The Risks of Coherence[1]

PAMELA A. MOSS

The major premise underlying this volume is that enhanced coherence between classroom assessment and system-level accountability is a good thing—assuming, of course, that the assessments are consistent with sound learning principles. Working from this assumption, four of the five source chapters provide promising assessment and accountability tools that enable and support enhanced coherence; the fifth source chapter, by Black and Wiliam, appears to enact a somewhat different relationship between assessment and accountability that I will address later in my commentary. Although acknowledging the potential benefits of coherence, I draw on critical and sociocultural theory to explore the risks of enhanced coherence, even when the standards, curriculum, instructional practices, and assessments through which it is enacted represent sound pedagogy. Are there ways in which enhanced coherence can undermine important outcomes? In what ways might there be too much of a good thing?

What Kind of Tools Do the Source Chapters Provide Toward What Kind of System?

The focus of my criticism is not on the usefulness of the individual sets of tools but rather on how they might be incorporated into an educational system at the district, state, or national level. What kind of system are they likely to enable, and with what effects?

Three of the five source chapters offer rich examples of assessments that can be simultaneously used in the classroom and aggregated to provide system-level indicators (Foster & Masters, Chapter 3; Frederiksen & White, Chapter 4; Wilson & Draney, Chapter 6). These

Pamela A. Moss is Associate Professor, 4220 School of Education, University of Michigan, Ann Arbor, MI 48109-1259; e-mail: pamoss@umich.edu. Her areas of specialization are at the intersections of educational assessment, validity theory, and interpretive social science.

assessments share many strengths: they offer opportunities for students' engagement in active learning around meaningful activities; they scaffold teachers' (and students') judgments of students' work; they provide the means of documenting progress (or development) over time within classrooms; and they can be aggregated and compared over time and location to provide system-level information about status and progress on the particular learning outcomes. If enacted as intended, there is no question that the information available about student learning outside the classroom would be richer and that the press to perform well on these curriculum-embedded assessments would be far less detrimental than with conventional state-sponsored paper-and-pencil tests. A fourth chapter (Smithson & Porter, Chapter 5) provides an intentionally more generic set of indicators for examining the degree of alignment among standards, curricula, assessments, instructional content, and so on. Although initially intended for policymakers to address questions about policy implementation in the classroom, the system can be and has been used by teachers and local administrators at various levels of the educational system to assist with decisions about curriculum and instruction. Summary graphs and statistics display the degree of alignment between different parts of the system. The fifth chapter, by Black and Wiliam (Chapter 2), focuses more on a professional development model that supports teachers in working collaboratively to develop the practice of formative assessment in their classrooms. Although one could imagine these professional development practices working in conjunction with the tools described in the other four chapters, as Black and Wiliam enact them they support different assessment practices in different classrooms, and coherence with indicators at other levels of the system is indirect and post hoc.

The sort of coherent system Wilson imagines in the opening chapter of this volume involves "the integration of assessment frameworks and methods across all levels of the assessment system, from the classroom to the system level" (p. 3). He called for system-level assessments that are directly usable in the classroom, that could serve as a model for teachers designing their own assessments, and that could "give an appropriate place in the accountability system to the professional knowledge and standing of teachers" by incorporating "teacher judgments of student performance that are the classroom reflection of student learning" (p. 3). Wilson pointed us toward the National Research Council's *Knowing What Students Know* (Pellegrino, Chudowsky, & Glaser, 2001; hereafter KWSK) as a theoretical resource, drawing on psychometrics and cognitive science, for coordinating our various efforts. As the

authors of KWSK describe it, "a vision for the future is that assessments at all levels—from classroom to state—will work together in a system that is comprehensive, coherent, and continuous" (p. 9). KWSK, along with four of the five source contributions to this volume, appears to be consistent with this vision. I will call this vision of a "comprehensive, coherent, continuous" system *coherence-through-alignment*. Coherence-through-alignment appears to promote commonality in curriculum frameworks, language, and methods across classrooms and throughout the levels of the educational system (albeit with differing degrees of specificity). The brunt of this commentary, then, explores the risks of coherence-through-alignment. Then, I will sketch out some principles for an alternative vision of coherence—*coherence-through-negotiation-of-meaning*—a phrase appropriated from Wenger (1998). Although most congenial to Black and Wiliam, I imagine how the tools provided in all five of the source chapters might work within a vision of coherence-through-negotiation-of-meaning, a vision with its own risks and benefits that supports more diversity in classroom practice. I do not mean to portray alignment and negotiation of meaning as antonyms. In fact, negotiation of meaning can result in alignment and alignment can encourage negotiation of meaning such that local actors come to "own" the concepts provided in the assessment system. The distinction between alignment and negotiation of meaning is intended to signal a difference in emphasis in the way in which coherence is promoted.

It is important to note that the part of the vision Wilson described— as "harnessing this flood of assessment information . . . of learning *within* the classroom, and as the source of crucial information flowing *out* of classrooms"—is not illustrated in any of the five source chapters (p. 2). Although it appears that teachers were actively and productively involved in the design of the three classroom-assessment systems illustrated, subsequent use of these systems in other classrooms provides a far more limited role for teacher input. Thus, for most classrooms, these assessments will arrive already developed by others and will be implemented by the classroom teacher, whose productive role in the assessment will likely be restricted to activities like judging student work using the guidelines provided, perhaps developing tasks consistent with a predetermined template, choosing the time of implementation, and of course, supporting students in making progress toward the goals. Smithson and Porter are "skeptical" about the possibility of using classroom-developed assessment for accountability purposes: "The challenges to building such a system in a way

that guarantees some standardization of results across classrooms and teachers seem insurmountable to us" (p. 128). Earlier work by Wilson (1994, reprised in chapter 1 of this volume) offers teachers the possibility of considerably more flexibility in designing assessments. Even here, however, to be viable, these assessments must enact a common language and curriculum framework across classrooms.

Whether the classroom assessments are externally imposed or locally shaped and harnessed for the larger system, they make classroom practice externally visible in a way—comprehensive, coherent, and continuous—that has not been previously possible. What are the potential consequences of such a system of routine visibility?

A Critical Response to Coherence-Through-Alignment

Critical theory and the research practices it supports illuminate a number of important questions that might be raised about the vision of coherence-through-alignment coupled with potential for "continuous" visibility. At the heart of these questions is an overarching concern with the effects of the assessment and accountability system on the people involved: "*What kind of people do they foster?*" (Scott, 1998, p. 348).

What Is Critical Theory?

Critical theory, as I am using it here, encompasses a diverse constellation of practices and theories that share some common features.[2] An important aim of critical research is to illuminate the dialectical relationship between social structures (like assessment and accountability systems) and local practices (like interactions among principals, teachers, and students or among school board members or legislators)—how each constructs, shapes, or challenges the other. Critical researchers typically ask questions about how social structures and local practices influence conceptions of knowledge or progress, conceptions of self and others, and differences in access to society's good. Analyses involving *different* perspectives, practices, and contexts are central to critical research because they illuminate the categories of thought and action "we" take for granted, situate them in the social historical conditions in which they arose, and allow us to imagine how things might be otherwise. A primary aim of critical research is social change. Although critical research shares this aim with many conceptions of social research, it privileges practices that enable change, not primarily through external controls but rather through action orienting self-reflection. In other

words, by helping us become aware of the social forces that shape our categories of thought and action, we can make better informed decisions about how and whether to change our practice (which in turn can affect the social structures within which we live). Critical theory reminds us, as well, that these sorts of questions are as important to ask about social scientists as they are about the people we study (Bourdieu, 1988, 1991; McCarthy, 1994): "social researchers are themselves engaged in socially situated forms of action," and "bringing this to consciousness and examining its implications" is a central role of critical theory (McCarthy, pp. 14-15).

What sorts of questions might be raised from a critical theoretic perspective about a system that privileges coherence-through-alignment, and what evidence might be gathered to address them?

The Genesis and Effects of an Indicator System

I begin with a historical narrative of the use of an indicator system in quite a different context. James C. Scott opens his book *Seeing Like a State* with a narrative of the development of scientific forestry in late eighteenth-century Prussia and Saxony—a narrative that also serves as a metaphor for examining state-sponsored indicator and planning systems, which make local practice visible, and hence manipulable (Scott, 1998, p. 2), from afar. Below, I have cobbled together excerpts from Scott's narrative. I invite readers to imagine the questions that this narrative might prompt us to ask about assessment and accountability programs.

The development of scientific forestry in Prussia and Saxony served the state's need to monitor and manage its resources:

The early modern European state, even before the development of scientific forestry, viewed its forests primarily through the fiscal lens of revenue needs [although] . . . other concerns—such as timber for shipbuilding, state construction and fuel for the economic security of its subjects—were not entirely absent from official management. These concerns also had heavy implications for state revenue and security. (pp. 11-12)

Increasingly precise measurements in scientific forestry enabled the state to engage in systematic planning by achieving a synoptic view of the forest:

The new forestry science was a subdiscipline of what was called cameral science, an effort to reduce the fiscal management of a kingdom to scientific principles that would allow systematic planning. . . .

The final result of such calculations was the development of elaborate tables with data organized by tree size and age under specified conditions of normal growth and maturation. By radically narrowing his vision to commercial wood, the state forester had, with his tables, paradoxically achieved a synoptic view of the entire forest. (pp. 14-15)

Scott notes, however, that much was missing from this abstracted vision of the forest:

Missing, of course, were all those trees, bushes, and plants holding little or no potential for state revenue. Missing as well were all those parts of trees, even revenue-bearing trees, which might have been useful to the population but whose value could not be converted into fiscal receipts. Here, I have in mind foliage and its uses as fodder and thatch; fruits, as food for people and domestic animals; twigs and branches, as bedding, fencing, hop poles, and kindling; bark and roots, for making medicines and for tanning; sap, for making resins; and so forth. (p. 12)

The set of state-sponsored indicators then suggested strategies for forest management:

The fact is that forest science and geometry, backed by state power, had the capacity to transform the real, diverse, and chaotic old-growth forest into a new, more uniform forest that closely resembled the administrative grid of its techniques. To this end, the underbrush was cleared, the number of species was reduced (often to monoculture), and plantings were done simultaneously and in straight rows on large tracts. These management practices . . . produced the monocultural, even-age forests that eventually transformed the . . . abstraction to reality. (p. 15)

This transformation of the diverse, old growth forest to the monocultural, even-age forest made centralized management more viable:

The more uniform the forest, the greater the possibilities for centralized management; the routines that could be applied minimized the need for the discretion necessary in the management of diverse old-growth forests. (p. 16)

And, it resulted in spectacular short-term success:

In the short run, this experiment in the radical simplification of the forest to a single commodity was a resounding success. . . . The productivity of the new forests reversed the decline in the domestic wood supply, provided more uniform stands and more usable wood fiber, raised the economic return of forest land, and appreciably shortened rotation times (the time it took to harvest a stand and plant another). (p. 19)

It was not until after the second round of planting that negative consequences of the system became apparent:

The negative biological and ultimately commercial consequences of the stripped-down forest became painfully obvious only after the second rotation of conifers had been planted. . . . An exceptionally complex process involving soil building, nutrient uptake, and symbiotic relations among fungi, insects, mammals, and flora . . . . was apparently disrupted, with serious consequences . . . [including] thinner and less nutritious soils, . . . more vulnerab[ility] to massive storm felling, . . . and a favorable habitat for all the "pests" which were specialized to that species [of trees]. (p. 20)

Scott attributes the negative consequences in large part to the radical simplicity of the "scientific forest" (p. 20).

Any unmanaged forest may experience stress from storms, disease, drought, fragile soil, or severe cold. A diverse, complex forest, however, with its many species of tress, its full complement of birds, insects, and mammals is far more resilient—far more able to withstand and recover from such injuries—than pure stands. Its very diversity and complexity help to inoculate it against devastation. . . .
     The simplified forest is a more vulnerable system, especially over the long haul, as its effects on soil, water, and "pest" populations become manifest. Such dangers can only partly be checked by the use of artificial fertilizers, insecticides, and fungicides. (pp. 21-22)

This simplified, vulnerable forest led to the development of "restoration forestry" and the need for sustained outside intervention to attempt to remedy the consequences:

"Restoration forestry" attempted with mixed results to create a virtual ecology, while denying its chief sustaining condition: diversity. . . . Given the fragility of the simplified production forest, the massive outside intervention that was required to establish it—we might call it the administrators' forest— is increasingly necessary in order to sustain it as well. (pp. 20, 22)

This narrative of the effects of centralized planning coupled with routine visibility is not an isolated case. Scott's book takes us through case after case—collective farms, planned cities, and so on—that trace the genesis and effects of large, state-sponsored systems that make local practice "legible—and hence manipulable from afar" (p. 2). Although acknowledging the egalitarian and emancipatory intent of such systems and the unjust social orders they attempted to replace,

Scott documents a consistent pattern of failure for social planning that also does not nurture local knowledge and practical skill.

Without denying the incontestable benefits either of the division of labor or of hierarchical coordination of some tasks, I want to make a case for institutions that are instead multifunctional, plastic, diverse, and adaptable—in other words, institutions that are powerfully shaped by metis [that is, "forms of knowledge embedded in local experience" (p. 311)]. . . .

   To any planned, built, or legislated form of social life, one may [ask]: to what degree does it promise to enhance the skills, knowledge and responsibility of those who are a part of it? On narrower institutional grounds, the question would be how deeply that form is marked by the values and experience of those who comprise it. (pp. 353, 355)

   Scott's narrative raises questions about some potential consequences of a system that privileges coherence-through-alignment and routine visibility: it does not illuminate (at least as currently illustrated) features of the local context that are essential to decision making; it provides opportunity (whether intended or not) for a large-scale *intervention* into local social systems (districts, schools, and classrooms); it supports a single set of centrally determined learning goals across the system (albeit at different levels of specificity for different approaches, with Wilson, 1994, offering the most flexible alternatives for teachers) and thus limits diversity; it enacts (whether intended or not) a particular view of the social world of schools with different responsibilities for different actors and thus reduces opportunity for local actors to influence or shape the system by which they are judged. I will consider each of these issues in turn. In so doing, I will draw heavily on Bryk and Hermanson's (1993) benchmark chapter on educational indicator systems, which uses different theoretical perspectives to raise many of the same issues.

Assessment Systems Are, at Best, Partial Representations

   With the exception of Smithson and Porter, there is not much attention paid in the source chapters to how the available information about student learning will be used by decision makers outside the classroom or about what additional information is needed to enable sufficient understanding to make good decisions. In this sense, the set of indicators illustrated in any one of these chapters is incomplete. This is not, in itself, a criticism, as none of the authors intended to illustrate a "complete" indicator system. Questions of the purposes to which the system will be put, however, are crucial to understanding and anticipating its effects; and so I consider them here even though

they may reflect concerns about issues not addressed in the source chapters.

Bryk and Hermanson (1993) worry that "although discussions about educational indicators may acknowledge the diverse aims of education, it is nonetheless common to presume that one can focus on the "core of schooling"—academic achievement and the processes instrumentally linked to it—while ignoring everything else" (p. 456). For instance, they note, "there is growing evidence . . . that the social structure of schools influences student engagement and teacher commitment, both of which are linked to students' academic achievement" (p. 456). These concerns raise issues of fairness as well as sufficiency: "Since the information system will shape the nature of . . . contests [over schools' aims and methods], concern must focus on issues of fairness toward all interests in these debates. Without care and sensitivity, seemingly technical decisions can advantage some interests and disadvantage others" (pp. 473-474).

In the classroom, the partial nature of an indicator system is less of a problem, since the teacher has ongoing and intimate knowledge of the social world of the classroom. A student's performance on a particular assessment can be interpreted in light of the teacher's goals, students' interests, established norms and routines, the past experiences of students inside and outside the classroom, the situated meanings of words in the classroom (Gee, 1999), the learning opportunities the assessment follows, all that was said as the assessment was undertaken, the particular uses to which the assessment will be put, and so on.

Harnessing classroom assessment for use outside the classroom entails stripping away the context. Once removed from the classroom to another social context, these indicators cannot be understood in the same way; they require "recontextualization" in the new social context in order for their salience to be adequately understood (Bryk & Hermanson, 1993, p. 458). Bryk and Hermanson outline what a more complete model might look like:

In outline form, a complete model would require a multilevel formulation, which includes at a minimum, classroom-level concepts about student learning, teacher pedagogy, and classroom practices; school level concepts about curriculum organization, academic and disciplinary policy, quality of social relations, adequacy of available resources, and school leadership; and, similarly, key concepts that capture the major support and administrative functions at the district and higher levels of government. . . . The educative influences of larger cultural forces and other social institutions . . . would also have to be included. In addition, the model component for student experiences and outcomes

would need an explicit developmental dimension [something, I should note, that three of the source chapters in this volume provide]. The interrelationships among student experiences in the first five years of life, learning in elementary and secondary schools, and adult outcomes including active citizenship, workplace productivity, and personal well-being would also have to be specified. (p. 462)

Bryk and Hermanson caution, however, that "our scientific knowledge about schooling is partial" (p. 462) and that it is "nothing less than a seductive delusion to presume that we could develop a comprehensive indicator model of sufficient intelligence to support instrumental use" (p. 462). By "instrumental" they refer to the use of information "to externally control schools through instruments such as rule writing, administrative sanctions, and incentives" (p. 453). They call, instead, for an "enlightenment" model of indicator use (an idea I will return to in the conclusion).

In this view indicators are of value in that they can broaden our understanding of problems and catalyze new ideas. They can signal new problem areas, offer conceptual frames in which to discuss these issues, provide some useful information for initial brainstorming about possible solutions, and, more generally, inform the broader public. (p. 465)

This use of indicators appears more consistent with what is imagined by the editor and the authors of the source chapters in this volume. Wilson calls for assessment and accountability systems that "give an appropriate place in the accountability system to the professional knowledge and standing of teachers" (p. 3); Smithson and Porter note that "successful data use . . . appear[s] to match many of the characteristics associated with professional learning communities. . . . These characteristics have less to do with the technical skills of interpreting and using data and more to do with the professional culture of the school and the ability of faculty to engage in collegial discussions about practice" (p. 123).

Indeed, practices envisioned in these chapters stand in stark contrast with the instrumental vision of assessment and accountability enacted, for instance, in the federal No Child Left Behind Act of 2001 (Pub. L. 107-110) (see Linn, 2003, and Linn, Baker, & Bettebenner, 2002, for critical reviews). There, schools that do not meet their states' goals for annual yearly progress, in terms of overall and subgroup achievement gains, face increasingly severe requirements and sanctions. Clearly the source chapters in this volume intend a far more

enlightened vision of reform. However, even assuming the assessment and accountability tools were *intended* for "enlightenment" rather than "instrumental" purposes, important questions remain about the effects of an assessment system that permits comprehensive and continuous visibility from outside and above. That is the issue I turn to next.

Externally Imposed Assessment and Accountability Systems Entail Interventions into Local Social Systems

Scott's case studies compel us to begin with the testable assumption that any externally imposed assessment and accountability system is an intervention into the social system it seeks to describe. The assessment system highlights certain activities or outcomes as important and worthy of particular attention; it offers a particular vision of progress; and it gives people a language to use in understanding themselves and others. It is not and cannot be a neutral language that simply describes reality: symbolic representations, once incorporated into local discourses, shape the social reality they describe in theory- and value-laden ways.

Let's look for the moment at what are arguably the two most general sets of conceptual tools this volume has to offer: Smithson and Porter's language for describing instructional content and KWSK's assessment triangle. Smithson and Porter characterize their language for collecting descriptions of classroom practice as a "theory neutral tool" (p. 125) with respect to pedagogical and curricular orientation. However, the resulting language, as they describe it, is "based on a two-dimensional model of instructional content consisting of topics and cognitive demand" (p. 105). To what extent can this language be "theory neutral"? Wilson (Chapter 1) asked all contributors to use the assessment triangle from KWSK in order to "consider this complex issue in a coherent way":

According to KWSK, assessment consists of 1) a cognition aspect (the model one has of a student's cognition); 2) an observation aspect (the methods one uses to assess the student's cognition); and 3) an interpretation aspect (the methods one uses to relate the observations to the cognition model). (p. 2)

Although neither Wilson nor the authors of KWSK portray this as a theory-neutral tool, they propose it as a conceptual tool with general relevance for assessment designers. In fact, the authors of KWSK are somewhat more prescriptive: "These three elements," they assert, "must be explicitly connected and designed as a coordinated whole. If

not, the meaningfulness of inferences drawn from the assessment will be compromised" (Pellegrino, Chudowsky, & Glaser, 2001, p. 2).

In what ways might these conceptual tools shape the social contexts in which they are used? Perhaps the best way to address this question is to contrast it with a vision of learning that this language ignores. From a sociocultural perspective,[3] learning involves not only acquiring new knowledge and skill, but taking on a new identity and social position within a particular discourse (Gee, 1999; Gee, Hull, & Lankshear, 1996) or community of practice (Wenger, 1998).[4] As Wenger puts it, learning "changes who we are" (p. 5) "by changing our ability to participate, to belong" (p. 227) and "to experience our life and the world as meaningful" (p. 5). Thus, learning is perceived through changing relationships among the learner, the other human participants, and the tools (material and symbolic) available in a given context (Beach, 1999; Chaiklin & Lave, 1993; Cole, 1996; Gee; Gee, Hull, & Lankshear; Mehan, 1993; Wertsch, Del Rio, & Alverez, 1995). From this perspective, evidence of learning cannot just focus on the cognition of the student—what takes place inside the head of the learner, as the assessment triangle would imply; rather it must focus on the *interaction* among the learner, the other actors (e.g., teachers and students), and the symbolic and material resources available. The situation rather than the individual becomes the unit of analysis. As Mehan (1998) notes, "By moving beyond the states and traits of individuals to social situations as the unit of analysis . . . [students' performances can be] recast as collaboratively constructed and continuously embedded in face-to-face interaction in social environments" (pp. 251, 254).

Seen from this perspective, both the assessment triangle and the language for describing instructional practice ignore the social dimensions of learning, the ways in which learning is shaped and demonstrated in interaction with other human beings and with the symbolic and material resources available in the social context. Even Smithson and Porter's notion that cognitive demands and topics can be crossed in a two-dimensional model, such that the meaning of a cognitive demand remains constant across applications to different content, is controversial. (See Beach, 2003, Gee, 2003a, and Greeno & Haertel, 2003, for responses to KWSK from a sociocultural perspective.)

There is, I would argue, no escape from the theory- and value-ladeness of any conceptual tool or indicator system or from the responsibility (acknowledged or not) for having intervened in a social system. When coupled with state authority, and associated rewards and sanctions, as with No Child Left Behind, the power of the assessment

system to shape the social realities it describes is greatly enhanced. However, *even in the absence of such tangible incentives*, the power to shape practice remains: "numbers [that] create and can be compared with norms . . . are among the gentlest and yet most pervasive forms of power in modern democracies" (Porter, 1995, p. 45). "They provide legitimacy for administrative actions, in large part, because they provide standards against which people judge themselves" (p. 45). And so it is appropriate to ask about any set of indicators: What does it permit or privilege and what does it exclude or ignore? What does it illuminate, invite, encourage and what does it marginalize or relegate to the background as less important in demonstrating progress or competence? In the next two sections, I focus first on the potential effects of the privileged learning outcomes—knowledge and skills—and second on the identities and social positions they offer actors (teachers, students, principals, district administrators, policymakers, legislators, and so on).

The Indicators Reflect a Single Set of Learning Outcomes

As with any sound curriculum framework, embedded in each of the assessment systems proposed (Foster & Masters; Frederiksen & White; Wilson & Draney) is a coherent set of learning outcomes. The vision of assessment and accountability privileged in this volume—of "the integration of assessment frameworks and methods across all levels of the assessment system, from the classroom to the system level" (Wilson, Chapter 1, p. 3)—implies that the same set of learning outcomes will be expected of all teachers and students within a given system. Presumably this will occur at the state level, but given development costs, the same programs may well be used in many states. While this will likely enhance the quality of education in many schools, it will also likely decrease diversity in learning experiences within the educational system. Of course, all standards-based reform efforts promote common learning goals within the systems in which they are used. However, the comprehensive and continuous visibility the envisioned systems permit will likely give them far more power to reshape the local environment in their own image.

Will we—educators, students, policymakers, and the public at large—be well served by a system that promotes a single set of learning goals? The National Academy of Education Panel on Standards Based Education Reform suggests not (McClaughlin & Shepard, 1995). The panel proposes that sets of standards serve as exemplars: Rejecting the assumption that there is "one best way to define and structure knowledge"

(p. 24) in a field, the panel argues for the value of multiple sets of "coherent, professionally credible" standards in a given domain that could serve as exemplars to state and local education agencies.

Perhaps the greatest long-range risk of centralized planning is that, as the social reality in different contexts comes to resemble the planner's model, we lose opportunities to experience alternative practices and so too our collective memory that things could be otherwise. Our understanding of how people learn is shaped by the way we foster learning. And it is the students in our classrooms who will become the educators, researchers, and policymakers of tomorrow. In fact, a number of sociocultural theorists argue for the importance of studying how learning occurs outside of school and in different cultural contexts in order to gain some purchase on how routinized practices in schools have shaped conceptions of learning.

The Indicators Entail Particular Identities and Social Positions for Different Actors

Moving to the second dimension—the way in which an assessment and accountability system positions those who use it—a related set of concerns arises. As Gee and colleagues describe it, "discourses [like assessment and accountability systems] create, produce, and reproduce opportunities for people to be and recognize certain kinds of people" (Gee, Hull, & Lankshear, 1996, p. 10). Focusing for the moment on teachers, what identities and positions do these assessment systems offer them? Although all of the assessment systems involved teachers in their production, and all seemed intended to honor their professional judgment, they nevertheless limited the role that teachers could (or would likely) play when the system is fully operational in the classrooms of teachers who did not participate in its development. As I noted at the beginning: for most classrooms, these assessments appeared to limit the classroom teacher's productive role to activities like judging student work using the guidelines provided, perhaps developing tasks consistent with a predetermined template, choosing the time of implementation, and supporting students in making progress toward the goals.

Shepard (2003), in a recent response to KWSK, worries that "reliance on technological examples gives an implicit message that good assessment depends on the computer's statistical modeling and data-management capacities. It suggests that teachers need to receive information about typical errors and learning progressions from cognitive experts; and it seems to limit practical implementation to a few . . . modules while we wait for more subject areas to be codified" (p. 172).

Wenger (1998) raises questions about "'ownership of meaning'—that is, the degree to which we can make use of, affect, control, modify, or in general, assert as ours the meanings that we negotiate" (p. 200).

When, in a community of practice, the *distinction between the production and adoption of meaning* reflects enduring patterns of engagement among members—that is, when some always produce and some always adopt—the local economy yields very uneven ownership of meaning. (p. 203; italics mine)

Those whose role it is, consistently, to adopt meaning, Wenger argues, have less opportunity to learn from experience:

A split between production and adoption of meaning thus compromises learning because it presents a choice between experience and competence: you must choose between your own experience as a resource for the production of meaning and your membership in a community where your competence is determined by your adoption of other's proposals for meaning. (p. 203)

Porter (1995, 2003) raises a related issue about the impact of quantitative indicators and predetermined standards on administrators and policymakers: "reliance on numbers and quantitative manipulation minimizes the need for intimate knowledge and personal trust" (p. ix), and it removes from decision makers the need to attend to the unique features of each case and to take personal responsibility for their decisions. As Scott suggests, like the monoculture forest, centralized designs for social systems

tend to diminish the skills, agility, initiative, and morale of their intended beneficiaries. . . . Complex, diverse, animated environments contribute . . . to producing a resilient, flexible, adept population that has more experience in confronting novel challenges and taking initiative. Narrow, planned environments, by contrast, foster a less skilled, less innovative, less resourceful population. *This population, once created, would ironically have been exactly the kind of human material that would in fact have needed close supervision from above.* In other words, the logic of social engineering on this scale was to produce the sort of subjects that its plans had assumed at the outset. (1998, p. 349; italics mine)

Clearly, this is no thoughtful educator's intent, and it is certainly not the intent of the authors of the source chapters, who seek to honor the professional judgment of teachers. It is, however, a potential effect of a single "comprehensive, coherent, and continuous" assessment and accountability system, backed (if adopted at the state level) by state power, that provides a language for local actors to use in understanding

themselves and others, and that permits local actions to be visible from afar. Critical theory reminds us always to seek out and learn from the effects of our actions, which may well differ from our intent.

An Alternative Vision of Coherence

The effects of the rich set of tools the source chapters and KWSK offer will ultimately depend on the ways they are incorporated into the social contexts in which they are used. Will they be imposed by the state as requirements that everyone must implement? Or, will they be used as exemplars that teachers, schools, or districts might adopt, adapt, combine with other practices, or reject for alternatives? "How [will] the power to define, adapt, or interpret the design [be] distributed?" (Wenger, 1998, p. 235). The vision of coherence in KWSK—"the integration of assessment frameworks and methods across all levels of the assessment system, from the classroom to the system level" (p. 3)—suggests the press for alignment. In the short term, it is likely that this press for alignment will produce the more widespread effect. It will likely increase the quality of education in many schools and decrease inequities in learning opportunities. These are noble goals we all share. In the long run, however, the potential effects of narrowing both the range of learning experiences and the opportunities for local actors to contribute meaningfully to the way those experiences are conceptualized and judged will, I believe, be more consequential. What will be the disjunctions between system developers' good intentions and what they, in fact, effect? What kinds of people will they foster?

In closing, I'll sketch out some principles for an alternative vision of coherence—one that takes advantage of the rich tools the source chapters offer us but that skirts some of the risks of promoting coherence-through-alignment coupled with comprehensive, coherent, continuous visibility. Following Wenger (1998), I will call this coherence-through-negotiation-of-meaning, or more precisely, coherence through *mutual engagement* in the negotiation of meaning, a vision that comes with its own set of risks. Wenger defines *negotiability* as "the ability, facility, and legitimacy to contribute to, take responsibility for, and shape the meanings that matter within a social configuration" (p. 197). "*Mutual engagement in the negotiation of meaning* involves both the production of proposals for meaning and the adoption of these proposals" (p. 202; italics mine).

A central principle of critical theory is that we learn by encountering perspectives, practices, and social contexts that are different from

our own; such encounters make us aware of the categories of thought and action we take for granted, allow us to imagine how things might be otherwise, and encourage us to reconsider our perspectives and practices in light of this knowledge. From this perspective, *diversity is a resource for learning and for social change*. Further, learning and social change occur, not primarily through external controls but, rather, because people have reflected on their experience, come to understand its genesis and effects, and, as a result, have chosen to act differently.

The chapter by Black and Wiliam comes closest to illustrating practices consistent with what I am calling coherence-through-negotiation-of-meaning. Teachers come together around some common issues and resources, supported and challenged by knowledgeable colleagues from *outside* their social context, and they adapt and implement those resources as they choose in their own classrooms. This is similar to a role that Shepard (2003) envisions for the assessment systems illustrated in KWSK and for the teachers who use them:

A different role for teachers might make more sense—one that focuses on teacher learning and increases teachers' repertoire of assessment skills. . . . Why not use existing technology-based models as examples to scaffold teachers' developing understandings of formative assessment? If teachers were helped to analyze and attend to the salient features of technology-based assessments, they could learn to generalize these features to other content areas and instructional units. (p. 173)

For accountability purposes at the school level, teachers might be routinely expected to share with administrators and with one another evidence of their students' learning, their teaching practices, and their reflections on them. Collaborative planning might lead productively to local forms of alignment that permit locally relevant indicators of the form that Wilson imagines in the introduction or to the enthusiastic adoption of any of the programs proposed here. From this perspective, however, the strongest technology-based assessments will be those that not only serve as rich models but also permit local actors to alter the parameters of the program to suit their needs. Indeed, Gee (2003b), who documents the learning potential of computer games, notes that the better programs are those that give players the opportunity to customize their identities.

There are multiple examples of school communities that engage collaboratively in these sorts of practices (see, for example, Darling-Hammond, Ancess, & Falk, 1995; Meier, 1995; Rogoff, Turkanis, & Burtlett, 2001). District administrators might develop policies and practices that

support (or at least do not undermine) schools in developing these sorts of learning communities. Darling-Hammond (2001) represents the policy dilemmas as "developing top-down supports for bottom-up reform" (p. xv). Consistent with critical theoretic perspectives, the role of knowledgeable outsiders can be crucial in illuminating limitations of local practices, and district policies can institutionalize such roles. In these ways, accountability at the school level may be more appropriately contextualized with a deep understanding of the social reality students, parents, teachers, and administrators experience. This does not preclude the use of system-wide indicators at the district, state, or national level; it does, however, imagine a smaller, less commanding role, and more "prudent aspirations" (Bryk & Hermanson, 1993) for their use. Darling-Hammond (1994), for instance, envisions

carefully targeted . . . assessments at a few key developmental points that will provide data for informing policy makers about program successes and needs, areas where assistance and investment are needed, and assessment models for local schools. Meanwhile, locally implemented assessment systems—including portfolios, projects, performance tasks, and structured teachers' observations of learning—will provide the multiple forms of evidence about student learning needed to make sound judgments about instruction. (p. 20)

Conceptual tools that assess the nature of alignment between the centralized and local forms of evidence, of the sort Smithson and Porter illustrate, could be productively framed as helping local educators understand the differences in information between the two sources, leaving questions about whether to further alignment to local choice. We should not forget that the National Assessment of Educational Progress (NAEP) was expressly designed and implemented as a low-stakes assessment—to discourage explicit forms of alignment—so it could be used to monitor progress across educational systems with different curriculum frameworks (NRC, 2000). Interestingly, among the concerns represented in the summary of the NRC workshop on reporting district-level NAEP results is that "use of NAEP results at the district or school level has the potential to discourage states' and districts' use of innovation in developing their own assessments" (NRC, 2000, p. 10). Comprehensive and continuous assessment is not necessary for district, state, or national purposes. As Bryk and Hermanson (1993) note, "more information is not always better. . . . The ultimate long-term test of this system is not whether we are better informed but whether we act more prudently. In the shorter term, the best 'test' may

be found in the answer to the question 'Is our public discourse enriched (or impoverished) by this new information?'" (p. 476).

One of the most important roles for an indicator system is to help those at one level of the educational system make good decisions that support educational practice at other levels of the system for which they are responsible.[5] Teachers, principals, district superintendents, curriculum coordinators, school board members, state superintendents, state and federal department of education staff, legislators, and so on have different kinds of decisions to make, and the information they will need is likely different as well. Designing an appropriate set of indicators and related practices will depend on a rich, contextualized understanding of what information is needed, how it is used, and what the effects of this use are. Even decision makers at the federal level work within their own immediate social contexts. Those of us who develop assessment and accountability systems need to acquire deeper understandings of the social practices that surround them, not just at the classroom level, but at all levels of the educational system where they are used. (Spillane's program of research on distributed leadership provides a rich, evidence-based example of the sort of research I have in mind; see Burch & Spillane, 2003 and Spillane, Diamond, & Jita, in press).

From this alternative perspective, coherence is achieved, not through actions and structures that promote widespread alignment, consistency, or commonality, but rather *through concerted efforts to understand and learn from our differences* (Moss & Schutz, 2001). "By seriously attempting to understand the insights . . . other perspectives provide, we can begin to educate ourselves and to revise or develop our understanding. . . . Each perspective may develop along its own lines . . . yet in concert with the others, each perspective can refine itself and become more differentiated and more aware of the internal difficulties with which it must deal" (Warnke, 1994, p. 131-132). Of course, this approach has risks, too: the impact of these practices on the larger system will be much more diffuse and likely slower to yield widespread progress on some valued learning outcomes; further, diversity in learning experiences may not as quickly ameliorate inequities in learning opportunities. Administrative structures that support and challenge local learning communities will need to be in place to mitigate these risks. However, as I have argued, the long-term risks of privileging commonality, and thus diminishing local responsibility for design decisions, seem far more consequential and less easily reversible if problems arise.[6]

As Scott (1998) asks, "to what degree does [the system] promise to enhance the skills, knowledge, and responsibility of those who are a part of it? . . . How deeply is [it] marked by the values and experience of those who comprise it?" (p. 355). Under a model of coherence-through-negotiation-of-meaning, the policy goals are to make rich and diverse learning opportunities available; to encourage learning at all levels of the system through ongoing dialogue across different perspectives and contexts about the means and ends of education; and to promote evidence-based critical reflection on the effects of our actions. Diversity together with the meaningful productive engagement of actors at all levels of the educational system represents, in my judgment, our best resources for social change. As Scott reminds us in the conclusion to *Seeing Like a State*, "Diversity and certain forms of complexity, apart from their attractiveness, have other advantages. . . . They may not be as productive, in the short run, . . . but they are demonstrably more stable, more self-sufficient, and less vulnerable" (p. 353).

Notes

1. As I write this commentary, I am engaged in ongoing discussions with other members of the Idea of Testing Project, an interdisciplinary initiative funded by the Spencer Foundation, focused on expanding the foundations of educational assessment. These discussions with King Beach, Jim Gee, Jim Greeno, Ed Haertel, Carol Lee, Bud Mehan, Bob Mislevy, Fritz Moser, Diana Pullin, and Lauren Young have challenged and advanced my thinking about educational assessment as it is reflected here. A brief description of the project appears in Spencer's 2003 annual report (Spencer Foundation, 2003, p. 18). I am also grateful to Mark Wilson for his insightful comments on an earlier draft of this commentary.

2. Some reserve the term *critical* for perspectives informed by the work of a particular community of theorists known as the Frankfurt school. Here, I follow the lead of other theorists (e.g., Calhoun, 1995; Hoy & McCarthy, 1994) who use the term more broadly to encompass the range of perspectives that share these features, including not only members of the Frankfurt school (like Horkheimer, Adorno, and Habermas), but also drawing on hermeneutic, poststructural, and feminist perspectives (like those of Gadamer, Bourdieu, Foucault, and Harding).

3. While some limit the term *sociocultural* to research that derives from the work of Vygotsky, others use the term more broadly to refer to a constellation of perspectives that attend to the dialectical relationship between social structure and local practice of individuals in context, which is the perspective I use here.

4. This paragraph is adapted from Moss, Pullin, Gee, & Haertel (2002) and Moss (2003). It is the relevance of these ideas to assessment that the Spencer Idea of Testing Project (see reference in Spencer, 2003) is intended to explore.

5. King Beach made a similar point at the October 2003 meeting of the Idea of Testing Project.

6. Even Popper (1944/1985) recommended a "piecemeal" approach to social engineering: "it is difficult enough to be critical of our own mistakes, but it must be nearly impossible for us to persist in critical attitudes toward our actions which involve the lives of many men. To put it differently, it is very hard to learn from very big mistakes" (p. 315).

REFERENCES

Beach, K. (1999). Consequential transitions: A sociocultural expedition. In A. Iran-Nejad & P.D. Pearson (Eds.), *Review of Research in Education*, *24*, 101-140. Washington, DC: American Educational Research Association.

Beach, K. (2003). Learning in complex social situations meets information processing and mental representation: Some consequences for educational assessment. *Measurement: Interdisciplinary Research and Perspectives*, *1*(2), 149-153.

Bourdieu, P. (1988). *Homo academicus*. Cambridge, UK: Polity Press.

Bourdieu, P. (1991). The peculiar history of scientific reason. *Sociological Forum*, *6*(1), 3-26.

Burch, P., & Spillane, J.P. (2003). Elementary school leadership strategies and subject matter: The case of mathematics and literacy instruction. *Elementary School Journal*, *103*(5), 519-536.

Bryk, A.S., & Hermanson, K.M. (1993). Educational indicator systems: Observations on their structure, interpretation, and use. In *Review of Research in Education*, *19*, 451-484.

Chaiklin, S., & Lave, J. (Eds.). (1993). *Understanding practice: Perspectives on activity and context*. Cambridge: Cambridge University Press.

Calhoun, C. (1995). *Critical social theory*. Oxford: Blackwell.

Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Harvard University Press, Belknap Press.

Darling-Hammond, L. (2001). *The right to learn: A blueprint for creating schools that work*. Somerset, NJ: Jossey-Bass.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, *64*(1), 5-30.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students that work*. New York: Teachers College Press.

Gee, J.P. (1999). *An introduction to discourse analysis: Theory and method*. London: Routledge.

Gee, J.P. (2003a). Practice, participation, and tools: An alternative starting point. *Measurement: Interdisciplinary Research and Perspectives*, *1*(2), e12-e15. Available: http://bear.soe.berkeley.edu/measurement/pubs/toc12.html

Gee, J.P. (2003b). What video games have to teach us about learning and literacy. New York: Palgrave Macmillan.

Gee, J.P., Hull, G., & Lankshear, C. (1996). *The new work order: Behind the language of the new capitalism*. Boulder, CO: Westview Press.

Greeno, J., & Haertel, E.H. (2003). A situative perspective: Broadening the foundations of assessment. *Measurement: Interdisciplinary Research and Perspectives*, *1*(2), 154-161.

Hoy, D.C., & McCarthy, T. (1994). *Critical theory*. Oxford: Blackwell.

Linn, R.L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, *32*(7), 3-13.

Linn, R.L., Baker, E.L., & Betebenner, D.W. (2002). Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, *31*(6), 3-16.

McCarthy, T. (1994). Philosophy and critical theory: A reprise. In D.C. Hoy & T. McCarthy, *Critical theory* (pp. 5-100). Oxford: Blackwell.

McClaughlin, M.W., & Shepard, L.A. (1995). *Improving education through standards based reform*: *A report of the National Academy of Education panel on standards-based education reform*. Stanford: National Academy of Education.

Mehan, H. (1993). Beneath the skin and between the ears: A case study in the politics of representation. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (pp. 241-268). Cambridge: Cambridge University Press.

Mehan, H. (1998). The study of social interaction in educational settings: Accomplishments and unresolved issues. *Human Development*, *41*, 245-269.

Meier, D. (1995). *The power of their ideas: Lessons for America from a small school in Harlem*. Boston: Beacon.

Moss, P.A. (in press). Rethinking validity for classroom assessment. *Educational Measurement: Issues and Practice, 2*(4).

Moss, P.A., & Schutz, A.M. (2001). Educational standards, assessment, and the search for consensus. *American Educational Research Journal, 38*(1), 37-70.

Moss, P.A., Pullin, D.P., Gee, J.P., & Haertel, E.H. (2002). The idea of testing: Expanding the foundations of educational measurement. Unpublished manuscript.

National Research Council (NRC). (2000). *Reporting district-level NAEP: Summary of a workshop*. Washington, DC: National Academy Press.

Pellegrino, J.W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know*. Washington, DC: National Academy Press.

Popper, K.R. (1944/1985). Piecemeal social engineering. In D. Miller (Ed.), *Popper selections* (p. 304-318). Princeton, NJ: Princeton University Press.

Porter, T.M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Porter, T.M. (2003). Measurement, objectivity, and trust. *Measurement: Interdisciplinary Research and Perspectives, 1*(4), 241-255.

Rogoff, B., Turkanis, C.G., & Burtlett, L. (2001). *Learning together: Children and adults in a school community*. New York: Oxford.

Scott, J.C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. New Haven, CT: Yale University Press.

Shepard, L.A. (2003). Commentary: Intermediate steps to knowing what students know. *Measurement: Interdisciplinary Research and Perspectives, 1*(2), 171-177.

Spencer Foundation (2003). Annual report: April 1, 2002-March 31, 2003. Chicago: Author. Available: http://www.spencer.org/publications/index.htm

Spillane, J., Diamond, J.B., & Jita, L. (in press). Leading instruction: The distribution of leadership for instruction. *Journal of Curriculum Studies*.

Warnke, G. (1994). *Justice and interpretation*. Cambridge, MA: The MIT Press.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.

Wertsch, J.V., Del Rio, P., & Alverez, A. (1995). Sociocultural studies: History, action, and mediation. In J.V. Wertsch, P. Del Rio, & A. Alverez (Eds.), *Sociocultural studies of mind* (pp. 1-36). Cambridge: Cambridge University Press.

Wilson, M. (1994). Community of judgment: A teacher-centered approach to educational accountability. In Office of Technology Assessment (Ed.), *Issues in educational accountability*. Washington, DC: U.S. Congress Office of Technology Assessment.