

## Classroom Writing Portfolios: Old, New, Borrowed, Blue

Robert C. Calfee  
*Stanford University*

Sarah Warshauer Freedman  
*University of California, Berkeley*

Catchy labels are attention-grabbers. Whoever first expressed the idea of portfolios of student work might have conjured up images of capable professionals opening impressive binders filled with polished displays of art, a far more appealing vision than dreary worksheets and multiple-choice tests.<sup>1</sup> Writing instruction seems especially suited to portfolio assessment and programs have sprung up around the country based on the metaphor. Like most metaphors, this one must be handled with care. Students are not professionals, and placing assignments into a manila folder does not guarantee a basis for assessment.

The chapters in this volume explore the current status of the portfolio concept—theory, research, and practice. The focus throughout the volume is on the tension between classroom assessment and externally mandated testing. This chapter places the writing portfolio in historical context (old), examines the mesh between portfolios and recent developments in cognition and learning (new), looks at the linkage between portfolios and other innovative ideas (borrowed), and then reminds the reader of the especially tumultuous state of today's American public schools (blue). We conclude the chapter by proposing a conceptual framework for the role of portfolios

---

<sup>1</sup>Sweet (1976) is the first reference to "writing portfolios" we have discovered in ERIC. In his research, the experience portfolio is a one-page checklist where students can indicate their interest in various writing topics. Times have certainly changed.

in the assessment of student writing, our effort to provide a coherent linkage among the chapters in this volume.

### THE OLD

They say that WYTIWYG—"what you test is what you get." For more than 50 years, testing has come to mean the standardized, group-administered, multiple-choice test. Critics have argued that we have gotten low-level outcomes, rote memorization, and mindless practice. Originally designed for cheap and efficient selection of soldiers during the World Wars, the multiple-choice technology came to play an increasingly important role in public education from the 1950s onward. The concurrent emphasis on efficient management was well-served by standardized testing; it satisfied accountability requirements, allowed placement of students with special needs, and provided data to evaluate competing programs. Standardized tests meshed with the concept of behavioral outcomes, and textbook publishers began to align their materials with objectives-based scope-and-sequence charts. Alignment came into its own with the evolution of criterion-referenced tests explicitly designed to determine curriculum goals (Bloom, Hastings, & Madaus, 1971; a counterargument was made by Glaser, 1981). Textbook publishers incorporated worksheets and end-of-unit-tests into their materials, linking external tests to the daily routines of classroom instruction.

Writing instruction was not easily assessed by the multiple-choice technology, and so writing fell by the wayside. Literacy became identified with reading, more specifically with an image of reading as the acquisition of basic skills. As for writing, what should be taught? Many of the answers took a negative slant. Studies appeared to show that teaching grammar was not only boring but ineffective. How should writing be taught? Paraphrasing model paragraphs and copying summaries from book covers was not very inspiring. How should writing be assessed? The obvious answer was that teachers should grade student compositions. Although standardized reading tests could boast reliabilities in the .9 range (high levels of item consistency), research showed that teachers disagreed on writing performance, with reliabilities around .5 (low inter-rater consistency). Surveying the situation in America's schools during the 1970s, Applebee (1980) found that relatively little time was spent on writing instruction, most student compositions were a few sentences or perhaps a paragraph, and the first draft was generally the only draft.

### THE NEW

The founding of the Bay Area Writing Project in 1972 and the National Writing Project in 1974 began to dramatically alter this state of affairs. The Project brought together classroom teachers who were interested in the

teaching of writing to a summer institute where they could share their knowledge, practice their own writing, and talk about how to revitalize the field of composition instruction. From the outset, the project, drawing on the experience of successful teachers and the research of the time (Britton, Martin, McLeod, & Rosen, 1975; Calkins, 1986; Emig, 1971; Graves, 1983; Perl, 1979; Sommers, 1980) emphasized the concept of *writing as a process*, the notion that students need to be taught to think through their ideas as well as revise their writing, and that teachers needed to provide space and time for the thinking and reworking of ideas that meaningful writing requires. The Writing Project also pushed for the assessment of whole pieces of writing rather than piecemeal multiple-choice snippets, and led the way in popularizing holistic scoring for state, district, and school-level assessment programs (Myers, 1980).

Following the summer institutes, participants served as teacher-consultants in local school districts, passing on their knowledge to other teachers through formal projects and informal interactions. At Writing Project sites, the yearly cycle just described quickly created a network of teachers committed to the importance of writing, confident in their professional status, and convinced of the importance of teacher-based assessment. These teachers depended on neither textbooks nor tests, and were rather distrustful of external mandates.

The Writing Project is now active in more than 150 sites in every state and several foreign countries. During its growth over the past 20 years, the model has sustained the centrality of the classroom teacher, and the stimulation generated by professional exchanges. It has offered teachers a unified voice for speaking to the importance of writing in the literacy curriculum.

In the late 1970s, policymakers hurried to the front of the parade. Legislators throughout the nation mandated writing tests as part of state assessments. To be sure, these on-demand tests bore little resemblance to the practices being promulgated by Writing Project teachers. Process writing emphasized student-initiated topics, whereas mandated tests employed predetermined prompts. In process writing, student work was scaffolded by instruction, whereas mandated tests were standardized with no support allowed by the teacher. Process writing allowed students adequate time to plan, compose, and revise, whereas on-demand tests were restricted to a single session and a prescribed amount of time. Process writing was a social event, whereas state testing placed the individual on his or her own. In process writing, evaluation was a collaboration between student and teacher, whereas mandated tests were scored by external judges using predetermined rubrics. Nonetheless, the introduction of mandated writing tests meant that many more teachers began to pay attention to writing—especially teachers at the "test" grades (usually fourth and eighth grades, along with high school juniors or seniors).

The most recent episode in the history of writing assessment in this country emerged in the 1980s under a variety of labels—authentic assessment, performance-based testing, portfolios. The initiating goal in this movement was the urge to link assessment policies with what many teachers saw as a more authentic curriculum—in both reading and writing. What should be taught as reading? From the perspective of a 1960s objective-based curriculum, the answer was basic skill in decoding and answering questions. The Whole Language movement of the 1980s (Goodman, Goodman, & Hood, 1989) emphasized instead the reader's engagement in a story. Remembering the facts about *The Diary of Anne Frank* was part of the process, but reliving the experience was the more critical outcome. What should be taught as writing? An objectives-based answer emphasized the surface features of performance: grammar and spelling could be quickly and reliably judged from a fill-in-the-blank exercise. Writing Project teachers talked instead about purpose, audience, voice, development, and coherence.

The language of this movement had a novel ring to it, but it also incorporated classical elements. Reading instruction began to emphasize critical analysis. The Greek *kritikos* was a person chosen to judge merits and faults, to get to the root of the matter. Attention switched from decoding to comprehension. Speed and accuracy in oral reading are no guarantee of genuine understanding from this perspective. *Comprehend* comes from the same root as *prehensile*, with the sense of grasping, struggling, wrestling, "getting it" by rebuilding a passage. The focus turned to the student's reactions, reflections, and personalization of a piece of literature, a classical concept.

Writing instruction also employed elements of the classical rhetoric. Planning and development were important elements in writing, as important for the teacher as it was for the final product. Students once more had to explain what they were doing and why. Ideas like thematic development and persuasive argumentation returned to the curriculum. The ancient Greeks would recognize the concepts.

The emphasis on comprehension and rhetorical coherence was supported by an emerging line of innovative research on situated, social, and strategic cognition as the foundation for powerful learning. A mouthful, to be sure. The behavioral learning of the 1940s and 1950s was concrete and observable, attractive both scientifically and administratively. Functional relations between stimulus and response sufficed for this model, which provided powerful techniques for behavior control. The cognitive revolution of the 1960s and 1970s moved inside the mind, the computer as metaphor, to explore intellectual capabilities and thought processes. Information from the sensory channels entered short-term memory, where it was translated for storage in long-term memory. Problem solving, linguistic competence, semantic networks—these concepts reopened the study of the mind (Berliner & Calfee, in press).

By the 1980s, the strengths and limits of the computer metaphor had become apparent, and a new revolution emerged in our understanding of thought and learning. Human beings are not machines. We are social, we communicate, we are adaptive, and we can plan. Computers, in contrast, do not perform any of these activities unless a human being has programmed them to do so. They do not communicate unless they are connected by cables. They work the same no matter where they are plugged in. We now have a rich understanding of social cognition, of group thinking, and of the effect of context on these processes. But despite more than a decade of research studies spanning a variety of disciplines to explore these concepts, these ideas have yet to take root in today's classrooms.

What might it mean for curriculum, instruction, and assessment to be situated, social, and strategic? It certainly means more than simply exposing students to textbook content and then testing whether the content has been stored in memory. Situated learning happens when learning is connected with prior experiences and beliefs. If learning is not situated, it is less likely to have any genuine impact on students' perceptions and understandings. Most graduates know that the earth turns relative to the sun, causing sunrise and sunset; they have studied this topic in textbooks and tests from fourth grade on. Nonetheless, students see the sun set in the west in the evening, and cannot imagine standing on a globe that is actually rotating away from the sun. They study literary works that convey messages of enormous thematic import—*The Grapes of Wrath*, *To Kill a Mockingbird*, and *Freedom Road*—but appear to leave these studies affected little by the thematic values. They memorize textbook accounts about the hazards of poor diet, inadequate exercise, and drugs like tobacco and alcohol, but statistics and anecdotes suggest that these learnings are not thereby connected to daily reality for many people. School is one thing, the real world is another. It is easy to administer a multiple-choice test to assess content coverage, but judging the degree of situated learning is more difficult.

Social learning means working together toward genuinely shared ends. More is needed instructionally than occasional cooperative learning sessions. Learning in groups can be more effective and more satisfying than working alone. We sometimes must work as individuals, and too many cooks can indeed spoil a soup. But the democratic principle of *e pluribus unum* is more than an ideal; today's world has moved beyond the assembly-line era to a time when our lives as citizens and workers require teamwork. Yesterday's norms depend on examinations in which an individual's accomplishments are judged by how well he or she can work in isolation in competition with others. In fact, the student's ability to get along with thirty other individuals is critical if the classroom is to promote learning, and this capability is equally important in the world beyond school. Report cards sometimes address these issues by including a behavior category, but this is typically

used to identify troublemakers rather than peacemakers. There are no formal tests of getting along together.

Finally, cognitively strategic learning emphasizes the importance of transfer to new situations. The only constant for tomorrow is change, and today's schooling is worthwhile only as it prepares graduates for circumstances that cannot be predicted, that builds on content, but that goes on to explore broader meanings and deeper extensions. Strategic learning encompasses three distinctive principles: it is *active*, it is *reflective*, and it is *expansive*. Students can learn by rote practice, but they are unlikely to learn about learning unless they are invested in a purposeful endeavor. A youngster may take a test or even write a passable essay about the causes of the Civil War, then wipe the slate clean a day later. Being able to explain yourself is the essence of reflectiveness; "Why?" is largely neglected in classroom discourse. Authentic assessment asks the student to "show your work." Learning is expansive when it moves beyond the immediate context for application in new and unpredictable situations. In life outside of schools, the answers can seldom be found at the back of the book. For instance, how does analysis of the Civil War help understand the several other civil wars going on around the world today? The daily newspapers suggest that policymakers do not have clearcut answers about either causes or remedies in this matter, and so the question is a real one. Judging students' responses to the question is not easily relegated to a Scantron machine.

Only recently have the innovative concepts of cognitive learning begun to influence assessment, and even now only indirectly. Some of the groups hard at work on authentic assessment build on cognitive foundations (e.g., Resnick, 1987). At the level of classroom practice, however, teachers are predominately oriented toward activities more than concepts, toward demonstrating mastery of learned tasks more than transfer to new situations. Performance is more important than thinking, doing more than explaining.

### THE BORROWED

Central to all of the movements just mentioned—whole language, process writing, authentic assessment—has been the establishment of professional communities, the borrowing of ideas among teachers. The conceptual foundations sketched provide a starting point for a common language, in the sense that they seem to share a common ideology, partly grounded in Deweyian philosophy, partly based on cognitive psychology. Unfortunately, the various movements use different words, so that connections are not easily established.

The problem is that education is a practical enterprise, and as such requires a blending of ideas, engineering, and art. Scientists and scholars can

contribute to the ideas, but it is in the field of practice that the engineering and art must be worked out. In an earlier time, the technology was simple and did not require professional judgment. Most schools in the U.S. could follow a factory model, each laborer at his or her individual workbench, teachers isolated within their individual classrooms. No need for the continuing dialogue that is the hallmark of other professions. Education that guarantees a high level of intellectual and social development for all students cannot be prepackaged, but calls for professional decision making, which in turn requires a professional language as the foundation for interaction.

And so it is understandable that where programs like whole language, process writing, and authentic assessment are having substantial impact, teachers have managed to network with one another. Sometimes the linkages are school-wide, more often they take shape as mentor or "buddy" systems. The school seems the natural unit for establishing these connections, but cross-school alliances have a unique potential as seen in the National Writing Project and other professional networks.

Several observers have commented on the value of local ownership, the idea that teachers must develop their own understanding of the concepts and practices. This strategy carries the risk of reinventing wheels, but we think that it makes sense for several reasons. First, if today's teaching rested on a more clearly established professional foundation, then teachers would encounter preservice experiences in thinking through comprehending and adapting new ideas at a conceptual level during their college careers and induction into the vocation. In fact, teaching during the past 50 years has been "managed," in the sense that teachers are generally told what to do. Now, when suddenly expected to make significant decisions on their own initiative, teachers are understandably taken aback and reluctant. It is only through collegial opportunities that support this process from beginning to end that genuine professionalism can develop. These experiences require sustained support and feedback over a matter of years; it is not enough to attend a summer institute and then return to the isolation of the individual classroom.

Unfortunately, opportunities for collegial interaction remain rare in today's schools. Unlike Japanese teachers, for whom as much as half the day is spent in collegial activities, American teachers spend most of the working day with children. Faculty meetings, after-school workshops, and committee sessions provide scant time for serious discussion, and are in any event typically occupied with mundane matters disconnected from serious discussion of instructional practice. If meetings were really substantive, they could be quite worthwhile. Freedman (1994) in her comparative study of U.S. and British schools found that British teachers routinely participated in department meetings where they discussed and debated curriculum orientations and theories about student learning, and where they analyzed the needs of

particular students. Their most important exchanges came from opportunities to discuss details of practice against their specific contexts, to design and experiment together toward genuine instructional projects. Despite the barriers to collaboration, a repeated theme from innovative programs in the U.S. is the revitalization that comes from interaction with other teachers around matters of genuine importance to education (Lieberman, 1992; Darling-Hammond, 1994). The event may be as apparently mundane as a two-week summer session spent scoring portfolios; nonetheless, it is a chance to share ideas about student learning, and teachers value the experience. In fact, teachers involved in writing portfolio assessment, even when the assessment task is externally mandated and controlled, routinely praise the opportunity for collegial interaction around student accomplishments—the opportunity to borrow.

### THE BLUE

The blue reflects the sad reality that the innovations sketched here come at a time when support for education is ebbing throughout many parts of the U.S. The federal role has focused increasingly on national standards and tests. To be sure, the Office of Education has adopted a policy that supports school-wide programs under some conditions. This policy offers possibilities for teachers to interact, and even the promise of funds for professional development. But the encouragement for school-wide programs is encumbered in regulations, the funding is threatened with cutbacks, and administrators are understandably skeptical about a program that offers greater freedom and responsibility, but also offers the risk of less control. At the state level, policies and practices for school reform vary widely, and schools no longer top the priority list in many places; crime, prisons, the economy, health care, and welfare have displaced attention from educational issues, even though effective education offers the only long-term remedy to the former problems.

We are not saying that concerns about schools have declined; politicians and the media continue to decry declines in public education (a claim not supported by the evidence; cf. Berliner & Biddle, 1996; Calfee & Patrick, 1995), and the importance to our country's future of providing higher quality education (a claim that is clearly on the mark). The federal and state policy initiatives to provide leadership and establish control are forward-looking; challenging curriculum frameworks, teachers capable of centering instruction around student needs, and authentic assessment are all found in the rhetoric. Although standardized multiple-choice tests remain the primary technology for gauging achievement, several state and national programs have explored the feasibility of alternative methods, including performance tasks and port-

folios (e.g., Gentile, Martin-Rehrman, & Kennedy, 1995). The daunting challenge is to link these mandated efforts at reform to the realities of schools and classrooms scattered around the many ecological niches throughout our nation. We confront the dilemma of shaping a common vision in a situation where one size does not fit all. Moreover, as we place these words on paper, several states that had taken leadership in assessment innovations have precipitously terminated the programs and returned to conventional multiple-choice tests (e.g., Arizona, California, North Carolina).

### THE CLASSROOM WRITING PORTFOLIO

This volume explores the efforts of elementary and middle school teachers to employ portfolios as a way to promote student writing. It springs from a project of the National Center for the Study of Writing and Literacy (NCSW) that investigated current practices in portfolio assessment of writing in the elementary and middle grades. The goal of the project was to understand the ways in which portfolios serve the classroom teacher for instructional decisions. A secondary purpose was to examine the value of classroom portfolios for other accountability purposes, including grades, parent reports, and evaluation of school programs. The focus was on locally determined portfolio practices, and the project did not explore portfolios that resulted from state or federal mandates.

We proposed four tasks in the project: (a) development of a conceptual framework for analyzing student portfolios, (b) analysis of current classroom practice through surveys, interviews and focus groups, intensive case studies, and review of various publications, (c) examination of evaluation techniques and standard-setting, and (d) investigation of aggregation strategies. Three project activities are covered elsewhere in this volume: an innovative survey strategy based on techniques familiar to teachers operating from a whole language perspective (chapter 3), a technique that we dubbed the "video visit" for acquiring contextually grounded information about portfolio methods (chapter 9), and a working conference that brought together researchers and practitioners around the topic of writing portfolios (the remaining chapters in the volume).

Several discernible themes run throughout this complex array of data: (a) commitment to student portfolios has an energizing effect on teachers and students, (b) portfolios in practice encompass a wide variety of methods and purposes, (c) standard-setting, grading, and aggregation receive little systematic consideration when teachers initiate portfolios, and (d) portfolios mean a lot of work for teachers (and presumably for students). It also appears that, whatever this "portfolio thing" is, it seems to have arrived on the scene, at least

for the time being. As one of the practices that fall under the broader category of authentic assessment, portfolios are important in the part of the reform movement that emphasizes teacher autonomy and more genuine instructional experiences. Exactly what is a portfolio within this context? A metaphor, for one thing (Bird, 1990). Our mind's eye easily envisions an architect, artist, or model striding down Madison Avenue with an enormous binder under his or her arm—an aspiring and competent individual on the way to success! Portfolios are clearly positive, whatever they may be.

Metaphors offer models and allegories, but they are not the real thing. Students are not accomplished professionals. They are not competing for top spots. We are talking about kids in the elementary and middle grades learning about their interests and their potential. A portfolio—*port*: to carry; *folio*: paper—has a clear meaning on New York's Madison Avenue, but what about Sacramento's Mack Elementary School? From one perspective, student portfolios have a long history in this country—binders with notes, folders with assignments, backpacks with scribbles. But a portfolio seems clearly more than a collection of papers; after all, a wastebasket could fulfill this definition. And today's activities require more than transferring student assignments from a three-ring binder to a manila folder with an accompanying name change.

### “Real” Portfolios

Looking more closely at the metaphor may be helpful. What about professional portfolios? These collections differ from notebooks (or wastebaskets) in several respects. They are prepared with a particular *purpose* in mind. An artist's portfolio aims to send a thematic message. An architect's portfolio is designed to convince a client. An investment portfolio should promise a profit, the more the better. A poet's portfolio is a collection of efforts, both published and unpublished, with quality more important than quantity. The academic's curriculum vita and the professional's “bio” highlight job accomplishments over a life span.

The portfolios of experts are *selective*. An architect may include a progression of drawings in a folder, demonstrating how she approaches a problem, but the final drawing is presented as the best work. Pianists spend enormous amounts of time practicing scales, but these are not included in their recordings. Scientists conduct numerous experiments that may be well-intended but on reflection not very thoughtful, and these excursions do not show up in their curriculum vita. Learning, even when it goes relatively smoothly, entails trial and error. Portfolios seldom contain any “outtakes.”

Professional portfolios call for *judgments*. They are designed to withstand the rigors of evaluation by informed judges. In this sense, the gymnast's acrobatics and the skater's performance are portfolios-in-action. Performer

and judge are on the same page, in the sense that both share a common perspective on contents and criteria for assessment. Both understand that the primary purpose of a portfolio is for evaluation. Professional portfolios are graded, both absolutely and relatively. In Olympic competitions, the panel of judges lifts its cards to display the ratings of a diving performance; television commentators explain why this splash merits a 9.2 whereas a previous one received a 9.8. The panel's assessment is subjective, but it must justify its ratings against a set of criteria. The ratings are also relative; one performance is of greater merit than another.

Finally, these portfolios are *transportable*. They are designed to be of similar value no matter where they are go. The demands on an architect may be different in San Diego and Duluth, but certain fundamental elements transfer from one location to another. An equestrian who triumphs in Long Island is likely to be respected in Long Beach. This facet of portfolios is arguably the least defensible, because context is always a factor. The New York fashion model may be welcomed in San Francisco but not in Los Angeles, and the weather in Duluth poses different architectural challenges than the balm of San Diego. But the basic idea is to establish consistency within certain boundary conditions.

### Student Portfolios

Learners are not professionals; novices are not experts. How does the metaphor connect with classrooms and learning? What else should students “collect” other than the massive accumulations that they already aggregate? When and how should evaluation proceed? In what ways does portfolio assessment supplement or replace prevailing practices? How does it connect with standardized tests, grades, retention, or assignment to special programs? These and related questions are not to be found in the metaphor. The contexts are very dissimilar, so the analogies quickly fail. New answers must be devised for what are genuinely new questions.

Nonetheless, the metaphor does offer a plan for approaching these matters through the four categories introduced previously. The most sensible *purpose* for portfolios in the elementary and middle grades, it seems to us, is the documentation of student learning and growth. We are talking not about experts but about novices moving toward expertise. The emphasis should therefore be on progress rather than accomplishment.

Student portfolios should thus be *selective*, but now the choices should reflect student development from the beginning of the school year toward definable curriculum goals at year's end. Because this activity is taking place in classrooms, the record should clearly document opportunity to learn, and the teacher's comments about instructional scaffolding. Objectives-based testing often alludes to skill mastery; third-graders are unlikely to demonstrate

mastery in any significant domains, but they can reasonably be expected to display progress on the continuum from novice to expert. It therefore makes sense to *select* work samples that show movement along this continuum.

If student portfolios are to serve for assessment, then they must include *judgments*, whether in the form of grades, narrative comments, or measures. Several studies have revealed that today's teachers are uneasy about rendering judgment when the evidence is complex. They often speak of portfolios as an opportunity for dialogue with students; grades are based on tests. This discomfort springs from several sources. One is the lack of professional preparation for evaluating student work (Stiggins, 1994). Another is the difficulty of the task (Wiggins, 1993). But this unease also reflects a classical tension: Should school grades reflect progress, effort, or accomplishment? Rubrics for mandated portfolio assessment, which serve as models for classroom portfolios, frequently portray olympian qualities rather than describing either progress or effort. These rubrics lay out a fixed standard of expertise and accomplishment rather than a developmental scale showing movement from beginner to expert. None of the rubrics with which we are familiar incorporates mention of students' attitudes and motivations. Most call for a holistic score, which is poorly suited for either documenting growth or for identifying strengths and areas where help is needed. In a word, teachers are probably correct in their uneasiness about the role of portfolios in formal assessment. At certain points in schooling it makes sense to make summative assessments; when students leave elementary school for middle or junior high school and when they move on to high school, the faculty should be able to gauge students' level of accomplishment. But the main job throughout most of these formative years is formative evaluation, for which portfolios may be ideally suited. But evaluation is needed in any event.

Which brings us to the issue of *transportability*—what elements in the student portfolio need to be carried across grades, schools, and subject matters? Some educators see the portfolio as a steady accumulation from kindergarten through—life? Asked about the practicality of such advice, some rely on technological remedies; scan everything onto a CD-ROM! Although this strategy may be possible, it begs the question: What is important to preserve for assessment purposes? Neither students nor teachers have time or interest to review everything that a student does. The problem becomes more challenging if the portfolio also incorporates instructional information. We suggest that transportability is actually a curriculum issue. Suppose the aim is to document the significant developmental mileposts marking a writer's progress from kindergarten novice to eighth-grade expert. Eighth graders have learned a lot, and individuals differ significantly in style and interest. Where are commonalities to be found? We think that the answer lies in constructing a genuine alignment of assessment with curriculum—a coherent description of the course of study.

Portfolio assessment brings all of the issues sketched earlier into bold relief. In particular, portfolios closely complement the emerging concepts of cognitive learning. If the aim is to document students' progress in learning that is situated, social, and strategic, then the assessment record must include contextual information that informs each of these dimensions. The record of curriculum opportunities and instructional support is certainly a significant part of this context. The portfolio, from this perspective, is not an accumulation of writing samples, but a reflective account of a student's development as supported by his or her teachers. By studying such a folder, teachers and parents should be able to see evidence of growth, but also the conditions that supported learning, the interaction with classmates (and others, including parents), and the blueprint that makes sense of the activity. Portfolios meeting this criterion may vary considerably on the surface, but require a coherent design if they are to transcend anarchy. A major goal of the NCSW project was to construct a framework for such a design, and we turn next to that matter.

### A CONCEPTUAL FRAMEWORK FOR PORTFOLIO ASSESSMENT

We approached the conceptualization of classroom writing assessment portfolios with four practical questions in mind: (a) what purposes and audiences are served by portfolios; (b) what do portfolios look like; (c) how does a teacher "do" a portfolio, and (d) how are portfolios used for evaluation? These four questions provide relatively independent categories for thinking about portfolio design, for classifying different strategies and designs, and for reviewing the tensions mentioned earlier.

#### Purposes and Audiences

Although the question of why and for whom might seem obvious, the answers are frequently tacit at the classroom level. A teacher attends a reading-writing conference and returns excited about the concept. The leader of a process-writing or whole-language workshop shows how portfolios link with these movements. A principal or district administrator, eager to be on the leading edge, recommends portfolios, perhaps even mandating their use.

None of these questions directly addresses the why question. Enthusiasm, connections with other popular programs, administrative dictate—none of these are genuinely purposeful. Too often the answer is often "because it is there." For teachers searching for innovation and independence, portfolio techniques are attractive: they are flexible, they are student-centered, they

are *not* standardized, and they highlight the teacher's professional role. At one level, "why" is legitimately answered with "Because I like them and they are mine!" The exclamation mark is intentional, because teachers can be passionate in their response. Over the long haul, however, purpose requires more than passion.

What other answers are possible? One might suggest that portfolios are more appropriate for gauging higher-level learning outcomes, for connecting instruction with assessment, for giving students more responsibility for assessment, for more adequately informing parents and authorities. These suggestions appear on occasion, but are relatively rare in surveys. Mandates are straightforward: "I'm doing portfolios because the district told us to." Although this answer is honest and not uncommon, it lacks any commitment to genuine purpose.

What about audience—for whom are portfolios constructed and displayed? Some of the most heartfelt answers to this question point to the student as the focal audience. The practical literature in this field offers numerous romantic anecdotes about student involvement in and celebration of their portfolio productions. At one level this idea makes sense; William Zinsser (1990) was convincing when he argued that the most important reader of any composition is the writer; you should write for yourself. But this idea is circumvented when students are mandated to compile portfolios, and when they have no say in the purpose and design of the folders. Some students may find pride in their accomplishments, and portfolios offer many possibilities not to be found in multiple-choice tests. But other students are likely to be just as happy if they can conceal their shortcomings behind the anonymity of random marks with a #2 pencil.

Teachers are clearly an important audience for portfolios, whether or not assessment is an explicit outcome. But the teacher's role as audience can vary markedly: interested browser, harried commentator, engaged dialoguer, formal evaluator. To further complicate matters, the same portfolio may attempt to serve all of these functions, so that one moment the teacher is a supportive and engaging partner, only then to become an unforgiving judge. The teacher also plays a central role in linking portfolios to other audiences: parents, principal, district and state accountants. A unique feature of writing portfolios, unlike on-demand writing samples, is the enormous range of roles that the teacher can play in connecting the student through the portfolio to more remote audiences. We have seen situations in which students were left to their own devices, others in which teachers virtually prepared the portfolio, and some in which teachers set the stage for students to present themselves effectively but genuinely to other audiences. If portfolio assessment is to support the concept of learning that is situated, social, and strategic, then the challenge is to guide students in dealing with the concept of audience, providing instructional support to undergird this essential but demanding task.

### Substance and Form

What is a portfolio? What does it look like? What is in it? How is it organized? Detailed answers to these questions depend partly on audience and purpose, but some cross-cutting issues can be identified, and within any given situation a variety of choices offer themselves. As a practical enterprise, the literacy portfolio usually comprises a folder with situated samples of student reading and writing performance (Calfée & Hiebert, 1991; Harp, 1991; Tierney, Carter, & Desani, 1991; Valencia & Calfée, 1991). Practitioners frequently talk about three categories of writing samples: showcase (examples of accomplishment), documentation (records of progress), and evaluation (assignments submitted for grading). These models overlap considerably, but they identify distinctive characteristics and technical demands.

From a curriculum perspective, portfolio contents are driven by a different set of issues. What genre should be included? Personal narratives, in their simplest form, are an easier task for the young writer than technical expositions. Poetry can pique the student's imagination with relatively few constraints (assuming that free verse is permitted), but evaluation is a challenge. What elements of the writing process should be included? Early drafts with technical deficiencies and unrefined development? Final but unpolished drafts? Personal journals with random reflections? Answers to these questions require curriculum guidance; what are the instructional outcomes of the exercise?

A second set of substantive issues centers around the reading-writing connection. These issues are important in portfolio design for several reasons. First, the value that comes from integration of reading and writing has been cogently argued by others (e.g., Moffett & Wagner, 1992). Nonetheless, a long history of separation continues to thwart efforts to blend these two domains both conceptually (the disciplinary handbooks have failed completely to join reading and writing) and practically (basal readers now offer snippets of writing exercises but not a coherent curriculum; writing programs say little about text comprehension or the reading-writing connection; Clifford-Jonich, 1987). The instructional languages are different in the two domains. Portfolio contents often sustain the division; the reading section has a book log, while the writing section has compositions (Flower, 1994; Spivey & King, 1989).

Thirdly, it is becoming increasingly clear that portfolios are markedly enhanced by the inclusion of reflective comments by students and teachers, and even parents. Current techniques may ask the student to complete a cover sheet for each writing sample in which they comment on the strengths and limitations of the piece. Teachers sometimes write notes or "captions" on student work. Reflection is a difficult task, and responses often tend toward the mundane. Students are understandably pleased when they finish

an assignment—it is finally done! Reflection is just one more piece of busy-work. When asked about ways to improve a piece, students naturally follow the model of their teacher's remarks—write more, fix the grammar, and correct the spelling. One can imagine more substantial dimensions for critiquing a work: organization, development of ideas, personal voice, thematic integrity. These dimensions reach beneath the surface toward the wellsprings that make writing a personally meaningful part of the curriculum in the elementary and middle grades. The reflective element of the portfolio opens the way to develop these facets of literacy acquisition, but realizing this potential is a substantial and largely unmet challenge.

We have placed the substantive issues at the forefront of this section to emphasize their importance, because we think that function precedes form. Nonetheless, we recognize the need to organize a portfolio, to establish its form, shape and style. In practice, no single format is likely to serve all purposes, so that different portfolios require different structures at different times. A progressive format makes sense for a working portfolio, organization by genre is appropriate for evaluation, and a project-based arrangement offers opportunities for constructing a showpiece. The preceding references offer numerous examples of organizational patterns, and we do not develop the matter further. Our main caution is to warn against developing the portfolio as random accumulation. The circular file approach to portfolios serves neither instruction nor assessment, neither students nor teachers.

### Process

The conceptual question under this heading is how to do portfolios. The issues here revolve partly around the student's activities in assembling the various collections described earlier. But they also entail the work of the teacher in constructing a portfolio environment within the classroom, and the efforts of the entire school in supporting the classroom environment.

Several features of portfolio assessment are significant practical guides in shaping these activities:

- Production is more important than recognition: students must demonstrate that they can actually do something, rather than simply picking the right answer.
- Projects are more important than items; the emphasis should be on depth over breadth, on validity over reliability.
- Informed judgment is more important than mechanized scoring; the teacher replaces the Scantron as the central character in the assessment process.

For the student, the process of building a portfolio is ideally directed by a coherent combination of curriculum goals, strategic understandings, and

personal interests. Many have written about student-centered learning; portfolios offer a technology for helping this slogan to become a reality. But this goal requires that the portfolio process begin in the early years of schooling with substantial support by the teacher, evolving in the later grades toward increased student responsibility.

For the teacher, assessment becomes a task of applied research: planning, collecting the data, and interpreting the evidence (Calfée & Hiebert, 1991). This perspective differs markedly from the activity-driven approach to daily instruction. Portfolios can easily become little more than a decorative addition to business as usual, but they have the potential to become a lever for strategic and reflective teaching. A paradigm shift of this order of magnitude is likely to require that teachers spend time in consultation and development, that policymakers recognize that genuine reforms will be measured in years rather than months, and that transformations in mindset be valued more than changes in behavior.

### Evaluation

Portfolios do not have to be submitted to formal judgment. Professional portfolios usually serve this purpose, but some collections are for the fun of it, for sustaining personal memories, for tradition. The family photo album is a kind of portfolio, but we would not ordinarily think about subjecting it to a formal evaluation ("Aunt Martha seems to have gained some weight and Uncle Fred lost some hair" is not the same as a formal assessment). Indeed, some educators have proposed that portfolios provide a unique opportunity for genuine interaction between student and teacher, or for a unique student experience—efforts to evaluate these artifacts will only undercut the foundational benefits, and intrude on the intrinsic merit of the experience.

Evaluation can be tough. Teachers often report on how difficult it is to assign grades. Wiggins (1993) noted the inherent tension between support and criticism, between buddy and judge. This tension is especially high during the early years of schooling. Kindergarten teachers often object to the turmoil of formal assessments, and many educators question the validity of formalized testing for first graders. The critical question, still lacking a clearcut answer, is how to achieve a reasonable balance between the role of the elementary teacher as advocate versus authority at different grades and in different situations.

But assuming that student evaluation is required and that portfolios are to be part of this process, several questions come to the fore. What parts of the collection should serve for assessment? Who should serve as the judges? When should evaluations be conducted? What standards should be applied? Based on our previous analysis, we offer the following recommendations. First, the classroom teacher is the person in the best position to

gauge student achievements in the elementary and middle grades, although it is probably important for these judgments to be moderated by other colleagues inside and outside the school. Second, where assessment of learning—growth, progress, change—is the focus, then continuous formative assessment makes most sense, and is quite feasible when evaluation is local. Occasional summative judgments can then build on this base.

The third issue—standards—merits special attention. The touchstone is the assumption made in the previous paragraph—that the aim of portfolio assessment in the elementary grades is to measure learning. We argued earlier in the chapter that absolute standards of achievement make sense at some points in the school experience. When a student leaves elementary school for junior high or middle school, that is a reasonable time to assess the student's accomplishment of curriculum goals against a set of benchmarks or rubrics. The form of these standards of accomplishment is something of a mystery, in our judgment. They are not really all that clear in the curriculum standards being prepared by various professional organizations. These latter seldom specify with adequate clarity what comprises marginal, adequate, and excellent performance at a given grade level. The general approach is to lay out a qualitative dimension ranging from poor to good to better, to exhort students toward higher achievements.

Lacking in these specifications are substantive criteria for defining the dimensions or for describing the boundaries between one level and another. Scales developed by the National Assessment of Educational Progress (NAEP) for authentic assessment illustrate the current procedures for establishing standards. Level 300 in the reading scale is described as "finds, understands, summarizes, and explains relatively complicated information," whereas level 350 requires the reader to "synthesize and learn from specialized reading materials." In the recent redesign of reading assessment, NAEP (Langer, Campbell, Neuman, Mullis, Persky, & Donahue, 1995) describes levels of extended response to a passage (an important task, because it calls for students to respond in writing to a passage that they have just read). On this scale, *Unsatisfactory* reflects "little or no understanding, repeated, disjointed, or isolated bits from the passage," *Partial* "demonstrates some understanding, but is incomplete, fragmented, and unsupported by appropriate argument or evidence," and *extensive* "includes enough detail and complexity to indicate that the student has developed at least generally appropriate understandings of the passage and the question." A final level, *essential*, continues the pattern of a scale that moves steadily upward, but with features that are diffuse and shifting. Here are two samples (slightly edited to eliminate mechanical differences) from the NAEP report; the question is "If she were alive today, what question would you like to ask Mandy about her career? Explain why the answer to your question would be important to know."

I would like to ask Mandy about how did she feel to usually not be able to participate in sports like baseball so I would know how Mandy felt.  
I would ask her how did it feel back there. Because I would want to know how it felt back there at that time.  
Why did you like being umpire? It would be important because we wouldn't know why she liked being umpire, when girls weren't supposed to play sports. Unless someone asked her that question, we would never find out.  
Did you really like basketball? Did you have any friends or fans? Were you ever at any basketball games? The reason I would ask these questions is because I like basketball too. Were you ever a cheerleader? What color is your hair? Because if you ever get lost people would have to know what color your hair is.

Two of these compositions received a rating of *Unsatisfactory*, one a rating of *Partial*, and one of *Extensive*. The point here is not whether you are able to match these assigned ratings (although that is by no means unimportant), but the way in which you would justify your judgments. And notice, we are taking for granted in this example that it makes sense to judge the performance of fourth graders against a criterion of accomplishment rather than progress—how well could each of these four students write the year before this assignment?

To the degree that standard setting seems unclear for constrained assessments designed by national experts with substantial funding, we find reason for concern about the prospects for locally developed portfolios that are more open-ended and that must rely on limited resources and expertise. Establishing absolute standards of accomplishment probably demands the best that can be achieved by the combined efforts of professionals, practitioners, and academics. It probably calls for a combination of situated assessments like portfolios and on-demand examinations like the California Learning Assessment System (CLAS), now unfortunately defunct. It certainly requires that we continue to struggle with the challenge of establishing more effective procedures for ensuring judgments that are consistent, generalizable, and valid.

The current enamored with national standards may be a matter of looking for lost keys under the nearest lamppost. Evaluation is often associated with control, and it is tempting for those in power to resort to centralization and standardization, which offers the promise of consistent rubrics and procedures. We agree with Wiggins (1993) on the overriding importance of validity, which leads us to the conclusion that the locals are in a far stronger position to decide how to assess learning and to gauge student progress toward established standards. Our rationale here springs from the goal of ensuring that instruction allows all students fulfill their potential (Howe, 1994). If the assessment aim is selection, then it may make sense to establish summative standards at the outset of schooling, thereby eliminating clients without obvious promise as early as possible. Test children

on entry to kindergarten and decide whether it is worth trying to educate them. Adaptive, formative standards make more sense when the goal is to optimize progress for every individual. The challenge in this position is to ensure that these local instructional activities, idiosyncratic on the surface, can be certified as valid efforts toward common summative achievements. Portfolios offer opportunities for both formative and summative assessment; they can offer evidence about learning and instructional support for learning, while establishing the degree to which the student has met summative performance criteria. For formative evaluation, standards gauge learning and instruction conjointly; for summative evaluation, standards gauge the level of accomplishment. Although these two sets of standards need to mesh, they are very different in character.

### SOME MESSY MATTERS

Conceptual frameworks like the one just described have a neat academic appearance. The real world is considerably more cluttered. The editors of this volume suggested that the authors keep the framework in mind as they prepared their chapters, but their works also reflect their own perspectives as well as the snarly realities with which they have chosen to wrestle.

There is chaos in order, and vice versa. In this final section we highlight several tensions that reappear throughout these chapters as well as the broader literature in the field of portfolio assessment. The first of these tensions centers around the contrast between internally based and externally mandated assessments. What best serves the classroom teacher does not mesh nicely with the needs of administrators and policymakers. Some scholars think that this gap is unbridgeable (e.g., Cole, 1988). We are inclined to think otherwise, and see in portfolios the possibility for spanning this chasm (Freedman, 1993). The proof will eventually come from the pudding, but we think that the incorporation of teacher judgments within public accountability is not only possible but also critically important.

A related tension relates to efforts to standardize the contexts and conditions of assessment. To what degree can or should portfolios and portfolio assessment be kept constant across different contexts, and what are the costs and benefits for students and teachers from such constancy? Bureaucrats are generally more comfortable when things go by the book; even if reality is otherwise, they are reassured through standard operating procedures. Teachers on the front lines often call for flexibility in adapting instruction and assessment to local conditions and diverse student needs. As just indicated, we think it is possible to develop a common standard design that allows—indeed, demands—flexibility.

The third tension arises from the commitment to educational equity (Astin, 1990; Darling-Hammond, 1994; Nettles & Nettles, 1995). When assessment must conform to fixed methods and when developing students must meet fixed levels of performance, both equity and quality can be undermined. Again, some policymakers have emphasized the virtue of high standards for all. We agree on the merits of this policy at certain points in the process of education, but question the imposition of high standards of accomplishment during the early stages of learning, given that students vary considerably in their academic preparation for school. An analogy to the early preparation of athletes makes our point. One approach is to decide to begin by asking a group of 6-year-olds to attempt the high hurdles, and provide further tutelage to those who do not fall down. They probably have natural ability, and someone has probably helped them develop whatever talent they possess. But athletic prowess takes a variety of different forms, and varies with the child's age. A swimming pool may be more appropriate than hurdles for 6-year-olds. Moreover, it probably makes more sense with young children to focus on quality performance with lowered bars than to set the bars at a level where half of the racers trip on their first attempt.

A fourth tension appears when we consider the differential affects of new assessment methods on both students and teachers. These changes are coming at a time of increased demands on teachers because of deteriorating family demographics and declining support for public education. State testing policies, for reasons of efficiency, typically focus on one or two grade levels per school. "Portfolios—I think they do them in fourth grade" is an understandable response of by a harried third-grade teacher. As we noted earlier, the full potential of portfolio assessment is likely to depend on linking this technique to a developmental curriculum, and to procedures and expectations that cut across all grades. This means a schoolwide effort. In England, teachers have a tradition of professional responsibility for educational matters. They have steadfastly resisted efforts to federalize curriculum and assessment at the elementary and middle school levels. Although respectful (and knowledgeable) of national reports and recommendations, they view decisions about *exactly* what to teach and how to assess the learning of young children as local responsibilities. Each school faculty makes time to discuss methods and standards for assessment. Educational advisors, moderators, circuit-riding mentor teachers, ensure a degree of commonality across schools. Their role is not to direct or dictate, but to connect the insights and problems in one school with the work of other faculties.

This list does not exhaust the tensions by any means, nor do we attempt to provide resolutions. The chapters that follow raise other issues and suggest possible accommodations. What seems clear is that portfolio assessment, which arose as a largely grassroots movement, has become a cutting-edge instrument for fundamental reforms in U.S. education. Portfolios are unlikely

to achieve this end, however, unless substantively linked to other reform elements. We see in teacher-based classroom assessment—for which portfolios are one source of information about learning outcomes—the potential for genuinely systemic reform, in which all parts of the educational endeavor are conjoined. But in this version of systemic reform, the impetus comes not from policymakers at the top of the hierarchy, but from professional communities at the local school level.

### PERSPECTIVES ON THEORY AND PRACTICE

Someone—it may have been Kurt Lewin—said that nothing is so practical as a good theory. The gaps between theory and practice, academics and practitioners, and ideas and activities remain substantial in the field of education. The situation seems quite different in other professions, for whatever reasons, but in education the linkages are much more tenuous. One of the primary aims of this volume is to help bridge the gap. This goal is appropriate for an assessment volume, given the tension between internal and external mandates in the testing domain. We surely do not claim to have achieved success, but the project has managed to bring together individuals representing a variety of perspectives. We have organized the volume to reflect the connections and contrasts between theory and practice. We lead off with chapters that are more conceptual and research-like, moving then toward contributions that are more practical and teacher-like.

To frame these segments, the chapters by Herman, Gearhart, and Aschbacher and Belanoff provide conceptual “bookends.” Herman, Gearhart, and Aschbacher anchor new developments in the field with serious questions about the technical foundations for classroom-based assessment and Belanoff reflects on the common themes from the chapters. Herman’s background has put her face-to-face with the practicalities of elementary and middle school classrooms, whereas Belanoff shifts us to a very different perspective—equally practical, but focused on the bottom line for the high school graduate entering college.

Achieving coherence in an edited volume is always a challenge. We have approached this task in three ways. First, we have offered the preceding conceptual framework as a road map for readers. Second, literacy—reading and writing—is the focus for all the authors. Finally, one focal question can be posed for each of the chapters: How can classroom reading-writing portfolios enhance curriculum, instruction, and assessment for both local and external accountability? This question is likely to be of increasing importance in years to come. The roadway for educational reform is still murky in parts, but new and improved assessments will almost certainly part of any successful design.

### REFERENCES

- Applebee, A. N. (1980). *A study of writing in the secondary school*. Urbana, IL: National Council of Teachers of English.
- Astin, A. W. (1990). Educational assessment and educational equity. *American Journal of Education*, 98, 458–478.
- Berliner, D. A., & Biddle, B. J. (1996). *The manufactured crisis*. Reading, MA: Addison-Wesley.
- Berliner, D. A., & Calfee, R. C. (Eds.). (in press). *Handbook of educational psychology*. New York: Macmillan.
- Bird, T. (1990). The schoolteacher’s portfolio: An essay on possibilities. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation* (pp. 241–256). Newbury Park, CA: Sage.
- Bloom, B. S., Hastings, J. T., & Madans, G. F. (1971). *Handbook of formative and summative evaluation*. New York: McGraw-Hill.
- Britton, J., Martin, N., McLeod, A., & Rosen, H. (1975). *The development of writing abilities*. London: Macmillan.
- Calfee, R. C., & Hiebert, E. H. (1991). Classroom assessment of reading. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of research on reading* (2nd ed., pp. 281–309). New York: Longman Publishers.
- Calfee, R. C., & Patrick, C. (1995). *Teach our children well*. Stanford, CA: Portable Stanford.
- Calkins, L. M. (1986). *The art of teaching writing*. Portsmouth, NH: Heinemann.
- Clifford-Jonich, G. J. (1987). *A Sisyphus task: Historical perspectives on the relationship between writing and reading instruction*. Berkeley, CA: Center for the Study of Writing, University of California.
- Cole, N. S. (1987). A realist’s appraisal of the prospects for unifying instruction and assessment. In C. V. Bunderson (Ed.), *Assessment in the service of learning* (pp. 103–117). Princeton, NJ: Educational Testing Service.
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64, 5–30.
- Emig, J. (1971). *The composing processes of twelfth graders*. Urbana, IL: National Council of Teachers of English.
- Flower, L. (Ed.). (1994). *Making thinking visible: Writing, collaborative planning, and classroom inquiry*. Urbana, IL: National Council of Teachers of English.
- Freedman, S. W. (1993). Linking large-scale testing and classroom portfolio assessments of student writing. *Educational Assessment*, 1, 27–52.
- Freedman, S. W. (1994). *Exchanging writing, Exchanging cultures: Lessons in school reform from the United States and the United Kingdom*. Cambridge, MA: Harvard University Press.
- Gentile, C. A., Martin-Rehrman, J., & Kennedy, J. H. (1995). *Windows into the classroom: NAEP’s 1992 Writing portfolio study*. Princeton, NJ: Educational Testing Service.
- Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist*, 36, 923–936.
- Goodman, K. S., Goodman, Y. M., & Hood, W. J. (1989). *The whole language evaluation handbook*. Portsmouth, NH: Heinemann.
- Graves, D. H. (1983). *Writing: Teachers and children at work*. Portsmouth, NH: Heinemann.
- Harp, B. (1991). *Assessment and evaluation in whole language programs*. Norwood, MA: Christopher-Gordon Publishers.
- Hiebert, E. H., & Calfee, R. C. (1992). Assessment of literacy: From standardized tests to performances and portfolios. In A. E. Farstrup & S. J. Samuels (Eds.), *What research says about reading instruction* (pp. 70–100). Newark, DE: International Reading Association.
- Howe, K. R. (1994). Standards, assessment, and equality of educational opportunity. *Educational Researcher*, 23, 27–33.

- Langer, L. A., Campbell, J. R., Neuman, S. B., Mullis, I. V. S., Persky, C., & Donahue, P. L. (1995). *Reading assessment redesigned*. Princeton, NJ: Educational Testing Service.
- Lieberman, A. (1992). *The changing contexts of teaching*. Chicago, IL: National Society for the Study of Education.
- Moffett, J., & Wagner, B. J. (1992). *Student-centered language arts* (4th ed.). Portsmouth, NH: Boynton/Cook.
- Myers, M. (1980). *A procedure for writing assessment and holistic scoring*. Urbana, IL: National Council of Teachers of English.
- Nettles, M. T., & Nettles, A. L. (1995). *Equity and excellence in educational testing and assessment*. Boston: Kluwer.
- Perl, S. (1979). The composing processes of unskilled college writers. *Research in the Teaching of English, 13*, 317-333.
- Resnick, L. B. (1987). *Education and learning to think*. Washington, DC: National Academy Press.
- Sommers, N. (1980). Revision strategies of student writers and experienced writers. *College Composition and Communication, 31*, 378-388.
- Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly, 24*, 7-26.
- Stiggins, R. J. (1994). *Student-centered classroom assessment*. New York: Merrill.
- Sweet, J. (1976). The experience portfolio: An approach to student writing. *English Journal, 65*, 50.
- Tierney, R. J., Carter, M. A., & Desani, L. (1991). *Portfolio assessment in the reading-writing classroom*. Norwood, MA: Christopher Gordon.
- Valencia, S. W., & Calfee, R. C. (1991). The development and use of literacy portfolios for students, classes, and teachers. *Applied Measurement in Education, 4*, 333-346.
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass.
- Winfield, L. F., & Woodward, M. D. (1994). Assessment, equity, and diversity in reforming America's schools. *Educational Policy, 8*, 3-27.
- Zinsner, W. K. (1990). *On writing well* (4th ed.). New York: Harper.

## Portfolios for Classroom Assessment: Design and Implementation Issues

Joan L. Herman

Maryl Gearhart

Pamela R. Aschbacher

*Center for Research on Evaluation, Standards,  
and Student Testing (CRESSST)*

*UCLA Graduate School of Education and Information Studies*

Portfolio assessment has made a dramatic entry into the 1990s assessment scene offering great promise: Assessments that will right the wrongs of traditional measurement practice; assessments that will help, not hurt or subvert, the instructional process; assessments that will provide bridges, not barriers, to student accomplishment. The appeal and potential benefits of portfolios are many, particularly when compared to traditional standardized testing. Because portfolios contain the products of classroom instruction, by definition, they should be integrated with it, not an intrusive add-on. In contrast to the focus of traditional testing on discrete skills, well-designed portfolios contain student work reflecting students' accomplishments toward significant curriculum goals, particularly those that require complex thinking and the use of multiple resources. Because the assessment of student performance on these tasks can provide evidence of students' accomplishments and thereby serve as a tool to support the instructional process, portfolio assessment can bolster the efficacy of teachers, encouraging them to consider deeply how students are progressing. Portfolios, in addition, invite students to reflect on and take responsibility for their own progress, the assessment process, and, ultimately, their own learning. Finally, portfolios provide parents and the wider community with credible evidence of student achievement, and inform policy and practice at every level of the educational system, from individual students to the nation as a whole (Freedman, 1993).

The promises of portfolios are enticing, yet claims alone do not assure their realization. In this chapter, we consider issues in moving beyond the