

PORTFOLIOS AS SITES OF
 LEARNING: RECONCEPTUALIZING
 THE CONNECTIONS TO
 MOTIVATION AND ENGAGEMENT

- Caroline Clark
 OHIO STATE UNIVERSITY
In this qualitative study, we examine how students' participation in the process of constructing portfolios may affect their motivation for and engagement in literacy learning. Research that is based on the "wisdom of practice" continues to emphasize the value of portfolios and portfolio-based assessment. Few studies, however, articulate an explicit theoretical frame for viewing portfolios and considering their effectiveness. This study grounds portfolio practices in 2 bodies of literature that lend theoretical support to their importance, and then extends outward to include literature that addresses the social construction of meaning and the situated nature of learning. Using vignettes of students from an urban high school, we lay out our initial understandings of portfolios and our a priori assumptions about the concepts of motivation and engagement; we articulate our emerging theory around process portfolios, and trace our processes of reading and interpreting the data; we explain how our theory has evolved and lead us to a new set of literature that broadens the focus of motivation and engagement to include not only the individual student but also other sociocultural factors and, finally, we discuss the implications of this work for practitioners, researchers, and policy makers.
- Todd K. Chow-Loy
 PUNAHOU SCHOOL, HONOLULU, HI
 California J. Ficker
 CALIFORNIA POLYTECHNIC STATE UNIVERSITY
 Pamela A. Moss
 UNIVERSITY OF MICHIGAN

J L R

V. 33 NO. 2

2001

pp. 211-241

THE STUDY PRESENTED HERE examines how students' participation in the process of constructing portfolios may affect their motivation for and engagement in literacy learning. The emphasis on process is important, for it is our argument that the action of constructing, negotiating, compiling, documenting, sharing, revising, reflecting on, and assessing one's own work in a portfolio – rather than the portfolio as simply a product or outcome – provides the link to motivation and engagement. We see portfolios, then, as more than simply folders that hold the physical evidence of students' growth and achievement. Rather, we believe that portfolios – both materially and through their social surround – are actually sites of learning.

It is important to note that, from our perspective, learning and engagement are wholly connected with the process of identity formation. As many researchers have found, issues of identity are central to students' engagement in schooling and, subsequently, to their academic success (Davidson, 1996; Finn, 1989; Newmann, 1981, 1985, 1989; Nystrand & Gamoran, 1991). Class, race, ethnicity, and gender are obviously implied in the term identity. However, issues of language, literacy and culture are equally part of these intersecting categories which make up this complex concept. Students' identities are shaped – for better or for worse – by messages in the media (Cortés, 1995; Giroux, 1996; Thompson, 1990), by their sense of the perceptions of others (Markus & Nurius, 1986; Fordham, 1996), and by the multiple community contexts in which they are embedded (McLaughlin & Heath, 1993). For some theorists, learning itself is seen as the construction of identities as individuals take up and take on different social practices in different contexts with different social communities at different times (Lave & Wenger, 1991). Hence, to engage in learning is to engage in the project of constructing and reconstructing identity.

Our goal, in this study, is to begin to reconceptualize how portfolios – including the social milieu surrounding them – can enable student motivation and engagement. This emphasis on process and portfolios as social and material sites of learning contrasts with portfolios that are viewed as end products or outcomes that may be disconnected from students' ongoing learning and development and may not require students' critical reflection on the contents. Furthermore, the concept of portfolios as in-process works is distinct from the idea of the writing process (Calkins, 1986; Elbow, 1981; Graves, 1983), where students are encouraged to plan, draft, and revise their written pieces. Instead, we see portfolios as sites of meaning construction where students are actively involved in producing multiple texts, critically examining these pieces, reflecting on what and how they are coming to know, and sharing these learnings with one another. In this sense, we see portfolios as not only in-process, but also as distinctly social.

J L R
Clark et al.

Our text is organized into five parts. In the first section, we describe the research context for this study, followed, in section two, by a review of the literature that informed our understandings of portfolios and the concepts of motivation and engagement. Next, in section three, we describe our research methods and modes of data collection and analysis. In section four, we interpret our findings and discuss how we have come to reconceptualize portfolios and their links to motivation and engagement, using students' words and work to illustrate these new understandings. And, finally, in section five, we discuss the implications of this work for practitioners, researchers, and policy makers.

The Research Context

The School Site and Collaborators

Our look at portfolios grows out of a long history of collaboration. For 4 years, we worked collaboratively with teachers at Henry Ford High School, a comprehensive secondary school in Detroit, Michigan serving about 2,600 students in grades 9 through 12. Approximately 99% of the students are African American, whereas the other 1% are predominantly Latino and White.¹ This group engaged in examining collections of student writing for assessment and accountability purposes, as well as chronicling students' writing experiences and teachers' professional practices through written case studies. Some of us from the university participated and observed in the English classrooms of our teacher colleagues, facilitating their implementation of student writing folders and portfolios for instructional and assessment purposes. In addition, all of us participated in monthly meetings at both the school and university sites to discuss student writings, and to write and share case studies of individual students. The study presented here draws on the data from this work. The decision to implement portfolios in this school context was not incidental. It had much to do with the fact that this was a large, urban comprehensive high school with a highly transitory student population. The teacher collaborators agreed that having a portfolio would give students "a constant" in their lives, a place to be in a new school and something they could "take with them" to another. Although issues of race, class, gender, and ethnicity certainly affected stu-

1. Our work, funded by grants from the Metropolitan and Ponting Foundations, has involved, over time, five classroom English teachers (including Herter), one computer writing/English teacher, one guidance counselor, two university professors (including Moss), and five graduate-student researchers (including Clark, Chow-Hoy, and Herter). Opinions expressed reflect those of the authors and do not necessarily reflect those of the granting institutions.

dents' participation in these portfolio classrooms, the teacher collaborators did not see these as central issues at the time, and hence, these issues are not foregrounded in this study.

The authors of this paper took on different roles and held different positions in relation to this project. Caroline Clark, a White female, was a graduate research assistant who spent 3 days a week or more in each of the classroom sites across the span of the collaboration. Todd Chow-Hoy, an Asian-Hawaiian male, was a graduate research assistant for 1 year of the collaboration. Roberta Herter, a White female, had been teaching at the school for over 25 years and was concurrently enrolled in a doctoral program at the university. Pamela Moss, a White female, was a professor at the university. An occasional visitor to the classroom sites, she regularly attended monthly research group meetings at the school. Along with Roberta, she was a founding member of the collaborative group. As a collaboration between individuals with different relationships to and histories in the school site, this particular study brings together our range of understandings and experiences with portfolios. In addition, we also bring our respective positionings - as White, Asian-Hawaiian, female, male, professor, teacher, and student - to bear as we explore connections between portfolios, motivation, and engagement.

Assumptions and Research Questions

The questions in this study grew out of our experiences with portfolios, and reflected our shared assumptions about their value in terms of assessment and learning. Already, we believed that portfolios could better inform multiple stakeholders about the nature of students' learning and could wed assessment and instruction in more seamless ways within the classroom. We also shared the assumption that portfolios could not be dictated, that there is no single approach to portfolio implementation and assessment. Rather, our group believed that portfolios would grow out of the classroom contexts in which students, with their teachers, created them. Given this, we were curious to see what else portfolios might have to offer. Examining portfolios with an eye toward motivation and engagement would enable us to see their dynamic potential in terms of learning.

Initial Understandings of Portfolios, Motivation, and Engagement

Given our shared history around implementing portfolios, we began this study with a fairly comprehensive understanding of their place in the literature, and a developing sense of their possibilities and limitations. We

know that in the past decade, portfolios have become an educational buzzword (Elbow & Belanoff, 1997). What began as a grassroots, classroom- or teacher-based assessment move aimed at basing evaluations on the daily classroom work that students produce over time (Carini, 1973, 1986) has grown into schoolwide (Goldman, 1989), state (Koretz, Stecher, & Deibert, 1992), national (Gentile, 1990), and other large-scale assessments based on portfolio evidence (Simmons, 1990). Although, on the one hand, we welcome such authentic assessment moves, we worry, as do others (Lucas, 1992), about the emphasis such moves place on portfolios as outcomes. As more and more schools, districts, and states require portfolios from individual students (often including standard guidelines for their form and content), less consideration is given to the process of their construction. At times, "portfolios" are simply a last-minute compilation of the required pieces placed in a folder not far different from the typical student permanent record. Consequently, the opportunities for enhancing student motivation and engagement are glossed over or lost completely in the rush toward a finished product. Given that "students' engagement in education and school has increasingly emerged as the bottom line of dropout prevention and school completion" (Grannis, 1991, p. 2), it seems wise to pursue all feasible opportunities for enhancing student engagement. We believe that a renewed emphasis on portfolios as material and social sites of learning may provide one such route.

Proponents of portfolios have written widely about their benefits in terms of individual student success and teacher enhancement (Belanoff & Dickson, 1991; Burnham, 1986; Gill, 1993; Glazer & Brown, 1993; Graves & Sunstein, 1992; Herter, 1991; Howard, 1990; Murphy & Smith, 1991; Paulson & Paulson, 1991; Smith & Ylvisaker, 1993; Tierney, Carter, & Desai, 1991; Wolf, 1989; Yancey, 1992), as well as the opportunities for reflection that portfolios can provide (Camp, 1990; Camp & Levine, 1991). Some researchers have even begun to look more closely at the social aspects of portfolio classrooms (i.e., Crumpler, 1996; Fenner, 1995). For us, these works have been germinal and have fostered much of our thinking and work around portfolios. At the same time, we have noticed that some of these studies do not articulate an explicit theoretical frame for viewing portfolios and considering their effectiveness. Much of the theory that is brought to bear on the concept of portfolios tends to be from the measurement realm (Archbald & Newmann, 1988; Partridge, 1990; Paulson & Paulson, 1990; Stiggins, 1994; Wiggins, 1989a, 1989b), focusing on technical issues of validation and reliability (Koretz, McCaffrey, Klein, Bell, & Stecher, 1992; Nystrand, Cohen, & Dowling, 1993). Although psychometricians are beginning to emphasize consequences of assessment as one criterion for validity (e.g., Moss, 1992), few studies emphasize understand-

ing the connections between portfolio assessment, student motivation, and engagement, and wider school reform, and fewer still consider the benefits and consequences of portfolio assessment for low-achieving and culturally and linguistically diverse learners (Bauer, 1999).

As we continued our study, we discussed what our beliefs about and understandings of portfolios were, and their potential connection to student motivation and engagement. We read articles and chapters on motivation and engagement to flesh out our basic understandings of these issues (e.g., Maehr, 1984; Maehr & Archer, 1987; McCombs, 1991; Nichols, 1983; Stipek, 1993), and found that we did not necessarily have a shared vision of these concepts. Our different experiences – as teachers, researchers, students, and learners more generally – caused each of us to bring our own autobiographically based understandings to these concepts. Roberta Flerter, for example, pointed out that in her experiences as a teacher for over 25 years, “direction” as an indicator of motivation – that is, students choosing among some behavioral alternatives – was not as clear cut as it appeared in some of our group’s readings. Although the literature on motivation suggests that a student working alone at his or her desk is an example of a motivated student, and a student chatting to a neighbor is an example of an unmotivated student (Maehr & Archer, 1987), Roberta’s observations as a classroom teacher suggested to her that such interactions are often signs of student engagement and opportunities for social support.

As we continued to read, we found that theories of student motivation and engagement could provide one possible lens through which to view the portfolio process. Research on motivation suggests that when individuals are working toward goals that are personally meaningful, they perform at their highest, most effective levels (McCombs, 1991; Schunk, 1990). Often times, “performance problems” which are attributed to a student’s lack of motivation may have much to do with a student’s sense of personal investment (Maehr, 1984; Maehr & Archer, 1987). According to Maehr and Archer (1987), personal investment is directly linked to the meaning of the situation in which an individual is involved. Meanings are formed based on an individual’s perceived options, sense of self, and personal incentives. Unfortunately, meaningful student goals and traditional school practices are often incompatible (Marzano, 1994). Nichols (1983), for example, has found that traditional classroom goals for students are often unrelated to goals for which students have intrinsic motivation.

2. Phillion (1997) is one exception to this. In his essay, he draws on the work of Csikszentmihalyi (1990) and Csikszentmihalyi, Rathunde, and Whalen (1993) to consider the roll of the motivational concept of “flow” in the context of portfolios.

Research on motivation has direct links to theories of student engagement (Newmann, 1981, 1985, 1989; Seeman, 1972). Newmann (1985) for example, draws on research in psychology and sociology in developing his theory of student engagement. He defines engagement as “more than motivation or the general desire to succeed in school. It involves participation, connection, attachment, and integration in particular settings and tasks” (p. 51). He describes five factors that are important for enhancing student engagement: student need for competence, extrinsic rewards, intrinsic interest, social support, and sense of ownership. Each of these conceptual frames provides an overlay through which to view, consider, and critique portfolio practices. Linking portfolio processes to theories of motivation and engagement allows us to begin to understand their usefulness based on more than simply stories of success.

Method

Data Collection

All of the data for this study were collected during the 1993–1994 school year. The student participants were 9th- through 12th-grade students enrolled in five different English/language arts classes at Henry Ford High School. Teachers in these classrooms, including Roberta, were all involved in an ongoing school/university collaboration exploring teachers’ approaches to implementing portfolios. There were approximately 120 students enrolled in these classes, which included one general English 9, two general English 12, one honors English 12, and a specific course for struggling writers, “Writing Improvement,” comprised of 9th through 12th grade students. Classes met 5 days a week for approximately 50 minutes, and attendance, as in many large urban districts, was spotty – some days, classes were full and well attended; other days, as few as 10 students might be present in one class.

In each classroom, teachers chose to implement portfolios. Simplistically, this meant that all student writing was kept in file folders for the entire year. In each classroom, however, portfolios looked somewhat different, depending on the negotiated goals of the individual teacher and the students. In the “Writing Improvement” class, students selected pieces of writing to include in their portfolios based on genre and best work. Students were encouraged to try writing a variety of pieces (poems, non-fiction/essays, short stories, and plays) and to select their best examples of these forms to include in their portfolio. In three of the classrooms, the general 9th- and 12th-grade classes, students selected their best works to include in their portfolios. And in the 12th-grade honors English class,

gator triangulation, having more than one researcher involved in the collection and analysis of the data (Denzin, 1978).

Analysis of Data

Using a grounded approach and a method of constant comparison (Glaser & Strauss, 1967), each of us read all of the data separately. Initially, we made tentative notes in the margins to describe what we were each seeing in the data. Next, we engaged in what Strauss and Corbin (1998) call "open coding" to arrive at tentative descriptions of concepts that seemed to fit the data. Through "axial coding" (Strauss & Corbin, 1998), similarly coded data were clustered together to develop more refined categories or themes. At this point, we came together to compare what we were seeing in the data and to discuss and account for differences in our interpretations, a hermeneutic, dialogic process (Moss, 1994, 1996). Together, we examined the data to see where new codes, categories or themes might be needed to comprehensively account for the available evidence. Using this process, we developed three main categories: self, task, and fit, and a list of sub-themes (see Appendix). The self category involves issues of student identity and students' attempts to answer the implicit question "Who am I?" The task category involves issues related to students' perceptions of the activities in which they engage. The fit category involves issues related to students' perceptions of their place within various groups in classrooms, in school, in family settings, and in the community. Although we teased these categories apart to distinguish the defining features of engagement and motivation, much of the data showed that these three concepts were intrinsically linked and negotiated in the portfolio context.

Interpretations and Discussion

Although our iterations of text reading, interpretation, and dialogue led us to the three main conceptual categories of self, task, and fit, we found that it was impossible to discuss or describe one of these without attending to the others. Each aspect of the portfolio process – be it choosing topics for inquiry and writing; setting goals; collaborating with peers; or selecting pieces for display, discussion, and reflection – entailed a shift among these concepts. And, we would argue, these shifts among these concepts of self, task, and fit were the mechanisms that triggered motivation and engagement among students. Rather than organize our interpretations around our three conceptual categories, we provide illustrative vignettes focused on representative students from among the 10 teacher-selected case students that

students included a combination of pieces to show how their writing had changed over time and to demonstrate their development as writers. In all cases, multiple drafts of students' texts and all related writings were kept in their writing folders as supplements to their portfolios. In addition, all students engaged in some form of reflective writing about their portfolio sections. Across classrooms, then, portfolios shared the common elements that Lewitt (1995) has outlined, affording:

1. an ongoing, tangible record of accomplishment
2. constant evidence that the student's work is valued
3. an opportunity for the student to reflect on his or her work
4. an excellent tool for parent or faculty conferences, as well as for student conferences.

Given the large number of students, the range of teaching situations and approaches, and our particular interest in broadly characterizing portfolios and their links to motivation and engagement, we collected an array of data using a range of data collection and engagement techniques. Data included researchers' field notes of their observations and interactions with teachers and students, and copied artifacts, such as handouts, worksheets, and student writings in order to capture the scene. Caroline (Clark was a participant/observer in each focal classroom 3 days a week from September through June, keeping field notes on teacher-student interactions, student-student interactions, and her own interactions with students around their writing portfolios. Caroline also conducted both formal and informal interviews with students. Informal interviews grew out of participant/observations in the classrooms. They were captured via written field notes and chronicled Caroline's interactions with students around their writing folders and portfolios. Formal, audio-taped interviews occurred in May of 1994, with two teacher-selected students from each of the classrooms. These interviews aimed at understanding the impact of portfolio assessment on the lives and learnings of students, both within and outside the classroom, and at discerning how students used writing in their academic and everyday lives. Additionally, students completed two surveys about their writing practices in January and May of 1994. Although it is impossible to eliminate bias in research, we worked to achieve multiple forms of triangulation to minimize bias in our findings and to increase the trustworthiness of our conclusions. Types of triangulation included data triangulation, using several sources of data such as students' brainstorming activities, notes, final drafts, and reflective pieces; methodological triangulation, collecting data in multiple ways, including through interviews, observations and collected student work; and investi-

of practice," in their terms, "is a set of relations among persons, activity, and world, over time and in relations with other tangential and overlapping communities of practice" (p. 98). This conception of learning in a "community of practice" made up of relations among "persons, activity, and world" closely approximates our conception of portfolios as sites of learning made up of relations among students developing a sense of self; students constructing and engaging in tasks; and students developing a sense of where they fit in the classroom community and in other communities in which they inhere. In these classrooms, the portfolio process becomes one of the central shared practices as students negotiate the relationships between self, task, and fit. Learning, then, takes on a different end as students seek to become "writers" (among other "possible selves") and thus fully participating members in the communities of practice which are available in the portfolio context. This shift in the ends of learning entails a shift in the nature of motivation and engagement. Learning, in such contexts, results in increasing participation in communities of practice. Viewing himself as a writer in his classroom, for example, transformed how Larry viewed other literacy tasks in this context, such as reading. Reading, now, took on a specific purpose, providing a model for Larry of how professional novelists write and enabling him to learn from their practices in order to engage more fully as a member of this classroom community. Rather than being motivated by the "exchange value of learning" or "learning to display knowledge," students like Larry are motivated by the "use value of increasing participation" (Lave & Wenger, 1991, p. 112).

Classrooms like Larry's, Angeli's, and Kaitlin's, where portfolios are central, provide a context for students to construct and explore conceptions of self and identity, and the relations among self, tasks, and fit. Their stories, and the work of Lave and Wenger (1991), highlight the role of community in shaping students as individuals. Reciprocally, they also compel us to consider student agency in these community contexts. Here, the work of Markus and Nurius (1986) provides a complementary lens, allowing us to focus on the importance of the individual self. Portfolios enabled engagement and motivation because of their connection to self. The emphasis on *process* is important. It implies a *becoming*. As students write, revise, and reconstruct assignments, they are often, similarly, constructing and reconstructing themselves. In this context, students are not simply the static preconceptions that teachers or society may hold of them. Instead, they can explore possibilities. Through their writings and sharings with others, students construct "possible selves":

Possible selves are the ideal selves that we would very much like to become. They are also the selves we could become, and the selves we are afraid of

J L R
Clark et al.

becoming.... An individual's repertoire of possible selves can be viewed as the cognitive manifestation of enduring goals, aspirations, motives, fears, and threats. Possible selves provide the specific self-relevant form, meaning, organization, and direction to these dynamics. As such, they provide the essential link between the self-concept and motivation. (Markus & Nurius, 1986, p. 954)

The motivating power of possible selves is evident in much of the talk among students from these classrooms. Students, within the material and social sites of portfolios, were able to play out, across time and space, their identity constructions or possible selves. At times, this was akin to what Eisenhart (1995) has termed "the self story":

[T]elling stories about self is not only a way to demonstrate membership in a group or to claim an identity within it. Telling stories about self is also a means of becoming; a means by which an individual helps to shape and project identities in social and cultural spaces; and a way of thinking about learning that requires the individual to be active, as well as socially and culturally responsive. (p. 19)

Here, again, constructing self is tied back into the community as self stories are linked back to fit and task. Like the newcomers in Eisenhart's (1995) study of a nonprofit conservation corporation, students, in the context of their portfolios, "contributed to forming themselves within the [classroom] as they interpreted their past and present personal experiences" for their teachers and for one another. "The claims they made publicly about themselves had consequences for how they learned (e.g., from whom, for what reasons) and what they came to know (e.g., to demonstrate, to be recognized for)" (p. 19).

We believe these reconceptualizations of motivation and engagement are significantly linked to portfolios. Portfolios become the material and social sites for the shifts among self, task, and fit that enable student motivation and engagement. Materially, as in most portfolio practices, they are repositories of student work. More critically, however, as physical sites, they serve as anchors for students, places to call their own, to point to in proclaiming, "I am a writer," or in making a knowledge claim based on their writing. We believe that having such a material site to call ones own is especially important in secondary classrooms where students are transient throughout the day with little sense of connection to one another or school. Would engaging in process writing assignments alone be enough to trigger these shifts in self, task, and fit? Might this be enough to enable student motivation and engagement? We would argue no, that process writing alone is not enough. Portfolios as dynamic sites of learning, portfolios as process, are distinctly different from simply process writing. They offer us the

... how they are thinking about themselves with respect to new activities or categories in the school or beyond. Even more rare is an attempt on the part of school adults to act on the basis of such expressions. This omission is especially striking when one considers findings such as Holland (1992), who demonstrates that resistance, avoidance, and lack of interest – when newcomers do not “have their hearts in it” or, in other words, when they do not identify themselves as agents or actors in a cultural or social system – are associated with strikingly less expertise. (p. 21)

In providing these opportunities for students to become members of a community of learners and to define themselves within portfolio sites and via the portfolio process, schools (and teachers) provide positive school-experiences that can potentially lead to student engagement. As Phelan, Davidson, and Cao (1992) note, in classrooms where personalities are allowed to show, students respond more fully – academically and personally. Along the same lines, Gamoran and Nystrand (1992) point out that when students are taken seriously, it sends a message to them that they are important members of a learning community, and, as a result, the work becomes more authentic as students can actually influence the course their work will take. In this sense, portfolios as sites of learning in which students are asked to seriously think about and shape who they are, the task at hand, and their fit within a learning community, become an enabler of student engagement.

Teachers play a critical role in developing supportive portfolio contexts within which students can explore issues of community and identity. In order for portfolios to enable motivation and engagement in the particular ways we have described, teachers need to understand and allow for the multiple communities of practice which will best support student learning. In addition, we recognize that portfolios are a reflection of not only students’ performance choices, but also of teachers’ instructional choices. In our group, for example, examining Kaitlin’s entire corpus of portfolio writings allowed her teacher, Ms. Bryant, to see with us how the particular career I-Search task fit with the rest of Kaitlin’s writing. For Ms. Bryant, and teachers like her, examining the portfolio and its contents in context opens a window on communities of practice in the classroom, as well as individual student identities. Students like Kaitlin are allowed to appear as individuals, and instructional choices can then be tailored to their needs (Dyson, 1994). At this juncture, practice becomes more accessible for reflection, more open to critique, and potentially more negotiable (Carr & Kemmis, 1986).

Linking portfolios to students’ motivation for achievement and engagement in school has implications for school reform and policy makers as well. As many educators, researchers and policy makers concede, student

window for seeing and chronicling how students engage in writing and learning. They afford an ongoing, tangible record of accomplishment, constant evidence that work is valued, opportunities for student reflection, and tools for facilitating how teachers learn from and about students (Camp, 1990; Hewitt, 1995). Without the context of the portfolio itself, pieces of writing are disconnected. They occur one at a time, perhaps with no bearing on or relationship to one another. Looking through the window of portfolio process, and employing the lenses of motivation and engagement theory, we can see much more. Portfolios as sites of learning enable a sense of connection by enabling a long look back at what students have accomplished to date, which in turn enables a looking ahead at a future self as writer in a particular classroom and beyond.

Implications

Research that is based in the “wisdom of practice” continues to emphasize the value of portfolios and portfolio-based assessment. This study grounds portfolio practices in two bodies of literature that lend theoretical support to their importance, and then extends outward to include literature that addresses the social construction of meaning and the situated nature of learning. This extension invites a rethinking of motivation and engagement. Here, we have asserted that motivation and engagement are both manifest in classrooms where portfolios are central, arguing that portfolios as material and social sites of learning provide a space within which students can negotiate their constructions of self, task, and fit and a unique window through which we can view and understand these phenomena. Conceptualized in this way, motivation and engagement are best understood as social rather than individual phenomena, realized when students become fully participating members in communities of practice. Students in such situations, are motivated by and engaged in the use value of learning, rather than by more product-oriented display or exchange values of learning.

Our rethinking of the nature of learning, motivation, and engagement supports maintaining an emphasis on the process aspects of portfolios, rather than simply the products or outcomes. For teachers and researchers alike, it is important that we continue to explore these connections. Reconceptualizing the connections between portfolios, motivation, and engagement can provide a space for students to explore issues of identity and negotiate the construction of self, task, and fit. Currently, few such spaces exist in schools. As Eisenhart (1995) explains:

Rarely ... are students in schools given opportunities to express themselves

engagement is a crucial component in preventing school leaving and promoting student achievement (Firestone, 1989; Hamby, 1989; Wehlage, 1989; Wehlage & Rutter, 1986; Wehlage, Rutter, Smith, Lesko, & Fernandez, 1989; Wehlage, Rutter, & Turnbaugh, 1987). Nonetheless, some school reform efforts continue to emphasize portfolios as products, a practice that may undermine student engagement, and, we would argue, the positive effects of portfolios as sites of learning. Our work invites administrators and policy makers to be at least aware of, and to potentially examine closely, the impact of their reform choices on the learning and teaching of students and educators in particular contexts. Furthermore, we suggest that more policy moves be initiated at the classroom and school level, encouraging environments where teachers can work together to develop supportive communities of practice within schools that will allow them to investigate, reciprocally, the best ways to create similar supportive environments in their classrooms.

J. L. R.
Clark et al.

REFERENCES

- Arbald, D.A., & Newmann, F. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school*. Reston, VA: National Association for Secondary School Principals.
- Bauer, E.B. (1999). The promise of alternative literacy assessments in the classroom: A review of empirical studies. *Reading Research and Instruction*, 38, 153-168.
- Belanoff, P., & Dickson, M. (Eds.). (1991). *Portfolios: Process and product*. Portsmouth, NH: Boynton/Cook.
- Burnham, C. (1986). Portfolio evaluation: Room to breathe and grow. In C. Bridges (Ed.), *Training the teacher of college composition*. Urbana, IL: National Council of Teachers of English.
- Calkins, L.M. (1986). *The art of teaching writing*. Portsmouth, NH: Heinemann.
- Camp, R. (1990). Thinking together about portfolios. *The Quarterly: National Writing Project/Center for the Study of Writing*, 12 (2), 8-14, 27.
- Camp, R., & Levine, D. (1991). Portfolios evolving. In P. Belanoff & M. Dickson (Eds.), *Portfolios: Process and product* (pp. 206-214). Portsmouth, NH: Boynton/Cook.
- Carini, P.F. (1973). The Prospect School: Taking account of process. *Childhood Education*, 49 (7), 350-356.
- Carini, P.F. (1986). Building from children's strengths. *Journal of Education*, 168 (3), 13-24.
- Carr, W., & Kemmis, S. (1986). *Becoming critical: Education, knowledge, and action research*. London: Falmer Press.
- Cortés, C.E. (1995). Knowledge construction and popular culture: The media as multicultural educator. In J. Banks & C.A. McGee Banks (Eds.), *Handbook of research on multicultural education* (pp. 169-183). New York: Macmillan.
- Cramler, T. (1996). *Exploring a culture of assessment with ninth-grade students: Convergences of meaning within dramas of assessment*. Unpublished doctoral dissertation, The Ohio State University, Columbus, OH.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.
- Csikszentmihalyi, M., Rathunde, K., & Waielen, S. (with contributions by M. Wong). (1993). *Talented teenagers: The roots of success and failure*. Cambridge, UK: Cambridge University Press.
- Davidson, A.L. (1996). *Making an I molting identity in schools*. Albany, NY: State University of New York Press.
- Denzin, N.K. (Ed.). (1978). *Sociological methods: A source book*. New York: McGraw Hill.
- Dyson, A.H. (1994). Confronting the split between "the child" and children: Toward new curricular visions of the child writer. *English Education*, 26 (1), 12-28.
- Eisenhart, M. (1995). The fax, the jazz player, and the self-story teller: How do people organize culture? *Anthropology & Education Quarterly*, 26 (1), 3-26.
- Elbow, P. (1981). *Writing with power*. New York: Oxford University Press.
- Elbow, P., & Belanoff, P. (1997). Reflections on an explosion: Portfolios in the 90s and beyond. In K.B. Yancey & I. Weiser (Eds.), *Situating portfolios: Four perspectives* (pp. 21-33). Logan, UT: Utah State University Press.
- Fenner, L. (1995). *Student portfolios: A view from inside the classroom*. Unpublished doctoral dissertation, The Ohio State University, Columbus, OH.
- Finn, J.D. (1989). Withdrawing from school. *Review of Educational Research*, 59 (2), 117-142.
- Firestone, W.A. (1989). Beyond order and expectations in high schools serving at-risk youth. *Educational Leadership*, 46, 41-45.
- Fordham, S. (1996). *Blacked out: Dilemmas of race, identity, and success at Capital High*. Chicago: University of Chicago Press.
- Gamoran, A., & Nystrand, M. (1992). Taking students seriously. In F.M. Newmann (Ed.), *Student engagement and achievement in American secondary schools*. New York: Teachers College Press.
- Gentile, C. (1990). *Exploring new methods for collecting students' school-based writing: NAEP's 1990 portfolio study*. Washington, DC: National Center for Education Statistics.
- Gill, K. (Ed.). (1993). *Process and portfolios in writing instruction (Classroom Practices in Teaching English, Vol. 26)*. Urbana, IL: National Council of Teachers of English.
- Giroux, H.A. (1996). Hollywood, race, and the demonization of youth: The "kids" are not "alright." *Educational Researcher*, 25 (2), 31-35.
- Glaser, B., & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Hawthorne, NY: Aldine de Gruyter.
- Glazer, S.M., & Brown, C.S. (1993). *Portfolios and beyond: Collaborative assessment in reading and writing*. Norwood, MA: Christopher Gordon.
- Goldman, J.P. (1989). Student portfolios already proven in some schools. *School Administrator*, 46 (11), 11.
- Grannis, J.C. (1991). Meeting the goals of school completion. *ERIC/CUE Digest*

- Graves, D.H. (1983). *Writing: Teachers and children at work*. Portsmouth, NH: Heinemann.
- Graves, D.H., & Sunstein, B.S. (Eds.). (1992). *Portfolio portraits*. Portsmouth, NH: Heinemann.
- Hambly, J.V. (1989). How to get an "A" on your dropout prevention report card. *Educational Leadership*, 46, 21-28.
- Hertler, R.J. (1991). Writing portfolios: Alternatives to testing. *English Journal*, 80 (1), 90-91.
- Hewitt, G. (1995). *A portfolio primer*. Portsmouth, NH: Heinemann.
- Howard, K. (1990). Making the portfolio real. *The Quarterly: National Writing Project/Center for the Study of Writing*, 12 (2), 4-7, 27.
- Koretz, D., McCaffrey, D., Klein, S., Bell, K., & Stecher, B. (1992). *The reliability of scores from the 1992 Vermont Portfolio Assessment Program: Interim report* (Center for the Study of Evaluation, c.s.e. Technical Report No. 355). Santa Monica, CA: Rand Institute on Education and Training, National Center for Research on Evaluation, Standards, and Student Testing.
- Koretz, D., Stecher, B., & Deibert, E. (1992). *The Vermont Portfolio Assessment Program: Interim report on implementation and impact, 1991-1992 school year* (Center for the Study of Evaluation, c.s.e. Technical Report No. 350). Santa Monica, CA: Rand Institute on Education and Training, National Center for Research on Evaluation, Standards, and Student Testing.
- Lave, J. (1993). The practice of learning. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (pp. 3-32). Cambridge, UK: Cambridge University Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lucas, C. (1992). Introduction: Writing portfolios - changes and challenges. In K.B. Yancey (Ed.), *Portfolios in the writing classroom: An introduction* (pp. 1-11). Urbana, IL: National Council of Teachers of English.
- Macrorie, K. (1988). *The I-Search paper*. Portsmouth, NH: Boynton/Cook.
- Maehr, M.L. (1984). Meaning and motivation: Toward a theory of personal investment. In R.E. Aines & C. Ames (Eds.), *Research on motivation in education*, Vol. 1: *Student motivation* (pp. 115-144). Orlando, FL: Academic Press.
- Maehr, M.L., & Archer, J. (1987). Motivation and school achievement. In L. Katz (Ed.), *Current topics in early childhood education* (Vol. 7, pp. 85-107). Norwood, NJ: Ablex.
- Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, 41, 954-969.
- Marzano, R.J. (1994). Commentary on "Literacy portfolios: Windows on potential." In S.W. Valencia, E.H. Hiebert, & P.P. Allierbach (Eds.), *Authentic reading assessment: Practices and possibilities* (pp. 41-45). Newark, DE: International Reading Association.
- McCombs, B.L. (1991). Motivation and lifelong learning. *Educational Psychologist*, 26 (2), 117-128.
- McLaughlin, M.W., & Heath, S.B. (1993). Casting the self: Frames for identity and

J L R
Clark et al.

- dilemmas for policy. In S.B. Heath & M.W. McLaughlin (Eds.), *Identity and inter-city youth: Beyond ethnicity and gender* (pp. 210-240). New York: Teachers College Press.
- Moss, P.A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62, 229-258.
- Moss, P.A. (1994). Can there be validity without reliability? *Educational Researcher*, 23 (2), 5-12.
- Moss, P.A. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher*, 25 (1), 20-28.
- Murphy, S., & Smith, M.A. (1991). *Writing portfolios: A bridge from teaching to assessment*. Markham, Canada: Pippin.
- Newmann, F.M. (1981). Reducing student alienation in high schools: Implications of theory. *Harvard Educational Review*, 51, 546-556.
- Newmann, F.M. (1985). *Student engagement and achievement in American secondary schools*. New York: Teachers College Press.
- Newmann, F.M. (1989). Student engagement and high school reform. *Educational Leadership*, 46, 34-36.
- Nichols, J.G. (1983). Conceptions of ability and achievement motivation: A theory and its implications for education. In S.G. Paris, G.M. Olson, & H.W. Stevenson (Eds.), *Learning and motivation in the classroom* (pp. 211-238). Hillsdale, NJ: Erlbaum.
- Nystrand, M., Cohen, A.S., & Dowling, N.M. (1993). Addressing reliability problems in the portfolio assessment of college writing. *Educational Assessment*, 1 (1), 51-70.
- Oldfather, P., & Dahl, K. (1991). Toward a social constructivist reconceptualization of intrinsic motivation for literacy learning. *Journal of Reading Behavior*, 26, 139-158.
- Partridge, S. (1990). Assessing students' writing in the 1990s: A discussion. (ERIC Document Reproduction Service No. ED 322512).
- Paulson, E.L., & Paulson, P.R. (1990). *How do portfolios measure up? A cognitive model of assessing portfolios*. Paper presented at the annual meeting of the Northwest Evaluation Association, Union, WA. (ERIC Document Reproduction Service No. ED 324329)
- Phelan, P., Davidson, A.L., & Cao, H.T. (1992). Speaking up: Students' perspectives in school. *Phi Delta Kappan*, 73, 695-704.
- Phillips, T. (1997). Portfolios and flow. In K.B. Yancey & I. Weiser (Eds.), *Situating portfolios: Four perspectives* (pp. 176-181). Logan, UT: Utah State University Press.
- Schunk, D.H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist*, 25 (1), 71-86.
- Seeman, M. (1972). Alienation and engagement. In A. Campbell & P.E. Converse (Eds.), *The human meaning of social change*. New York: Russell Sage.
- Simmons, J. (1990). Portfolios as large-scale assessment. *Language Arts*, 67, 262-268.
- Smith, M.A., & Ylvisaker, M. (Eds.). (1993). *Teachers' voices: Portfolios in the class-*

- room. Berkeley, CA: National Writing Project.
- Stiggins, R.J. (1994). *Student-centered classroom assessment*. New York: Merrill.
- Stipek, D.J. (1993). *Motivation to learn: From theory to practice*. Boston: Allyn & Bacon.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.
- Thompson, J.B. (1990). *Ideology and modern culture*. Stanford, CA: Stanford University Press.
- Tierney, R.J., Carter, M.A., & Desai, L.E. (1991). *Portfolio assessment in the reading-writing classroom*. Norwood, MA: Christopher Gordon.
- Wehlage, G.G. (1989). Dropping out: Can schools be expected to prevent it? In L.W. Weis, E. Farrar, & H.G. Petrie (Eds.), *Dropouts from school: Issues, dilemmas, and solutions* (pp. 1-19). Albany, NY: State University of New York Press.
- Wehlage, G.G., & Rutter, R.A. (1986). Dropping out: How much do schools contribute to the problem? *Teachers College Record*, 87(1), 374-392.
- Wehlage, G.G., Rutter, R.A., Smith, G.A., Lesko, N., & Fernandez, R.R. (1989). *Re-dicing the risk: Schools as communities of support*. Philadelphia: Falmer Press.
- Wehlage, G.G., Rutter, R.A., & Turnbaugh, A. (1987). A program model for at-risk high school students. *Educational Leadership*, 44, 70-73.
- Wiggins, G. (1989a). Teaching to the (authentic) test. *Educational Leadership*, 46(7), 41-47.
- Wiggins, G. (1989b). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.
- Wolf, D.P. (1989). Portfolio assessment: Sampling students work. *Educational Leadership*, 46(9), 35-39.
- Yancey, K.B. (Ed.). (1992). *Portfolios in the writing classroom: An introduction*. Urbana, IL: National Council of Teachers of English.

Manuscript received: August 10, 1999
First revision requested: January 25, 2000
Final revision received: July 17, 2000
Accepted for publication: July 20, 2000

APPENDIX A Categories and Sub-themes

Self: Who am I?

1. reflection (only as it relates to understanding self and identity)
2. self as learner
3. self as literate actor (reader, writer, etc. - this seems to involve agency)
4. acknowledges personal success
5. sense of personal history

Task: How do I view tasks?

6. fun

7. ownership
 - a. choice
 - b. control over some aspect of the task (timing, evaluation, etc.)
8. prior success
9. competence
10. cognitive challenge
11. extrinsic orientation; sense of means to ends
12. intrinsic orientation; interest beyond the task
13. personally meaningful; having a personal purpose

Fit: Where do I fit in?

14. classroom-level membership
15. school-level membership
16. community/family-level membership
17. classroom-level support
 - a. teacher
 - b. peers
18. school-level support
19. community/family-level support
20. classroom-level collaboration
 - a. teacher
 - b. peers
21. school-level collaboration
22. learning from others at the classroom-level
 - a. teacher
 - b. peers
23. learning from others at the school-level
24. learning from others at the community/family-level

Wolf, D., Bixby, J., Glenn III, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. *Review of Research in Education*, 17, 31-74.

To Use Their Minds Well: Investigating New Forms of Student Assessment

DENNIE WOLF, JANET BIXBY, JOHN GLENN III, and HOWARD GARDNER
Project Zero
Harvard University

The irony of social inventions is that one-time innovations turn to habit. There is perhaps no better illustration of how insight turns to reflex than what has happened to the practice of educational testing in American schools. Currently, American educators, the most determined designers, advocates, and consumers of standardized testing, are confronting the limits of the testing technologies they perfected between 1900 and 1950—at least in instances like the College Board examinations—as a way of providing more equal access to higher education. In the face of demands to teach thinking to all students and to open the curriculum to more than recall and simple rules, the shortcomings of multiple-choice formats as a model or as a singular probe for thought have become stunningly clear. There is growing, if far from universal, impatience with student assessment that addresses chiefly facts and basic skills, leaving thoughtfulness, imagination, and pursuit untapped. There is equal impatience with testing treated as a matter of pure measurement rather than an evolving discussion of what versions of excellence will be encouraged. Researchers and educators, families and students want assessment that offers rigorous and wise diagnostic information rather than the rankings of normal curves. In this climate, the possibilities of performance assessments borrowed from fields as disparate as business and the arts have become increasingly, perhaps even romantically, attractive.

This discussion of how we will measure educational progress cannot afford to be sheerly contemplative. In the spring of 1990, the president

This chapter was supported by grants from the Rockefeller, Mellon, and Panasonic Foundations. The ideas in it were substantially refined in the context of editing a series of essays on thinking, supported by the Office of Academic Affairs at the College Examination Board. We would like to thank several readers for their comments and critiques: Lawrence Frase, Ross Traub, and Kate Wilson. The opinions are solely those of the authors.

of the United States met with state governors to put forth a set of national education goals—broad, even sweeping, pointers to change in the educational success of a nation. The third, and longest, of these goals reads:

By the year 2000, American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so that they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy. (cited in Walker, 1990, p. 16)

The word *ensure* should not be missed. Although the words *testing* and *assessment* never surface explicitly in this statement or any of the other six points, undoubtedly both will have a major role in defining and documenting what it is to use a mind well (DeLoughry, 1990; Goodlad & Keating, 1990; Wolf, in press). Moreover, the search for these forms of insurance will occur at an urgent speed: Only 6 months after the presidential goals appeared, there were already proposals for national examinations (Rothman, 1990c).

But we are in no sense ready to "ensure" student learning. Teachers, alone in their classrooms, observe, comment on, and grade students' actions, homework, and projects almost largely without professional training in assessment or ongoing critique by colleagues (Stiggins, 1988, 1989a, 1989b). They work off of solitary intuitions and received traditions; they are frequently conflicted about whether assessments should index behavior, effort, or achievement. Outside of classrooms, we have just the opposite—technically elegant testing systems designed to focus clearly on the achievement of overspecified and particular learner outcomes measured with respect to national norms. However, if we scrutinize the practices and results of this technically elegant system, we find that it distorts instruction (Raizen et al., 1989; Romberg, Zarinnia, & Williams, 1989; Zessoules & Gardner, in press), underscores inequities in access to education (Chachkin, 1989; Hilliard, 1990; O'Connor, 1989), and forecloses on students and teachers becoming active participants in signal debates over the standards that will be applied to their work (Schwartz & Viator, 1990; Wolf, in press). In effect, many current tests are psychometrically accurate and efficient means of data collection—at a high price. Multiple-choice items are, at least in their average and widely used forms, exercises in detection and selection rather than generation. They often enforce a view of single correct answers at the expense of recognizing culturally variegated forms of excellence or contrasting approaches to displaying understanding (Cummins, 1986). Finally, the technical demands of item construction, test design, and psychometrics discourage, if not outright prevent, vigorous discussions of what will count as evidence

that all students are able to use their minds well (Schwartz & Viator, 1990).

To counter this technically elegant but demonstrably distorting testing system, educators are beginning to examine and experiment with alternative forms of assessment. If nurtured and made rigorous, these alternatives might permit the assessment of thinking rather than the possession of information. They might also enable teachers to develop sophisticated clinical judgments about students' understanding of significant ideas and processes and encourage educators to discuss, rather than simply measure, educational progress. However, the design and implementation of these new forms of assessments will entail nothing less than a wholesale transition from what we call a *testing culture* to an *assessment culture*. This transition will have to be wholesale because the observable differences in the form, the data, and the conduct of standardized testing and its alternatives are in no way superficial matters or mere surface features. They derive from radical differences in underlying conceptions of mind and of the evaluation process itself. Until we understand these differences and their network of consequences, we cannot develop new tools that will allow us to ensure that a wide range of students use their minds well.

In this chapter, we sketch out two major dimensions of what is involved in such a shift. These include changes in the basic epistemology of learning and equally far-reaching alterations in our approach to monitoring and evaluating educational progress. In so doing, we discuss the design of several approaches to assessment that seem promising as productive means of not only monitoring but also promoting student learning, including developmental assessments, performance tasks, exhibitions, and portfolio-like processes. But in closing, we apply the demand for mindfulness to those who would move beyond testing to assessment, insisting that they face a number of difficult questions that stand between promise and realization (Rothman, 1990b).

A VIEW OF LEARNING AND ASSESSMENT

Any discussion of changing views of student assessment has to be situated in the larger framework of views on learning and education. At present, we have a national curriculum—a course of study that yields low-level basic skills for a large and diverse population of students (Smith, 1990). But we simply cannot afford schools that train only a very few students in more than decoding level reading, calculation arithmetic, or the kind of writing required for filling out unemployment forms. Everything we know about early literacy learning stresses the power of families (chiefly mothers) to convey early lessons in the pleasures and powers of language for invention and inquiry (Heath, 1983; Snow & Chall, 1982). Therefore, we cannot think that young girls who will "only" be wives

and mothers need nothing more than meager survival skills; if enabled, they first become teachers, passing along curiosity and belief (Comer, 1988). Nor can we think about young minority males who fail in school as falling into a safety net of jobs as manual laborers. As Berryman (1990) points out:

In machining, computerized numerical control (CNC) machiners radically alter the processes of set-up, control, and operation, replacing manual set-up and control by skilled hand and skilled eye, with set-up by symbolic command. Such systems depart in significant ways from the traditional systems of knowledge that reflect accumulated production wisdom. They are content-free, formal, closed conceptual systems that have many of the characteristics of school subjects, such as mathematics or grammar. Individuals who elected traditional machining were usually, while in school, not thought to have to function within such systems. Now they do. (p. 37)

But if we are determined to teach all students more than decoding or calculation, there remains the hard work of determining just what constitutes "being able to think." Many of the aspects of what we need to teach beyond basic skills can be captured if we imagine thinking as a performance (Wolf, in press). First, serious thinking, like any performance, is a combination of humility and risk: It takes on noisy, ill-defined problems, alternately collecting data, observing, and hazing guesses (Holt, 1990; Lakotas, 1976; Lampert, 1990; Polya, 1954). It involves large projects that combine invention and investigation with craft and insight and embedded accuracy. Second, like other performers, someone engaged in thought sustains a long arc of work over time and across obstacles. Thinking involves rehearsals, revisions, criticisms, and new attempts arranged in nothing like the straightforward orderings we offer in discussions of the scientific method or the directions for writing a term paper (Greene, 1988; Perkins, 1982, 1986; Schoenfeld, 1988). Finally, thought, like performance, involves interpretation. Like an actor or a musician, a historian or a scientist has to decide how to make sense of information and beliefs.

The difference between a performance of thought and school learning as we currently teach and test it is evident in what a 17-year-old biology student says about two episodes in his biology class: what amounts to a rote verification of an "experiment" and a final field project in which he had to formulate and sustain an investigation. His accounts of these two experiences are shown in Appendix 1.

Courtesy of the respirometer experiment (see Appendix 1), this student would certainly be able to match the word *respirometer* to a definition. He would be able to write a short essay stating what "happens" when you test for oxygen with dried and living material. He might be able to use the equipment to measure oxygen in another task. But he would

know very little about acting on the critical verbs of science: wonder, experiment, and infer. The results of the field project are promisingly different. To begin, the student learns the facts concerning the history of earlier approaches to forest fires as a part of understanding the larger question of the shifting nature of ecological theory building. He also learns about sustaining work over time—zigzagging from an original notion of fire as damage to a revised conception of fire as a necessity in a healthy ecosystem. He acquires skill in provoking information from diverse sources, including books, interviews, and on-site observations. In this way, he learns about the social construction of knowledge: His formulations emerge from talks with the ranger, from reading, and from sharing early drafts with his family and his teacher. Not incidentally, he confronts the diversity of opinions about fire damage that emerge across time and among different commentators. If only implicitly, he realizes that human knowledge has kinds and varieties rather than solitary correct answers.

To assess this kind of performance demands both a conception of learning and an assessment quite different from what standardized testing currently offers. First, a useful assessment of this student's performance would have to be multidimensional; it ought to capture his craft (his control of relevant information and his skills such as interviewing, researching, and writing) as it is exercised *in the context* of his larger undertaking. It should be longitudinal enough to inquire into the processes through which he developed his understanding. It should offer information about his ability to amplify his own thought by connecting it to tools, resources, and other thinkers. Moreover, it ought to be keen enough to index the student's depth of understanding; whether he acts only as a correct summarizer or whether he develops a point of view—with all the risk of having to meld values and experience with information and data. Second, any district, state, or national assessment of this student's learning should be complemented and supported by classroom practices in which students reflect on the quality of their own work and in which teachers and students openly discuss the standards for good work (Schwartz & Viator, 1990; Wiggins, 1989a, 1989b). Third, any powerful assessment should provide a response to the student, offering both a critique and the possibility of returning to improve the work. Finally, any worthwhile assessment would ask teachers to consider this and other students' work, engaging them in a discussion of what that work suggests to them about science learning and asking them to debate where it stands in their vision of excellence, and what its implications are for changing what is taught.

BETWEEN TWO EPISTEMOLOGIES

But currently any educator who would raise, never mind answer, questions about how we would use testing or assessment to "ensure" that

128

students think hears two competing messages. The first is that salvation lies in shoring up and pushing harder on familiar and well-articulated forms of testing: refining the tools we have, such as commercially available achievement tests, state tests, and the National Assessment of Educational Progress, and testing more frequently with higher stakes for success and for failure. This is the message of the president's recent educational summit, many state education reform acts of the early 1980s, and numerous business roundtables. The second message is that we must turn away from these well-rehearsed measures and practices, in part reclaiming and in part inventing quite different modes of assessment: observations, performances, and portfolios (Collins, 1988; Frederikson, 1990; Resnick & Resnick, in press). In this second view, we must also redraft, if not invent, the assumptions and procedures of traditional psychometrics. This debate is often phrased as if it were a matter of turning away from some dark chapter in earlier learning theories and models of measurement. But perhaps it is more a matter of standing on the shoulders rather than the faces of another generation. We can never know what we should demand of the design of alternative modes of assessment without grasping what we have learned about the limitations of the earlier epistemology of intelligence and the testing practices to which it gave rise.

The Epistemology of Intelligence

Preliminary investigations indicate that an I.Q. below 70 rarely permits anything better than unskilled labor; that the range of 70-80 is pre-eminently that of semi-skilled labor; from 80-100 that of skilled or ordinary clerical labor; from 100-110 or 115 that of the semi-professional pursuits; and that above all these are the grades of intelligence which permit one to enter the professions or the large fields of business. Intelligence tests can tell us whether a child's native brightness corresponds more nearly to the median (or one or another of these classes). This information will be of great value in planning the education of a particular child and also in planning the differentiated curriculum. (Terman, 1922)

The epistemology that informs Terman's views has deep roots going back as far as the Renaissance conception of the great chain of being and early-19th-century notions of progress (Nisbet, 1980). But, it took definitive shape between 1850 and 1900 in the wake of developing evolutionary theory and the design of comprehensive public schooling for what many educators experienced as an abruptly multiplied and heterogeneous population. The view that emerged had several linked components. First, intelligence was a unitary and immutable trait. It had no kinds or varieties, only ranks. Intellect was also envisioned as rare, predictably located (varying with race, class, and gender), and fixed. Therefore, as Terman so confidently described, individuals could be ranked in stable ways according to their mental capacities. Second, education should be organized in a scalar fashion so as to prepare individuals for their inevitable lot in

later life. Consequently, the curriculum should be a ladder: Entry-level learning, and schooling for the less than talented student, should be simple; only at later levels or in classes for those headed toward the "larger fields" could complex skills be introduced. Finally, as a major marker of individual difference and the chief metric for sorting students, it was imperative to assess learning in the most individualistic fashion, apart from collaboration, resources, or tools. No one unfit should slip into the pool of the average or the gifted. These were not a random trio of beliefs. They were a formidable, mutually reinforcing network of presuppositions. To appreciate the almost seamless and unquestionable justification they offered for who could learn what, they must be examined in closer detail.

Intelligence as Ranked, Fixed, and Predictably Located

Evolutionists were fascinated by what set man apart from beast and, by extension, what distinguished "primitive" from civilized society and worker from manager. As these questions were taken up, particularly by Americans, three putatively scientific but highly debatable stances about intelligence appeared. The first, and perhaps most pernicious, proposed to treat intelligence as a matter of relative position on a normal curve, not as a matter of criterion-referenced achievement. But that is only an ordinal description—one that carries no necessary predictions about anything but relative capacity. Yet both early studies of brain size and later studies of intelligence assumed that to be below the mean was to be stupid, and to be above average was to be bright. The result has been an enduring confusion between rank and accomplishment. In a second equally dangerous stance, intelligence was likened to other heritable characteristics such as skin pigmentation and, thus, was taken to be immutable. The final stance concerned predictable associations between levels of intellect and group membership. If race and gender are regularly associated with certain heritable physical features, why shouldn't there be a genetically mediated tie between the pattern or amount of intellect and group membership? In an effort to get at these associations, mid- and late-19th-century evolutionary scientists developed comparative studies in which children, retardates, and women, as well as different races and classes, were used as proxies for "less-developed" individuals based on a profound error in which naturally occurring within-group variation was used as a model for across-group differences (Fallows, 1989; Gould, 1981). In this way, social access and familiarity with the knowledge valued in Western, middle-class culture became fundamentally confused with ability. This stance provided the basis for the conflation of difference and deficit that has dogged the comparative study of intelligence ever since (Cole & Means, 1981).

In the late 19th and early 20th centuries, as the school population

swelled with visibly different immigrants from southern and eastern Europe, as students were being pulled from the work force by emerging child labor laws, and as many African-American families migrated northward, these notions of a ranked, fixed, and predictably located intellect filled a critical need in American public schooling (Callahan, 1962; Katz, 1987; Tyack, 1967). Under these conditions, the conceptions of ranking, fixedness, and predictability provided the "scientific" basis for two enduring institutional responses to the diversity of styles, cultures, and academic backgrounds of students: universal testing and the systems of tracking students.

Prior to compulsory schooling, educators relied on sheer numbers of years in a common curriculum to carry educational distinctions. Once the school-leaving age became 16, however, educators armed with the notion of fixed cohorts of differentially skilled students designed different kinds of instruction and distinct sorts of educational materials for the students who tested at what they interpreted as different levels. This solution was so appealing and so singular that even progressive educators called for sorting, ranking, and the consequent implementation of different courses of study for different students (Fass, 1989). By the period of the Depression, the notions of individual differences in intellect and equal educational opportunity were dangerously linked in a pattern of differential access to the curriculum:

The two concepts were intended to describe a school system that was at once pedagogically enlightened, progressively democratic, and protected the individual from the tyranny of sameness threatened by mass education. The linking of these two concepts was seriously misleading . . . it . . . provided the foundations for systematic differentiations in curricula, in performance, in expectations, and ultimately in school rewards. (Fass, 1989, pp. 69-71)

But particularly in the face of democratic beliefs, there had to be a strong rational basis for sorting students. Test developers like Terman and Thorndike, eager for just such a tool, borrowed and transformed Binet's fundamentally clinical procedure for diagnosing retardation into paper-and-pencil batteries that could be cheaply reproduced and efficiently administered to entire school populations. With this change, intelligence testing became virtually universal, providing scores that were to dominate the work of distinguishing different types of students and virtually creating the language of ranks we still live with: "gifted," "academically talented," and "low-achieving" (Fass, 1989; Oakes & Lipton, 1990; Rose, 1988). Although the rhetoric surrounding IQ has had to soften, the practices it enables remain largely intact. Entry into most gifted and talented or special education classes is set by test scores. Similarly, access to classes in general math or algebra is typically established via scores

on achievement tests that are descendants from Terman and Thorndike's tests in impulse and form.

These tracking practices are not antique: They still keep the majority of American students cordoned off from the best teachers, resources, and instructional practices. (In 1982, for instance, 62% of students fell below the highest tracks and, of that majority, a disproportionate number were females or poor and minority children [Oakes, 1985; Oakes & Lipton, 1990].) This remains so even though we have widespread evidence concerning the destructive rather than remedial character of tracking (Esposito, 1973; Good & Marshall, 1984; Noland, 1985; Persell, 1977; Rosenbaum, 1980; Slavin, 1986). We know that the longer students stay in lower tracks, the more their performances decay. We also know that tracking prevents both less prepared and academically skilled students from making improvements that occur when they participate in heterogeneously grouped classrooms. But the underlying beliefs about intelligence and resulting possibilities for learning bind us still to these destructive practices (Rand Corporation, 1990; Rothman, 1990a).

A Scalar View of Learning

The epistemology of intelligence not only ranks students, but exhibits a privileging of orders over kinds of knowledge as well. This is evident in several ways—in an evaluative weighing of kinds of knowledge, in the conviction that there is a fixed progression linking simple entry-level skills through a long chain of hard-earned but molecular steps leading to mastery, in the view of literature as "soft" and physics as "hard," and in the displays of knowledge deemed appropriate to different gradations of learning (e.g., multiple-choice formats for special education classification and essays for advanced placement candidates).

As far back as Aristotle, philosophers have drawn a sharp distinction between the superior work of thinkers who ask "why" and the inferior work of artisans who make and do. Subsequent theories of knowledge in Western culture have continued to privilege acts of pure thought in this way. According to this view, theory building, the acquisition of concepts, and symbolic manipulations are more worthy and valuable than practical, situated, or commonplace problem solving (Resnick, 1987b). In this view, learning is the acquisition of information or skills that generate decontextualized knowledge (Resnick, 1987a; Resnick & Resnick, in press; Scribner, 1984). Translated into school terms, individuals know nothing until—or unless—they take up the particular brand of formal knowing prescribed by school. For instance, children who can merely say "he runs" or "I run" do not understand the concept of subject-verb agreement. Only those who can state the rule and define the terms grasp the concept. In short, those who have the competence, but not the terms,

the definitions, the algorithms, and the orthodox procedures, cannot be viewed as knowledgeable. Thus, there is clearly a distinct hierarchy of knowledges: At the bottom is practical problem solving and at the upper reaches is theoretical speculation.

This hierarchy of kinds of knowledge was complemented by the scalar view of learning that envisioned a fixed and long chain of small steps that separated entry-level skills from expertise. This view of learning was evident in the very structure of the first generation of intelligence tests developed by Binet, Terman, and Yerkes; all began with simple questions and led gradually to more complex ones in a highly ordered progression. Once conjoined with strict age rather than ability grading, educators had what has since become the bread and butter of public school curricula: a scope and sequence organization that purveys information and low-level algorithms not as concomitants but as prerequisites to concepts, powerful strategies, or independent thinking. In the textbooks of the early 20th century, beginning reading became phonics practiced on single words rather than the attempt to match sounds with remembered texts taken from familiar books. Early mathematics instruction was translated into the deductive application of rules to closely controlled instances (such as "doing the three's tables"). Moreover, successful performance on these kinds of simple and rote tasks became the gatekeeper to more complex and demanding work.

As the curriculum for comprehensive schools was formed, educators sought new ways to convey the requisite skills to large numbers of average students seen by educators as headed for semiskilled labor (Cuban, 1984). The models they drew on included application forms, shipping invoices, and business letters rather than experiments, essays, or journals. Out of these choices have come the media of most public education: fill-in-the-blank exercises, multiple-choice items, sets of preestablished chapter-end questions, five-paragraph essays, book reports, and science labs that entail no more than verification of someone else's hypotheses and procedures in a highly specified order. The verbs that dominate directions for sent work, class instruction, or tests are *complete*, *choose*, and *match*, not *ask*, *analyze*, *investigate*, or *revise* (McNeil, 1986; Olson & Astington, in press).

This parceled approach to cognition was taken up and ratified first by the efficiency movement within schools (Callahan, 1962) and then later by the appearance of explicit psychologies of learning that stressed the molecular and incremental nature of cognition. In these frameworks, learning was formally and explicitly envisioned as a linear sequence of acquisitions in which complex understandings are the result of "the accretion of elemental, prerequisite learnings" (Shepard, 1990, p. 15) that, *it is assumed*, eventually add up to larger achievement. Thus, Thorndike

explained the learning of arithmetic as the formation of local bonds between elements. Later, B. F. Skinner described learning in this way:

The whole process of becoming competent in any field must be divided into what is not a very large number of small steps, and reinforcement must be contingent upon the accomplishment of each step. (Skinner, cited in Shepard, 1990, p. 15)

The conviction was that the answer to efficient and successful teaching was identical—dissolve all learning "all the way back to chains and simple discriminations" (Gagné, cited in Shepard, 1990). Again, these scalar views of learning are not antique; Benjamin Bloom's (1956) taxonomy describes learning as progressing in linear fashion from basic forms of recognition and imitation to higher activities such as analysis and synthesis. Moreover, the bulk of instructional and assessment materials produced as if this description held, despite abundant evidence, that even sophisticated thought follows a "zigzag" course between craft and vision (Lakotas, 1976; Polya, 1954).

This stratification of learning became and remains pervasive, even in fields like the visual arts where one would least expect it. For example, the following excerpt comes from the section on teaching and evaluating student artwork in a state curriculum framework. In it, the writers struggle, but essentially fail, to break out of the language of small behavioral objectives:

As suggested . . . the instructor can specify (the) product or outcome for a particular lesson: visible easily read and identifiable art work where certain criteria is (are) evident. It is important that a specific idea is measurable. An example would be: With appropriate instruction, students will be taught information about body movement, specifically, the bending of joints; the outcome would be that the student will show evidence that the figure which they are asked to produce will demonstrate bent joints; the measurement or evaluation will be clear and simple . . . bent joints. But evaluation can go further. An expressive component can be added. By asking the students to demonstrate bent joints and adding that the figure should be doing an activity, the teacher can get expressive artwork. (Arizona Department of Education, 1988, p. 69)

The distortion is large: in the course of making drawings, the skill of rendering bent joints is not an end in itself; rather, it is a kind of accuracy that is embedded in a much larger endeavor of making meaning. Moreover, expression is no more something you add on to rendition than a point of view is something you tack onto an essay. Here an 18-year-old and her teacher reflect on the integral relation between rendition and expression in the context of portraying a relationship between herself and her father.

Student: This (portrait) is of my dad and myself. . . I drew it pretty much as his likeness, his high cheekbones, the Indian part of him, and I like that. I felt (at first) that I didn't draw

the arm right. (She looks at a later version of the earlier portrait.) It's the same arm. I guess what changed my opinion of it was the fact that, hey, it's not too bad. Because I have pictures in art magazines where the arms are somewhat distorted. . . .

Teacher: I never thought about it . . . but the significance of the arm really encompassing, wrapping around . . . does it add something to the drawing for you?

Student: My dad is just like that arm. I mean, he's really protective. . . . Yeah, I think that arm wrapping around shows his personality, his character a lot. (Wolf & Pistone, in press)

The Assumed Autonomy of Intelligence

Nineteenth-century scientists and educators embraced still a third set of values—one that claimed increasing autonomy and individuality of thought as one went from laborer to thinker. Wage earners had to operate in conjunction with others, and their work depended on tools. By contrast, a philosopher, an artist, or a scientist was an isolated individual—Descartes alone in his bed, Rodin's *The Thinker* lost in contemplation, Eve Curie alone at midnight in her laboratory. Moreover, these individuals used tools, but were not limited to them, dealing primarily in the probes of inspiration, insight, and imagination. Consider how the French mathematician Poincaré framed his struggle to prove that a particular variety of mathematical phenomena, Fuchsian functions, could not exist. As he reports it, he worked for more than 2 weeks steadily, in long bouts, on the proof without success. Yet, one day while climbing aboard a bus, he was struck by the realization that Fuchsian functions were identical to another set of well-known functions, the transformations of non-Euclidean geometry. At this point he recalls:

The incidents of travel made me forget my mathematical work. Having reached Coutances, we entered an omnibus to go some place or other. At the moment when I put my foot on the step, the idea came to me, without anything in my former thoughts seeming to have paved the way for it. . . . I did not verify the idea; I should not have had time, as, upon taking my seat in the omnibus, I went on with a conversation already commenced, but I felt a perfect certainty. (Poincaré, 1955, p. 37)

What Poincaré offers us is a picture of an unheralded revelation discontinuous with his own or others' earlier framing of the problem. Yet we know from his own account that the solution came in the wake of deep involvement in the domain. The account also stresses the private and isolated nature of creative work. Even though it was a time of rapid exploration in abstract mathematics in Europe, we get no mention of colleagues whose views he took to be wrong or insufficient, articles he had read, or debates he had been a part of. Even his account of his insightful interior monologue and his apparently ordinary exterior dialogue underscores the essentially private, and untooled, nature of his mathematical creativity. It ends by privileging individual processes such as meditation

or writing to the exclusion of group processes such as collaborative research or debate. To this tradition, the epistemology of intelligence has added the conviction that individual intellect is one of the major dimensions of individual difference, a characteristic that distinguishes one person from another, just as names or fingerprints do. Thus, it has taken cognitive research from Soviet and Asian cultures to suggest that a critical determinant of cognitive growth may reside in the individual's ability to imitate, eavesdrop, borrow, or seek help (Hatano & Inagaki, 1987; Yegorov, 1978).

Such is the epistemology of intelligence: It is tightly woven around the fundamental image of a unified scale of worth, ratified in biology, and verified in the search for enduring group differences. Intellect, like height, is a trait. Individuals arrive and die in this world with fixed amounts of intelligence. Where this epistemology reigns, the only hope for public schooling is to organize and be informed by those realities and to develop tracks and instructional sequences in which the acquisition of conventions and basic skills acts as a gatekeeping mechanism. And the only reasonable way to make the system rational is for each student to take the same scientifically designed tests and to be educated according to the results.

The Culture of Testing

In the earliest investigations of intelligence, certain physical characteristics such as brain weight, head circumference, stature, and limb length were taken as clues to an individual's intellectual development, just as gills, circulatory systems, or skeletal patterns were taken as evidence for an animal's position in the evolutionary order (Gould, 1981). Later, in the work of psychologists like Terman and Thorndike, calipers and scales gave way to a series of test items that were believed to measure not the brain itself, but the individual's capacity for intellectual work. In this way, the *concept* of intelligence was reified, and testing became the instrumental realization of the many presuppositions endemic to the epistemology of intelligence. Although the effects of this epistemology on the form of standardized tests are well documented (Collins, 1988; Gardner, in press; Resnick & Resnick, in press; Wiggins, 1989a, 1989b), the culture of testing extends far beyond the specifics of item design and test construction. It involves a network of activities that extend to the conduct of testing, the forms of reporting data, and the uses to which those data are put. Chief among these practices are the emphasis on relative ranking above actual accomplishment; the privileging of easily quantifiable, rather than messy and complex, displays of skill and knowledge; the assumption that individual performances, rather than collaborative forms of cognition, are the most powerful indicators of educational progress; and an image of evaluating educational progress as a matter of scientific measurement.

The Dominance of Ranking.

The view of intelligence as rare, fixed, and predictably associated with group membership has meant that we have designed tests to yield data that will rank students' performances rather than describe their level of learning against some fixed and external criteria. Thus, the dominant tool or image for testing is the normal curve, which presumes some poor performers, a bulge in the middle, and a few high performers. This tool also assumes that the most useful form of information is the comparison of students or cohorts. Hence, in major tests, there are many items composed of only those types and contents that distinguish between students. Consequently, the vast majority of test results are reported in terms of percentiles or ranks that describe where a student stands with respect to other comparable students (by age, grade, type of school, etc.). The resulting pressure is to move up in rank rather than in accomplishment.

But if closely examined, this is a troublesome legacy. Given our history of ranking and tracking, the mathematical necessity of a bottom quartile translates into a readiness to believe that there must be a group of students who cannot learn. Moreover, information about where a student actually stands with respect to achieving the standards for good work in a domain is obscured. A student may be calculating better than 50% of a national sample of sixth graders and still be much less than halfway toward the kinds of mathematical knowledge that will allow him or her to function in a job that technology may change three times in his or her working life (Berryman, 1990; Cannel, 1988, 1989). Conversely, a student may consistently score in the 30th percentile across 3 years and still have acquired significant understandings in mathematics or reading—news that the student, her or his family, and teachers might use to fuel belief and further efforts (Comer, 1980, 1988; Dweck, 1986). In essence, for all the sophistication of our testing system, the concern for ranking and classifying has led to the acceptance of a significant proportion of failure or poor performance as "natural." The attention to such relative information overshadows the responsibility to see that all students learn and the necessity to provide explicit information about students' current levels of achievement.

Testing Artifacts

Current forms of standardized testing are largely paper-and-pencil collections of individual items with single correct answers, requiring chiefly the detection of errors or selection among alternatives. Given this structure, many tests are without analogues in the actual conduct of problem solving, writing, or experimentation. Consider a sample item that a class-

room teacher modeled on one that appears in the language arts section of the standardized test her children take:

Find and correct the error in each section of the sentence.

/the old dog/good home/slow/
A. B. C.

A should be:

- the Old Dog
- The old dog
- The Old Dog
- it is correct

In writing or speaking, one is guided by a sense for the intent and meaning that develops over the long run of the previous discourse. The above item is a single sentence with no surrounding context. To write or speak is to produce—to have ideas, to map them into chosen words, and to reflect on the power and accuracy of what you have produced. To pick out and to correct is only to detect. At its best, this item might be a model for the skill of copyediting or paper correcting. At its narrowest, it is a predictor of how a student will do on other such items. It is as incoherent a sample of a student's ability to generate or understand written English discourse as a written test for the administration of cardiopulmonary respiration would be. Because they are based on these kinds of items, standard achievement tests inquire only into answers as end products. They offer no way to sample the wondering, investigation, data collection, or reflection that are essential to serious work well done (Collins, 1988; Fredrikson, 1990; Resnick & Resnick, in press). Moreover, tests are timed collections of first-draft responses. In this way, tests promote a model of knowing that privileges speed and efficiency above contemplation and accuracy above risk (Calthan, 1962). Thus, as a result of this kind of testing and the instruction that flows from it, students who can do quite well on the items may, in fact, be unable to perform in the domain those items are meant to sample (Schoenfeld, 1987).

In this way, we have yoked the high stakes and publicity of testing to the strange work of teaching students how to engage in cold-start, often artifactual forms of learning and to understand that speed and correctness outweigh overall quality and risk. Not surprisingly, these messages overwhelm the boundaries of standardized testing. Teachers' own tests frequently mimic these structures and underlying values. Textbook exercises customarily use the item types that appear on tests, and in many schools where districtwide end-of-course exams are required, there are often explicit rules that the exam must contain at least one half to one third multiple-choice questions.

Tests as a Portrait of Solo Learning

Many tests also enforce a view of cognition as individual and isolated, insisting that the correct measure of achievement is the most naked: No collaboration and no tools are to be present. The instructions that any teacher reads to students before they break the seal on their California Test of Basic Skills or Metropolitan test booklets are indicative. They bid the teacher to respond to questions only by rereading or paraphrasing the original test instructions. Those same directions also bid students to do their own work, not to converse or look at another student's paper. Those same test directions often insist that students put away all books, papers, and calculators. It is the absolutely individual, solo performance that is sought—not what a student can accomplish in concert with others or with the tools or resources of a domain or his or her own preparations based on those tools. So deep do these beliefs run that it was front page news when, in the fall of 1990, the College Board announced that students might use calculators in the SAT. Moreover, we have been so convinced of the notion that intellect is an isolated, individual quality that we utterly lack the procedures or the psychometrics to study students' performances in group situations. We regularly assay the quality of a school or a district via aggregated individual student scores, never imagining that perhaps we ought to be examining the quality of what occurs in classroom discussion or the range of performances in groups of students working together. In this way, our belief in the individuality of intelligence is peculiarly out of touch with the nature of most serious adult work, occurring as it does in the context of co-workers and tools designed to amplify human performances (Resnick, 1987b).

Tests as Scientific Instruments

Finally, we presume that standardized tests are instances of scientific measurement acting like thermometers, fathometers, or scales. They are meant to slip into an educational setting (e.g., for a brief period), collect relevant data, and silently withdraw, causing no perturbation. They are presumed to be neutral, treating students just like the ocean bottom or any substance to be weighed—as if the students who tick off items were inert matter to be assayed and as if all the agency and inquiry belongs to those doing the measurement. The same might be said for teachers: The tests they administer are frequently manufactured by outside experts on the grounds that only these outsiders have the expertise to make such precision instruments. Tests arrive sealed in plastic pouches and, when completed, must be resealed and signed—unread by teachers. Ahead of time it is difficult to examine entire tests, to question them, or to get access to sample performances that were scored as passing, failing, or

superior. In most cases, neither students nor teachers will see test results, or at best they will see *only* results without the opportunity to look at errors and successes. Consequently, there is little opportunity to contest, discuss, or learn from their students' performances. Through these routines, we teach students and teachers not to value their past performances as a source of learning.

The nature of scores on standardized achievement tests is equally revealing. The number of items correct, not the overall quality of response, determines the score. This accretion model obscures precisely those dimensions and qualities that inhere in no one answer but characterize larger performances of thought: pursuit, research, imagination, and point of view. It is as if the number of completed sentences in an editorial mattered more than the overall power of its argument or the integrity of its perspective. In this way, the technology of scoring has become one of the most powerful realizations of behaviorist views of learning and performance.

Thus, the epistemology of intelligence, realized in the culture of testing, has given us an efficient and highly quantifiable way of assaying student learning. The difficult discovery has been that both what is assessed and the manner of assessment are problematic. Standardized tests, the chief technology of this epistemology, are constructed so that only a few can score high; they have yielded artifacts rather than authentic samples, promoted an isolated view of learning, and treated assessment as a matter of pure measurement. Out of these realizations has developed a growing conviction that we must find alternatives: first, an epistemology that is concerned with accomplishment rather than rank and, second, modes of assessment that sample what it is we want students to know rather than what it is possible to score cleanly.

Epistemology of Mind

From its earliest applications, the epistemology associated with measuring intelligence and the ensuing culture of testing galvanized an opposing point of view. Its proponents argued against the rare and fixed nature of intelligence, and perhaps no one spoke out more vigorously or more elegantly than Charles Eliot, at least during his tenure as the chairman of the Committee of Ten on Secondary Schools at the close of the 19th century. During those years, his was a deep-running belief in the potential of virtually all learners, and consequently his opposition to a tracked and segregated curriculum based on teachers' intuitions or on test scores was heated:

It is a curious fact that we Americans habitually underestimate the capacity of pupils at almost every stage of education from the primary school through the university. . . . It seems

37D DC

to me probable that the proportion of grammar school children incapable of pursuing geometry, algebra, and a foreign language would turn out to be much smaller than we imagine. (cited in Madaus, 1990)

In recent years, Eliot's early values and hunches have gathered empirical support with the emergence of quite a different view of learning, what we term the *epistemology of mind*. In this epistemology, the capacity for thoughtfulness is widespread, rather than the exclusive property of those who rank high, and our views of students' abilities are susceptible to change. Not only may students' capacities leapfrog our predictions, but our cultural conceptions of skill and learning inevitably develop (or at least change). In this alternative theory, learning at all levels involves sustained performances of thought and collaborative interactions of multiple minds and tools as much as the individual possession of information.

The Widespread Presence of Mind

Eliot spoke out of a long tradition reaching back to Jefferson that envisioned a broadly educated populace, regardless of whether they worked as farmers and cabinetmakers or physicians and scholars. The underlying tenet is that the capacity for thoughtfulness is widespread: To make the best use of expensive mahogany requires knowledge of geometry. Currently, both developmental and cognitive psychology offer empirical and philosophical grounds for entertaining this more generous view of the distribution of thoughtfulness (Gardner, 1989; B. Rogoff, 1989; Scribner, 1984). Together, these perspectives emphasize a notion of mindfulness as an essential human quality that looks much the same across infants and adults, blue-collar workers and professors. Across ages and classes, all learners construct (rather than merely absorb) knowledge, because inference, observation, rule generation, and theory building are open to all. Studies of young children learning their native languages indicate that even 15-month-old children are able to generate powerful grammatical rules based on the diverse input they receive (R. Brown, 1973). Young children reinvent the categories that underlie their uses of male and female pronouns and the cognitive architecture that makes counting such a powerful way to organize and compare phenomena as different as pennies and number of teeth lost (Armstrong, 1980; Beth & Piaget, 1966; Carey, 1985; Paley, 1986). Similarly, if we acknowledge rather than denigrate the everyday forms of thought, we find that it is not just engineers and mathematicians who solve problems in sophisticated ways. If we look, for example, at the arithmetic practiced by members of Weightwatchers and by workers loading dairy cases (Lave, 1988; Scribner, 1984), or at the rhetorical and linguistic skills of oral narrative and personal arguments (Cummins, 1983, 1986), we find that individuals who perform poorly on

academic tasks can recognize and resolve problems where the situation is motivating, the materials concrete, and the stakes high and clear.

If mind is not rare, neither is it as fixed as Terman and his colleagues would have had us believe. Researchers working with retarded and at-risk children, for example, have demonstrated that it is possible to teach what were at one time construed to be immutable and inaccessible constituents of intelligence, such as basic strategies for problem solving and reading comprehension (Feuerstein, 1979; Hilliard, 1990). Moreover, these same teaching efforts have led not to temporary or memorized strategies, but to the lasting and apparently generative acquisition of more demanding standards for questioning, comprehension, and remembering (A. Brown, 1988; Palinscar & Brown, 1984). Research of this kind upsets the usual presumption that only achievement, not underlying ability, can be modulated by experience and instruction. Instead, these studies suggest that the apparent immutability of intelligence may have more to do with what we have been willing or able to *teach* as capable of being taught.

At the same time, it is clear that intelligence cannot be fixed in at least one other sense: Our cultural concept of what counts as thoughtfulness inevitably evolves. At one time, to be literate was to be able to read out of highly familiar texts, whereas we now expect readers to decode anything at first glance (Wolff, 1988). At an even more pervasive level, a number of theorists have called into question the most basic presuppositions underlying the epistemology of intelligence. Although there remains debate (Lohman, 1989), many researchers doubt that intelligence is the unitary dimension of human mental functioning that Binet and Terman proposed. They argue instead that the human ability to process and understand is the result of the interaction of multiple intelligences such as an acute understanding of verbal, spatial, or kinesthetic information or differing kinds of aptitudes such as "street smarts," creativity, or the capacity for academic inquiry (Gardner, 1982; Sternberg & Wagner, 1985).

Performances of Thought

Long before Skinner and Gagné gave us visions of student learning as a matter of gradually accruing basic skills, educators raised questions about the image of understanding as a linear sequence of small skills. Again, in writing for the Committee of Ten in 1892, Eliot insisted that it was not enough for students to possess what he called "the furniture of the mind." Students must also have "the discipline" of the mind, that is, the capacity to generate and apply that information in thoughtful ways (Eliot, 1961).

Like Eliot's earlier reservations about a stratified curriculum, this view has been borne out a century later by the research of developmental and

cognitive scientists. If that work makes a major point, it is that learning does not take place in small, linear increments but is better described as occurring in qualitative and uneven shifts in understanding. A 5- and a 10-year-old, therefore, are dramatically different thinkers not only because of the amount of information they control, but also because the older child understands how to organize a task, how to think in conventional categories, and how to take directions. However, whereas the 10-year-old may be able to organize a game or a narrative, she may have a hard time setting up a rigorous experiment where she isolates a cause. Similarly, the differences in a novice and an expert photographer, or an introductory and an experienced chess player, do not reside so much in the information they control as in their ability to organize that information all at once, in a vision of the finished work or the projected game. Thus, learning might be better described as passing major milestones in the nature and organization of knowledge and skills (Chall, 1983; Fischer & Bullock, 1984). Learning is the individual's understanding of how to apply what she or he knows (e.g., when to estimate and when to count, when to make a drawing and when a map is more to the point, when to use standard English and when to speak in a home dialect). As Messick (1984) puts it:

At issue is not merely the amount of knowledge accumulated but its organization or structure as a functional system for productive thinking, problem solving, and creative invention in the subject area as well as for further learning. (p. 1)

If developmental research has a second point, it is that there is intelligent activity at all levels. A dyslexic student who invents alternative routes to spelling and reading exhibits high levels of cognition, despite what appear to be low levels of literacy. Young children, who lack conventional world knowledge or certain structures of logical thought, nevertheless display ardent forms of questioning and investigation (Tizard & Hughes, 1984).

Mind as Extending Beyond the Skin

Thinking, like other performances, is also profoundly social—it faces outward. It occurs often at the instigation of others (B. Rogoff, 1989). Understanding becomes deeper or more complex with the opportunity to witness other minds at work and under the pressure of intrusions, challenges, and differing versions of others (Collins, Brown, & Newman, 1989; Halano & Inagaki, 1987; Palinscar & Brown, 1984; B. Rogoff, 1989; R. Rogoff & Lave, 1984; Vygotsky, 1978; Steyer, Kilpatrick, & Schlieinger, 1990; Stenmark, 1990). Thinking occurs in conjunction with resources and tools: pencils, resource books, or computer programs that

contain both data and algorithms that will organize and manipulate that data (National Council of Teachers of Mathematics, 1989; Pea, 1987). Moreover, it is with the use of such tools that people with limited education or novice-level skills are able to participate in cognitively complex activities that turn out to be powerful learning environments. For instance, keyboards and word processing permit dyslexic students to write. Audiotapes accompanying children's books permit 3-year-olds to "read." To invent or to use cognitive prostheses, whether they are diagrams or software programs, is to understand how to overcome the limits of one mind and one memory. This leads to a radically different view of ability; any individual has a range of knowledge and competence rather than some fixed level of performance. Depending on how much support and familiarity with the materials at hand she or he has, an individual's performance will be greatly affected. It may be just as crucial to measure the quality of that supported performance—or the gap between solo and supported thought.

The Culture of Assessment

Even in its earliest formulations, the culture of testing, like the epistemology of intelligence, provoked dissenting responses. No one was a more articulate doubter of the technology of tests and their underlying presuppositions than William James in his *Talks to Teachers*:

No elementary measurement, capable of being performed in a laboratory, can throw light on the actual efficiency of the subject; for the vital thing about him, his emotional and moral energy and doggedness, can be measured by no single experiment, and becomes known only by the total results in the long run. . . . Be patient, then, and sympathetic with the type of mind that cuts a poor figure in examination. It may, in the long examination which life sets us, come out in the end in better shape than the glib and ready reproducer, its passions being deeper, its purposes more worthy, its combining power less commonplace, and its total mental output consequently more important. (James, 1913, pp. 135-143)

Implicit in what James wrote were key critiques of the practices that were to amplify and deepen the effects of testing. James worried that assessment would become solely a matter of one-time measurement. He argued instead that if we want rigorous evidence concerning educational progress, we must describe large-scale accomplishments and we have to think developmentally, collecting longitudinal data that follow the growth of a student "in the long run" against the background of major accomplishments in a field or subject. But James went further still. In his use of words such as *passion* and *purpose*, James also points at another characteristic of assessment. As compared to measurement, assessment is inevitably involved with questions of what is of value, rather than simple correctness. Questions of value require entry and discussion. In this light,

assessment is not a matter for outside experts to design; rather, it is an episode in which students and teachers might learn, through reflection and debate, about the standards of good work and the rules of evidence.

From Relative Ranks to Developmental Assessments

If William James was right, then assessment ought to document *accomplishment* rather than the relative positions of Terman's "normal" distribution of individuals within a population concept of intelligence. Key in this change is a move from norm-referenced to criterion-referenced evaluations of student learning in which what students can and cannot do is clearly stated. These descriptions have to be anchored at one end in the capacities most children bring to school and at the other end in the capacities all high school graduates should possess. Between these endpoints, moments of major conceptual reorganization have to be cited and described. The point of these developmental sequences is that a student's real, rather than relative, skills can be assessed both for the adequacy and the fullness of his or her learning. The materials in Appendix 2, taken from the current national assessments in Great Britain, sketch out what adequate or expected levels of accomplishment are for young adolescents in the domain of design and technology. They also suggest directions and extent of progress, and, in this way, indicate explicit educational goals and standards.

To document accomplishment, we must also, as James hinted, design assessments that are longitudinal, sampling the baseline, the increment, and the preserved levels of change that follow from instruction. Building on notions quite close to those suggested by James, Ralph Tyler has long argued for the power of plotting such growth curves:

One is not able to evaluate an instructional program by testing students only at the end. . . . Without knowing where the students were at the beginning, it is not possible to tell how far changes have taken place. . . . However, it is not enough to have only two appraisals in making an educational evaluation because some of the objectives aimed at may be acquired during an educational program and then be rapidly dissipated or forgotten. In order to have some estimate of the permanence of the learning, it is necessary to have still another point of evaluation which is made sometime after the instruction has been completed. (Tyler, 1949, pp. 106-107)

Many teachers working in primary school classrooms engage in this kind of developmental and longitudinal assessment of emerging literacy. On the basis of extensive naturalistic research on early reading and writing, the major steps to independent literacy have become clear. Using checklists, teachers observe children at work, keeping "running records" of students' current level of writing skill as evidenced by both independent and group work. They index these skills at a number of points during

the year, monitoring whether children have passed the milestones that would allow them to change the level of their current reading and writing activities. A sample of one teachers' checklist generated for kindergarten children in a school assessment in Madison, Wisconsin, is shown in Figure 1.

Even though it is simple, this approach to assessment provides concrete descriptions of what children can and cannot do posed within a full range of accomplishments characteristic for the earliest phase of literacy development. Also, through its columns for repeated observations, this approach permits a look at whether or not the child's literacy skills are changing substantially or only marginally during a period of time when reading and writing are key school tasks. In accompanying comments, a teacher can index the robust or thin nature of these acquisitions by including comments, observational notes, and samples. As simple as this approach appears, along with regional and national standard-setting and adjudication sessions, it forms the basis of some aspects of national assessments in countries as large and as diverse as New Zealand and Great Britain and states as diverse as California and Connecticut (Baron, 1990; California Assessment Program, 1990a, 1990b).

Although such assessments are a beginning, they are only a beginning. For instance, they fail to capture the depth or richness of student learning. Is it accomplished with simple and meager forms of understanding, or is it done in a way that exhibits invention, transfer, and further inquiry? Appendix 3 describes how teachers working for the Northern Examining Board in England have tried to capture the leanness versus richness of student work. Quite possibly, then, we need to break step with the usual two-dimensional models of charting student progress that simply map age against level. In addition, we have to find ways to index newer dimensions such as the "thickness" or richness of accomplishment (Traub, personal communication, October 15, 1990).

There are significant consequences here for how achievement data are reported and educational goals are set. Major questions become ones of levels of accomplishment, longitudinal change, and depth of understanding. In thinking about accomplishment, we have to do away with the ranking and the relativity carried by the technology associated with normal curves. The real point is, How many children, and at what ages, are there at each milestone? In thinking longitudinally, we have to ask, How much genuine movement does a child (or a group of children) make in a given period? That is, how steep is the learning curve from baseline to year-end for middle- and working-class children or African-American and Caucasian children? In thinking about depth of understanding, we have to ask whether we are getting small or large numbers of children to un-

FIGURE 1
Teachers' Checklist of Early Steps in Literacy

Writing Development Record
for _____ yr. _____ teacher _____

	Occasions for Observation											
Forms of Writing												
Conventional Spelling												
Invented Spelling - transitional												
Invented Spelling - beginning												
Copying Environmental Print												
Letters - random												
Letter-like symbols												
Scribble												
Drawing												
Forms of Rereading												
Conventional												
Attempt to decode/read sight words												
Recalls intended content: written language-like												
oral language-like												
Label/Describe/Expand												
Dialogue with Teacher												
Refusal												
Eyes on Print?												
Yes												
Occasionally												
No												

Note. Adapted by permission from working materials of the Madison Metropolitan School District, 1989. The first page of that text was, in turn, adapted from "Developmentally Appropriate Assessment of Reading and Writing in the Early Childhood Classroom" by William H. Teale, 1988, *Elementary School Journal*, 89 (2).

understand what they are learning in generative ways. The point is simple: The purpose of assessment must be to demonstrate where students are in their educations, not where they are with respect to a constantly changing population of peers (Cannell, 1988, 1989). Within this framework, the index of educational success becomes not the will-o'-the-wisp chase for all students and schools to clamber above normal, but the demonstration that with each successive class of first or fifth graders increasing numbers of students from all segments of the population attain later milestones, exhibit steeper growth curves, and acquire and hold on to a deeper understanding. To do so, we have to replace the hegemony of normal curve and two-dimensional statistics with longitudinal and criterion-referenced measures of more than age and level.

Sampling Performances of Thought

At the Waterford crystal workshops, in order to graduate from apprenticeship, an individual has to cut the "apprentice bowl"—one that bears samples of each kind of cut a master would have to produce (Madaus, 1990). Similarly, auditions, driving tests, foreign language proficiency exams, and athletic competitions all directly sample the behaviors valued (Hirsch, 1989; Wiggins, 1989a, 1989b). Thus, if we care about the performance of thought, that is what we have to sample and assess.

Increasingly, in an effort to break free of the artifactual knowledge evaluated in many standardized tests, schools, districts, and states are instituting what is often termed *performance assessment*. In these evaluations, students are asked to write, to read, and to solve problems in genuine rather than artificial ways. The example of a performance assessment shown in Appendix 4 is taken from the Pittsburgh Public Schools Syllabus Examination Project and was designed as the final exam for individual students in a required high school course in Shakespeare based on *A Midsummer Night's Dream*. A centerpiece of this course is the work of viewing and interpreting three distinctive filmed versions of the one scene as part of the larger effort to understand that plays are essentially scores with many interpretations. The essay portion of that exam is presented in Appendix 4.

The assessment pushes beyond what might well be considered a familiar aim of traditional literature exams by acknowledging the place of film and performance as much as text. It insists explicitly that students are to take a point of view and support that stance with examples from vision, audition, and memory. It stresses the possibility of a range of possible answers. In these ways it models hunches and analyses of avid film viewers and readers.

Assessments of Distributed Cognition

If we want to acknowledge and understand what students know about the profoundly social character of mind, we have to alter the forms of assessment we design in major ways. For instance, if we take assessment events like graduate thesis orals or the review of scholarly papers as serious models, we would see that assessment, not just learning, can and perhaps ought to be a highly social experience. In thesis orals, students quite literally defend their research to a diverse audience that includes professors and often other students. They take on-line questions, and part of the assessment touches on their ability to present themselves as knowledgeable, alert, and able to think *in situ*. In the case of the journal review, the author's work is reviewed by more than a single reader and a range of opinions is returned. The expectation is that the spectrum of opinions faces the writer with making good use of criticism—understanding where reviewers all converge in their recommendations and where they offer divergent opinions that are a matter of point of view and taste.

These same qualities can be extended to many levels of education through what Sizer (1984) has referred to as exhibitions. Clear examples of such exhibitions include entry into athletic competitions, graduation recitals at a performing arts conservatory, and science fair displays. They share many of the qualities of performance assessments (attention to large issues, sustained work, etc.). However, they can be set off by an additional characteristic. Exhibitions demand intense attention to the social and public nature of understanding. The designers of the International Baccalaureate exams in foreign language, in order to inquire into students' competence as readers, listeners, and speakers in a second language, ask them to read a book or a series of books in a foreign language and to prepare a 20-minute talk on the author or a subject of interest. Furthermore, this talk must be delivered using only note cards. Following the talk, the student is asked questions by fellow students and by a panel of examiners. The student is graded not only on the prepared portion of the talk, but on the flexibility with which she or he entertained the inquiries from the floor—as someone traveling in a foreign country certainly would have to do.

It is equally clear that we need assessments that focus on what students can do with resources: Do they know how and when to call on texts, colleagues, or documents? The work of the national assessment center (CITO) in Holland provides a remarkable illustration of this kind of assessment. Each year in the fall, the committee of teachers who design the yearly and school-leaving exams announces the organizing theme or central problems for the national end-of-year examinations. In art history, for example, the committee selects a theme that will be the unifying sub-

ject for the exams, such as "the supernatural" or "revolution" as exemplified in the art of a diverse range of periods and cultures. In order to support the best possible performances by students, the government funds the development and distribution of packets of background materials (readings, slides, study questions, and lists of Dutch museums where relevant works can be found). Teachers are encouraged to use these materials and to work intensely with their students on the selected theme, modeling the process of visiting works, discussing them, and developing a point of view. Here, the occasion of year-end assessment is used, quite deliberately, to help students synthesize and extend their current knowledge through models, conversations, and demonstration of how to use cultural resources (Hermans, 1988).

Assessment as an Episode of Learning

Growing out of the epistemology of mind is quite a different conception of evaluating student learning. It presumes that all learners have minds that develop with experience and teaching; it underscores thinking rather than the possession of information; and it recognizes that minds function in concert with other minds and the tools their culture makes available (Baron, 1990). But, in addition, if testing is modeled on the measurement of learning with a neutral instrument, assessment practices are modeled on something closer to the workings of a visiting committee or the debate that goes on among film critics. At its heart are reflection, judgment, and discussion, in addition to quantification and reliability.

In the culture of testing, we presume that a week of Stanford Achievement Tests can slip in unnoticed, like a thermometer. But if we look at the letters, journals, and notebooks of productive and inventive individuals, it is clear that these individuals use occasions of both internal and external assessment to reflect on their own cognitive and creative footprints. These reflections are a way of deriving lessons and possibilities for future work (John-Steiner, 1987; Perkins, 1982). Thus, in assessments, we ought to recognize the perturbations caused by any evaluation and attempt to use them productively for both students and teachers. In this context, the point of collecting samples of student performances extends beyond the derivation of indicators of educational progress. Assessment then becomes an occasion for learning.

Portfolio-based assessment can provide a context where students can learn to regard assessment as an occasion for learning. Typically, portfolios include not only earlier and later works, but also a structured sampling of different kinds of works-characteristic of the domain (e.g., writing in multiple genres or math problems taken from different topics). The result of keeping portfolios is that teachers and students have access to a continuous body of work in which to discern acquisitions, characteristic

patterns of interest and style, or remaining difficulties. As work with portfolios has been pursued in greater depth in recent years, however, this traditional conception has begun to evolve. For example, in the context of Arts Propel, an assessment project in the arts and humanities conducted in the Pittsburgh public schools, students, teachers, and researchers have been developing what are called *process-folios*. These collections of work differ from familiar portfolios in a number of ways. The generation of these process-folios is embedded in a much larger classroom context where teachers and students frequently discuss what goes into creating worthwhile work, what makes for helpful critique, and how to plow comments back into ongoing work. In addition to finished works, these collections contain sample "biographies of work" — documentation of the various stages of a project. When collected at diverse points, these biographies permit a longitudinal look at a student's changing control of the processes for shaping a final piece. Students often keep journals and write reflections about their work (Seidel & Zessoules, 1990). Finally, the collections of work students build are anything but archival. They regularly return to earlier works to revise or make comparisons with later ones. At the close of the year, students reenter their collections to make a final selection of biographies, reflections, and final pieces that can serve as the basis for a course grade and/or part of a permanent record of their development (Camp, 1990a, 1990b; Howard, 1990; Wolf, 1989). In this sort of work, students have the opportunity to see samples of different levels of work and to discuss the criteria that distinguish strong performances. They also witness the multidimensional nature of such work (i.e., that it involves the ability to pose an interesting problem, to learn from and comment on someone else's work, or to revise an earlier draft). Finally, if students are permitted to reenter the assessment process by making new selections or improving earlier ones, they learn something about how to make use of criticism and reflection to improve their skills and understandings (Brewer, 1990; California Assessment Program, 1990a, 1990b).

If students benefit from being included rather than excluded from assessments, so do teachers. Although most teachers receive little or no training in assessment, based on experiences in district-level competency exams, advanced placement programs, and International Baccalaureate programs, we know that with the requisite training teachers can become skilled and reliable appraisers of student work (Berliner, 1989; Clark, 1988; Elbaz, 1983). Thus, when teachers have to set the questions for end-of-year exams, they must grapple with what is worth remembering and carrying forward. When they read district-level exams or portfolios, they have to confront very difficult questions concerning competency and excellence. One group of high school teachers reading writing tests, for

example, came across a well-written but highly racist essay. For several hours, they debated whether the paper was to be scored purely on the grounds of craft or whether the politics of the writer could enter into its rating. For the participating teachers, this was an important moment in working out the criteria on which they would judge the writing of students in their district.

At this juncture, the point of assessment becomes much more complex. Yes, it is a part of holding educators accountable to an outside community for good work well done. However, it is also designed to promote intense discussion of standards and evidence among all of the parties who are affected. This involves a move away from the presumptions of pure measurement to a model of clinical judgment (LeMahieu, personal communication, September 26, 1990). If teachers are going to take charge of instructional and diagnostic work, and if they are to grow and develop as professionals, they need the opportunity to debate standards and dimensions of performance (National Board for Professional Teaching Standards, 1990). Moreover, if students are going to leave school as competent assessors of their own work, they deserve sustained opportunities to internalize standards and ways of questioning and improving the quality of their work. This should and will have deep repercussions for how we think about traditional approaches to scoring and reliability.

CONCLUSION

As our conceptions of students, our definition of intelligence, and our sense for the point of assessment have changed, there has been a growing awareness that we need expanded and diversified models of assessment that offer alternatives to both epistemology and the practices associated with standardized testing. In particular, we need assessments that can illuminate the capacities of a wide range of students, model sustained thoughtfulness, and offer useful information about teaching and learning (Wiggins, 1989a, 1989b).

As the earlier quotes from Eliot, James, and Tyler suggest, the search for good measures of understanding is not a contemporary concern. Nearly a hundred years ago, observers wondered how testing would distort the curriculum. Teachers working in progressive schools as far back as the 1920s relied heavily on running records, projects, and folders of student work (Kilpatrick, 1918; Tyler & Smith, 1942). Many of the innovative science, mathematics, and social studies curricula of the 1960s collected ongoing samples of student work rather than imposing separate tests. Since its inception during the Cold War period, language proficiency testing has made thoughtful use of on-line student performances in conversation, impromptu translation, and realistic writing tasks rather than multiple-choice items (Hirsch, 1989). Therefore, what are often heralded

as "new modes of assessment" are much more accurately described as "rediscovered" modes of assessment. There is a critical lesson here: If the current interest in alternatives to standardized testing is to be anything but this decade's flurry, we have to be as tough-minded in designing new options as we are in critiquing available testing. Unless we analyze the workings of these alternatives and design them carefully, we may end up with a different, but perhaps no less blunt, set of assessment instruments (Linn, 1990). As with any form of assessment, the familiar, difficult, and nasty issues of efficiency, equity, and evidence persist. Whereas there is considerable criticism of the approaches taken by standardized tests, as yet we have no such critical tradition for new modes of assessment. And we cannot be without one.

Efficiency

New modes of assessment will inevitably take shape against the demand for accountability in schools, states, and the nation. The relationship between these multiple levels of assessment could easily become schizophrenic with local forms of student assessment taking the form of intensive sampling of performances, and other levels of assessment adhering to the proven efficiency, low costs, and familiarity of standardized testing. This could be disastrous. On one hand, it creates obvious discontinuities in putting together an articulated view of educational progress. On the other hand, such a situation would undercut students and teachers taking part in the discussion of the evidence and standards entering into high-stakes decisions. Even though exhibitions or portfolio readings are undoubtedly more expensive and unruly than standardized tests, we have to find the means to ensure that such forms of assessment become a part of district- and state-level assessments. We can call on matrix sampling (the selective reading of representative sets of portfolios), involve members of the community as readers, or argue in union negotiations that acting as readers is a part of teachers' professional training and responsibility.

Equity

There is considerable hope that new modes of assessment of the kind outlined above will provide one means for exposing the abilities of less traditionally skilled students by giving a place to world knowledge, social processes, and a great variety of excellence. This may eventually be the case, particularly if assessments are used to unify, rather than stratify, access to knowledge and strong educational practice. For example, teachers writing Pittsburgh's Shakespeare exam are insisting that all students, not only the best and the brightest, be enrolled in a literature course that permits them to think about and interpret what they read. And clearly

the hope is that it draws instruction in its wake, so that the exam is not the first occasion when students are asked to read and think like actors and directors.

There is a second dimension to the work of using performance assessments to expose mindfulness in a wide rather than a constricted range of students. We can deliberately honor real-world knowledge, expressly scaffold the relevant steps of questions in the process, permit students the stimulation of collaborative work (Brown, 1988; Hatano, 1982; Hatano & Inagaki, 1987), allow for a range of forms and formats in responses (e.g., graphs, essays), and provide for revision (Frederikson, 1990). The following example, taken from the Connecticut Core of Common Learning, demonstrates what this kind of assessment might look like (Baron, 1990):

Many local supermarkets claim to have the lowest prices. But what does this really mean? Does it mean that every item in their store is priced lower, or just some of them? How can you really tell which supermarket will save you the most money? Your assignment is to design and carry out a study to answer this question. What items and prices will you compare and why? How will you justify the choice of your "sample"? How reliable is the sample, etc.?

Here students are being asked to apply their understanding of basic mathematical operations (such as averaging, units, and comparing quantities) to a familiar situation in which the goal of the construction of a sample, the scope of inferences to be drawn from the sample, and the applicability of the model have evident consequences. Its designers have effectively broken through the presumption that only calculus students can or should practice demanding mathematics.

However, new modes of assessment such as writing samples or mathematics portfolios could demonstrate equal or greater gaps between the performances of minority and majority children. If so, we have to be prepared to investigate why, rather than cursing or abandoning the project. Is it because the tasks and scoring systems entail subtle presumptions about single, correct approaches (Chachkin, 1989; O'Connor, 1989)? For instance, will a scoring rubric designed to rate the essays of white, middle-class students be able to distinguish what is remarkable, albeit different, about this African-American student's writing?

It was a phrase by Woodrow Wilson, former president of the United States that surprisingly gave the "negro race" hope in 1917. It was the beginning of World War I. The reason given by this president for the American involvement in the war was "to make the world safe for democracy". This phrase had special meaning for the "Old Negro" because for so many long, incredibly long years they were deprived the promised, documented, and signed freedom rights for "man". So when the white president of that time came right out and said safe for democracy most Black Americans of that time thought the president meant them

also. In fact he did not. I say surprisingly gave hope because Black Americans and women are people too and when the constitutions and amendment and bills said "... for all" they found out that these excluded them. So after years of knowing that, some white man said "world democracy," and again this statement excluded them. (Woodard, 1990)

Alternatively, new modes of assessment could reflect the fact that some students have been taught how to enter into and make use of new modes of assessment, whereas others have been left to flounder. Like any measure of student learning, performance assessments entail a number of presuppositions. They assume a student who knows that it is all right to collaborate, that revision is likely to be helpful, or that the test is not culling for a regurgitation of what was done in class. Access to these presuppositions has to be widely available or alternative assessments will reproduce, rather than challenge, old patterns of performance across boys and girls and majority and minority children. Finally, if, in our best judgment, valid and reliable forms of performance and portfolio assessment index gaps in the learning of more demanding thought processes, we have to be willing to examine such pernicious habits as tracking and differential resources for poor schools.

Evidence

If we want to pursue these new modes of assessment, we cannot do so on the mere conviction that they are better. We cannot use the notion of developmental accomplishments or holistic scores to excuse us from developing rigorous standards and thoughtful rules of evidence that will offer candid pictures of what students are learning. We have to establish clear expectations about what a high school student should be able to do in a field like mathematics or history, and equally explicit descriptions of what portion of that should be accomplished by the end of middle school: if we move to developmentally based scoring, are we going to acknowledge that students in the fifth year of school will range in their abilities from those characteristic of second grade to those characteristic of eighth grade? We must think through whether it is enough for students to pass performance assessments with lean, rather than generous, understandings. When we insist that all students be able to engage in performances of thought, what do we actually mean? What do we take as evidence that a student grasps concepts such as "friction" or "heat"? Is the traditional essay or problem set enough, or do we need evidence from experiments, demonstrations, and diagrams? If we are going to look at student performances, we have to learn to design tasks that actually tap the skills in question. If we want teachers as clinical judges and want to honor multiple opinions, are we going to demand averaged scores, or do we want to

preserve the diversity of opinions by means of a range of scores or the collection of comments?

A New Psychometrics

If we are able to design tasks and modes of data collection that permit us to change the data we collect about student performance, we will have still another task in front of us. This is the redesign or the invention of educational psychometrics capable of answering the much-changed questions of educational achievement. In place of ranks, we will want to establish a developmentally ordered series of accomplishments. However, this will entail hard questions in two directions. First, in stepping away from the 19th-century notion of a single chain of being, we are opening up the possibility of multiple paths to excellence. Yet the vast majority of scoring rubrics for performance assessments, exhibitions, or portfolios still presume a single progression from novice to expert. We live in an increasingly multicultural society, learning-disabled students are more frequently mainstreamed, and students enter school with widely varying backgrounds in literacy and numeracy. In this setting, as we move to different modes of assessment, the contest between idealized, universal descriptions of progress and differentiated but potentially divisive rubrics will be fierce.

Second, if we indeed value clinical judgment and a diversity of opinions among appraisers (such as certainly occurs in professional settings or postsecondary education), we will have to revise our notions of high-agreement reliability as a cardinal symptom of a useful and viable approach to scoring student performance. We will have to find a previously uncharted course between insisting on uniform judgments and mayhem. Possibly, we will have to seek other sorts of evidence that responsible judgment is unfolding—that participants agree on the relevant categories for describing performance, that scores fall within a certain range, or that recipients can make thoughtful use of the range of opinions offered to them.

Third, we will have to break step with the drive to arrive at single, summary statistics for student performance. In line with more diversified notions of intelligence, it is critical to develop ways of looking at student profiles, both across and within domains. After all, it is critical to know that a student can arrive at an idea but cannot organize her or his writing or cannot use the resources of language in any but the most conventional and boring ways. It is equally important to know whether students can be authors but cannot read the writings of others critically. Unless we develop these kinds of differentiated portraits of student performance within a domain, it is difficult to envision student assessment ever informing, rather than merely measuring, the educational process.

Finally, we have to consider different units of analysis. At present, we deal in single students or aggregates of individual performances. However, because so much of learning occurs either in social situations or in conjunction with tools and resources, we need to consider what performance looks like in those more complex units. For instance, we need measures of the quality of classroom discussion or the functioning of a heterogeneous group of students working out a problem in economics or history. We also need to understand the kind of scientific modeling students can accomplish given computer support or the kind of writing a student can do given word-processing training and support.

A Reconceptualization of Assessment

Assessments of student learning are exceptionally human documents. Our conceptions of learning and what is worth being learned evolve at the high speed of culture rather than the gradual speed of biology. Only a century ago, to be literate was to be able to sign your name or read a highly familiar text, but neither of these definitions is sufficient today. Similarly, if we are to ensure student learning, we will have to conduct ourselves as learners in developing alternatives to standardized testing. We have to push beyond generating engaging alternatives by listening to critiques and revising and improving our own portfolio of approaches. It will take what the painter Ben Shawn saw as the heart of good artistry: "the capacity to be the spontaneous imaginer and the inexorable critic"—not once, but iteratively—as our culture shifts and our understanding deepens.

REFERENCES

- Arizona State Department of Education. (1988). *Arizona visual arts essential skills*. Tucson, AZ: Author.
- Armstrong, M. (1980). *Closely observed children*. London: Writers and Readers.
- Baron, J. (1990, July). *Blurring the edges of assessment, curriculum, and instruction*. Paper presented at the Institute on New Modes of Assessment, Cambridge, MA.
- Berliner, D. (1989). Implications of studies of expertise in pedagogy for teacher education and evaluation. In *New directions for teacher assessment: Proceedings of the 1988 ETS International Conference* (pp. 39-78). Princeton, NJ: Educational Testing Service.
- Berryman, S. (1990). Sending clear signals to school and labor markets. In J. Schwartz & K. Viator (Eds.), *The prices of secrecy: The social, intellectual, and psychological cost of current assessment practice* (pp. 33-46). Cambridge, MA: Educational Technology Center, Harvard Graduate School of Education.
- Beth, E. W., & Piaget, J. (1966). *Mathematical epistemology and psychology*. Dordrecht, Holland: Reidel.
- Bloom, B. (Ed.). (1956). *Taxonomy of educational objectives. Handbook I: Cognitive domain*. New York: McKay.
- Brewer, R. (1990, July). *The development of portfolios in writing and mathematics for state-wide assessment in Vermont*. Paper presented at the Institute on New Modes of Assessment, Cambridge, MA.
- Brown, A. (1988). Motivation to learn and understand: On taking charge of one's own learning. *Cognition and Instruction*, 5, 311-321.
- Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.
- California Assessment Program. (1990a). *The California Assessment Program: A position paper on testing and instruction*. Sacramento, CA: Author.
- California Assessment Program. (1990b). *Guidelines for the mathematics portfolios*. Sacramento, CA: Author.
- Callahan, E. (1962). *Education and the cult of efficiency*. Chicago: University of Chicago Press.
- Camp, R. (1990a). *Arts Propel: Ideas for stimulating reflection*. Unpublished manuscript, Educational Testing Service.
- Camp, R. (1990b, Spring). Thinking together about portfolios. *The Quarterly of the National Writing Project*, pp. 8-14, 27.
- Cannell, J. J. (1988). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: Issues and Practice*, 7, 2, 5-9.
- Cannell, J. J. (1989). The "Lake Wabegone" report: How public educators cheat on standardized achievement tests. Albuquerque, NM: Friends for Education.
- Carey, S. (1985). Are children fundamentally different kinds of thinkers and learners than adults? In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 2, pp. 487-517).
- Chachkin, N. (1989). Testing in elementary and secondary schools: Can misuse be avoided? In B. Gifford (Ed.), *Test policy and the politics of opportunity allocation: The workplace and the law* (pp. 163-187). Boston: Kluwer Academic Publishers.
- Chall, J. (1983). *Stages of reading development*. New York: McGraw-Hill.
- Clark, C. M. (1988). Asking the right questions about teacher preparation: Contributions of research on teacher thinking. *Educational Researcher*, 17(2), 5-12.
- Cole, M., & Means, B. (1981). *Comparative studies of how people think*. Cambridge, MA: Harvard University Press.
- Collins, A. (1988). Reformulating testing to measure learning and thinking. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafiq (Eds.), *Diagnostic monitoring of skills and knowledge acquisition* (pp. 75-87). Hillsdale, NJ: Erlbaum.
- Collins, A., Brown, J. S., & Newman, D. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Erlbaum.
- Comer, J. (1980). *School power: Implications of an intervention project*. New York: Free Press.
- Comer, J. (1988). *Maggie's American dream: The life and times of a black family*. New York: New American Library.
- Cuban, L. (1984). *How teachers taught*. New York: Longmans.
- Cummins, J. (1983). Functional language proficiency in context: Classroom participation as an interactive process. In W. J. Tikunoff (Ed.), *Compatibility of the SBIS features with other research on instruction for LEP students* (pp. 109-131). San Francisco: Far West Laboratory.
- Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Educational Review*, 56(1), 18-36.

- DeLoughry, T. (1990, March 7). Governors approve education goals with little mention of colleges. *Chronicles of Higher Education*, p. A-24.
- Department of Education and Science and the Welsh Office. (n.d.). *National curriculum: Task group on assessment and testing*.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040-1048.
- Elbaz, F. (1983). *Teacher thinking: A study of practical knowledge*. London: Croom Helm.
- Eliot, C. W. (1961). *Charles W. Eliot and popular education* (E. A. Krug, Ed.). New York: Teachers College Press.
- Esposito, D. (1973). Homogenous and heterogeneous ability grouping: Principal findings and implications for evaluating and designing more effective educational environments. *Review of Educational Research*, 43, 163-179.
- Fallows, J. (1989, May). What's wrong with testing? *The Washington Monthly*, pp. 12-24.
- Fuss, P. (1989). *Outside in: Minorities and the transformation of American education*. New York: Oxford University Press.
- Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device*. Baltimore: University Park Press.
- Fischer, F., & Bullock, D. (1984). Cognitive development in school-age children: Conclusions and new directions. In W. A. Collins (Ed.), *Development during middle childhood: The years six to twelve* (pp. 70-146). Washington, DC: National Academy Press.
- Fredrickson, J. (1990, August). *New approaches to assessment*. Paper presented at the annual meeting of the American Psychological Association, Boston.
- Gardner, H. (1982). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Gardner, H. (1989). Zero-based arts education: An introduction to Arts Propel. *Studies in Art Education*, 30(2), 71-83.
- Gardner, H. (in press). Assessment in context: The alternative to standardized testing. In B. Gifford (Ed.), *Report of the commission on testing and public policy*. Boston: Kluwer Academic Publishers.
- Good, T. L., & Marshall, S. (1984). Do students learn more in heterogeneous or homogenous groups? In P. Peterson, L. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction* (pp. 15-28). New York: Academic Press.
- Goodlad, J., & Keating, P. (1990). *Access to knowledge*. New York: The College Entrance Examination Board.
- Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.
- Greeno, J. (1988). *Situations, mental models, and generative knowledge* (IRL Report No. 5). Palo Alto, CA: Institute for Research in Learning.
- Hatanog, G. (1982). Cognitive consequences of practice in culture specific procedural skills. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition*, 4, 15-18.
- Hatanog, G., & Inagaki, K. (1987). A theory of motivation for comprehension and its application to mathematics instruction. In T. A. Romberg & D. M. Stewart (Eds.), *The monitoring of school mathematics: Background papers, Volume 2. Implications from psychology, outcomes for instruction* (Program Report No. 87-2, pp. 27-66). Madison: Wisconsin Center for Educational Research.
- Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. New York: Cambridge University Press.
- Hermans, P. (1988, July). *The development of art history exams in Holland*. Paper presented at the AICA Institute in Art Education, New York.
- Hilliard, A. G. (1990). Misunderstanding and testing intelligence. In J. Goodlad & P. Keating (Eds.), *Access to knowledge: An agenda for our nation's schools* (pp. 145-157). New York: The College Entrance Examination Board.
- Hirsch, B. (1989). *Languages of thought*. New York: The College Entrance Examination Board.
- Holt, T. (1990). *Thinking historically: Narrative and imagination in history*. New York: The College Entrance Examination Board.
- Howard, K. (1990, Spring). Making the writing portfolio real. *The Quarterly of the National Writing Project*, pp. 4-8, 27.
- James, W. (1915). *Talks to teachers*. Boston: Riverside Press.
- John-Steiner, V. (1987). *Notebooks of the mind*. Albuquerque: University of New Mexico.
- Katz, M. (1987). *Reconstructing American education*. Cambridge, MA: Harvard University Press.
- Kilpatrick, P. (1918). The project method. *Teachers College Record*, 19, 319-335.
- Lakotas, I. (1976). *Proofs and refutations: The logic of mathematical discovery*. New York: Cambridge University Press.
- Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, 27, 29-64.
- Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. New York: Cambridge University Press.
- Linn, R. (1990, June). *Cautions about performance assessments*. Paper presented at the Education Commission of the States, Boulder, CO.
- Lohman, D. (1989). Human intelligence: An introduction to advances in theory and research. *Review of Educational Research*, 59, 333-373.
- Madaus, G. (1990, July). *The history of testing in American schools*. Paper presented at the Institute on New Modes of Assessment, Cambridge, MA.
- McNeil, L. (1986). *The contradictions of control: School structure and school knowledge*. New York: Routledge & Kegan Paul.
- Messick, S. (1984). Abilities and knowledge in educational achievement testing: An assessment of dynamic cognitive structures. In B. S. Plake (Ed.), *Social and technical issues in testing: Implications for test construction and usage* (pp. 155-172). Hillsdale, NJ: Erlbaum.
- National Board for Professional Teaching Standards. (1990). *Request for proposals for cross-cutting research to develop instruments for assessing accomplished teaching*. Washington, DC: Author.
- National Council of Teachers of Mathematics, Commission for Standards on School Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- Nisbet, J. D. (1980). *The impact of research on policy and practice in education*. Aberdeen: Aberdeen University Press.
- Noland, T. (1985). *The effects of ability grouping: A meta-analysis of research findings*. Unpublished doctoral dissertation, University of Colorado, Boulder.
- Northern Examining Association. (1988). *Standards for English literature*. Sheffield, UK: Author.
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.
- Oakes, J., & Lipton, M. (1990). Tracking and ability grouping: A structural barrier to access and achievement. In J. Goodlad & P. Keating (Eds.), *Access to knowl-*

- Tyack, D. B. (1967). *The one best system: A history of American urban education*. Cambridge, MA: Harvard University Press.
- Tyler, R. (1949). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.
- Tyler, R., & Smith, E. (1942). *Appraising and recording student progress*. New York: Harper Bros.
- Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Walker, R. (1990, February 28). Governors set to adopt national education goals. *Education Week*, p. 16.
- Wiggins, G. (1989a). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.
- Wiggins, G. (1989b, October). *Questions and answers on authentic assessment*. Paper presented at Beyond the Bubble: Curriculum Assessment/Alignment Conference, Sacramento, CA.
- Wolf, D. (1988). *Reading reconsidered: Literature and literacy in high school*. New York: The College Board.
- Wolf, D. (1989). Portfolio assessment: Sampling student work. *Educational Leadership*, 46(7), 35-39.
- Wolf, D. (1990, Spring). Toward a theory of creativity. *Newsletter for the Society for Research in Child Development*, pp. 1-2.
- Wolf, D. (in press). *Presence of minds, performances of thought*. New York: College Entrance Examination Board.
- Wolf, D., & Pistone, N. (in press). *Taking full measure*. New York: The College Entrance Examination Board.
- Woodard, T. (1990). *Harlem renaissance*. Unpublished high school essay.
- Zessoules, R., & Gardner, H. (in press). Authentic assessment: Beyond the buzzword and into the classroom. In V. Perrone (Ed.), *Assessment in schools: Issues and possibilities*. Washington, DC: The Association for Supervision and Curriculum Development.

APPENDIX 1*

A 17-Year-Old's Description of Two Biology Experiences

An In-Class Biology Experiment

Once we were doing an experiment with a respirometer to measure how much oxygen was being used by germinating peas and dry peas. And we knew that the germinating peas should be using more oxygen because they are conducting cellular respiration, and the dry peas are dormant. So we spent about 45 minutes trying to get numbers that would show that. The respirometers were leaking, so we were having trouble getting the kinds of numbers we knew we should. So we just used someone else's data that was right.

An Independent Biology Project

Once I was in a state park and saw these people standing around watching a fire burn. Someone from the park told me it was a "controlled burning." I couldn't get it: "Let the forest burn?" So I began asking questions.

*From *Presence of Minds, Performances of Thought* by D. Wolf, in press, New York: College Entrance Examination Board. Adapted by permission.

See, originally forest fires were seen as bad. You can see how that happens: You look at all the black trunks and the black ground and think about all animals without a place to live. So everyone tried to protect the forests from fire—Smokey the Bear and rules against campfires and all that.

Then what people came to understand is just the opposite. When you suppress fires, it leads to the build-up of fuels. In a way it makes for something like a traffic jam in the forest. So when a fire does happen, it's giant.

If you look at trees in a healthy forest, you find fire scars. The burns allow for an old forest to give up to a new generation of forest. The old one is just a plateau. The fire changes the forest into a new environment. It allows for new species or more species of plants and animals to emerge and survive.

So in the 1960s, biologists and state park people began to argue that conditions should be restored to what they were before the white man came. They wanted to let fires burn as a part of the natural way things work. That was really tested in 1988 when there were giant fires in places like Yosemite. People were raging that the park service was letting the park burn. But it was really much more of a political fire than a real disaster. What you see is that it took people a hundred years to figure out that part of what they ought to be conserving was fire.

APPENDIX 2*

Developmental Descriptions from a Design and Technology Assessment

Statements of Attainment

Level 4 (Average 11-year-old)

Pupils should be able to:

1. Use hand tools and simple equipment appropriate to the materials to be worked, safely, and with a broad degree of accuracy (for example, use scissors to cut cardboard, a saw to cut wood, a computer keyboard to enter data);
2. Use under supervision making and assembling procedures appropriate to the range of tools and materials;
3. Use simple plans, drawings, and diagrams to assist making (for example, use their working sketches and diagrams to make a vehicle that moves); and
4. Suggest a possible solution to a problem that arises during making (for example, suggest different means of dealing with wheel spin in a vehicle they have designed).

Level 5/6 (Average 14-year-old)

Pupils should be able to:

1. Use a range of hand and powered tools and equipment (for example, plane, air-brush, cooker, database and spreadsheet software package, sewing machine) with due regard to function, safety, and the need to leave them in a fit condition for future use;
2. Prepare tools and equipment for use (for example, checking routines for powered tools and equipment, setting tension on a sewing machine, and setting for depth of cut);
3. Use these tools and equipment to a level of precision and finish appropriate to the working characteristics of the materials and the function of the artifact or system justifying any departure from the design brief (for example, setting hinges to a box lid, inserting a concealed zip);
4. Use an increased range of making and assembling procedures required by the range of tools and materials, including techniques such as preparing and marking out timber, preparing and cutting out a fabric to a pattern, and assembling several components (for example, an electrical circuit, using a spreadsheet and database together in an IT system);

*From *National Curriculum: Task Group on Assessment and Testing* by the Department of Education and Science and the Welsh Office.

5. Interpret plans, drawings, and diagrams (for example, a switching circuit, a working drawing, a dress pattern) in order to achieve the outcome described; and
6. Alter planned procedures to cope with unforeseen problems arising from the materials or tools being used (for example, power failure, component breakage).

APPENDIX 3*
Northern Examining Board Draft Criteria To Be Used as Grade Indicators in the Process of Assessing English Literature

3.18 In describing the process of reading and engaging with literature, we have tried to emphasize the degree to which it is a creative task (i.e., the reader's individual construction of the text). Accordingly, any attempt to describe performance and specific levels must take account of the range of factors that may influence that performance (e.g., the text, the task set, the stage in the process of reading the text reached by the candidate, supplementary information offered to the candidate, and the candidate's previous reading experiences). In the descriptions that follow, we attempt to define the qualities that we think are characteristic of the work of candidates who will achieve grades F and A. We do not wish to suggest that all these qualities will be present in the work of an individual candidate. Equally, we recognize that there are many other qualities that may be present but that we have not described.

The Work of Candidates Achieving Grade F

3.19 F grade candidates show glimpses of autonomy as readers, in relation to chosen appropriate texts. They identify the main contents of texts, following narrative threads, seeing the main issues and themes. They describe characters, their motives, and their relationships, and recognize similarities and differences between them. They begin to show awareness of particular effects and uses of languages. In reading and rereading texts they can give personal reactions and select elements they like or dislike and relate the issues and characters in texts to their own experience. They empathize with characters who are facing choices and problems and for whom they have sympathy. During the process of reading or rereading they modify their views in the light of new evidence that bears on previous meanings and judgments. They make use of personal knowledge and experience to inform their reading. In their work, written or oral, these candidates convey their ideas and feelings directly and use evidence to support their views showing some awareness of the nature of the task specified.

The Work of Candidates Achieving Grade A

3.20 A grade candidates convey a sense of autonomy, as readers and as writers, about their reading. Whether work is speculative, creative, or expository, they have a confident grasp of the text as a whole, tracing developments in the plot, appreciating crucial moments, and connections between various strands of the text, particularly between central and peripheral issues. These candidates recognize that the world created and conveyed by the text is shaped by the ways it is constructed, and so they comment on such matters as characterization, plot, theme and intention, layers of meaning, effects, and the functions of these in the text. They are aware of the viability of alternative interpretations of particular language use and of overall structures in texts. In responding to particular tasks A grade candidates draw on a wide reading experience, both explicitly and implicitly using their views of what they have read.

3.21 For these candidates, feeling "at home" with a text suggests that they can give personal reactions to what they read, selecting elements that they like or dislike, and that they can reflect upon these reactions. They recognize clues to meanings and speculate on their significance during the process of reading and rereading; they tol-

erate ambiguities, exploring possible alternative meanings; and when considering alternative interpretations of texts, they construct viewpoints and versions of events for a range of characters and situations, whether these are sympathetic or antipathetic to the reader.

3.22 In their work, both written and oral, these candidates convey a sense of coherence and in work of various kinds they are forceful and convincing in using evidence and in conveying their ideas and feelings.

Intermediate Grades

3.23 In the grade indicators for F and A there are elements that are clearly connected and that indicate areas of progress and development for candidates. The incremental progression that we envisage will be an individual process, so that there will be unevenness of achievement across these elements. Candidates achieving the same grade between F and A are likely to present quite different profiles of performance, and the grade awarded to them must be a recognition of what each has done. When we analyzed candidates' work within this range, we found that elements within the work were similar, and so the profiles of individual candidates seemed inevitably to be expressed in terms of quantity—"more" or "less," "often" or "on occasion." Such quantities cannot be readily defined, and we did not feel able at this point to develop distinctive statements of the performance of candidates achieving particular grades between F and A. It may be that the analysis of individual candidates' work could in time lead to agreement as to how some such candidates' grades can be indicated, but our work did not reveal such methods.

3.24 We noted, however, that between grades F and A candidates became increasingly able to use narrative of plot, comments about character, and explanations of meaning to support an argument, rather than to act as a substitute for one. Their ability to discriminate between key events, issues, and features of a text can also be seen to improve.

3.25 With more reading experience, they develop familiarity with their chosen material, or with genres, and with the ways in which the literature works, and are able to make comparisons between texts or references to other relevant material that they have read.

3.26 They develop the confidence and willingness to persist with the interpretation of literature, and the explanation of what it means to them, moving from glossing the surface (of a story or poem) to more probing exploration of the text, and making better use of evidence from it to illustrate their comments.

3.27 There is an increasing sense of themselves as readers, and writers about their reading, and of their ability to convey their knowledge, understanding, and enjoyment of literature.

APPENDIX 4*
**SEP English 10, Drama Unit Examination, Essay Section:
A Midsummer Night's Dream**

General directions: Read the assignment carefully. Then take a few minutes to plan your essay. Write your essay on pages 3 and 4.

From your study of this drama unit you have developed an awareness of the choices a director has when he or she is staging a play. As you viewed three different interpretations of the Pyramis and Thisbe scene from *A Midsummer Night's Dream*, you watched carefully for the various choices the director made and took notes that you later transferred to a graphic organizer. In a discussion that followed, you shared per-

*From "Midsummer Night's Dream": Tenth Grade Drama Unit of the Syllabus Examination Project by the Pittsburgh Public Schools, 1990. Adapted by permission.

*From *Standards for English Literature* by Northern Examining Association, 1988. Sheffield, U.K. Updated materials are available from the Joint Matriculation Board, Manchester, U.K., M15 6EU.

ceptions and information with other students. Use this knowledge and the information on your graphic organizer to write a composition on the following topic:

Choose the interpretation of the Pyramids and Thisby play in which you feel the director has made the most effective staging choices. In a multiparagraph essay, explain, using details from your graphic organizer, why the interpretation you chose is effective.

Remember to include the following elements in your composition:

- Develop a *focus* for the essay.
- Use *details* from the *selected scene* to develop your focus.
- Provide an *appropriate conclusion*.

Chapter 3

To Capture the Ineffable: New Forms of Assessment in Higher Education

PETER T. EWELL

National Center for Higher Education Management Systems

Assessment in higher education has become, by any estimate, a curious phenomenon. Begun modestly as a series of isolated institutional experiments in the late 1970s, assessment has become, for many institutions, a condition of doing business. A concomitant result has been the conversion of a scholarly measurement technology into a major and controversial tool of public policy. Thus, on one hand, assessment in higher education has the characteristics of a social movement—driven by political forces and constrained by societal demands (Ewell, 1989). On the other hand, through its language, symbolism, and technology, it is equally rooted in measurement practice. Any attempt to come to terms with it, therefore, must equally partake of both.

In this review, I attempt to do so in the following manner. First, in order to fully understand current assessment practice, it is critical to establish its historical and political context. Regardless of its technical guise, assessment in higher education cannot be divorced from a visible reform agenda. I begin by laying out this background with particular emphasis on the evolution and origins of assessment's contradictory imperatives: academic improvement and external accountability. Interaction between these two imperatives has produced a bewilderingly complex pattern of "assessment" activities on college and university campuses. Indeed, the term itself has rapidly acquired multiple meanings, each of which carries with it a particular referent and history. In the second portion of the review's initial section, therefore, I attempt to sort out some of these complexities in terms of a commonly applied taxonomy of assessment approaches.

This preliminary discussion sets the stage for the second section: a critical review of current practice. Like many other areas of applied research, assessment in higher education has rapidly moved from a sophisticated measurement technology emphasizing prediction and precision to-