

Social justice and educational measurement: a thumbnail sketch¹

Zachary Stein

Lectica, Inc.
Harvard University Graduate School of Education

John Rawls's (1971; 1996) philosophical methods involve the use of complex *representational devices*, which can be thought of as structured models or thought experiments. More broadly, model-based reasoning in the sciences involves the deployment of a variety of “useful fictions” that simplify phenomena and exemplify the properties or processes of interest (Elgin, 1996). One of the most common kinds of models is the *miniature*, such as the scale-models used in engineering and systems biology that represent large structures or long timelines in ways that “shrink” them down so they can be seen at a glance. This paper is just such a miniature; it aims to bring a large and complex manuscript into view by shrinking it down and distilling its most important properties. This means stepping back from the particular arguments to gain a view of the whole; its a map of the forest, whereas the body of the full manuscript deals with the various trees.

The simplest way to miniaturize my overall argument is to consider the design of standardized testing infrastructures as if from behind a Rawlsian “veil of ignorance.” The “original

¹ The following is based on: Stein, Z. (2014). *Tipping the scales: social justice and educational measurement*. (doctoral dissertation). Harvard University Graduate School of Education. Cambridge, MA. Please address all correspondence to: stein.zak@gmail.com

position” is the central representational device deployed by Rawls. It is intended to clarify the objectivity and universality of the “moral point of view.” The original position is basically a set of decision-making constraints that support reasoning about the nature of justice; it simply asks us to consider the basic structures of a society as if ignorant of our eventual place in it. The archetypal case is drafting a constitution without knowledge of whom or where you will be in the society it creates. Thus it would be irrational to draft a constitution supporting slavery or limiting voting rights to landowning males, because there is no guarantee you would not end up enslaved or disenfranchised. Engaging Rawls’s thought experiment means that instead of viewing social structures from *my* perspective (i.e., that of a well-educated white male), I am forced to consider society from *everyone’s* perspective (e.g., that of a woman, of a minority, etc.). A social system that can be viewed as reasonable from this meta-perspective is one that provides justice for all. Rawls intended this thought experiment for use only in adjudicating between different philosophical principles of justice, and thus not for thinking about more specific social structures and institutions. However, for the purpose of miniaturizing the overall argument, the thought experiment serves as a valuable heuristic and allows us to cut directly to the chase.

The overarching themes of this work can be distilled into a single question: what kind of standardized testing infrastructure could be agreed to in the original position? This question is at the center of the *theory of just educational measurement* that is the overall focus of the work. Of course, there is much more to it than that. As explained shortly, a theory of just educational measurement requires related theories about the nature institutionalized measurement in general, as well as a supplemental philosophy of education, and a system of distinctions concerning the dynamics of testing-intensive educational reforms, which centers on the difference between justice-oriented testing and efficiency-oriented testing. But before adding these details to the minia-

ture being built here, it is worth exploring the simplest way of thinking about testing infrastructures and social justice.

Several issues are clarified immediately through the “ideal role taking” exercise of imagining that one could end up anywhere in the systems affected by testing infrastructures. Key stakeholder groups emerge, each with their own institutionalized relationship to testing and each containing a range of individuals (from least well off to most well off). Students and their parents are one group, and their range can be viewed both in terms of socio-economic factors and in terms of learning abilities. Teachers are another group, again including a range of individuals who vary according to their skills and socio-economic positions. Then there are the administrators at various levels within the school system (from principle to superintendent), who are also differentially positioned. Policy makers and politicians constitute another group, as do psychometricians and others representing the interests of testing companies. There are of course other stakeholders (e.g., parents, educational researchers, college admissions officers, etc.), but the heart of the argument resides with these main groups, including students and their parents, teachers, administrators, policy-makers, and test providers.

The most vulnerable individuals in the social structures created by testing are the least-well-off students and teachers (e.g., learning disabled students in an inner city school and their special education teacher); the least vulnerable are the politicians and those representing the interests of testing companies (e.g., Arne Duncan and Educational Testing Service executives). As in many cases where injustices occur, the most vulnerable—those who are potentially most seriously affected—are also the least empowered and the farthest away from influence over the systems that profoundly shape their lives. The task of building a just standardized testing infrastructure (as with any basic social structure viewed from the original position) requires taking as pri-

mary the perspectives of the least well off, because this is the social position that is of greatest concern from behind the veil of ignorance (e.g., its the place you would least like to end up in the system). Justice requires maximizing benefits to the least well off while maintaining overall fairness within the system.

Broadly speaking, this means a testing infrastructure that awards those who are already advantaged while punishing those who are already disadvantaged is unjust because it further disadvantages the least well off. I argue that the history of testing from the early IQ-testing movement to recent policies supporting test-based accountability has mostly fit this unjust pattern of differential reward and punishment.

Social justice is also implicated at the level of test design and administration. For example, objectivity and standardization are required by justice—this is the moment of truth expressed by those who tout the social justice benefits of testing. Indeed, many of the most egregious cases of injustice due to testing have involved a *pretense* of objectivity that has disguised the existence and impact of overt bias and errors in test design and scoring.

I argue that all individuals have a right to objective measurement. However, even truly objective tests, when they are used in high-stakes contexts or focus on a narrow range of constructs and item types, create injustice. Justice demands that individuals have access to measures that are *relevant* and *beneficial*.

Social justice and institutionalized measurement

A society's basic structures are those institutions and legal codes, such as its taxation mechanisms, judiciary processes, and educational system, which constitute the society as a shared social world. Individuals participate in these basic structures as a result of participating in society at all; and because these basic structures shape the lives of every person they are the sub-

ject of theories of justice. Measurement infrastructures are basic structures. In fact, measurement practices were some of the first basic structures to be perceived in terms of their impact on social justice. From the systems of scales and bushels used in ancient marketplaces to the precisely calibrated instruments used in modern science, measurement has always been a crucial component of shared social life, literally structuring and facilitating mutual understanding and consensus. Because measurement infrastructures are basic structures Rawls has lessons to teach about the way they ought to be designed.

Rawls's principles of justice state that a society's basic structures should be designed so that all individuals are granted the same rights and freedoms, all inequalities result from conditions of fair opportunity, and the distribution of benefits always advances the position of the least well off. The lesson of the *unjust miller*—who changes the size of the village bushel by *fiat* to serve his own needs—teaches that *objective* measurement is a prerequisite for social justice. The creation of the metric system in Revolutionary France, which corrected this kind of injustice, was as much an ethical undertaking as a scientific one.

But objectivity is not the only key to a just measurement infrastructure. Tracing the spread of the metric system and the advance of large-scale measurement-intensive organizational structures reveals that objectivity and its accouterments (e.g., scientific technologies and expertise) can create injustices. The lesson of the *unjust bureaucracy*—which unilaterally imposes objective measures that reshape the practices of individuals and groups—teaches that measurement infrastructures must also be *relevant* and *beneficial* to those who are most affected by them. We now have three principles of just institutionalized measurement, all of which follow from Rawls's principles of justice: all individuals have a right (1) to objective measurement whenever

possible, (2) to measures that are relevant to their needs, and (3) to measurement practices from which they benefit.

These are general principles that apply to all measurement infrastructures. They must be supplemented with a theory about what makes for a just educational system, for which Rawls again can provide inspiration.

Social justice and education

The first insight Rawls offers concerning the philosophy of education is about the nature and allocation of *educational primary goods*. These are the educational experiences that all individuals are entitled to—the “amount” of education owed to everyone. More specifically, educational primary goods are defined as those educational experiences that reliably provide for the skills and dispositions that enable full participation in a society’s civic culture and public sphere, as well as those that enable individuals to pursue a self-chosen conception of the good life.

Related to the idea that there are certain educational goods owed to all is the idea that schools ought to function as part of a system of institutions that secure fair equality of opportunity. Importantly, Rawls does not shoulder the educational system alone with the task of securing equality of opportunity, but instead positions schools in relation to both economic and political institutions. This complex system of institutions, in which schools play a critical role, are intended to provide for a kind of “pure procedural justice.”

Finally, school systems in a just society must provide for the possibility of self-actualization. This idea follows from a key concept in Rawls’s moral psychology, known as the *Aristotelian Principle*. It posits that all individuals have an inborn preference for learning and exercising increasingly complex skills in contexts over which they have control (e.g., a preference for non-alienated learning and labor).

These are the three lessons Rawls teaches about the nature of just educational institutions: (1) they provide for educational primary goods, (2) contribute to a system of institutions that secures equality of opportunity for all, and (3) make possible individual self-actualization. When this minimal philosophy of education is combined with the aforementioned principles of just institutionalized measurement the basic outlines of a theory of just educational measurement come into view.

Social justice and educational measurement

Tests are value laden in ways that physical measures are not. Tests necessarily create interpersonal relationships in which there are both epistemic and social power differentials—the one who gives the test is, by definition, both more knowledgeable and more powerful than the one taking it. Power differentials of this kind are not necessarily a part of physical measurement practices, which *involve* people, but are not inherently *about* people. In other words, treating testing practices as if they were physical measurement practices ignores the power relationships inherent in testing. Nonetheless, analogies to medical diagnostics and engineering (both of which involve the use of physical measures) have been present throughout the history of testing (Brown, 1992). Today the most common form of this conflation is referred to as *the education commodity proposition*.

The education commodity proposition is the simple yet powerful idea that education can be treated like any other commodity. This turns educational measurement into a means for putting a number on the value of educational processes, which can then be converted into monetary terms, usually in the form of cost-benefit analysis. The classic statement of this ubiquitous idea is: *how much education are we getting for our tax dollars?* Testing is seen as a necessary part of answering this kind of question and thus for monitoring changes in the financial value of educa-

tional processes. The reasoning is that if you cannot measure it then you cannot monetize it, and if you cannot monetize it then you have no way of knowing if your investments have paid off. This is the basic dilemma facing those who invest in educational institutions and then must prove return on investment (e.g., governments, philanthropies, venture capital). This same way of thinking impacts administrators, teachers, and students and their parents, all of who at different times and for different purposes deal with the financial meaning of test results—and in doing so run the risk of reducing the value of education to the terms of the education commodity proposition.

Importantly, this testing-enabled representation of educational value is not wrong; it is true but partial. Economic efficiency is necessary for any viable enterprise. Indeed, efficiency-oriented testing in general is necessary for the maintenance of most large-scale educational institutions. Efficiency-oriented testing transcends but includes the education commodity proposition, contextualizing it in terms of broader social goals, usually codified in terms of educational standards. Efficiency is always determined relative to some goal, and efficiency-oriented testing is always carried out relative to some definition of what education ought to be. Social justice concerns arise when the quest for objectivity and efficiency, in themselves necessary and good, results in testing infrastructures that distort social relationships within the schools by limiting what counts as good education to what can be measured. Again the reasoning is simple yet powerful: efficiency is a non-controversial good for all schools to pursue, objective measurement is required for determining and improving school efficiency, therefore only that which can be measured can be included in considerations about what makes a school good.

Because they are designed for system-level surveillance, efficiency-oriented testing infrastructures are often irrelevant and harmful to those most affected by them (e.g., students and

teachers)—thus violating the second and third principles of just institutionalized measurement. Moreover, efficiency-oriented testing tends to undermine the possibility that schools can provide the kinds of educational primary goods justice requires, let alone foster fair equality of opportunity and individual self-actualization. So, while much of the discourse about testing and justice has historically centered on bias, cheating, and the damage done when tests *lack* objectivity, most contemporary testing-induced injustices are the result of an *excess* of objectivity. In a sense efficiency-oriented testing provides too much of a good thing. The necessary and reasonable pursuit of objectivity and efficiency are taken too far, overriding the metrological rights of students and teachers while drastically truncating the scope of what is perceived as educationally valuable.

Justice-oriented testing infrastructures, on the other hand, are built to assure that objectivity and efficiency are achieved, but not at the expense of relevance and benefits to those most affected by them. A testing infrastructure that honors the metrological rights of students and teachers actualizes the commitment of an educational system to social justice by providing for the full range of educational primary goods. It contributes to a system of background institutions that promote equality of opportunity and facilitate the self-actualization of all students. Needless to say, there has never been a pure instance of justice-oriented testing (just as there has never been a pure instance of efficiency-oriented testing). But attempts at justice-oriented testing have been in evidence since the first testing infrastructures were built in the early decades of the 20th century. So it is that testing infrastructures have served a variety of functions that are a necessary part of an educational system committed to social justice, including administering a kind of pure procedural justice in the allocation of opportunity, assuring the equitable distribution of educational primary goods, and identifying the unique learning needs of each student.

But the differences between efficiency-oriented testing and justice-oriented testing cannot be grasped in the abstract. Examples from the history of testing must be examined to determine the worth of any theory of just educational measurement. The full manuscript treats three historical case studies: (1) the origins of educational measurement, (2) the founding of the Educational Testing Service, and (3) the recent history of test-based accountability following in the wake of No Child Left Behind. For the purposes of this paper I will discuss only a part of the first case study.

Conclusion: a road not taken

Alfred Binet is famous as the inventor of what would become known as the IQ test. What is less well known is that Binet intended his invention to be used according to a vision of how testing could contribute to social justice. His ideas touch on all three principles of just institutionalized measurement and position testing as a part of a just educational system. Binet understood the essential need for objectivity, but he worried about the potential harm that could be done when test scores took on an institutional life of their own. He foresaw that tests could be used in ways that would negatively mark the child for life and he warned against forms of testing that would never be of any direct benefit to the child. Binet believed that IQ-tests should be used *only* as a means for helping the least well off children in attaining the educational primary goods to which they are entitled. Gould best summarizes Binet's three cardinal principles for the use of his tests:

1. The scores are a practical device for objective classification; they do not buttress any theory of intellect. They do not define anything innate or permanent. We may not designate what they measure as "intelligence" or any other reified entity.

2. The scale is a rough, empirical guide for identifying mildly retarded and learning-disabled children who need special help. It is not a device for ranking normal children.
3. Whatever the cause of difficulty in children identified for help, emphasis shall be placed upon improvement through special training. Low scores shall not be used to mark children as innately incapable (Gould, 1996 p. 185).

These are fully congruent with the principles of just institutionalized measurement, positing each child's right to be objectively measured in ways that are both relevant and beneficial. IQ-tests implemented in schools according to these principles would be insulated from co-optation for use as a part of efficiency-oriented testing practices. Of course, during the first decades of the 20th century IQ-tests came to be used almost entirely in the name of efficiency, typically as a part of practices wherein whole student bodies were ranked and sorted so that resources could be funneled away from the least well off and towards those with greater “innate” abilities. But if Binet had his way IQ-tests would be instruments used only to identify and help the least well off. Moreover, having only this pragmatic use, test scores would never serve as enduring labels for children.

This last point—that the test score must not become a permanent label—speaks to the power of testing to impact how students understand themselves and are understood by others. The benefits of objectivity and quantification must be weighed against the direct effect of testing on the social relations of the school. Testing practices, like other forms of measurement, set the terms of mutual-understanding and facilitate coordinated interpersonal activities. The meaning of the test for the student, teacher, and administrator structures their relationships; it establishes a shared sense of “what is the case.” Because of this tests should be designed and used in ways that

assure they do not create mutual-understandings that are systematically distorted—as in, for example, when test scores are understood predominately as markers of “innate” or “inherited” differences. As the history of the IQ movement shows, testing can create schools in which there are “first class and second class citizens”—schools that do not provide every student with the educational primary goods that are their right.

It is one of the great ironies in the history of testing that the man who invented the first and most widely used standardized test understood the social justice implications of his invention, only to be completely ignored by his most enthusiastic and ambitious followers. I am not saying that Binet had it all figured out. He did not confront the complexities of post-modern school systems, with their pressing needs for large-scale efficiency-oriented testing. Neither did he envision advances in the field of psychometrics that would allow for a vastly greater range of approaches to objective psychological measurement. But he nevertheless displayed a profound moral sensitivity and offered a vision of testing as a means for social justice. This vision reflects the theory of just educational measurement we have distilled from Rawls, arguing for the right of all students to objective measurement that is relevant and beneficial, and positioning testing as a part of an educational system that provides educational primary goods for all, secures equality of opportunity, and promotes self-actualization. I leave you with Binet’s words, passionately pleading on behalf of the learning-disabled and protesting against the use of tests in ways that stigmatized children:

If we do nothing, if we don’t intervene actively and usefully, [the learning disabled child] will continue to lose time.... and will finally become discouraged. The situation is very serious for him, and since his is not an exceptional case (since children with defec-

tive comprehension are legion), we might say that it is a serious question for all of us and for all of society.... [Shame on those] teachers who are not interested in students who lack intelligence. They have neither sympathy nor respect for them, and their intemperate language leads them to say such things in their presence as ‘This is a child who will never amount to anything... he is poorly endowed... he is not intelligent at all.’ How often have I heard these imprudent words... Some recent thinkers seem to have given their moral support to these deplorable verdicts by affirming that an individual’s intelligence is a fixed quantity, a quantity that cannot be increased. We must protest and react against this brutal pessimism; we must try to demonstrate that it is founded upon nothing (Binet 1909, quoted in Gould, 1996 p. 183).

Sources:

Brown, J. (1992). *The definition of a profession: The Authority of a metaphor in the history of intelligence testing*. Princeton University Press.

Elgin, C. (1996). *Considered Judgment*. Princeton, New Jersey: Princeton University Press.

Gould, S. J. (1996). *The mismeasure of man*. New York: Norton.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Rawls, J. (1996). *Political Liberalism*. New York: Columbia University Press.