expansions, and contractions, are in the end more important, and that it is dangerous to take the part for the whole. Nor is it satisfactory to say that the part must be mastered before the whole can be attacked. For, by the nature of the case, the whole enters into the part, that is, it is a determining factor in the *way* in which one learns to read (Dewey, 1929, p. 64).

Cavell's "serious communication" or "turning toward" and Schwab's deliberation

each manifest, because presupposing, the thinking *at once* of the part and the whole in

unstable superposition – this view accounts for the simultaneous and paradoxical fact that

deliberation and serious communication both express one's own stake, one's authority, with

respect to the community and simultaneously express the *need* for response, augmentation,

acknowledgement.

Precisely this paradoxical situation calls for *courage*. Cavell has it right when he

says, in imagining the position of something like the representative of one of the

commonplaces in Schwab's deliberative model, "The anxiety in teaching, in serious

communication, is that I myself require education" (Cavell, 1996, p. 45). Cavell's name for

this simultaneous expression of teaching and the need for learning is simply "philosophy."

He explicitly considers the relation of the part to the whole in expressing the nature of this

sense of philosophizing:

> In philosophizing, I have to bring my own language and life into imagination. What I require is a convening of my culture's criteria, in order to confront them with my words and my life as I pursue them and as I may imagine them; and at the same time to confront my words and life as I pursue them with the life my culture's words may imagine for me: to confront the culture with itself, along the lines in which it meets in me.
> This seems to me a task that warrants the name of philosophy. It is also the description of something we might call education (Cavell, 1996, p. 45)

Joseph Schwab's deliberation, taking place locally, among representatives of the range of the disparate commonplaces involved with one another, offers the possibility of restoring the *seriousness* to the serious task of knowing, doing, and telling anything with regard to the processes and contexts of education and their evaluation. This seriousness, I have argued, is merely dodged or deferred in abdicating to the machinations of statistical analysis predicated upon frozen and fallacious *indicators* of flexible and full *concepts*, which, by virtue of claiming sole pride of place in the judgment of an expansive and trans-human undertaking, in effect absolve individuals and the community of all responsibility in the matter, a fact that a fortiori renders the proceedings decidedly and manifestly *non*-serious, although the consequences are anything but.

Schwab's proposal that in *doing* things relative to the educative endeavor, varying viewpoints must be brought into direct contact with one another, in an exchange that Cavell might call the "convening of criteria," represents a radical departure from the impersonal and, in the multitudinous senses of the word, *irresponsible* use of student achievement data in undertaking these same efforts. Schwab's deliberation thus acknowledges the necessity of conversation, of turning toward the members of one's community, around an issue of common Heideggerian concern. The resulting conversation, this deliberation, might then be seen as the response to the need expressed in this concern. Such a conversation around such needs is precisely what Cavell calls "philosophy" itself. The inherent connection that Cavell perceives between philosophy and education should be lost on no one.

I suggest that we think of Schwab's model of deliberation as the appropriate means for intervening in educational processes from both (and at once) within and without. The

convening of criteria, the expressing of a conviction that exposes one's own inadequacies, requires courage. Genuine exchange, serious communication, stakes claims to authority and acknowledges weaknesses at once. Such a means of settling and finalizing judgments, in which individual and collective responsibility for evaluating teachers goes along with the judgments themselves, is, as Cavell says, "a practice worth having." In fact, one might think specifically of Schwab's deliberation when, in the double implication of the genitive for which he is famous, Cavell declares that, "In this light, philosophy becomes the education of grownups" (Cavell, 1996, p. 45).

REFERENCES

Anderson, Jenny. (2012, February 20). States Try to Fix Quirks in Teacher Evaluations, *New York Times,* p. A1.

Andersson, Christian, Johansson, Per, & Waldenström, Nina. (2011). Do you want your child to have a certified teacher? *Economics of Education Review, 30*(1), 65-78. doi: 10.1016/j.econedurev.2010.07.003

Baker, Eva L., Barton, Paul E., Darling-Hammond, Linda, Haertel, Edw ard, Ladd, Helen F., Linn, Robert L., . . . Shepard, Lorrie A. (2010). Problems with the Use of Student Test Scores to Evaluate Teachers. *EPI Briefing Paper #278*.

Berliner, David C., & Biddle, Bruce J. (1997). *The manufactured crisis : myths, fraud, and the attack on America's public schools* (College ed.). White Plains, NY: Longman Publishers USA.

Bound, John, Hershbein, Brad, & Long, Bridget Terry. (2009). Playing the Admissions Game: Student Reactions to Increasing College Competition. *The Journal of Economic Perspectives, 23*(4), 119-146.

Bourdieu, Pierre. (1980). *The Logic of Practice*. Stanford: Stanford University Press.

Brill, Steven. (2009, August 31). The Rubber Room. *The New Yorker*.

Brooks, David. (2008, December 5). Who Will He Choose, *New York Times,* p. A39.

Brügelmann, Hans. (2011, January 13). Pisa macht die Schulen nich besser, *Die Zeit*. Retrieved from http://www.zeit.de/2011/03/C-Pisa-Kritik

Cahill, Kevin. (2011). *The Fate of Wonder: Wittgenstein's Critique of Metaphysics and Modernity*. New York: Columbia University Press.

Cannell, John Jacob. (1988). Nationally Normed Elementary Achievement Testing in America's Public Schools: How All 50 States Are Above the National Average. *Educational Measurement: Issues and Practice, 7*(2), 5-9. doi: 10.1111/j.1745-3992.1988.tb00424.x

Cavell, Stanley. (1979). *The Claim of Reason*. New York: Oxford University Press.

Cavell, Stanley. (1996). The Normal and the Natural. In Stephen Mulhall (Ed.), *The Cavell Reader*. New York: Blackwell.

Cavell, Stanley. (2010). *Little Did I Know: Excerpts from Memory*. Stanford, CA: Stanford University Press.

Cavell, Stanley. (2012). Philosophy and Education. In Naoko Saito & Paul Standish (Eds.), *Stanley Cavell & the Education of Grownups*. New York: Fordham University Press.

Census. (2012). *Bachelor's Degree Attainment Tops 30 Percent for the First Time, Census Bureau Reports*. Retrieved from http://www.census.gov/newsroom/releases/archives/education/cb12-33.html.

Chetty, Raj, Friedman, John N., & Rockoff, Jonah E. (2011). The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood. *NBER Working Paper No. 17699*.

Chetty, Raj, Friedman, John N., & Rockoff, Jonah E. (2012). Great Teaching: Measuring Its Effects on Students' Future Earnings. *Education Next, 12*(3), 58-64.

Chingos, Matthew M., & Peterson, Paul E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review, 30*(3), 449-465. doi: 10.1016/j.econedurev.2010.12.010

Craig, Aimee. (2011, October 12). US Secretary of Education Arne Duncan Meets with CLASS Project to Hear Progress Report Retrieved March 5, 2012, from http://chalkboardproject.org/press-room/us-secretary-of-education-arne-duncan-meets-with-class-project-educators-to-hear-progress-report-on-24-4-million-federal-investment-in-oregon-teachers-and-students/

Crease, Robert P. (2011). *World in the Balance*. New York: W. W. Norton & Company.

DeMaio, Joseph, Parkinson, Stanley, Leshowitz, Barry, Crosby, John, & Thorpe, Jack A. (1975). *Visual Scanning: Comparisons between Student and Instructor Pilots*.

Dewey, John. (1929). *The Sources of a Science of Education*. New York: Horace Liveright.

Dillon, Sam. (2010, December 4). Teacher Ratings Get New Look, Pushed by a Rich Watcher, *New York Times,* p. A1.

DOE. (2011). Obama Administration Sets High Bar for Flexibility from No Child Left Behind in Order to Advance Equity and Support Reform, from http://www.ed.gov/news/press-releases/obama-administration-sets-high-bar-flexibility-no-child-left-behind-order-advanc

Dreyfus, Hubert. (1967). Alchemy and Artificial Intelligence. *Rand, Paper P3244*.

Dreyfus, Hubert, & Dreyfus, Stuart. (1992). What Artificial Experts Can and Cannot Do. *AI & Society, 6*(1), 18-26.

Dreyfus, Hubert L. (1988). The socratic and platonic basis of cognitivism. *AI & Society, 2*(2), 99-112. doi: 10.1007/bf01891374

Dreyfus, Hubert L. (1991). *Being-in-the-world : a commentary on Heidegger's Being and time, division I*. Cambridge, Mass.: MIT Press.

Dreyfus, Hubert L. (1992). *What Computers Still Can't Do: a Critique of Artificial Reason*. Cambridge, MA: M.I.T. Press.

Dreyfus, Hubert L., & Dreyfus, Stuart. (1986). *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.

Dreyfus, Hubert L., & Dreyfus, Stuart. (2005). Peripheral Vision: Expertise in Real World Contexts. *Organization Studies, 26*(1), 776.

Duncan, Arne. (2009a). *Partners in Reform*. Retrieved from http://www2.ed.gov/news/speeches/2009/07/07022009.html.

Duncan, Arne. (2009b). *Robust Data Gives Us the Roadmap to Reform*. Retrieved from http://www2.ed.gov/news/speeches/2009/06/06082009.html.

Duncan, Arne. (2009c). *States Will Lead the Way Toward Reform*. Retrieved from http://www.ed.gov/news/speeches/states-will-lead-way-toward-reform.

Duncan, Arne. (2009d). *Teacher Preparation: Reforming the Uncertain Profession*. Retrieved from http://www.ed.gov/news/speeches/teacher-preparation-reforming-uncertain-profession.

Duncan, Arne. (2009e). *Turning Around the Bottom Five Percent*. Retrieved from http://www.ed.gov/news/speeches/turning-around-bottom-five-percent.

Duncan, Arne. (2012a). *Statement*. Retrieved from http://www.ed.gov/news/press-releases/statement-us-secretary-education-arne-duncan-1.

Duncan, Arne. (2012b). *Teachers Get R-E-S-P-E-C-T*. Retrieved from http://www.ed.gov/news/speeches/teachers-get-r-e-s-p-e-c-t.

Eisner, Elliot. (1984). No Easy Answers: Joseph Schwab's Contributions to Curriculum. *Curriculum Inquiry, 14*(2), 201-210.

Elmore, R. F. (1996). Getting to Scale with Good Educational Practices. *Harvard Educational Review, 66*(1), 1-26.

Ercikan, Kadriye, & Koh, Kim. (2005). Examining the Construct Comparability of the English and French Versions of TIMSS. *International Journal of Testing, 5*(1), 23-35. doi: 10.1207/s15327574ijt0501_3

Ercikan, Kadriye, & McCreith, Tanya. (2002). Effects of Adaptations on Comparability of Test Items and Test Scores

Secondary Analysis of the TIMSS Data. In David F. Robitaille & Albert E. Beaton (Eds.), (pp. 391-405): Springer Netherlands.

Evans, Rodney. (2007). Existing Practice Is Not the Template. *Educational Researcher, 36*(9), 553-559.

Fehr, Mario. (2010, June 3). Schweizer Berufsbildung -- wirksam gegen die Krise, *Neue Zurcher Zeitung*. Retrieved from http://www.nzz.ch/nachrichten/startseite/schweizer_berufsbildung__wirksam_gegen_die_krise_1.5865921.html

Fogelin, Robert J. (2009). *Taking Wittgenstein at his word : a textual study*. Princeton, N.J.: Princeton University Press.

Gabriel, Trip. (2011, October 9). G.O.P. Candidates Take an Anti-Federal Stance, *New York Times,* p. A28.

Gardner, David Pierpont. (1983). *A Nation At Risk*. Retrieved from http://www2.ed.gov/pubs/NatAtRisk/index.html.

Glazerman, Steven, Goldhaber, Dan, Loeb, Susanna, Raudenbush, Stephen, Staiger, Douglas, Whitehurst, Grover J., & Croft, Michelle. (2011). Passing Muster: Evaluating Teacher Evaluation Systems (Brown Center Task Group on Teacher Quality, Trans.): Brookings Institution.

Glendinning, Simon. (1998). *On Being with Others*. New York: Routledge.

Glennan, Thomas K., Bodilly, Susan J., Galegher, Jolene R., & Kerr, Kerri A. (Eds.). (2004). *Expanding the Reach of Education Reforms*. Santa Monica, CA: RAND Corporation.

Godofsky, Jessica, Zukin, Cliff, & Horn, Carl Van. (2011). Unfulfilled Expectations: Revent College Graduates Struggle in a Troubled Economy: John J. Heldrich Center for Workforce Development.

Goggin, M. L., O'M Bowman, A., Lester, J. P., & O'Toole, L. J. Jr. (1990). *Implementing Theory and Practice Toward a Third Generation*. Glenview, Il: Scott, Foresman/Little, Brown.

Green, Elizabeth. (2010, March 2). Building a Better Teacher, *New York Times Magazine*.

Healy, F. Henry, & DeStefano, Joseph. (1997). *Education Reform Support*. Education Resources Information Center.

Heidegger, Martin. (1962). *Being and Time* (John MacQuarrie & Edward Robinson, Trans.). New York: Harper & Row.

Heidegger, Martin. (1971). *On the Way to Language* (Peter Hertz, Trans.). New York: Harper and Row.

Heidegger, Martin. (1982). *The Basic Problems of Phenomenology* (Albert Hofstadter, Trans.). Bloomington, IN: Indiana University Press.

Hill, Heather, & Ball, Deborah Loewenberg. (2009). The Curious — and Crucial — Case of Mathematical Knowledge for Teaching. *The Phi Delta Kappan, 91*(2), 68-71.

Hill, Heather C., Rowan, Brian, & Ball, Deborah Loewenberg. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal, 42*(2), 371-406.

Hlebowitsh, P. (2012). When best practices aren't: A Schwabian perspective on teaching. *Journal of Curriculum Studies, 44*(1), 1-12. doi: Doi 10.1080/00220272.2011.637184

Hlebowitsh, Peter S. (2005). Generational Ideas in Curriculum: A Historical Triangulation. *Curriculum Inquiry, 35*(1), 73-87.

Hlebowitsh, Peter S. (a gazillion different occasions).

Hollabaugh, Jaliene. (2011). Review: Teach Like A Champion. *International Journal of Instruction, 4*(1).

Kane, Thomas J., Rockoff, Jonah E., & Staiger, Douglas O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review, 27*(6), 615-631. doi: 10.1016/j.econedurev.2007.05.005

Kane, Thomas J., Taylor, Eric S., Tyler, John H., & Wooten, Amy L. (2011a). Evaluating Teacher Effectiveness. [Feature Article]. *Education Next, 11*(3), 54-60.

Kane, Thomas J., Taylor, Eric S., Tyler, John H., & Wooten, Amy L. (2011b). Identifying Effective Classroom Practices Using Student Achievement Data. *Journal of Human Resources, 46*(3), 587-613.

Keller, Bess. (2010). Content Seen Lacking Specificity. [Article]. *Education Week, 30*(11), S17-S19.

Kenaw, Setargew. (2008). Hubert L. Dreyfus's Critique of Classical AI and its Rationalist Assumptions. *Minds and Machines, 18*(2), 227-238. doi: 10.1007/s11023-008-9093-7

Kisiel, Theodore. (1993). *The Genesis of Heidegger's Being and Time*. Berkeley, CA: University of California.

Koretz, Daniel. (2008). *Measuring Up: What Educational Testing Really Means*. Cambridge, MA: Harvard UP.

Koretz, Daniel M. (2002). Limitations in the Use of Achievement Tests as Measures of Educators' Productivity. *The Journal of Human Resources, 37*(4), 752-777.

Kristof, Nicholas. (2007, May 1). Gold Stars and Dunce Caps, *New York Times,* p. A23.

Kristof, Nicholas. (2009, February 15). Our Greatest National Shame, *New York Times,* p. WK11.

Kristof, Nicholas. (2012a, January 22). How Mrs. Grady Transformed Olly Neal, *New York Times,* p. SR13.

Kristof, Nicholas. (2012b, January 12). The Value of Teachers, *New York Times,* p. A27.

Laurillar, Diana. (2002). Rethinking Teaching for the Knowledge Society. *EDUCAUSE review, 37*(1), 16-25.

Lee, Lois. (2012, March 12). Admit Rate Holds Steady at 12.3 Percent, *Daily Pennsylvanian*.

Lee, Okhee, & Luykx, Aurolyn. (2005). Dilemmas in Scaling up Innovations in Elementary Science Instruction with Nonmainstream Students. *American Educational Research Journal, 42*(3), 411-438.

Lemov, Doug. (2010). *Teach Like a Champion*. San Fancisco: John Wiley & Sons.

Lenat, Douglas. (2008). The Voice of the Turtle: Whatever Happened to AI? *AI Magazine, 29*(2), 11-23.

Lenat, Douglas, & Feigenbaum, Edward. (1991). On the Thresholds of Knowledge. *Artificial Intelligence, 47*(1-3).

Lenat, Douglas, & Guha, R. V. (1990). *Building Large Knowledge-Based Systems*. Reading, Mass: Addison-Wesley.

Leonhardt, David. (2010, July 28). The Case for $320,000 Kindergarten Teachers, *New York Times,* p. A1.

Levin, Jessica, Mulhern, Jennifer, & Schunck, Joan. (2005). Unintended Consequences: The Case for Reforming the Staffing Rules in Urban Teacher Union Contracts. *ERIC Clearinghouse*. Retrieved from http://www.eric.ed.gov/PDFS/ED515654.pdf

Levy, Jonathan. (2012). *Freaks of Fortune: The Emerging World of Capitalism and Risk in America*. Cambridge, MA: Harvard University Press.

Linn, Robert L., Graue, M. Elizabeth, & Sanders, Nancy M. (1990). Comparing State and District Test Results to National Norms: The Validity of Claims That "Everyone Is Above Average". *Educational Measurement: Issues and Practice, 9*(3), 5-14. doi: 10.1111/j.1745-3992.1990.tb00372.x

Martini, Edwin. (2012). *Agent Orange: History, Science, and the Politics of Uncertainty*. Amherst, MA: University of Massachusetts Press.

McDowell, John. (1981). Non-Cognitivism and Rule-Following. In Steven H. Holtzman (Ed.), *Wittgenstein: To Follow a Rule*: Routledge.

Medina, Jennifer. (2008, January 21). New York Measuring Teachers by Test Scores, *New York Times*.

Medina, Jennifer. (2010, April 24). Last Teacher In, First Out? City Has Another Idea, *New York Times,* p. A1.

Medvetz, Tom. (2012). *Think Tanks in America*. Chicago: University of Chicago Press.

NCTQ. (2010). Advisory Board Retrieved March 13, 2012, from http://www.nctq.org/p/about/advisory.jsp

NCTQ. (2011). State of the States: Trends and Early Lessons on Teacher Evaluation and Effectiveness Policies: National Council on Teacher Quality.

Newell, Allen, & Simon, Herbert. (1985). Computer Science as Empirical Inquiry: Symbols and Search. In John Haugeland (Ed.), *Mind Design*. Cambridge, Mass: MIT Press.

Nocera, Joe. (2012, May 22). Gates Puts the Focus on Teaching, *New York Times,* p. A27.

Obama, Barack. (2012). State of the Union Address Retrieved February 26, 2012, from http://www.whitehouse.gov/the-press-office/2012/01/24/remarks-president-state-union-address

Otterman, Sharon. (2010, December 27). Hurdles Emerge in Rising Effort to Rate Teachers, *New York Times,* p. A1.

Penskar, Carol. (2010, May 3). LETTERS; Deciding How to Grade the Teachers, *New York Times*.

Polyani, Michael. (1962). Tacit Knowing: Its Bearing on Some Problems of Philosophy. *Reviews of Modern Physics, 34*(4), 601-616.

Ravitch, Diane. (2010). *The Death and Life of the Great American School System*. New York: Basic Books.

. *Reading in the Content Area*. (2012). State of Tennessee Department of Education Retrieved from http://www.state.tn.us/education/ci/english/reading.shtml.

Ripley, Amanda. (2012). Why Kids Should Grade Teachers. *Atlantic*.

Rivkin, Steven G., Hanushek, Eric A., & Kain, John F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica, 73*(2), 417-458. doi: 10.1111/j.1468-0262.2005.00584.x

Rothstein, Richard. (2008). *Grading Education*. New York: Teachers College Press.

Schwab, Joseph J. (1959). The "Impossible" Role of the Teacher in Progressive Education. *The School Review, 67*(2), 139-159.

Schwab, Joseph J. (1969). The Practical: A Language for Curriculum. *The School Review, 78*(1), 1-23.

Schwab, Joseph J. (1971). The Practical: Arts of Eclectic. *The School Review, 79*(4), 493-542.

Schwab, Joseph J. (1973). The Practical 3: Translation into Curriculum. *The School Review, 81*(4), 501-522.

Schwab, Joseph J. (1983). The Practical 4: Something for Curriculum Professors to Do. *Curriculum Inquiry, 13*(3), 239-265.

Selinger, Evan M., & Crease, Robert P. (2002). Dreyfus on expertise: The limits of phenomenological analysis. *Continental Philosophy Review, 35*(3), 245-279. doi: 10.1023/a:1022643708111

Shepard, Lorrie A. (1990). Inflated Test Score Gains: Is the Problem Old Norms or Teaching the Test? *Educational Measurement: Issues and Practice, 9*(3), 15-22. doi: 10.1111/j.1745-3992.1990.tb00374.x

Shulman, Lee. (1986). Those Who Understand: Knowledge Growth in Teaching. *Harvard Educational Review, 57*, 1-22.

Shulman, Lee. (1987). Knowledge and Teaching: Foundations of the new reform. *Educational Researcher, 15*(2), 4-14.

Shulman, Lee S. (2007). Practical Wisdom in the Service of Professional Practice. *Educational Researcher, 36*(9), 560-563.

Sireci, Stephen G. (1997). Problems and Issues in Linking Assessments Across Languages. *Educational Measurement: Issues and Practice, 16*(1), 12-19. doi: 10.1111/j.1745-3992.1997.tb00581.x

Standish, Paul. (2001). Data Return: The Sense of the Given in Educational Research. *Journal of Philosophy of Education, 35*(3), 497-518. doi: 10.1111/1467-9752.00240

Stern, David. (1995). *Wittgenstein on Mind and Language*. New York: Oxford University Press.

Stewart, Jon (Producer). (2011, February 28th). The Daily Show. Retrieved from http://www.thedailyshow.com/watch/mon-february-28-2011/crisis-in-dairyland---angry-curds

Stewart, Jon (Producer). (2012, February 17). February 16. *The Daily Show*.

Strong, Michael, Gargani, John, & Hacifazlioğlu, Özge. (2011). Do We Know a Successful Teacher When We See One? Experiments in the Identification of Effective Teachers. *Journal of Teacher Education, 62*(4), 367-382. doi: 10.1177/0022487110390221

Taylor, Charles. (1989). *Sources of the Self*. New York: Cambridge University Press.

Taylor, Charles. (1993). To Follow a Rule. In Edward Lipuma, Moishe Postone & Craig J. Calhoun (Eds.), *Bourdieu: Critical Perspectives*. Chicago: University of Chicago Press.

Tishman, Shari, Jay, Eileen, & Perkins, David N. (1993). Teaching Thinking Dispositions: From Transmission to Enculturation. *Theory into Practice, 32*(3), 147-153.

Trigwell, Keith, Prosser, Michael, & Waterhouse, Fiona. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education, 37*(1), 57-70. doi: 10.1023/a:1003548313194

Tyre, Peg. (2011, September 18). Putting Parents in Charge, *New York Times,* p. SR8.

Vevea, Rebecca. (2011, November 13). Tough Questions on Changing Teacher Evaluations, *New York Times,* p. A29.

Welker, Robert. (1991). Expertise and the Teacher as Expert: Rethinking a Questionable Metaphor. *American Educational Research Journal, 28*(1), 19-35.

Winerip, Michael. (2012, February 27). Amid a Federal Education Inquiry, an Unsettling Sight, *New York Times,* p. A10.

Wittgenstein, Ludwig. (1958). *Preliminary studies for the "Philosophical investigations" generally known as the Blue and brown books*. New York,: Harper.

Wittgenstein, Ludwig. (1983). *Remarks on the Foundations of Mathematics*. Cambridge: MIT Press.

Wittgenstein, Ludwig. (2009). *Philosophical Investigations* (G.E.M. Anscombe & P.M.S Hacker, Trans. 4th ed.). Malden, MA: Wiley-Blackwell.