

# 3

## Rethinking Portfolios for Evaluating Writing *Issues of Assessment and Power*

Brian Huot  
Michael M. Williamson

### Introduction

ISSUES IN WRITING ASSESSMENT HAVE TRADITIONALLY REVOLVED AROUND our ability to construct procedures that represent the ways students write and at the same adhere to the guidelines set down by theories of educational measurement. Moss asserts that this tension between theoretical constraints of literacy education and assessment has been productive in promoting the many new and improved methods for assessing student writing (see Camp 1993a for a discussion of the relationship between the teaching and testing communities in creating writing assessment procedures). Moss also warns, however, that “Proposed solutions often reflect compromises between competing criteria rather than the fundamental rethinking that might push both fields forward” (Moss 1994b, 110). We concur with Moss’s admonition about relying solely upon compromises between teaching and testing. While these compromises have been a necessary part of the development of writing assessment, they are also responsible for much of the dissatisfaction educators feel about the continuing importance of interrater reliability and test-type conditions which constrain our ability to develop assessment practices sensitive to the ways people read and write.

To meet Moss's challenge to "rethink" solutions that are more than compromises, we focus in this chapter on portfolios because they are, perhaps, the most popular form of writing assessment ever.<sup>1</sup> As well, portfolios and other forms of performance assessment provide the most rigorous challenges to traditional notions of educational assessment (Moss 1992). Our "rethinking" demands broadening the discussion beyond a consideration of just assessment and pedagogy to include important but often forgotten issues of power. Moss's tension between competing criteria is framed in theoretical terms. We contend that oftentimes issues of power rather than theory drive important assessment decisions. While Moss cites tension between the two disciplines of literacy education and educational measurement, we believe that power is a third, important determinant in crucial decisions about how students will be tested and what impact this testing will have on student learning. To control testing is to control education, to control what will be valued and taught within the schools. Crucial decisions concerning assessment are often made by regulatory agencies and political and educational policymakers based on practical and political concerns of cost, efficiency, and public opinion.

This chapter discusses the relationship between assessment procedures and the underlying power structures which dictate and profit from their use. Examining the various theoretical and political pressures which influence what measurements are chosen and how they are implemented allows us to conceive of assessment procedures as instruments of power and control, revealing so-called theoretical concerns as practical and political. We challenge the notion that concepts like validity and reliability are unquestionable and theoretically necessary. In other words, the need to standardize assessment procedures to achieve reliability, validity, or some common standard can also be seen as a move to impose particular standards on large numbers of teachers and students. Our reconception of the tension Moss describes focuses on who will control assessment and curriculum.

We fear that unless we make explicit the importance of power relationships in assessment, portfolios will fail to live up to their promise to create important connections between teaching, learning, and assessing.

### Issues of Assessment

Newer approaches to writing assessment, such as writing portfolios, continue to be subjected to the routine scrutiny of the various theoretical approaches and political pressures all procedures undergo in the fight for

control over writing assessment in American schools and colleges (Messick 1989; Moss 1992). No matter what form assessment takes, tradition and accountability dictate a need for standardization. "Standardization refers to the extent to which tasks, working conditions, and scoring criteria are the same for all students" (Moss 1994b, 110). Primarily, standardization is used to compare different educational programs or institutions in terms of their relative effectiveness in student achievement (Moss 1994a).

In writing assessment the need for standardization has been central to its development. The scoring of essays was so unreliable (inconsistent) that writing ability was commonly tested indirectly through the use of multiple choice tests of usage and mechanics.<sup>2</sup> Although the debate between the implementation of direct and indirect measures of assessing writing was often cast in terms of the tension between the teaching and testing communities (White 1993), in fact this debate was always within the field of measurement since it involved the achievement of the psychometric concept of reliability. In direct writing assessment, consistency in scoring is achieved through a set of procedures developed explicitly to ensure agreement of independent raters on the same papers. These procedures which ensure rater consistency in scoring include having students write to common topics in a controlled environment. Readers are trained to agree with one another on scoring guidelines they may or may not have any control over. An acceptable rate of reliability in scoring is crucial because traditionally testing theory dictates it.

Moss (Moss 1994a) challenges the traditional notion that assessment has to be reliable in order to be valid. For Moss, the very concept of reliability as a consistent interchangeable series of judgments on discrete skills or test items privileges standardization, thus limiting the power of local, contextual, performative, and holistic forms of measurement and the curriculum they inform and justify. Moss advocates local, contextual reading of portfolios or other assessment instruments. She offers the example of the procedures commonly used to decide upon the best candidate in a job search, where a committee of colleagues convene and discuss their understanding of each candidate's qualifications based on a full dossier of material. Moss suggests that this discursive, communal, interpretive search for value and meaning makes more sense for performance measures like portfolios. She acknowledges the inability of the psychometric theory of traditional testing to support such procedures but advocates instead the theoretical umbrella of hermeneutics in which the shared search for knowledge and judgment are often considered appropriate. Moss calls for a shift from one conceptual

framework to another in order to create practices that are more firmly based on theoretical grounds which support the activity of reading and responding to literate activities. Delandshere and Petrosky invoke a similar switch from psychometrics to poststructuralism in the creation of assessment procedures for teacher performance and certification. Both Moss (Moss 1994a) and Delandshere and Petrosky contend that psychometric theory stipulates a limiting and inaccurate framework for interpretative and judgmental decision-making about complex human behavior.

In current psychometric theories of testing, individual achievement is decontextual and standardized, so that testers can draw generalized inferences about individual performances and compare particular students and groups based upon performance on a particular test. These types of comparisons delete the context of individual learning environments and student populations and assume that the ability to write is a universal, identifiable human trait that can be measured accurately and consistently. The emphasis is on the technical rigor of testing procedures and statistical operations and explanations rather than the complexity of student performance and judgments about that performance. The goal of large group and/or standardized assessment procedures is typically to assess substantial numbers of students and to provide a single numerical index that can be used to compare different groups of students within and among particular settings, assuming that the assigned numbers depict an adequate picture of student achievement and teacher effectiveness across various social, cultural, historical, and geographical contexts.

The losers in the high stakes assessment<sup>3</sup> game made possible by psychometrics are the students and teacher. (See Moss 1994b, "Validity in High Stakes" for a review of the literature on the deleterious effects of large-scale, high stakes testing on students' ability to learn.) Moss notes that large group, standardized assessment procedures present an inherent validity problem (Moss 1994a). Current theories of validity privilege the concept of construct validity in which a test must contain an adequate representation of the ability to be tested and the influence of this test on the teaching and learning of those who take it (Cronbach 1989; Messick 1989). Large-scale, high stakes testing requires standardization and tends to reduce the curriculum to what can be measured. At best, test scores obtained under these conditions are a very poor indicator of the range of learning fostered by a school curriculum. The value of these scores is often affected by the number of students tested and the diversity inherent in such large

populations of students. Furthermore, when tests are used for comparisons among students, the procedures have to be standardized. Moss's critique of standardized assessment procedures is that they sacrifice validity for the objectivity of reliability, often resulting in a trivialization of the goals of assessment itself (Moss 1994a). Wiggins contends that this focus on standardization is really a confounding of standards with standardization:

Standards are never the result of imposed standardization . . . Standards, like good assessment, are contextual. The standards at Harvard have little to do with standards at St. John's College or Julliard; the standards at all our best independent schools and colleges are determined by each faculty, not by policy-maker mandate. (Wiggins 1993a, 282)

Although we recognize the inevitability of assessment driving delivery of curricular goals, we do not see assessment as an inherent evil. If assessment procedures are developed from specific curricular goals, then the assessment will tend to influence teachers and students toward mastering those goals. If, however, the assessment is based upon only those goals that are easily measured, then curriculum will be limited to its assessment procedures (Berlak 1992; Moss 1994b). The crucial element in all these "ifs" and in the ability of assessment to be a positive influence on teaching and learning revolves around the degree of power local stakeholders like principals, teachers, parents, and students have over the many aspects of an evaluation program. Many assessment programs, including those associated with reform movements which advocate site-based decision-making (see Callahan, this volume, for a good review of portfolios and educational reform), mandate certain assessment procedures or euphemistically titled "conceptual frameworks" school districts, principals, and teachers are obliged to implement (Murphy, this volume).

The particular form of assessment creates much of what is considered relevant, valuable, and worthwhile by teachers, students, and parents; assessment is never separate from curriculum. Whether curriculum can drive assessment or whether assessment always drives curriculum is a matter for debate (also an issue upon which we, the authors of this chapter, do not agree). Murphy's recent review of various portfolio programs illustrates that there can be an interactive relationship between assessment and curriculum in which they exist as a dialectic, limiting, affecting, and informing each other (Murphy 1994b). Traditionally, high stakes writing

assessment has been handed down, reducing the amount of interaction and creating a situation where, indeed, assessment not only drives curriculum, it “subsumes” it (Elbow and Yancey 1994).

Much has been made about the diverse and individual nature of portfolios to best represent literate behavior in a school setting (Belanoff 1994; Berlin 1994; Graves and Sunstein 1992; and others). However, the move to standardize portfolios is an important aspect of the tradition in educational measurement since assessment instruments have always been standardized in some sense or another. This sets up a conflict, relative to Moss’s notion of competing criteria of two disciplines. In fact, the deck is slightly stacked on the side of standardization, for as Moss points out, “we are considerably less knowledgeable about how to design and evaluate nonstandardized assessments and about how to incorporate them into our ongoing assessment practices” (Moss 1994b, 124). What do we do with portfolios as assessment instruments is a legitimate and perplexing question. The problems occur, we believe, when we succumb to the knee-jerk answer “standardize them!” Moss and others would have us look beyond psychometrics to hermeneutics or poststructuralism for theoretical answers to address the tension between the disciplines involved with literacy education and those who assess that education (Moss 1994a). Nonetheless, we think it necessary to also consider issues of power which often appear to exist outside or be invisible within this tension. In fact, issues of control and political expediency ultimately often supply much of the pressure to standardize portfolios and other performance assessments.

## Power

If recent history in writing assessment has taught us anything, it has demonstrated that decisions about assessment ultimately involve decisions about where to locate power in educational and political institutions. For instance, the aspects of a writing curriculum that are chosen for evaluation through an assessment program and the procedures of the assessment itself control students’ learning and teachers’ instruction. The simple truth of educational assessment is that what we choose to evaluate in our students’ performances will determine what they attend to in their approach to learning. For example, Resnick and Resnick point to the need to evaluate students’ abilities to do independent and self-chosen tasks because they contend that what is not assessed often disappears from the curriculum

(Resnick and Resnick 1992). Those aspects of the curriculum for which we are held accountable will determine what we emphasize in our teaching. Furthermore, our approach to assessment can lead to some unexpected learning on the part of our students when we design an assessment that inadvertently cues them to attend to some aspect of our classroom that we had not intended.

The effects of testing are pervasive and at times surprising. In some instances, poor test results are better than strong ones because this might mean more funding to shore up the valiant but failing efforts of the schools who are seen to be struggling against the inherent problems that certain members of the community bring with them to school. In other instances, notably strong achievement test results can increase the value of property in a specific school district, information which is routinely used by realtors to sell homes to prospective buyers. Test scores can give a school or district the right to claim that it is winning the fight against educational sloth. Clearly, test results can carry with them strong and persuasive outcomes beyond the intended function of the tests themselves.

Another powerful influence of testing on our schools is that assessment often functions as a form of surveillance<sup>4</sup> (Berlak 1992): a way for administrators or other powerful stakeholders to assume and wield their power and influence. Testing in the public schools, for example, allows principals to check up on teachers, who are in turn watched by superintendents and school boards, who are checked up on by state agencies, who are ultimately responsible to the federal government.<sup>5</sup> Linn, in examining the influence of performance assessment instruments on testing practices notes that in the mid-90s we have entered an era of increased testing. Unlike past initiatives, however, “the role of the federal government is much greater than with previous test-based accountability and reform efforts” (Linn 1994, 4). This increased role of the federal government in assessment can also be seen at the postsecondary level in the form of the proposed National Assessment of College Student Learning (NASCL)<sup>6</sup> which will give the federal government more influence over higher education.

Kentucky, which is in the midst of massive and ambitious school reform, provides a good example of the many issues surrounding power, assessment, and portfolios as it moves toward a new statewide curriculum that calls for activity-based instruction and interactive classroom environments. In the Kentucky system, students attend ungraded primary classrooms their first three years in school and are given increased instruction and exposure to computers, and much of the curriculum centers on problem solving

and group projects. Also, individual schools have some say over the actual form and rate of change. However, another aspect of the reform is that all fourth, eighth, and twelfth graders are to submit learning portfolios in math and composition to be graded according to the same rubric and anchors generated by the state department of education.

Although there have been efforts by the state to involve teachers in the construction of the assessment program, the program itself has been mandated by the state, and the scores of the portfolios are used to make high stakes decisions. In an ethnographic study of one high school in Kentucky during the second year of the state's mandated assessment program, Callahan (this volume) observes that the use of portfolios increased both the amount of writing students do and the attention teachers give writing in the classroom. "However, since 'portfolio' and 'test' have become synonymous it [will be] difficult for Kentucky teachers to use portfolios for any other purpose . . . [because] they perceive the creation of a portfolio as a stressful activity performed only in response to an external set of demands."

Even though we may use portfolios to assess student writing performance, standardizing their contents and scoring works to locate the power centrally in the hands of the very few who control other sorts of power and decision-making. For example, in the case of portfolio assessment in Vermont, the low interrater reliability coefficients have been enough to raise the call for increased standardizing of the contents of portfolios, even though portfolios are already being viewed as having many positive, though immeasurable, effects on teaching and learning (Koretz et al. 1993). This move to standardize portfolios is based on traditional notions of reliability which claim it "a necessary but insufficient condition for validity" (Cherry and Meyer 1992; and others). In other words, if a measurement system doesn't produce consistent judgments among independent raters, then it cannot be valid. Within the measurement community, however, there is no consensus about the absolute necessity for interchangeable judgments from independent raters. New, emerging theories of assessment point to the problems with rigid and simple conceptions of reliability for measures which include sophisticated judgments about complex activity like that exhibited in a portfolio of student writing. A whole range of assessment specialists are in the process of developing alternative forms of assessment which conceive of reliability as a "critical standard" or "confirmation" (Berlak 1992; Guba and Lincoln 1989; Johnston 1989; Moss 1992, 1994b; and others). At the very least, current conceptions of validity require a consideration of the importance of a test's consequences (Cronbach 1989; Messick 1989).



However, these appeals to less rigid notions of reliability and the positive consequences of portfolio assessment in Vermont are not part of the decision to further standardize writing portfolios to achieve higher interrater reliability coefficients (Koretz 1994). In other words, decisions about portfolios in Vermont are not being based upon the theoretical developments which inform performative assessment procedures like portfolios. If the decision to standardize portfolios in Vermont is being based upon theory, we need to ask whose theory is being used and why?

It is not difficult to see where the power for assessment is located when portfolios or any other measurement instrument is mandated and standardized by a state department of education. The fact that students are compiling portfolios or writing in their classes with their teachers' and classmates' help is secondary. The ultimate authority in these situations has nothing to do with the activity in the classroom which produces the portfolios themselves. Instead, they are being used to generate scores which can support the reform movement. Like all such massive changes, the ones in Kentucky and Vermont require a huge investment from its citizens and politicians, and all of them want some proof that the effort is worth it. While all of this is understandable, we have no assurance that portfolios can encourage a learning environment in which the teachers and students have no say in how they are used, compiled, and scored. In these instances, it appears that the use of portfolios in high stakes assessment scenarios are predicated on political rather than educational rationale. While it is hoped that the wide-scale use of portfolios like that in Kentucky and Vermont can improve student writing ability, surely we increase the chances of this happening when we base decision-making upon educational rather than political premises.

This interweaving and confounding of politics and education is an ongoing dilemma in American schools. Part of the problem stems from the fact that in a very real sense schools are "agents of government to be administered by hierarchical decision-making and controls" (Darling-Hammond 1989, 63). This mixture of political policy and educational theory often creates an odd and ineffective marriage. For example, Berlak talks of how the educational policies of the Reagan and Bush era were contradictory and incoherent because on the one hand they called for increased local control while at the same time they advocated increased use of standardized assessment for increased accountability. According to Berlak, schools cannot attain autonomy when there is an emphasis on standardized assessment which takes the power for curriculum, accountability, and finances away

from localities and invests it in centrally located sites controlled by those without knowledge or investment in local contexts.

Alternatives to locating power centrally already exist. In the job search scenario we referred to earlier, Moss offers an example of the way hiring decisions are made at the college level. In her example, the power for judgment rests within the committee itself and the local community from which it is constituted and to which it is responsible. This type of arrangement is considered appropriate for making important decisions about hiring university personnel, and as Wiggins argues, similar localized procedures are used in private and independent institutions to make decisions about students. In discounting traditional notions of reliability as interchangeable consistency, Moss calls for a critical standard by which student performance can be assessed on a local level which honors the importance of contextual and community values necessary for students and teachers to perform at their best within a specific environment (Moss 1994a). Moss's position is similar to Wiggins's, who maintains, "Standards are not fixed or generic. They vary with a performer's aspirations and purpose . . . It is true we use the word standard as if there were a single excellence. But that hides the fact that different criteria and contexts lead to different single excellences" (Wiggins 1993a, 283-284). Citing Sizer, Wiggins maintains that the correct question is not "Which Standards?" but "Whose Standards?" (Wiggins 1993a, 283), similar in effect to our question about whose theory.

As we see it, ultimately, decisions and discussions about standardization or reliability are political since they are about where to locate the power in an assessment program. Traditionally we have disguised the political character of such issues by referring to the sanctity of technical terms like reliability or validity even though there is little consensus in the measurement community not only about what such terms mean but about their value as meaningful representations. In fact, there have been several calls for dismantling the very notion of validity itself (Berlak 1992; Guba and Lincoln 1989; Johnston 1989).

One way to approach the dilemma we have raised about rethinking the tension between the assessment and educational communities is to "rethink" the notion of accountability. Most initiatives to assess student ability and educational programs are based upon the need for administrators and teachers to be accountable for their programs, practices, and the performances of their students. While we wholeheartedly endorse the importance of education striving for, achieving, and documenting excellence,

we wonder how teachers and site-based administrators can be accountable to individuals and organizations who have little understanding of local problems and conditions. The problem, as we see it, is that the concept of accountability often assumes unequal power relations in an inverse relationship to the knowledge and understanding of the salient difficulties in providing a quality education. In other words, the least knowledgeable people often make the most important decisions, many times based upon assessment schemes that are so pared down by standardization that they produce information that has little meaning and importance for local contexts. Programs like those in Kentucky which advocate site-based councils recognize this inherent flaw in the power relationships of accountability. However, as we have already demonstrated, to control curriculum and other important factors in education, you must also control the assessment instruments.

Our “rethinking” of accountability is to replace it with the concept of responsibility. At first glance, there appears little difference between being accountable and being responsible. Like accountability, responsibility also involves providing evidence that local teaching and administrative decisions are based upon the ability of schools to provide quality educational experiences for their students. The difference lies in the relationship of power. Being responsible does not assume that local authorities have to account to higher authorities. The use of assessment for surveillance and other hierarchical functions diminishes as local assessment instruments focus on local programs and actually assist teachers and administrators in being responsible for the spending of public money, the design of educational program, and the education of its students. Changing the power relationships opens up a much more productive set of possibilities for assessment practices.<sup>7</sup>

### In Conclusion: Considering Portfolios

As portfolios are continually defined in terms of both their pedagogical value and measurement properties, it is important to remember that an assessment technique itself is not always of primary importance. Although we have some good examples of how portfolios can function in the classroom (see for example Belanoff and Dickson 1991; Paulson, Paulson and Meyer 1991; Yancey 1992a, 1992b), how portfolios are defined by the assessment procedures and how they are used and received by educational regulatory agencies, administrators, teachers, students, and parents will

determine their ultimate role in enabling or disabling teaching and learning in writing classrooms.

Although we have no commonly agreed upon definition of portfolios, certain characteristics seem constant. Portfolios contain not only a collection of student work but also the process of how the writing got to be included in the portfolio. Ideally, students learn to make decisions about their writing in terms of what to include and how to improve what they choose to work on. Portfolios can also contain the reflective work students do as they prepare a body of their work not only to be evaluated but to represent them as writers. In this sense each portfolio can be an individual record of a student's journey to understand herself as a writer. Efforts to standardize such a record cut into its ability to help the individual student make sense of herself as a literate person struggling not only to make meaning but to create a context within which she learns to read and write.

As Moss notes, there is an obvious tension between standardized assessment and the highly contextualized, individual nature of communication (Moss 1994b). The power struggle over portfolios is a result of this tension. Any form of assessment which is so individualized as to let students choose their own tasks will be extremely difficult to standardize, unless their individual and self-directed nature is controlled by outside criteria. To do this is to risk reducing portfolios to a specific number of papers on specified topics to enable scoring reliability and standardization that would permit comparisons among different schools. Furthermore, as we have demonstrated, this tension results from the pressure to locate power in a central regulatory agency such as the state education department rather than in the schools and school districts themselves. To preserve the integrity of portfolios and to harness their ability to truly alter the power relationships in assessment, it is necessary to maintain their localized character and to resist any attempts to centrally evaluate them. "Compromises" like statewide scoring guidelines and training sessions are merely disguises to enable standardization.

Many of the initial arguments for portfolio assessment were made in opposition to the standardization required for the reliable scoring of essays. Portfolios are an important juncture in the struggle between educational assessment and political forces. They represent a crossroads, of sorts, at which we need to decide if we will continue along current and traditional lines and standardize their use, so that regulatory agencies can maintain their grip on educational practices. It is important to recognize that this decision is not just about theoretical soundness but about political pressures. We can choose to serve political expedience and create portfolio systems that

produce numerical indices and allow for comparability. Or, we can resist such pressures, citing the importance of local control and the power of context in the creation of effective communication.<sup>8</sup> Our position in calling for a reassessment of the way power is located in assessment, especially in the use of writing portfolios, can be viewed, perhaps, as somewhat utopian, unrealistic, or unobtainable. However, there are ways to use portfolios and other assessments which allow them to retain their local character and allow for the kind of assessment which provides rich feedback to inform and enrich teaching and learning. These are already emerging (see Berlak 1992; Johnston 1989; Moss 1994b; and Murphy 1994b for a discussion of such methods). For example, instead of having portfolios compiled by students at various levels and having them read and scored according to mandated guidelines, portfolios could be read by a local board comprised of the teachers themselves, parents, school administrators, and students, who would decide what criteria most relates to their students and school. These portfolios would be discussed and the criteria could change from year to year as student populations and local concerns evolved.<sup>9</sup> Instead of complicated numerical scores, we might think of judging portfolios on the basis of whether a student is on track, ahead of the game, or needs additional help. These numbers could be used to report student progress to the school district or department of education. A central board composed from local constituents would look at a small number of student portfolios either randomly or at particular segments of the school's population, depending upon the purpose. It might be possible, because of the much smaller numbers, to look at portfolios from several grades each year. In terms of the positive effect of assessment on curriculum, this scheme dictates that students compile portfolios every year, and that they are locally read with the potential of being sampled beyond the school. Portfolios have the potential to be more than just what "you do" in certain grades for assessment. Instead, they have the ability to assume a positive role in influencing the curriculum and culture of the school.

Such examples do not, by themselves, provide the necessary reconceptualization we are suggesting; they do, however, acknowledge the critical importance of schools retaining power over their ability to assess and teach. Of course, there are no easy answers to this struggle between locating power for assessment within or outside the schools. Compromises in this struggle have traditionally been resolved in favor of standardization and central authorities, often in the guise of being theoretically sound. It is important that we begin to devise new schemes for assessment which recognize the power

relationships within our decisions for assessment and acknowledge the importance of context. It is also vital that individual teachers recognize the power struggles they and their students find themselves in as they attempt to use assessment instruments like portfolios to teach their students.

## Notes

1. We base our contention about the popularity of portfolios on the impressive number of volumes (more in the last five years than on all of writing assessment in the last two decades) and the four national conferences held between 1992 and 1994.
2. By the way, these indirect tests are still quite common. In a recent survey on placement practices of colleges and universities, half of the respondents report using indirect measures to place students (Huot 1994).
3. By "high stakes" we borrow a definition from Moss, to include any assessment used for "informing consequential decisions about individuals and programs" (Moss 1994b, 110).
4. There is a long standing concern for government agencies and policies assuming "big brother" roles. See Foucault for an historical review and critical discussion.
5. Although most testing for regulation takes place in the public schools, there is increasing pressure to extend this type of assessment to postsecondary institutions as part of the emerging National Assessment of College Student Learning (NACSL). For a review of the NACSL and its relationship to writing assessment, see Witte and Flach 1994.
6. See Witte and Flach, 1994 for a discussion of the NASCL and its influence on the assessment of writing at the postsecondary level.
7. We are indebted to Patricia F. Carini for discussing with us the differences between accountability and responsibility and their importance in education and educational assessment.
8. The importance of context in language use is arguably the most significant development to come out of the great changes in linguistics, rhetoric, and education during the last three decades. See Witte and Flach, 1994 for a review of the literature on context in communication and its importance to the construction of adequate measures of literacy.
9. Murphy (Murphy 1994b) describes such procedures already in use in her review of school districts and portfolios across the country.