1 Introduction

Edward H. Haertel, Pamela A. Moss, Diana C. Pullin, and James Paul Gee[1]

The most pressing issue facing U.S. education may be providing all students with a fair opportunity to learn (OTL). Although most would embrace the goal of enhancing OTL, there are fundamental disagreements about how best to accomplish this and different understandings of the meaning of "opportunity to learn." Historically, conceptions of OTL have been closely tied to the practice of testing. OTL has been conceptualized as opportunity to learn what is tested, and test-based accountability has been widely implemented as a means of enhancing OTL. In the United States, policy makers have embraced test-based accountability as a means of somehow forcing schools to bring "all children" to a "proficient" level of achievement. By law, tests must be "aligned" to rigorous "academic achievement standards." Thus, standardized tests are relied upon to provide both the definition of successful learning and the means to assure that OTL is extended to all learners. Against this vision, many have criticized the conception of learning underlying large-scale testing programs and have argued that test-based accountability has, in fact, undermined many students' opportunities to learn.

It is rare to find any productive dialogue between the critics and the proponents of test-based accountability systems. By and large, testing advocates embrace a straightforward account of educational improvement. It is taken as a given that schools are doing a poor job – the goal of schooling is to impart skills to students, and it is common knowledge that many students graduate without having acquired the skills they need. Moreover, learning opportunities are unequally distributed, as attested by large differences in test score distributions (read uncritically as indicators of skill distributions) for groups defined by race and ethnicity, poverty or parent education, language background, or disability status. Tests indicate which students, individually and collectively, have or have not acquired the skills expected, and thus provide a quantitative index of school performance. If "teaching to the test" turns out

to mean cheating – teaching children the answers to actual test questions, for example – then that is clearly a problem. However, if "teaching to the test" means instruction aimed broadly at the skills tested, then that is precisely what schooling ought to be about. Therefore, testing guides educators' efforts, providing both the means and the measure of schooling success.

The various critics of test-based accountability offer no such single, tidy account. Many are concerned that the accountability testing reform model is *incomplete*, that "reform" turns out to mean accountability testing and little else. If educators are to be held accountable for bringing "all children" to "proficient," then teachers must have more and better preservice preparation and inservice professional development; school facilities must be upgraded; more and better instructional materials must be provided. In short, capacity building and accountability must go hand in hand if educational reforms are to succeed.

Others are concerned that the accountability testing reform model has been *poorly implemented*. Definitions of "proficient" vary wildly from place to place; achievement targets and timelines are often grossly unrealistic; the special needs of particular groups of learners are ignored. Alternately, requirements are implemented in such a way that successful schools are penalized and failing schools escape sanctions, or regulations have weakened the legislation to the extent that there is no longer any real accountability at all.

Others contend that the accountability testing reform model must employ *better or different tests*. The multiple-choice tests used from one year to the next resemble each other too closely, permitting teachers to ignore material not on the examination. Another argument holds that to assess the "thinking curriculum" (Resnick and Resnick 1992), one must set aside multiple-choice tests in favor of high-quality performance assessments, so that "teaching to the test" means providing meaningful, engaging instruction. A closely related critique, grounded in modern cognitive psychology, found its fullest expression in the National Research Council's recent publication, *Knowing What Students Know* (National Research Council 2001). This report makes the case that testing practice is largely out of touch with contemporary psychology, particularly cognitive science. It calls for enhanced collaboration between psychometrics and cognitive science and offers a carefully theorized vision of how collaboration "among researchers and assessment developers working at the intersection of cognitive theory and educational measurement" (p. 13) might enhance learning and assessment. Existing tests offer snapshots of isolated elements of factual and procedural knowledge. A new and better generation of tests could provide rich diagnostic information about students' complex, evolving knowledge structures.

Our book could be read as a critique of the logic and the effects of high-stakes testing, but that is not its primary goal. By and large, the authors represented are skeptical about the benefits of contemporary accountability testing, but we offer no simple list of recommendations for improvements. We share deep concerns about the direction and effects of contemporary accountability testing and would probably agree with many of the criticisms just cited: Accountability and capacity building must go hand in hand; state and federal laws and regulations are technically flawed; the tests used are of insufficient scope and quality, are disproportionately multiple-choice, and are based on outmoded theories of cognition and learning. That said, most of the authors represented here view these concerns and criticisms as almost beside the point. These sorts of problems are serious but in principle remediable. Our book interrogates instead the fundamental premise that schooling is about imparting skills and that OTL is simply a matter of ensuring universal skills acquisition. In various ways, different chapters call into question the meaningfulness and the trustworthiness of test scores as evidence about learning that matters. If rising test scores cannot be taken as evidence of better student learning, then test-based reforms cannot be relied upon to extend OTL to all learners.

PSYCHOMETRIC AND SOCIOCULTURAL/SITUATIVE PERSPECTIVES

One could, somewhat unfairly, characterize prevailing conceptions of learning and OTL as locating knowledge inside the heads of individual learners, privileging symbolic representation over embodied experience, and relegating the social dimensions of learning, however important, to the role of background or context in the business of measuring learning outcomes. These views comport with a psychometric perspective, with its roots in mathematics and psychology, and with the conceptual and statistical tools of educational testing as currently practiced. To be sure, not all psychometricians would agree with this caricature (e.g., Brennan 2005). As Mislevy's chapter in this volume explains, the conceptual and statistical tools of psychometrics may also be quite compatible with very different conceptions of knowledge and learning.

Although a version of the psychometric perspective is represented in this volume, most of the chapters represent one or another variant of either sociocultural theory or situated cognition – collectively the *SC perspectives*. (This *SC* label glosses over important, some would say fundamental, theoretical differences. It is used here in contrast to the similarly simplified *psychometric*

label.) SC perspectives locate knowledge not primarily inside the head, but rather in the relationship between a person and an environment in which that person thinks and feels, acts and interacts. The SC approaches represented here are variously rooted in the learning sciences, sociolinguistics, sociology, and cognitive anthropology. They differ in focus but are generally compatible. Some highlight social interaction and participation, characterizing learning by individuals and groups through interactions within an *activity system* and conceiving learning and knowing from the perspective of distributed cognition. Others focus more sharply on notions of *embodied cognition*, attending to the ways in which perception and cognition are connected with human bodies embedded in a material and social world, including the social and cultural contexts of students' lives outside school. These chapters raise questions as to what OTL means, how it might be productively conceptualized at different levels of the educational system, what constrains and enables it, and how it can be assessed in a way that supports rather than undermines learning. Subsequent chapters pursue these questions, offering accounts of classroom and large-scale assessment and their functions in different activity systems for different purposes, informed by the SC perspectives developed in earlier chapters.

ORIGINS OF THE COLLABORATION

The idea for what was originally called "The Idea of Testing Project" first arose in a conversation between Pullin and Moss in a coffee shop in Montreal at the annual conference of the American Educational Research Association (AERA) in April 1999. With the year 2000 looming, a conference session on "Testing in the 21st Century" had set forth an optimistic vision, comfortably focused on the intersection between psychometrics and cognitive psychology. That prompted Moss and Pullin to wonder what "testing" might look like, and what its impact on schooling might be, if it were informed by conversations with scholars from other fields as well. They invited Haertel and Gee to join them in exploring that question. Following several desultory conversations during the ensuing months, the group of four "organizers" met for half a day the following year in Chicago at AERA 2000. At that point, the organizers developed a list of colleagues to invite to join in their exploration – all of whom said yes – and decided to seek funding. Young, from the Spencer Foundation, was enthusiastic and supportive, so the Spencer Foundation came to support the work of the group, which now grew to include Beach, Greeno, Lee, Mehan, and Mislevy. The four organizers met for the first time with Young at Stanford in early 2002. She became the fifth organizer;

and Mosher (consultant to the Spencer Foundation) also joined the project. A description of the project as it was originally conceptualized appears on the Spencer Web site; the manuscript that informed our original proposal to Spencer, "The Idea of Testing: Psychometric and Sociocultural Perspectives" is published in *Measurement: Interdisciplinary Research and Perspectives* (Moss et al. 2005). Then, like all truly collaborative efforts, the project evolved as our new collaborators joined the conversation.

The full group met for the first time in San Diego in April 2002. In addition to outstanding support from the staff of the Spencer Foundation, Andrew Ho, at that time a graduate student at Stanford University, offered his assistance as note taker and synthesizer. Given the group's multidisciplinary membership, a primary goal for the first meeting was to learn about each other's work. To that end, all participants were asked to share an example of their work that would introduce to a multidisciplinary audience the perspectives and practices of their field. The organizers also requested that everyone read and prepare to comment on a common text, *Knowing What Students Know* (National Research Council 2001). Participants were also asked to consider the following questions: How do you see your work or that of others in your field illuminating the conceptualization, practice, and use of assessment? What alternatives or analogues to assessments based in psychometrics/cognitive science do you imagine might be developed? How might your work or that of others in your field critically examine the way assessments shape and are shaped by the social contexts in which they are produced and received?

As a result of that three-day discussion, it became clear that OTL was at the heart of all of our interests. Thus, the group decided that exploring the concept of OTL and its relationship to teaching and assessment would provide an important and productive focus for our joint work. We wanted to develop a manuscript that would illustrate, for a broad educational audience, how the concept of OTL could be productively conceptualized, studied, and enacted from the intersections of our fields of work. Thus, although rethinking assessment has remained one important goal of our work, our interest in assessment has come to be framed in terms of the intersection between assessment and OTL.

The group met seven times between 2002 and 2005. Four of the meetings, like the first, were three-day meetings in settings where both formal and informal conversations could occur. Agendas for these meetings are posted on the Spencer Web site (www.spencer.org/publications). We also met just before each annual AERA meeting from 2003 to 2005 and presented our work at two of those meetings. The organizers relied on e-mail, conference calls, and additional meetings to plan experiences for the group that would foster the

collaboration. We became friends as well as colleagues and began to develop a language for talking across our different research perspectives that has served us well as we have worked to make our ideas available to a broader audience.

The early meetings followed the pattern of the first meeting, where participants shared their own work, read common "texts," and used those texts as springboards for sharing their perspectives on OTL. At the second and third meetings, we focused on a case prepared by Pullin about the Massachusetts student-testing program, on videotapes and other artifacts from classrooms studied by Lee and Greeno and on complex context-sensitive performance assessments studied by Mislevy. Conversations about these common texts were crucial to our developing a common language for talking about OTL. As the meetings progressed, we began to develop drafts of chapters for this volume. As these initial chapters evolved, the need for new chapters arose, and the manuscript came to take the form in which it now appears.

OUR ARGUMENT IN A NUTSHELL

How might our collaboration across sociocultural and psychometric perspectives extend conventional discourse in the educational research and policy communities about learning, OTL, assessment, and the relationships among them? Like all concepts, the meanings of learning, OTL, and assessment are deeply situated in people's experiences of them in educational practice. In chapter 2, Pullin and Haertel set the stage for our argument by situating conventional conceptions of OTL and its relationship with testing, historically, in terms of access to content, resources, and instructional processes, typically assessed via standardized tests and related indicators that students have been exposed to the content tested. We contend that these understandings and indicators of OTL are at best inadequate. Opportunities to learn do not exist for learners who cannot take advantage of them. Questions of OTL cannot be equitably addressed by looking simply at content, resources, or processes of instruction, or even by looking at all three. Although these are surely important, one must look further at the relationships among particular learners and these elements of their learning environments. Furthermore, neither tests in common use (as noted above) nor any particular assessments (including the more powerful ones described in subsequent chapters) should or could embody a vision of learning sufficient to orient educational practice. Much more is needed.

Our argument begins with a theory of learning informed by sociocultural, situative, and sociological perspectives alongside more familiar cognitivist ones. What is it (we hope) students are learning, how, and why (Engeström

2001)? Using the work of many contemporary educational theorists, we endorse a theory of learning that moves beyond an emphasis on acquiring information and skills to an emphasis on rich conceptual understanding, reasoning, and problem solving in a domain. However, sociocultural perspectives push us still further by emphasizing the deeply situated nature of learning in the social contexts and experiences in which it is put to work. This has multiple implications for conceptualizing and supporting learning. *First*, as Gee argues (this volume, chapter 8), "Any actual domain of knowledge, academic or not, is first and foremost a set of activities (special ways of acting and interacting so as to produce and use knowledge) and experiences (special ways of seeing, valuing, and being in the world). Physicists *do* physics. They *talk* physics. And when they are being physicists, they *see* and *value* the world in a different way than do non-physicists" (p. 200). The same goes for any other domain of knowledge. Of course, students "learn" to *do* school – to act, talk, perceive, and value in the ways they experience there – but unless these experiences are preparing them to act, and to learn, in ways that are valued outside school, schools are not serving them well.

Second, this emphasis on activities and experiences reminds us that learning entails *interaction* between learners and the other people and tools in their environment. By tools, we mean both physical tools (such as calculators and dictionaries, pens and paper, computers, and so on) and conceptual tools (like domain-specific language and representations), all of which mediate learning. Learning can only be developed, enacted, supported, or observed in interaction between learners and these elements of their environment. Sociocultural perspectives provide multiple lenses through which we can understand learning as the interaction, or relationship, between learners and their environments. In addition to asking questions about (changes in) what they know and can do or how they engage in reasoning and problem solving, we can also ask questions about (changes in) what they mean when they act and interact (Gee, this volume, chapter 4); about how they are using domain-specific language, representations, and culture (Gee, this volume, chapter 4; Lee, this volume); about how they are participating in the activities of their local (school and other) communities; about the positions they are enacting with respect to one another and the subject matter (e.g., raising questions or simply answering them? challenging interpretations or simply reproducing them?) (Greeno and Gresalfi, this volume); about the resultant identities they are developing; about the social networks in which they participate (Mehan, this volume); and so on. Of course, sociocultural theorists also privilege a particular set of answers to these questions. These answers involve knowledge useful in the world; complex, authentic, and domain-specific forms for

reasoning and problem solving; identities and positions that enact conceptual agency (choices in formulating questions, choosing methods, explaining solutions); critical reflection on multiple values and worldviews; a view of learning as important to one's growth; and so on (see, Moss, this volume, Figure 9.2, for one summary of the criteria valued by the authors).

Third, sociocultural studies draw our attention, explicitly, to what learners – with minds and bodies, home and peer cultures and languages, previous learning experiences, interests and values – bring to their learning environments and how that shapes their interactions with those learning environments. Thus, all of the questions about meaning, experience, language, culture, positioning, identities, and so on need to be asked about the interactions between particular learners and their learning environments as they evolve over time (their "learning trajectories").

Fourth, providing adequate OTL means scaffolding the dynamic interaction between students' "forms of knowledge and ways of using language [from their] everyday experiences in families and communities" (Lee, this volume, p. 136) and the resources of their learning environments. Furthermore, as Gee (this volume, chapter 8) notes, this can't be accomplished by simply "telling" students what to do or by "turning them loose" in the domain's activities; rather, what is needed is a combination of "immersion and guidance." Thus, a socioculturally informed theory of learning entails a well-designed curriculum with a coherent learning trajectory, connections that build on students' prior knowledge and experiences, explicit instruction that involves connections between academic and everyday language, just-in-time feedback as experience is unfolding (Gee, this volume, chapter 8), meta-conversations about how you know what you know (Lee, this volume), activities that permit meaningful participation in the group's work (Greeno and Gresalfi, this volume), and so on. It also entails social scaffolds beyond the classroom that support academically oriented friendships; productive connections among home, school, college, and business; explicit socialization in how to participate in these social networks; and so on (Mehan, this volume).

The chapters that follow are rich with extended concrete examples of practice consistent with this theory of learning – in classrooms, schools, and districts or other external organizations that support multiple schools – along with the theoretical tools for analyzing these examples of practice to support the design of learning environments.

How can we conceptualize assessment in light of the complex demands of a socioculturally informed learning theory? We consider classroom assessment as well as assessment that crosses boundaries from the classroom to the school, district, and beyond to inform professional learning, evaluation,

and accountability. This portion of our argument entails three underlying premises about assessment: assessment should support teachers and other educational professionals in learning how better to support students' and one another's learning; professionals working in different contexts (classrooms, faculty study groups, district offices, etc.) have different information needs; and assessment and the routines that surround it do far more than provide information – they shape people's understanding about what is important to learn, what learning is, and who learners are.

Our reconceptualization of assessment is deeply informed by "reading" accounts of educational practice, consistent with our learning theory, for how evidence is used to monitor and support learning. Such readings suggest that assessment can be most productively conceptualized around the particular questions or problems that evidence is needed to address rather than around a particular instrument or activity ("a test") and the circumscribed evidence it provides (Moss, Girard, and Haniford 2006). The types of questions addressed include classroom teachers' moment-to-moment questions about "what to do next"; their questions about how to plan and enact lessons, revise curricular routines, solve particular pedagogical problems, or inform parents and guardians about students' learning; and the questions school, district and other educational leaders face about allocating resources, planning professional development, selecting and refining curricula, developing policies, and evaluating the impact of these choices. Answering any particular question entails multiple sorts of evidence about the evolving interaction between learners and their learning environments (their learning trajectories). Thus, our conception of assessment includes formal assessments that we recognize as "*an* assessment," as well as informal evaluations and judgments, both tacit and explicit, that routinely occur in classroom interactions and in other educational settings (Jordan and Putz 2004).

This conception of assessment certainly recognizes the need for large-scale standardized[2] assessments to assist in addressing many questions that educators face. Mislevy (this volume) offers theory and examples of how large-scale standardized assessments, far more consistent with socioculturally informed theories of learning, might be developed, implemented, and evaluated. Other chapters provide examples of professional environments where large-scale assessments may be useful alongside other evidence and practices of interpretation, decision making, and learning (Lee; Mehan; Moss, Girard, and Greeno, all this volume). It is important to note, however, that standardized assessments, like other activities, entail social situations, including routines, roles, and responsibilities for different people, and conceptions of what counts as progress. Questions about the generalizability of knowledge from

students' participation in the social situation of the standardized test to their participation in the many other social situations where knowledge is used cannot be answered satisfactorily from within the framework of standardized testing alone (Gee, this volume, chapter 4).

Furthermore, it is important to note that when standardized assessments are intended to inform particular decisions in local contexts, they can be meaningfully interpreted only in light of locally relevant evidence. The results from a standardized assessment form, at best, a good hypothesis about a particular case (student, school) that must then be evaluated from the ground up by local interpreters who are using it in their own contexts of work. This means that *any* practice of assessment depends, in large part, on the capacity of local interpreters (e.g., teachers making instructional decisions; school and district leaders making decisions about allocating resources; students whose understandings of themselves may well be shaped by the assessment) to interpret it in light of other relevant evidence and in light of the questions or problems they are using it to address. The essential role of local interpreters illuminates the need for professional development and for a professional environment that supports this sort of inquiry into students' learning. Furthermore, it calls into serious question prominent practices of large-scale assessment that enforce particular decisions or actions based solely on the scores a particular individual or organization has received (Mislevy, Gee, and Moss, in press).

Our socioculturally informed learning theory reminds us that assessment practices are inevitably elements of learning environments that shape (enable and constrain) learning and opportunities to learn. Any robust practice of assessment, consistent with high-quality learning and OTL, must turn the lens of inquiry back on itself. All of this suggests that developing useful assessment practices – practices that function productively at different levels of the educational system – will depend on richly contextualized understandings of what information is needed, how it is used, and the effects of this use.

We hope that we have succeeded in providing readers with a new set of theoretical resources for conceptualizing, enhancing, and evaluating OTL and for rethinking the theory and practice of assessment through new disciplinary lenses. We hope to introduce educators, researchers, school leaders, and policy analysts who are new to sociocultural perspectives to the power of these theoretical resources for conceptualizing, enhancing, and assessing OTL; to encourage those who already draw on sociocultural resources to focus attention on OTL and assessment in ways that will impact educational policy and practice; and, more generally, to nurture dialogue and collaboration among