



The Gordon Commission
on the Future of Assessment in Education

Epistemology and Measurement: Paradigms and Practice

II. Social Epistemology and the Pragmatics of Assessment

Kenneth J. Gergen
Swarthmore College

Ezekiel J. Dixon-Román
University of Pennsylvania

Our major concern is with the function of assessment in education, and most particularly, the adequacy of current educational practices and policies. In Part I our chief focus was on traditions of measurement, as related to the process of large-scale assessment programs in education. As proposed, there are significant conceptual flaws inherent in the traditional empiricist orientation to measurement, and these raise important questions regarding the adequacy and relevance of "one-size fits all" assessment programs. The aim of Part II is first to sketch the contours of a promising alternative to the empiricist paradigm, namely a socially based epistemology. With this orientation in hand we are positioned to consider the more general practice of assessment - of which measurement is but constituent. Here we are especially concerned with socio-cultural outcomes, and especially the problematic results of large-scale programmatic testing. After scanning contributions to more formative practices of assessment, we offer recommendations for the future.

From Empirical Foundations to Social Epistemology

During the past several decades of scholarly inquiry, an enormous and far-reaching transformation has taken place in the concept of knowledge, and the attendant concepts of truth, objectivity, and validity. Briefly put, the transformation can be traced in its earliest phases to a number of insoluble conceptual problems inherent in attempts to establish rational foundations for scientific knowledge, and most notably, empiricists' accounts of such knowledge. Included among the insoluble problems are the challenges of matching words to world (Quine, 1960), accounting for the origin of theory (Popper, 1959), and sustaining falsification in light of the infinite plasticity of theory (Duhem, 1954). However, inspired in part by the growing critiques of the dominant orders - government, commerce, law, and military - and the resulting oppression and bloodshed of the past century, a new wave of critical work began to emerge. From this work also emerged an alternative to what Robert Mises (1997) now characterizes as a "discarded epistemology" of knowledge. Three of these critical movements bear particular attention:

Ideological Analysis

Central to the positivist/empiricist movement is the view that empirically grounded descriptions of the world carry no ideological biases. As proposed, properly supported scientific accounts of the world do not reflect the values, moral prescriptions, or religious beliefs of any particular group. This view met an early challenge from Marxist theorists arguing that capitalist economic theory – despite all the research and analysis in its support – was essentially a mystifying means of fortifying the existing class structure. As the critique suggested, all knowledge-seeking privileges certain interests over others, favors a certain political and economic configuration to the detriment of alternatives. Or more broadly put, scientific descriptions are not mirrors of the world; based on one’s particular interests, certain accounts are preferred over others. The scientist essentially observes from a particular perspective or point of view, and this perspective is never value-free. Such critique gained additional depth as scholars began to study the rhetoric of scientific accounts (see, for example, Gross, 1996). One could begin to see how social science terms such as “conformity,” “prejudice,” “obedience,” “aggression,” “altruism,” “development,” “mental illness,” and “intelligence,” were saturated with value, and how such values would not only color the interpretation of findings, but the way in which such findings were presented to and used by the public (Gergen, 1973). As scholars such as Emily Martin (1987) began to demonstrate, such colorings were not simply a problem for social sciences, but permeated the natural sciences as well.

This early critical work subsequently unleashed a broad and continuing critique of scientific and scholarly accounts in terms of their subtle biases in matters of gender, race, economic class, religion, culture, and more. Whose voices, they continue to ask, are being silenced, exploited, or erased? Many critics found their work is galvanized by the writings of Michel Foucault (1978, 1980). As Foucault argued, when authoritative claims to knowledge are circulated through the society, they act as invitations to believe. As people embrace these claims, so do they come to act in ways that support them. Or, in Foucault’s terms, claims to knowledge function to build and sustain structures of power. Thus, for example, when an authoritative group singles out certain behaviors and call them indicators of “reasoning ability,” develops measures that claim to be valid indicators of reasoning, and uses these to grant privileges to certain people and not others, they sustain a position of power in society. More broadly, these critiques raise

questions regarding the ideological and social outcomes of all forms of institutionalized assessment.

Linguistic and Literary Theory

A second major challenge to the empiricist account of knowledge emerged from linguistic and literary theory. To appreciate what is at stake, it is useful to consider two broad implications of Ferdinand de Saussure's classic work, *A Course in General Linguistics*. In simple terms, Saussure first proposed that the relationship between words and their referents is ultimately arbitrary. For example, each of us is assigned a name, and this assignment is useful in sustaining longstanding social conventions of identification. Yet, there is no inherent reason we could not have been given other names (or no name at all). Or, more generally, all naming is essentially arbitrary. In this light, it is interesting to consider the empiricist concepts of objectivity, validity, and truth. Each of these depends on the assumption that certain words correspond to or mirror what is the case. On this view, certain utterances are truth bearing, while others are exaggerated or untrue. If, however, the relationship between words and world is ultimately arbitrary, then in principle multiple possible utterances could be used to represent any state of affairs. What privileges any particular arrangement of words or observations as being "true" or "valid" is established social convention. In terms of observations, it is no more valid to say that objects are propelled to earth by the force of gravity than to say that they are thrust downward by God's will. Thus, when claims are made to "truth," "objectivity" or "validity" in reporting, we are being exposed to "one way of putting things," that is privileged by certain groups of people.

Saussure's second significant proposal was that words function within rule governed systems of usage. Put simply, our language functions in terms of various conventions, most particularly in terms of rules of grammar and syntax, but more generally in terms of idioms, definitions, and vernaculars. It may be more broadly said that all descriptions of the world will be significantly affected by conventions of writing. When this line of reasoning is extended, we find that all descriptions and analyses of the nature of the world will necessarily be driven by a system of language. They will be subservient to the demands of the logic of representation. This line of thinking has subsequently led to substantial study of the ways in which scientific accounts are governed by linguistic devices such as metaphor (e.g. Leary, 1990) and narrative (e.g.

Genette, 1980). In the latter case, for example, evolutionary theory is only intelligible by virtue of its drawing from narrative traditions of storytelling (Landau, 1993).

In sum, the world makes no necessary demands on description or explanation. Truth claims always issue from some particular tradition of representation, and once committed to a tradition, its linguistic conventions will guide our understandings of the world. This is to say that the world is not simply there to be measured, but that once a discourse of measurement has been adopted, it will largely determine what can be said about the world. Measurement effectively constitutes and creates its object. Following the previous discussion of ideology, questions must be addressed concerning the socio-cultural impact of any measure or assessment device.

Social Construction of Scientific Knowledge

These preceding critiques, emerging in quite separate domains of scholarship, come to a head in perhaps the third and perhaps most essential contribution to a viable replacement for empiricist epistemology. The movement here is essentially toward a social epistemology. Its origins may be traced to Mannheim's 1929 volume, *Ideology and Utopia*. As Mannheim proposed, (1) the scientist's theoretical commitments may usefully be traced to social (as opposed to empirical) origins; (2) scientific groups are often organized around certain theories; (3) theoretical disagreements are therefore issues of group conflict; and (4) what we assume to be scientific knowledge is therefore a byproduct of a social process. This seminal work was followed by a substantial number of influential contributions.¹ However, in terms of the evolution of ideas, the landmark volume is Thomas Kuhn's *The Structure of Scientific Revolutions*. Most importantly, this work represented a frontal challenge to the longstanding presumption that scientific knowledge is progressive, that with continued research – testing hypotheses against reality – we come ever closer to the truth. Thus, proposed Kuhn, the shift from a Ptolemaic to a Copernican account of the relation of the earth to the sun is not progress toward truth; nor is the shift from Newtonian theory to quantum mechanics in physics. Rather, Kuhn proposed, our propositions about the world are embedded within *paradigms*, roughly a

¹ These include, for example, Ludwig Fleck's *Genesis and development of a scientific fact*, Peter Winch's *The idea of a social science*, George Gurvitch's *The social frameworks of knowledge*, and Berger and Luckmann's *The social construction of reality*.

network of interrelated commitments (to a particular theory, conception of a subject matter, methodological practices, and the like). Thus, even our most exacting measurements are only sensible from within the paradigm. A look into a microscope tells you nothing unless you are already informed about the nature of the instrument and what you are supposed to be looking at. What we call progress in the above cases of astronomy and physics, is not then movement from a less to a more objectively accurate account of the world. They represent shifts in paradigm, different ways of thinking and observing.

In recent decades this social view of science has been buttressed by an enormous body of scholarship centered on the cultural and historical contingency of scientific knowledge. As broadly acknowledged, the philosophical search for universal foundations of empirical knowledge is now moribund. Rather, summarizing the three critical waves outlined above, we find that scientific knowledge is a byproduct of negotiated agreements among people concerning the nature of the world. Whatever exists makes no fundamental requirements regarding our attempts to describe and explain. But, once we have entered into a particular tradition or paradigm of understanding - as represented in a shared language - this tradition will provide both direction and limits on our explanations, descriptions, and observations. Further, all such traditions will be wedded to particular ways of life, which is to say, they will carry certain implicit or explicit values or desired goals.

This social constructionist conception of science is not at all fatal to the empirical tradition. Rather, it simply removes the foundations for such a tradition. Rather than using research to illuminate the truth about an independent world, the questions become pragmatic and political. In what ways is a given line of research useful, for whom, and with what socio-political implications?

Assessment as Social Construction

In light of the transformation from an empiricist to a constructionist view of knowledge, how are we to understand the practice of educational evaluation, and the function of testing and assessment as informed by measurement within this process? First it is important to reiterate a major theme in Part I, namely the initial lodgment of measurement practices in the empiricist paradigm. Within the paradigm the major questions are typically focused on the psychometric

properties themselves (e.g. validity, reliability, scaling), along with issues of sampling, random errors, and so on. However, from a constructionist standpoint, such questions are highly limited. That is, they issue from the assumptions shared within a particular social group. And, while there is nothing problematic about such issues within the relevant enclaves, a constructionist view opens the door to the full range of stakeholders concerned with education. For the constructionist, the voices of traditional testing and measurement communities, represent the interests of particular traditions, useful for certain purposes, but not for all. What stands as "objective assessment" for one, may for another be viewed as "prejudice in action."

From a constructionist perspective, the chief question concerns the pragmatic utility of testing, and simultaneously, for what parties are these outcomes valuable or not? Such questions must be addressed not only within the enclaves of science and government, but also across the spectrum of the population, and most especially within the schools, communities, families, and student participants affected by tests and the policies they support.

There is no doubt that traditional measurement practices have been useful for certain groups in terms of providing a vantage point for deliberating about educational standards and policies. More debatable is whether such tests have been successful in rendering the educational system effective in attaining its goals (see, for example, Ravitch, 2010), and if so, at what costs? When we open the door to the voices of other stakeholders, we find ample room for concern with the outcomes of our practices. In this context, we wish to give voice to a range of existing critiques. Some of these critiques are well known; others are harbored within significant sub-cultures. However, a broad scan will be useful as a preliminary to a subsequent discussion of alternatives to contemporary testing practices. For purposes of simplification, we focus on three realms of critique, concerned with the impact of testing on: cultural ideology, societal structure, and educational practice.

Assessment and Ideology

Practices of testing are far more than technologies for providing information. Rather, they carry with them a range of implicit values or ideologies. Through the broad institutionalizing of testing procedures, there is a shaping of cultural values. Practices of assessment do not so much reflect the nature of the individual as they construct the individual in their terms, and thus shape

the cultural milieu. In this light, it is important to consider some of the ideological values that are fortified by traditional testing practices:

Neoliberalism. Although the term "neoliberalism" has been used in many ways, one prominent usage denotes an ideological orientation to cultural life in which the metaphor of "the market" is dominant. Thus, a neoliberal view of education emphasizes market-based logics for evaluating the productivity of the educational system. In these terms, the prominent question to be asked about educational systems is whether the investments in the systems are paying off in terms of specifiable outcomes. Measuring devices are thus required to assess these outcomes in a systematic way. This logic is, of course, the dominant driver in national testing of students.

It is this logic, critics propose, that now travels across cultural life, reconstructing relations of trust, care, and nurturance in terms of costs and benefits. As critics lament, for example, when the logic of the marketplace is inserted into cultural life, people lose intrinsic worth, as do institutions, traditions, rituals, the arts, and so on. Worth is defined in terms of some form of "price." Many view this orientation as *instrumentalist*. That is, one begins to think of others in terms of costs and benefits. Having a child, for example, is robbed of intrinsic worth, and replaced instead by questions of "what do I gain by a child," "what will it cost me?" People count only in terms of costs and benefits to one's own pleasure or well-being

In important respects, institutional reliance on testing contributes to just such an orientation. Education is important primarily as it produces desirable test scores. In this sense, test scores reduce the student to a set of numbers, used by others to make decisions that will shape the student's behavior to their standards. This logic moves from high levels of policy making into the classroom, where teachers learn that whatever their students' particular characteristics, and regardless of their caring relations for them, the students are ultimately test performers whose scores may vitally affect their own careers. Their task as educators is reduced to producing higher test scores. Rather than helping or assisting students for intrinsic reasons, teachers find themselves using students to protect or help themselves. Students become instruments for the teacher's well-being. The market-place mentality expands its reach.

Individualism. Traditional testing practices are altogether focused on the performance of individuals, chiefly the individual student. Through the agglomeration of test scores, the behavior of individual teachers and individual school administrators are also evaluated. In part, this focus reinforces the presumption of *psychological causes* of individual behavior, as

described in Part I. More broadly, critics find this focus sustains and supports the longstanding Western value of individualism, prizing as it does, the autonomy of the individual. There is much to be said about the contribution of this ideology to the development of Western cultural life. However, within recent decades there has been growing concern with the detrimental implications of individualist ideology for cultural life (see, for example, Bellah, et al. 1985; Gergen, 2009). At the heart of many such critiques is the divisive impact of individualism, the way in which people come to understand themselves as fundamentally separated from each other, and necessarily on their own. Favored by this separation are tendencies toward selfishness, callousness, and a general suspicion of bonded relationships (e.g. marriage, family, friendships, community). The chief aim of life, in this tradition, becomes one of self-maintenance, with a maximization of self-gain and a minimization of personal discomfort.

This focus on the individual is all the more significant in a world in which technologies of communication bring people together in unprecedented degree. As noted in Part I, this transformation places an increasing demand on the capacities of people for collaboration. The "self-made man" is no longer functional, as complex decisions now require the participation and integration of many voices, many vantage points. As social theorists point out, the shift toward collaborative process also favors shifting from accounts of society that focus on individual units (e.g. the person, the institution) to visions of relational process (e.g. systems, networks, confluences, aggregates, synergies).

In this light, testing the performance of an individual student is myopic. Such performance will surely depend in part on the quality of teaching. But if a student is hungry, resides in a chaotic home environment, and participates in a peer group in which schooling is devalued, his or her performance will also be affected. Or, more systemically, the student's performance is the outcome of a confluence of processes, both within and outside the classroom. Test scores are mistakenly presumed to reflect an individual characteristic, when it is far more reasonable to view them as artifacts of system functioning. Attention should properly be directed to the larger processes from which failed performances are generated. Ultimately these could include issues of economic and social inequality.

Assessment and Societal Well-being

Although testing practices carry with them subtle forms of ideology, there are various ways in which testing more directly affects patterns of behavior. As Dahler-Larsen (2011) argues, testing practices not only reflect societal beliefs and values, but once set in motion, their effects are reverberating. Related practices of many kinds are affected. In this context it is essential to take account of what many see as problematic "side-effects" of current practices.

Social Division and Distrust. Assessment practices effectively establish a structure of power, with those who administer tests positioned to affect the outcomes of those under evaluation. This four-tiered structure - placing government over school administrators, over teachers, over students - also defines their relationship. Those in the lower echelons will fear those above, and those above will be suspicious of all below. The insertion of testing procedures into this mix fortifies the barriers between these groups, intensifying both distrust and antagonism. This is most obvious in the case of the teacher administering a test to students. The test itself positions the student as "in question," or in effect, constructs the student as one whose worth is not yet established. Yet, when students fail on tests, the teacher's capabilities are now thrown into question. They too become aliens to be judged by school administrators. In turn, with failing schools, the fate of school administrators becomes an issue for judgment. At this point, distrust becomes endemic to the entire system.

Theodore Porter (1996) argues that testing is typically used under conditions of distrust. One uses tests to ensure others can be trusted. Ironically, however, while tests should generate trust that the system is working properly, the result is more often the reverse. As students, teachers, and school systems are placed in jeopardy by the tests, the possibility for cheating becomes attractive. This is especially so, as the power structure has already defined relationships in terms of alienation. Thus, as evaluators have found, cheating on tests is widely practiced. Multiple instances of cheating further feed the fires of suspicion. As a case in point, an audit of Pennsylvania state exams recently flagged 38 school districts and 10 charter schools for possible cheating. Following such findings, the governor then proposed a 43 percent increase in funding for educational assessment. At the same time, school funding in general remained at its current level. We thus have a spiral in which tests generate the rationale for cheating, which then provides the grounds for more formidable means of testing.

Hierarchies of Worth. From a constructionist perspective, the properties or characteristics attributed to human beings are the outcomes of historically and culturally contingent negotiation. Thus, people have variously been defined as witches, pure in soul, or bipolar, depending on history and culture. Further, these attributions typically reflect group interests or values, with a positive value placed on a purified soul, for example, and a negative value on witchery and mental illness. In this context, critics are concerned with the way in which tests function in the classroom and society to create hierarchies of value - such as distinguishing between the "able" minded and the "disabled" (Dixon-Román, 2010; Gutiérrez & Dixon-Román 2011). By their mere participation in the school system, students come to learn that they are able or disabled in ways they never imagined. These evaluations will color the way in which they are treated by their teachers and influence, as well, their relations with peers. Some students learn to view themselves as "failures" and may fear voicing an opinion or indeed, attending school altogether. Others will come to see themselves as "superior" to their peers, which further invites distancing and disdain.

Pluralism in Peril. Much has been said in recent decades concerning the pluralist makeup of the society and the necessity of giving voice to the many vibrant minorities making up the whole. Such celebration of pluralism is consistent with the democratic ideals of the nation. Yet, the testing tradition in the U.S. carries vestiges of an earlier tradition of cultural univocality. Those determining what is to be mastered by students and how this mastery is to be reinforced through standardized tests represent the values and assumptions common to a once-dominant culture. Minority concerns and values are seldom reflected in the design and application of tests. For various minorities, for example, schools might optimally provide safety and support for their offspring, character training, or practical skills. The significant shift toward home schooling is only one indicator of these specialized concerns for education. In the worst case, one might say that standardized testing - subjecting all to a universal standard - functions to obliterate minority differences. At a minimum, current testing practices function oppositionally to the democratic ideal of a pluralist engagement in building the society's future.

The Loss of the Local. In the same way that standardized testing is unresponsive to the multiple ethnic, religious, and racial enclaves in society, so it is oblivious to the local particularities of school systems. School systems in upper middle class suburbs are evaluated on the same dimensions as schools in lower class urban settings, Hispanic communities in the rural

Southwest, and so on. Yet, in each of these settings the educational needs may be quite specific. In one case the investments are in preparing students for college, in another, merely finishing high school safely, and in another, fitting in effectively into one's multi-ethnic community, and so on. In some communities there may be companies seeking graduates with particular kinds of skills, and in others there may be no employment opportunities. For many, the issue of workplace skills is especially important. Here critics point out that national tests typically stress knowledge within structured systems (e.g. mathematics, grammar). Yet, it is argued, most decisions in the workplace now depend on one's ability to work with others in highly complex, ambiguous, and ever-changing circumstances. Traditional tests are largely irrelevant to such skills. In effect, the diverse needs and aims of students, teachers, schools, and communities are shoved to the margins by traditional large-scale testing practices.

Assessment and the Educational System

We finally turn our attention to the effects of assessment on the educational process. As pointed out, the extent to which traditional large-scale testing has improved the educational process remains in question. However, when we expand the range of voices beyond those concerned primarily with improving test scores, what may be said about the impact of assessment procedures on education? We touch here on four outcomes.

Curriculum and Pedagogy. One of the most widely voiced critiques of institutionalized testing comes from the teaching community. When student test scores become matters of public evaluation, their own capabilities are also implicated. As previously noted, poor student performance suggests poor teaching performance. And, when one's career is thus placed in jeopardy, it is natural to focus on means of boosting student test scores. For many teachers this means adjusting both curriculum and teaching practices. For example, the teacher may sacrifice compelling concerns with the interest value and local applicability of the materials to be taught, and narrow the curriculum to precisely those materials that are relevant to the test. Rather than fostering class discussion or developing group projects, the teacher focuses on improving students' capabilities in taking tests. As a study by the Center for Education Policy (2005) revealed, demands for accountability generated a major reduction in the time teachers devoted to science, social studies, art, music and physical education. Or, as commonly put, quality

education is sacrificed by "teaching to the test."

Teacher Motivation and Engagement. Teachers often enter the profession because they are drawn to the educational process. They enjoy working with the young and relating to them in ways that are mutually nourishing. Further, they develop ways of teaching they find enjoyable and effective in their particular classroom. As the character of the student body changes over time and new technologies become available, many teachers also enjoy the challenge of developing new teaching practices. Teacher motivation and engagement in the process importantly depend on the teacher's freedom to work creatively within the immediate context. In this sense, the imposition of extraneous requirements placed upon the teacher, runs the risk of reducing motivation and engagement. The teacher's individual engagement in the teaching process is subverted by demands to boost student scores. We have already mentioned the ways in which testing requirements invite an instrumental and distrustful orientation toward students. And, when personally nourishing forms of practice must be sacrificed in order to improve test scores, the teacher is reduced to a "machine part" within a system in which they have very little voice and constrained pedagogic agency.

Student Motivation and Engagement. Self-definition is importantly wedded to social process. Thus, how students come to understand themselves - their worth, abilities, and potentials - can be vitally affected by their school experiences. It is here that a number of preceding discussions come into sharp focus. As Bandura (1994) has put it, "Educational practices should be gauged not only by the skills and knowledge they impart for present use but also by what they do to children's beliefs about their capabilities, which affects how they approach the future." (p.76) With the insinuation of testing systems into the educational process, children rapidly begin to see themselves in terms of evaluative hierarchies. A small percentage will come to see themselves as superior to others; the majority will come to view themselves as "just average," and by virtue of scaling requirements, a number will come to see themselves as "inferior." And, as commonly demonstrated, children from minority or impoverished circumstances, along with children for whom English is a second language, are more likely to find themselves defined as undesirables. These negative effects are all the more accentuated by what becomes the instrumentalist orientation of teachers, for whom low student scores are threatening to career.

Closely related to this critique, educators have long been critical of the early tradition of

learning through intimidation, whether by the old-fashioned threat of a hickory stick or the more contemporary threat of failure and derision. As many believe, far better outcomes are obtained by tapping into intrinsic motives and collaborative processes. Even in the earliest years, students enter schools with interests and enthusiasms. It is when the school can tap into these proclivities and unite them with their educational aims that student engagement is maximal. Yet, as many see it, the imposition of large-scale, high outcome testing functions regressively. It redefines the classroom as a context of threat and reprisal. An abiding concern with high test performance subverts those educational practices not directly tied to such performance – namely those that are student-centered. The capacity of the curriculum and the pedagogical practices to absorb or resonate with student interests and motivation is eroded. In effect, teaching to the test comes at a cost of student engagement in the educational process. Schools and teachers become irrelevant and alien. In this context, the otherwise startling drop-out rates in inner-city schools should come as little surprise.

Parent-child Relations. One may properly surmise that the instrumentalist orientation fostered by the use of tests in the school also carries over to family relations. That is, parental definitions of their children are informed by their relations with the schools. They come to define their children in terms of the educational system and to see their children in terms of institutional categories. Among the most obvious cases are those in which they might be surprised to find that their otherwise happy child is mentally ill with ADHD. More subtly, however, they learn the evaluative hierarchies implied by testing practices and to judge their children in these terms. In all but the minority of cases they will learn that their children are less than prized – simply average or below. Often parents are enlisted by the teachers as a "support staff" in an effort to boost test scores, thus importing the force of negative judgment into the home, and expanding further the domain of alienation.

Alternatives to Traditional Assessment Practices

In many respects, the logic of traditional, large-scale testing practices is compelling. In a democratic nation we are committed to education for all citizens. We believe that the future well-being of individuals, communities, and the nation as a whole, depends on high quality and efficacious educational institutions. While the economic investments we make in education are

less than we dedicate to law enforcement and prisons, we do wish to assess the outcomes of our investments. Are these investments truly fulfilling the goals and aims of the nation? Are there groups of young people “left behind” by our current practices, and are schools succeeding in the task of educating? In search of standards-based education, we turn to scientifically based measurement to provide the answers and to serve as the basis of educational policy and action.

Yet, as James Scott (1999) has proposed, large-scale, top-down proposals such as these can be recipes for disaster. As we have described, for public education such logics are particularly perilous. It is not simply that we cannot be certain what is being measured by any given test, but there is no compelling evidence that testing actually contributes to the valued goals. Rather, if we take into account the realities of the multiple stakeholders involved, we find that in many respects our current approach is counter-productive. From within the enclaves of government the mandate for testing is clear enough; for the scientific researcher the logics of measurement and application are well understood. However, for those outside these enclaves, current policies and practices are found oppressive and destructive.

To summarize our views, at the ideological level our current policies and practices transform human relations into a marketplace; for others, they foster an alienating individualism, and invite an instrumentalist orientation to relationships. In terms of societal well-being, otherwise solid testing practices invite distrust and alienation, create divisive hierarchies of worth, undermine pluralistic values, and erode the capacities for local self-direction. In the local school setting, teachers find that standards-based accountability means narrowing the curriculum, denaturalizing pedagogical practices, and a loss in motivation. Students find their self-worth placed in question and little in school that excites their interest, with school drop-out rates resultantly increased. Parents find their nurturing orientation toward their children replaced by a performance based evaluative posture.

Although one could argue for adjusting current policies and practices (cf. Resnick and Zurawsky, 2005), the emerging world conditions invite us to open fully the range of possibilities for achieving commonly shared goals of education. In this context it is useful to consider innovative initiatives. Although there are many such initiatives extant, we outline here several that are illuminating in their avoidance of hierarchical structures. Each favors pluralism, localism, democratic process, qualitative understanding, and processes of improvement over product. Rather than striving for summative statements about existing conditions, they focus

more formatively on the continuous and collective processes of improvement (Dinesen, 2009).

Empowerment Evaluation

Perhaps the most widely known assessment alternative is empowerment evaluation. Developed by David Fetterman (2000, 2004, 2007) and his colleagues, the attempt is to shift the site of evaluation from the distant overseer to the local participants. And, rather than simply assessing students and teachers, the attempt is to enable the local community to become self-directing, to deliberate on its activities, set goals for itself, and take necessary actions. Empowerment evaluators, in this case, serve as coaches or facilitators, helping local communities to build ongoing practices of self-evaluation. Such practices may vary considerably from one community to another, but a typical empowerment procedure might take place in several stages, including, 1) a process of “taking stock” in which all members of the school community participate in identifying educational goals and surveying available resources, 2) broad and inclusive discussion of the ideal or dream school to which they might aspire, 3) the establishment of cadres that develop specific plans for moving toward the ideal, and 4) setting up administrative committees that implement change. The various groups engaged in implementation also establish yardsticks by which they can evaluate their progress. Over time educational communities are enabled to chart their future, evaluate their progress, and alter plans and programs on a continuing basis. This is not to say that outside testing procedures are precluded. Rather, standardized tests can provide information helpful in judging local progress. Rather than dictating policy, test scores become adjuncts to local school development.

Dialogic Evaluation

Seeking a more inclusive and democratic orientation to evaluation, a broad number of educational specialists have made strides in developing processes of dialogic evaluation (Ryan and Stefano, 2000; Schwandt, 2005). In brief, dialogic evaluation practices tend to emphasize egalitarian dialogue, equality and justice, multi-cultural intelligences, dialogic learning, and qualitative analysis as opposed to quantification - all with an eye to broad scale social

transformation. As Greene (2001) has put it, dialogic evaluation refers to, “engaged, inclusive and respectful interactions among evaluation stakeholders about their respective stances and values, perspectives and experiences, dreams and hopes, and interpretations of gathered data related to the value and its context.” (p. 182). Dialogic evaluation practices have been explored more visibly in the European context than the American. For example, the English National Board for Nursing, Midwifery and Health Visiting emphasizes the use of context-based dialogue, centering on competence in practice, along with portfolios that can be used to reflect, envision, and plan. Relevant work in the European Union has stressed the need for training practitioners in the use of dialogic evaluation. Much as in the case of empowerment evaluation, proper assessment is not simply a matter of reading percentile scores on national norms, but requires the weaving together of multiple voices, values, and potential indicators of efficacy. (See also Cousins and Whitmore, 1998, on participatory evaluation.)

Appreciative Evaluation

Appreciative evaluation practices have emerged largely in the context of large -scale organizations and in program evaluation. However, the relevance to the educational sphere is clearly apparent. Appreciation evaluation practices are lodged within social constructionist premise that we create our realities through dialogue. Thus, dialogues that center on *problems* – for example, the poor performance of students, teachers, or school systems - solidify the reality of the problems. And when fortified, this reality will lead to mutual blame, alienation, distrust, disrespect, lowered motivation, and more. Further, this focus is myopic in its search for a "solution to the problem" without taking into account the systemic processes in which it is embedded. The appreciative approach centers discussion on valued actions or performances, that is, what may be prized by the group (see Preskill and Catsambas, 2006; Coghlan, Preskill & Catsambas, 2003). Narratives of positive performance are shared and collected, and dialogue then turns to the common values represented in these narratives. Reflection on these values then moves to the means of building futures in which these values would be most fully instantiated. What concrete steps would be useful and promising? Groups are then established to monitor progress toward the shared goals. Appreciative practices attempt to bring all stake-holders together in future building. In the educational sphere, this might involve the student, his or her

teachers, parents, and possibly others. On a broader scale, the entire school system might participate in what is called an appreciative inquiry “summit.”

Socio-cultural/Situative Assessment

Within the past two decades new approaches to assessment have drawn from sociocultural and situative theories of learning and knowledge (Moss et. al., 2008). The socio-cultural/situative perspective concerns itself with the *affordances* (i.e., perceived possibilities within the learning environment) and the learners *effectivities* (i.e., the set of capacities for transforming an affordance into action) (Gee, 2008a). Learning and knowledge are not so much understood as inside the head of the learner as embodied in the relational actions and practices taking place in the learning environment. Assessment practices from a socio-cultural/situative perspective not only evaluate the actions and practices based on the local community’s understandings, but also attempt to account for the affordances and effectivities of the learning environment. We see examples of this in Myford and Mislavy’s (1996) Advanced Placement studio art portfolio assessment, and Gee’s (2008b) work on video games and assessment.

These four explorations into alternative ways of thinking about and practicing evaluation are salutary, but at the same time, are only exemplary of a movement of a far greater scope. Increasingly educators, program evaluators, and organizational change specialists – to name but a few - are turning to more egalitarian, reflective, dialogic, collaborative, and context sensitive practices of enhancing human performance. The interested reader might wish to explore, for example, developments in teacher research (Fishman and McCarthy, 2000), educational action research (Noffke and Somekh, 2009), and collaborative management research (Shani et al., 2007).

On the Future of Assessment

The nations of the world offer a broad panoply of approaches to educational testing. One might contrast, for example, the stringent and high stakes practices of testing in Japan and Korea with the low levels of standardized testing in Finland. By the same token, one may ask about the ways in which these differing practices function in society. Interestingly, while the strong testing

approaches taken in Japan and Korea are associated with high rates of adolescent suicide (along with laws to prevent excessive study), Finnish minimalism is co-terminal with what is regarded to be the most effective educational system in the world. Further exploration of such differences would be useful in expanding deliberation on alternatives to current practices.

As proposed at the outset of our initial paper, technology based shifts in global conditions appear to militate against current testing practices in the U.S. The attempt to impose a single set of standards on an enormously varied domain of practices fails to recognize the rapidly shifting and endlessly complex landscape of educational desiderata. What is needed in the way of education - the knowledge and the skills required for productive and meaningful participation - is ever changing. To lock in a highly circumscribed set of ideals and to shape the entire educational system by these specifications is to reduce the potential for productive participation in the future. Further, these same technologies of communication also foster multiple affinity groups, enabling them to generate rationales, agendas, and plans of action. In effect, there is an expansion in the array of voices that demand to be heard, voices that are set against otherwise totalizing agendas. In our view, the way must be prepared for more democratic, inclusive, and dialogic contributions to the ways in which local systems of education function.

Let us briefly summarize what we see as the major outcomes of our deliberations. In our view, we should move toward modes of evaluation that will support and enable the participation of all the nation's people in building effective educational systems and the flourishing of human capabilities more generally. Assessment should in no way drive these efforts, but should serve ancillary or supportive functions. Our views support a human development approach to achieving national goals, over and against the reliance on the economic marketing of education. In what follows, we first attend to what we believe assessment should *not be* and then shift the focus to what, for us, assessment *should become*. In planning for the future, we should move away from:

- *employing assessment as an instrument for administrative sorting, selecting, or predicting.*
- *assessing internal mental structures such as cognitive or affective knowledge, reasoning, or other mental abilities.*
- *using assessment results to establish policies that differentially reward and/or*

- penalize students, teachers, or school administrators.*
- forming policies demanding all educational systems to meet the same standards in terms of student performance.*

To be clear, we are *not* arguing against active and continuing deliberations on the performance adequacy of students, teachers, administrators, or school systems in general. Such reflections can make vital contributions to educational success in rapidly changing and highly differentiated world conditions. Assessment instruments or tests can contribute to such deliberations in significant ways. Thus, we propose that such tests be employed in evaluation practices that:

- are dialogic in nature and situated primarily in local knowledges and practices.*
- include multiple stakeholders, and importantly, those whose performance is under evaluation.*
- include multiple criteria, reflecting the needs and values of multiple stakeholders.*
- center on processes of continuous improvement of whole systems, including students, teachers, and surrounding communities.*
- provide for the education of teachers, school administrators, and other relevant stakeholders on the limits of measurement and in dialogic centered practices of evaluation.*

What could this mean for the future of testing and measurement instruments more generally? In our view the above recommendations would not mean a diminishment, but rather, an enrichment and expansion of such services. With respect to testing, there is good reason to move toward the following:

- Making standardized tests available (as opposed to mandatory) to all educational institutions.* Whether and how local school systems or districts employ test scores in their deliberations should be locally determined.

- Radically expanding the kinds of measures available to schools for evaluating students.* For example, depending on locale, schools might wish to have measures that would enable them to bench-mark students in terms of computer literacy, career fluency, civic and political

participation, bi-lingual capacity, dialogic skills, environmental knowledge, musical aptitude, physical competence, health, and so on.

- *Expanding the range of measures available to schools and school districts for evaluating their own development.* For example, schools might varyingly wish to benchmark themselves in terms of parental participation, excellence as a learning community, internal collaboration, civic contribution, relationships with business and government, and the like.

- *Offering educational services enabling local schools to generate effective practices of participatory evaluation.*

- *Developing means of exploring the relational validity of assessment devices.* Due to the locally situated nature of language, culture, and knowledge, assessment devices are understood in many different ways. In effect, they are relationally situated within local cultural communities. A better understanding of these cultural processes is much needed.

- *Developing more dialogic and systematic approaches to evaluating consequential validity (Messick, 1989).* This would include examining more rigorously and thoroughly the social consequences of score interpretation within local venues.

In conclusion, if we take into account the increasing development of communication technologies and the resulting shifts in demands and opportunities, it is imperative to explore new ways of practicing evaluation. Along with Nussbaum (2011), we argue here for evaluation in the service of creating capabilities as opposed to judging them.

References

- Bandura, A. (1994) Self-efficacy. In V.S. Ramachaudran (Ed.) *Encyclopedia of human behavior* (V. 4. 71-81.) New York: Academic Press.
- Bellah, R.N., Madsen, R., Sullivan, W. M., Swidler, A., & Tipton, S. M. (1985) *Habits of the heart: Individuals and commitment in American life*. Berkeley: University of CA Press.
- Berger, P. & Luckmann, T. (1966) *The social construction of reality*. New York: Doubleday/Anchor.
- Center on Education Policy (2005) *From the capital to the classroom: Year 3 of the No Child Left Behind Act*. Washington, DC: Center for Education Policy Report.
- Coghlan, A.T., Preskill, H. & Catsambas, T.T. (2003) An overview of Appreciative Inquiry in evaluation. *New directions for evaluation*. 100, Winter.
- Derrida, J. (1976) *Of grammatology*. Baltimore, MD: Johns Hopkins University Press.
- Dahler-Larsen, P. (2011) *The evaluation society*. Palo Alto, CA: Stanford Business School Press
- Dinesen, M.S. (2009) Systemic appreciative evaluation: Developing quality instead of just measuring it. *AI Practitioner*. 11, 49-56.
- Dixon-Román, E. (2010) Inheritance and an economy of difference: The importance of supplementary education.” in Lin, L., Gordon, E.W., & Varenne, H. (eds.). *Educating comprehensively: Varieties of educational experiences*. Lewiston, NY: Edwin Mellen Press.
- Duhem, P. (1954) *The aim and structure of physical theory*. Princeton: Princeton University Press. (originally published in 1914)
- Fetterman, D.M. (2000) *Foundations of empowerment evaluation*. Thousand Oaks, CA: Sage.
- Fetterman, D.M. & Wandersman, A. (Eds.) (2004) *Empowerment evaluation, principles in practice*. New York: Guilford.
- Fetterman, D.M. & Wandersman, A. (2007) Empowerment evaluation, yesterday, today, and tomorrow. *American Journal of Evaluation*, 28, 179-198.
- Fishman, S and McCarthy, L. (2000) *Unplayed tapes: A personal history of teacher research*. New York: Teachers College Press.
- Fleck, L. (1979) *Genesis and development of scientific fact*. Chicago: University of Chicago Press. (originally published in 1935).
- Foucault, M. (1978) *The history of sexuality, Vol. 1, an introduction*. New York: Pantheon)

- Foucault, M. (1980) *Power/knowledge*. New York: Pantheon.
- Gee, J.P. (2008a). A sociocultural perspective on opportunity to learn. In Moss, P.A., Pullin, D.C., Gee, J.P., Haertel, E.H., & Young, L.J. (eds.). *Assessment, Equity, and Opportunity to Learn*. New York NY: Cambridge University Press.
- Gee, J.P. (2008b). Game-like learning: An example of situated learning and implications for opportunity to learn. In Moss, P.A., Pullin, D.C., Gee, J.P., Haertel, E.H., & Young, L.J. (eds.). *Assessment, Equity, and Opportunity to Learn*. New York NY: Cambridge University Press.
- Genette, G. (1980) *Narrative discourse*. Ithaca, NY: Cornell University Press.
- Gergen, K.J. (1973) Social psychology as history. *Journal of personality and social psychology*, 26, 309-320.
- Gergen, K.J. (1994) *Realities and relationships: Soundings in social construction*. Cambridge: Harvard University Press.
- Gergen, K.J. (2009) *Relational being: Beyond self and community*. New York: Oxford University Press.
- Greene, J.C. (2001) Dialogue in evaluation, A relational perspective. *Evaluation*, 7, 181-187.
- Gross, A. (1996) *The rhetoric of science*. Cambridge: Harvard University Press.
- Gurvitch, G. (1966) *The social frameworks of knowledge*. New York: Harper and Row.
- Gutiérrez, R., and Dixon-Román, E. (2011) "Beyond Gap Gazing: How Can Thinking About Education Comprehensively Help Us (Re)envision Mathematics Education?" in Atweh, B., Graven, M., Secada, W. & Valero, P. (eds.). *Mapping Equity and Quality in Mathematics Education*. New York, NY: Springer.
- Hopson, R.K. (2009) Reclaiming knowledge at the margins: Culturally responsive evaluation in the current evaluation moment. In Ryan, K.E. & Cousins, J.B. (Eds.) *The Sage International Handbook of educational evaluation*. (pp. 431-448). Thousand Oaks: CA: Sage.
- Kuhn, T.S. (1970) *The structure of scientific revolutions*. 2nd ed. Rev. Chicago: University of Chicago Press.
- Landau, M. (1993) *Narratives of human evolution*. New Haven: Yale University Press.
- Leary, D. (1990) *Metaphors in the history of psychology*. New York: Cambridge University Press.
- Martin, E. (1987) *The woman in the body: A cultural analysis of reproduction*. Boston: Beacon Press.

- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd Ed.) (pp. 13-103). New York: American Council on Education/Macmillan.
- Mislevy, R.J. (1997) Postmodern test theory. In A. Lesgold, M.J. Feuer, and A.M. Black (Eds.) *Transitions in work and learning: Implications for assessment*. (pp. 280-299). Washington, DC: National Academies Press.
- Moss, P.A., Pullin, D.C., Gee, J.P., Haertel, E.H., & Young, L.J. (2008). *Assessment, Equity, and Opportunity to Learn*. New York NY: Cambridge University Press.
- Myford, C.M., & Mislevy, R.J. (1996). *Monitoring and improving a portfolio assessment system*. CSE Technical Report 402. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Noffke, S. and Somekh, B. (Eds.) (2009) *Sage handbook of educational action research*. Thousand Oaks, CA: Sage.
- Nussbaum, M.C. (2011). *Creating capabilities: The human development approach*. Cambridge MA: The Belknap Press.
- Popper, K. R. (1959) *The logic of scientific discovery*. London: Routledge.
- Porter, T. (1998) *Trust in numbers*. Princeton NJ: Princeton University Press.
- Preskill, H. & Catsambas, T.T. (2006) *Reframing evaluation through appreciative inquiry*. Thousand Oaks: CA: Sage.
- Quine, W.V.O. (1960) *Word and object*. Cambridge: MIT Press.
- Ravitch, D. (2010) *The death and life of the great American school system: How testing and choice are undermining education*. New York: Basic Books.
- Resnick, L. and Zurawsky, C. (2005) Standards-based reform and accountability: Getting back on course. *American Educator*, Spring. 1-13.
- Ryan, K.E. and DeStefano, L. (Eds.) (2000) *Evaluation as a Democratic process: Promoting inclusion, dialogue, and deliberation*. San Francisco: Jossey-Bass.
- Schwandt, T.A. (2005) A diagnostic reading of scientifically based research for education. *Educational Theory*, 55, 285-305.
- Schwandt, T.A. (2008) Educating for intelligent belief in evaluation. *American Journal of Evaluation*. 29, 139-150.
- Scott, J. (1999) *Seeing like a state: How certain schemes to improve the human condition have failed*. New Haven: Yale University Press.

Shani, A.B., Mohrman, S.A., Pasmore, W.A., Stymne, B., and Adler, N. (Eds) *Handbook of collaborative management research*. Thousand Oaks, CA: Sage.

Whitmore, E. (Ed.) (1998) *Understanding and practicing participatory evaluation*. San Francisco: Jossey-Bass.

Whitmore, E. et al. (2006) Embedding improvements, lived experience, and social justice in evaluation practice. In I. F. Shaw, J. C. Greene, & M. M. Mark, (Eds.) *The Sage handbook of evaluation*. London: Sage.

Williams, R. (1977). *Marxism and Literature*. New York NY: Oxford University Press.

Winch, P. (1946) *The idea of a social science*. London: Routledge.

Wittgenstein, L. (1953) *Philosophical investigations*. New York: Macmillan.