2007

Best Practices in Writing Assessment

Robert C. Calfee
University of California - Riverside

Roxanne Greitz Miller
Chapman University, rgmiller@chapman.edu

Recommended Citation

Calfee, R. C., & Miller, R. G. (2007). Best practices in writing assessment. In S. Graham, C. MacArthur & J. Fitzgerald (Eds.), *Best practices in writing instruction* (pp. 265-286). New York: Guilford Press.

13

~

Best Practices in
Writing Assessment

ROBERT C. CALFEE *and*
ROXANNE GREITZ MILLER

We begin this chapter with three snapshots to ground our discussion in the realities of practice. Each snapshot reflects several classroom experiences that we have shaped into an integrated portrait:

Samuel had delivered his first show-and-tell report earlier in the morning. Now he sat beside Ms. Hancock as she reviewed the notes she had made at the time. Sam was small for his age and a bit shy, and talking in front of the entire class had been a challenge, but he had made it! His topic had been his new baby sister. After announcing that she had come home from the hospital, he was at a loss about what to say next. Ms. Hancock prompted him with a few questions. What kind of hair did she have? What kind of noises did she make? What did she do? Samuel had something to say about each of these matters. Ms. Hancock has written four sentences from Samuel's words: "Martha is my new sister. She is bald. She gurgles. She mostly sleeps." Samuel has completed his first academic project, which will appear in the upper lefthand section of the weekly parent newsletter. His parents show delight as reads his report to them, and it sets the stage for the upcoming parent–teacher conference.

June has been a voracious reader since preschool and started a personal journal in second grade. As a fourth grader, she has already

written several brief papers, but now she faces a different challenge. Her teacher, Mr. Buchers, has announced that the March assignment will be a research paper. Students must first select a current events topic for background reading. Both the reading and writing will be expository rather than personal narrative. Mr. Buchers is enraptured by history and spent time during the fall and spring introducing students to historical analysis. Now the class is going to study history in the making! Mr. Buchers explains that he is pushing the class; this type of writing is generally introduced in fifth or sixth grade, but he thinks they are up to it, so he is giving them a head start. The assignment will take 2 or 3 weeks to complete; they will work in small groups. Mr. Buchers reviews some basics: how to find materials in the library and on the computer and how to take notes for the report. June is considering "Can a woman be President?" as her topic. Her parents have different positions; her father is inclined to support the idea, but her mother is less sure. June agrees with her father but knows that she must consider both sides of the issue. It will be a different kind of writing and reading for her.

Tom and Chizuko have been good friends since they met in ninth-grade math. As they near the end of high school, the SAT writing test looms large on the horizon. They both enjoy math and science but are less comfortable with composition assignments, and neither did especially well on the PSAT writing test. They now study together, using materials from the College Board website as a guide: "Brainstorm, collect information, organize, do a rough draft, revise and refine, read more, and write more" (*www.collegeboard.com*). Great advice, but how should one use it during an on-demand timed test? Math seems simpler to them—analyze the problem, work out the answer, and that's it. Writing is so mushy, with never enough time to make sure that everything is exactly right. And no one seems to teach writing! English class is about novels and plays, and their other teachers expect students already to know how to write.

These snapshots capture the range of writing scenarios that students experience during their school years. By the end of elementary school, students seldom write unless they have to and then only because it "counts." In high school, writing begins to count a lot, across the board and over the long run. The College Board advice about effective writing is certainly on target. Best practices should follow these guidelines, but too often neither students nor teachers can find the time. Standards have to be met, content has to be covered, and the textbook has to be finished. Writing is included in the standards, but the responsibility for acquiring skill rests largely on students' shoulders. In the elementary grades, reading has priority; in many classes, as much as half the aca-

demic day is devoted to the basal reader. By the middle school years, teachers deal with more than 100 students every day, which means little opportunity for individualization of any sort.

Our assignment in this chapter is to discuss best practices in writing assessment, a task that poses a twofold challenge for teachers—first, the task of providing authentic opportunities for students to acquire skill in writing while covering an ever-increasing array of other curriculum demands; second, the overriding pressures to ensure that students perform well on the standardized tests that have become the primary accountability index. As we complete this chapter, few state testing systems rely to any significant degree on performance tests for measuring student achievement. Multiple-choice tests dominate, and on-demand writing tests (including the SAT) generally contravene the counsel provided by the College Board.

We assume that we are writing for teachers who are writers and that you understand the importance of establishing both audience and purpose. We have framed our audience as "teachers who are writers" rather than "writing teachers." Ideally, every teacher, across all grades and subject matters, should incorporate writing as an integral part of instruction because writing reveals thinking (Miller & Calfee, 2004) and can serve as a critical source of information for both teacher and student. Writing takes time and patience, which can be a challenge in a daily curriculum packed with objectives and standards. We assume that our audience has some freedom, although it may be limited, to deviate from the official schedule and the patience required to help students reflect on their learning.

Our purpose is to survey assessment concepts and techniques supported by research and practical experience and to suggest ways to fit these ideas into the realities of policies that, although well intended, often conflict with best practices. The advice from the College Board illustrates this point; it captures many facets of best practices, but the real SAT assessment permits none of these elements. We have limited space for presenting how-to details, but we will provide selected references to help apply the ideas.

The chapter is organized around three topics. First, we describe the concept of embedded classroom writing assessments designed to inform instruction and provide evidence about learning. The bottom line here is the recommendation that writing tasks (instruction and assessment) be designed to support the learning of significant academic topics (Urquhart & McIver, 2005). Next, we present several contrasts that emerge from this perspective: process versus product, formative versus summative evaluation, and assessment versus testing. Finally, we review a set of building blocks that is essential to all writing assessments, especially

those that are classroom-based: the prompt, the procedures, and the rubrics. As you have probably realized from the scenarios and the discussion thus far, our focus will be on *composing* more than *mechanics*. Attention to spelling and grammar is eventually important, but it helps if the writer has something to say and has learned how to organize his or her ideas.

WRITING TO LEARN

When and how should students learn to write? In reading, the contrast is often made between learning to read and reading to learn (Chall, 1995). A similar distinction can be made for writing, but we think that the basic idea is flawed in both instances. From the earliest stages, both reading and writing should be grounded in the purposes of literacy: to think and to communicate. To be sure, students need to acquire skills and strategies for handling print, which requires time (and patience) from both teacher and student, but learning is more effective when motivated by a clear purpose. In short, learning to write is generally best grounded in writing to learn.

In a world where student achievement is often gauged by the application of a pencil to a multiple-choice bubble, it is important to remember that writing is a performance task that requires substantial effort, motivation, persistence, strategic planning, and skill, as well as knowledge about the topic. If any of these ingredients is missing during an assessment, then the student's potential can be substantially underestimated. Valid assessment needs to tap into both product and process, with probes that gauge each of the preceding elements.

For all of these reasons, it makes sense that writing and writing assessment should be linked to meaningful academic outcomes. Given the substantial costs to everyone, writing activities are best focused on significant matters, rather than on writing for the sake of writing. This recommendation does not mean that writing must center around dull schoolwork. For Samuel, the kindergarten show-and-tell report demonstrates that he can focus on a topic (his new sister) and elaborate with a few sentences. He can now apply the strategy to the rock in his pocket, the snake in the terrarium, and (later) the causes of the Civil War. June and her classmates are acquiring new skills and strategies, including the mechanics of the five-paragraph essay, but the focus is the topic of their historical research. Tom and Chizuko have, in the best of worlds, spent more than a decade learning to write as laid out by the College Board. The reality is probably quite different. Instead of approaching the SAT with experiences that leave them confident and self-assured, able to

adapt what they know to a broad range of situations, they suddenly confront a high-stakes task for which they feel poorly prepared. If they are lucky, their teacher (or tutor) will take this opportunity to help these two young people learn to deal with the SAT and to see the difference between SAT writing and real writing.

Best practices in writing assessment begin with an authentic task, where purpose and audience are clear and meaningful, where support and feedback are readily available, and where the final product has academic value for the student. Contemporary surveys of writing assessment (Black, Harris, Lee, Marshall, & William, 2003; Chappuis, Stiggins, Arter, & Chappuis, 2005; Harp, 2006; Stiggins, 2004) typically employ a conceptual framework with the elements shown in Figure 13.1.

Let us offer a few cautions before discussing the elements in Figure 13.1. The model suggests a fixed path not intended by the framers, but the basic elements generally make sense as a model of classroom assessment. The model needs to be filled in for a specific application, of course. In this instance, what is missing is "writing about what, and why?" At the risk of overstating the point, we recommend that you not teach writing for the sake of writing. Instead, think about ways in which writing can support learning of academic outcomes, including both content and process. From the lowly book report to the daunting research paper, subject matter provides opportunities for students to demonstrate learning at the same time that they acquire skill in communicating. For the teacher, the point is that writing (like reading) becomes an integral part of virtually every lesson. You may be asking yourself, "Who is going to grade all of this stuff?" We address this question later.

Three implications spring from embedding writing in subject-matter learning. First, this approach addresses issues of topic and purpose directly. Whether a check of background knowledge, a quick quiz to review an assigned reading, or an extended project, writing becomes an integral part of the learning process. Second, writing (like reading) varies substantially with developmental level and subject matter. What can be expected of a second grader describing a collection of fall leaves, a fifth
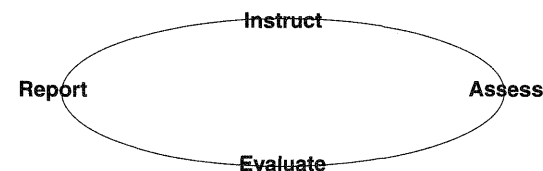

FIGURE 13.1. Conceptual model for classroom assessment.

grader preparing a report on why gas prices rise and fall, and an eighth grader developing an opinion piece about graduation ceremonies? Third, assessment techniques need to be shaped differently for each of these scenarios. In particular, valid and informative assessments must balance content with writing; to what extent does the composition reveal understanding of the topic, and how well written is the work?

Now a few words about each element in Figure 13.1. The action element is *instruct*, a word from the same root as *structure*, with the connotation of building, designing, framing, and completing. Teachers have the responsibility to help young people construct academic edifices during the school years—identifying essential parts of various structures, setting the stage for student projects, engaging them in the building task, checking the work along the way, and inspecting the final project. Literacy serves as an essential tool kit (or machine shop) for the construction process. Best practices for writing assessment check the status of the kit; the tools need to be in good shape, and the user must know how to use them effectively. Authentic curricular goals are the critical substance for the enterprise.

The joining of *assess* and *evaluate* is critical to best practices. We say more about *assess* in the next section, where we contrast it with *test*, but the core idea is the collection of evidence about student learning. *Evaluate* refers to the interpretation of the evidence. These two are interwoven, rather than sequential, but they require different activities and states of mind. When Ms. Hancock takes notes on Samuel's words during his report, these serve as evidence as does the scaffolding she provides along the way. Samuel was clearly eager to tell classmates about his new sister, but leading questions kept him going. He was still learning his ABCs and, from one perspective, was not able to read, but when Ms. Hancock wrote his sentences in the daily report, he could read them on his own. All these observations serve as evidence, which often takes shape as a story like Samuel's.

What does the story tell about what Samuel knows and can do? What might be the most useful next steps instructionally? Such questions exemplify the evaluation process, which requires reflection and debate. The point is not to decide whether the performance is good or bad but to consider alternative interpretations that suggest various instructional responses—to think like an experimenter (Calfee & Hiebert, 1990). The question is not "Can Samuel compose/write a show-and-tell report?" which implies a yes-or-no response. Rather, it is "Under what conditions can Samuel produce a show-and-tell report with particular characteristics?" (e.g., three ideas related to a central theme). The evidence in this case suggests that Samuel has not yet learned this task to the point where it has become automatic. On the other hand, with a bit of guidance he was able to complete the task reasonably well and seemed enthused

about his accomplishment. He was engaged, he could talk about the results, and the event set the stage for his future learning activities and provided a model for the entire class.

Recommendations for best practices typically include completion of a *report* as an element. The basic idea is to document the activity. In fact, classroom assessment is often on the fly, with the results recorded mentally; where the evidence is oral, memory may be all that is possible. An important feature of written material is that one does not have to rely on memory—there is a concrete record. The question is how to make effective use of the information.

The most important assessment record is the one that serves the teacher in documenting student learning and steering instructional decision making. The student is clearly an important audience for such information, which can provide feedback, encouragement, guidance, and sometimes grades. Other audiences include parents, administrators, and other teachers. Reporting, except for formal mandates like report cards, tends to receive relatively little attention in educational situations, which is somewhat strange when you think about it. If you visit your doctor or auto mechanic, you expect assessment and evaluation to be part of the exchange, typically as a basis for action. You also expect a record of the entire process—what was checked, what was found, and what was done. Best practices for schooling should be documented to provide a basis for reflection on student learning, and to guide the teacher in shaping the curriculum in practice.

June's teacher, Mr. Buchers, appears tuned in to this principle. The class assignment is to write a research paper about a significant curriculum goal, the analysis of historical happenings. The task requires both reading and writing, but, most important, it requires thinking. June will report regularly to the class about her project, and Mr. Buchers will gently but firmly model and shape questions during these discussions. What evidence does June offer for her claims about a woman president? What about other interpretations? lessons from previous decades? possibilities for the future? June's written record informs Mr. Buchers about her progress in dealing with these questions. Tom and Chizuko, in contrast, often feel that they are working in the dark. Their English teacher administers biweekly practice exams and offers suggestions for self-assessment. The opportunity to practice helps, but it is up to the students to review their progress and decide what they need to do to improve.

CONTRASTS IN WRITING-BASED LEARNING

When writing becomes a commonplace of daily life in the classroom, the teacher confronts interesting contrasts. Evidence of student learning is

everywhere, which allows assessment of both process and product. The teacher directly experiences the distinction between formative and summative evaluation, between growth and accomplishment. It clearly serves no purpose to grade every piece of student work, but neither can one ignore students' efforts; rather, the goal is to use compositions to guide growth. Students will eventually begin to take for granted the conditions of authentic writing, which allow time and offer support (from the teacher and other students), but they also need to learn about the realities of tested, on-demand writing. Each situation offers opportunities to review the distinctive features of best practices (or at least very good practices).

A student composition provides information about both *content* and *process*, about what students have learned and how well they can communicate it. One aim of this volume is to encourage teachers to consider writing as an integral part of learning for all curriculum domains, so the *what* is especially important.

We return to this matter shortly, but first we address the *how well* question, which centers on communicating, which is at its core a two-way process. Notice that, unlike reading and writing, schools do not schedule separate classes on listening and speaking. The point is not that students do not know how to listen and speak—they generally appear able to communicate informally with family and friends—but that virtually every student needs to learn about academic discourse (Heath, 1983). The usual assumption is that they will acquire this language register through participation in classroom conversations throughout the day. In fact, if you listen to such conversations, they frequently turn out to be rather one-sided, with teacher talk the dominant discourse and student talk rather sparse.

What are the alternatives to teacher talk? How can the teacher provide reasonable opportunities for students to engage in genuine academic discourse during classroom discussions? Time is limited. Only one student can hold the floor at any given time. Small-group techniques offer one option, but management poses a challenge, as does documentation (and, hence, assessment and evaluation). Writing provides a practical approach—students can all write at the same time, and the information does not disappear into the air—but (1) students must write well enough to capture what they have to say, and (2) someone has to handle evaluation tasks.

Separating content and process is an important first step for speaking as well as writing. A student may have produced a beautiful piece of writing (process), but the content may show little grasp of the topic or may simply be off topic. Another student may turn in a piece that is difficult to handle—poorly organized, misspelled, ungrammatical—but somehow one realizes that the student is deeply engaged with the message.

Surface features may stand in the way of appreciating what a student has to say. The challenge is to see beyond the mechanical flaws to an appreciation of the substance of the message.

We are not completely satisfied with our content and process labels. A contrast is also made between process and product, the difference between how a student writes a paper and the quality of the final work. Product often refers to organization and coherence, as well as mechanical details, all of which are important, but the substance of the composition is generally overlooked in product rubrics. Especially when writing is linked to subject-matter learning, it is important for assessment to give attention to the substance. Holistic approaches mush everything together, and most analytical approaches downplay or ignore substance. So we use process to refer to the student's activities in writing a composition as well as the characteristics of the written work, and product to refer to the substance of the work as a response to the assignment. Books and articles on writing and writing assessment deal generically with the writing part of the equation. Teachers are generally left on their own to figure out how to use a composition to judge student understanding of a topic.

Suppose, for example, that you have assigned eighth graders to write a paper explaining variation in the seasons. You emphasize that the work needs to be well crafted, and you have explained the rubrics for gauging coherence and conventions. How do you explain your plan for judging the substance of the papers? We assume that your instructional aim is something other than a collection of Web-based cut-and-paste pieces, that you expect students to go beyond a summary of resource materials to reconstruct or transform the ideas. In this case, you can expect individual variation in both structure and specifics. Within these variations, it is still possible to define the key concepts (e.g., the inclination of the earth on its axis) and essential relations (what happens at the North Pole as the Earth moves around the Sun). If these critical elements are missing or unclear, then the student presentation of the content is problematic, no matter how engaging or well structured the piece. In a study of Harvard graduates by Schneps (1989), students told enthralling stories about how the seasons changed as the Earth moved around the Sun. Unfortunately, the substance of their compositions (which were oral rather than written) was wildly wrong in many instances.

Assessment of content should allow writers considerable leeway, within reason, in how they approach the task. For example, suppose a student structures the essay as a narrative in which a space station crew recounts their observations as they circumnavigate the globe from 25,000 miles above the earth. This composition would be quite different from a more scientific piece that lays out the sequence of seasonal patterns in mechanical detail or a persuasive essay that describes how the

shifting seasons play a role in the development of holes in the ozone layer. The goal in assessing content is to judge students' mastery of the essential information as they play with different styles and audiences. Suppose the assignment is to explain the seasons to third graders. What considerations come into play when the challenge is to explain the content to a genuine novice? The power of including compositions in subject-matter curricula is that students are called upon to demonstrate their understanding of a topic in a range of settings. The assessment challenge is to develop a set of principles that accommodate a wide range of topics in a consistent manner, so that the rules of the game do not constantly change. Notice that more is involved than getting the facts right. Facts are part of the puzzle, but concepts and relations among the facts are even more important.

The second contrast is between *formative* and *summative evaluation* (Bloom, Hastings, & Madaus, 1971), which often amounts to the difference between judging growth versus judging accomplishment (Calfee, 1997). Formative evaluation entails relatively informal procedures for obtaining information that can guide improvement in student learning. The primary goal in the classroom setting is to establish the degree to which the student is making progress—and, if he or she is not, to find out how to help the student begin to move ahead. Formative evaluation is experimental in the truest sense of the concept; under what conditions can a student perform the task at a level adequate to meet prescribed standards? A student may fail a task for many reasons. Formative evaluation searches for the conditions that support success, which can include helpful advice from the teacher, can also open the way to explore interest and motivation, opportunities to cooperate, and various accommodations. Notice that in the formative mode of assessment and evaluation, documentation of the scaffolding conditions is an important part of evidence, along with student performance. Formative evaluation is richly qualitative, creating portraits that can be viewed from different perspectives.

An example may serve to illuminate the point. Martin, who is repeating third grade, spends part of the day in a special education class. The diagnosis includes poor decoding skills, weak vocabulary, limited comprehension skills, and a total lack of motivation when it comes to reading. Martin cannot read and does not want to learn. During a school visit, one of us (Calfee) was asked to take a look at Martin, an invitation for formative evaluation. Martin, the teacher, and I met in the teacher's office, where a stack of textbooks had been placed on the table. The teacher asked Martin to pick a favorite book to show me how he could read. The aim was to give Martin a choice, which made sense. The problem, of course, was that Martin had little interest in any of the books, as evidenced by an uplifted eyebrow. I moved the books to the

side and asked, "What do you think about them Bulls?" The school was in Chicago, and it was the heyday for the local basketball team. Martin brightened considerably and spoke at length about the team's exploits. The structure of his exposition was well formed, his vocabulary was rich and appropriate to the topic, and he seemed quite enthusiastic. When asked about his sources of information—"You must watch a lot of television?"—Martin responded that he watched the broadcasts, but was also an avid reader of the sports section in the local paper. How to evaluate this exchange? Formative evaluation pays attention to subtle clues such as raised eyebrows, tone of voice, and choice of words. Triangulation is critical; any single piece of evidence is limited, so the evaluator must put many pieces together. In Martin's case, I concluded that he could decode under certain conditions, that he exhibited significant capacity to handle complex vocabulary and comprehension tasks, that he was capable of constructing an engaging exposition, and that he was highly motivated. The obvious challenge was to move these capacities from the basketball arena to the classroom setting.

Summative evaluations address bottom-line issues, captured not by a rich portrait but by a few numbers or letters. Today's students must meet established levels of accomplishment on mandated tests at prescribed grade levels. The teacher needs to keep these requirements in mind for individual students, especially those who enter the school system at a disadvantage. The point is that the consequences of formative and summative evaluation are quite different; if a student fails a summative event, seldom does the evidence provide valid information about how to remedy the problem. In addition, summative assessments arise from external mandates and are beyond the control of the classroom teacher.

Which brings us to the third contrast, *assessment* versus *testing*. *Assess*, from French, means to "sit beside," to exchange thoughts, and so on—a positive context. *Test* is from Latin, and conveys the sense of pulverizing—less friendly. Most summative evaluations rely on a testing environment. The individual is placed under stress with no support and limited time. Assessment is critical for guiding instruction, but the teacher also needs to help students prepare for the demands of testing situations in modern life, from the SAT to the driver's license test.

As students move through the grades, it makes sense to help them manage a gradual transition from assessment to testing. Testing kindergartners is questionable practice and may even be unethical. It is probably reasonable to test most high school students to certify their capacity to demonstrate skill and knowledge in the academic arena and to provide evidence about work habits and the like. Test taking can and should be taught, including strategies for multiple-choice exams and for writing.

In today's high-stakes environment, it is especially important for young people to be clear about the distinction between testing and assessment and the gradations between these two points. Accordingly, the teacher needs to incorporate the distinction explicitly in classroom practice. Psychologists use the term *metacognition* to describe "thinking about thinking." It may be important to teach *meta-writing* to help students learn to describe how they approach various writing tasks, including their understanding of conditions, expectations, and criteria. Turning again to the College Board advice, we can imagine teachers leading their students through authentic writing exercises like those recommended but also introducing students to the realities of testing, such as how to apply skill and knowledge when taking the SAT. Older students are capable of handling these contrasts. They all know the difference between real driving and behavior on a driving test.

CONSTRUCTING EMBEDDED WRITING ASSESSMENTS

This final section focuses on building an assessment protocol that provides valid evidence to support the model of curriculum-embedded writing sketched earlier (Calfee & Miller, 2005). The facets covered here are important for any writing assessment. The purpose is to place these facets within the context of the classroom teacher's daily work. It is one thing when a testing company or state develops a large-scale writing assessment. It can call upon its teams of experts, conduct pilot runs, calculate complex statistics, and so on. It is another thing when the classroom teacher sets forth to prepare a writing task that is relatively casual, intended for a one-time, low-stakes formative assessment. The second scenario may actually be more critical, in that the teacher can use the information to make judgments about student learning—the stakes are not high, but they are significant.

The facets required to construct a writing assessment are similar for virtually any scenario: the prompt, the procedure, and the rubric. Next, we explore each of these basic constructs, emphasizing the application to formative assessment in classroom settings.

The *prompt* sets the stage for the writing task. Rather amazingly, there is relatively little research on how variation in prompt design affects the quality of student writing, and we have accordingly spent considerable time on this facet in our research and practice with teachers.

Constructing a prompt is almost like writing a separate passage. In a brief amount of space, the teacher has to cover the following points:

- Develop a *focus statement* that directs students' attention to the key topic for the composition, activates prior knowledge (including the target text), and directs students in thinking about the task.
- Present, as clearly as possible, the *purpose* of the composition. Words like *tell*, *describe*, *explain*, *convince*, and *illustrate* serve this purpose, especially if students have received prior instruction on these terms.
- Identify the *audience* for the work. This is a challenge in school writing because everyone knows that the teacher is the real audience, but students can learn, to good advantage, to imagine various audiences. Indeed, for a good deal of authentic writing, we have to rely on imagination. Freedman (1997) gives a delightful and informative account of the ways in which high school students in San Francisco and London handled audience in writing to one another. With a little creativity, local audiences can be identified—the principal, the mayor, the editor of the newspaper. Writing for nobody can be discouraging.
- Where appropriate, specify the *form* of the product, such as a paragraph (or more) or a letter (a favorite because of the style).
- Tell the writers as much as possible about the *criteria* to be used in judging the work. How important are supporting details? If a text is provided, how should it be used? Is the work a draft, or should the student attempt a polished product? Ideally, for classroom assessments, criteria have been defined early in the school year and practiced (with feedback) regularly.

This advice means that the prompt becomes a mini-essay in its own right, and this is a problem because it takes a lot of work for the teacher to prepare and a lot of time for the student to digest. For classroom assessments, however, the investment in prompt preparation can be worth the effort because it provides an opportunity to teach students how to comprehend a prompt, including those instances when much of the information is missing.

Consider the following two prompts:

- Describe the differences between evergreen and deciduous trees. List at least three examples of each type of tree and describe what they are like.
- Explain the differences between evergreen and deciduous trees. Give three examples for each type of tree, and describe how these examples illustrate the differences.

These prompts are fairly typical of what we see in writing assessments developed by districts for classroom assessments. They are silent about audience and form. Both are roughly equal in length, but we suspect that the second question would pose the greater challenge. *Describe* and *explain* place quite different demands on the writer. Examples that "illustrate the differences" require more thought than random examples presented without a purpose.

Think about the possible *answer spaces* for each prompt, the content that students might generate in response to the prompt. Think about ways in which students might set out on productive paths or how they might be stalemated or led astray. For example, consider the following prompt:

- Describe the differences between evergreen and deciduous trees. Based on your personal experiences, what is your favorite kind of tree for each type?

When students are asked to build a composition around personal experience, which appears frequently in writing prompts, the way is open for them to move in any of a wide range of areas or nowhere, depending on the topic. We are not suggesting that prompts never invite students to draw on personal experience, but rather that the ground rules for such invitations require careful attention.

Of course, none of these prompts may provide a suitable starting point for students from urban areas, where trees are rare. The most effective starting point for a writing task often provides students with background about the topic—an experience, a discussion, or a passage, preferably with graphics. A basic contrast can be drawn between *stand-alone* and *text-based* prompts, and we recommend the latter, especially for classroom purposes. Stand-alone assessments, the prototype of a writing test, are cold-turkey scenarios. Students are given paper, pencil, and question and are then on their own. It is hard to imagine a more challenging situation! This approach makes sense only as a way to immerse students in the experiences they will encounter in large-scale testing, but the teacher should also try to build a ramp from a scaffolded situation to more spartan test environments.

A reference passage can provide a resource for the writing task. Even when students are writing on a topic that they have just studied, it often makes sense to include a target passage with the assessment procedure, as it lessens the demands on memory and ensures all students have a common starting point for the task. Selecting the reference text requires care, of course, much as for a comprehension test. Indeed, in this situation, comprehension and composition become interwoven.

The *procedure* for a writing assessment builds, in the ideal situation, on what we know about the writing process (Gray, 2000). Students need time, information about the topic, scratch paper, support and advice, and a strategy.

Time is the most precious commodity for the classroom—there is never enough time, even for basic writing. Then there is reading and scoring, discussing and reviewing, and handling the needs of individual students. No wonder many teachers assign writing a low priority. We offer two suggestions in this arena. The first is never to ask students to write about nothing. The second is to engage them in the assessment process. By "writing about nothing," we mean exercises (including district assessments) that are solely designed for writing. The result is akin to taking a driving test; you drive to show that you can drive, but you do not go anywhere! School subjects provide a plethora of openings for students to demonstrate knowledge, reasoning ability, and communicative capacity by composing, both orally and in writing. Especially when embedded in an authentic project, writing tasks evoke imagination and force that is otherwise totally lacking.

Information about *topic* was mentioned earlier in the distinction between text-based and stand-alone writing. Most readers can remember the closed-book exams of days past (they can still be found, of course). The contrast with writing tasks in life after school is striking. Seldom does a professional work on a problem (or write about it) with a closed book. Imagine a doctor, about to operate on you, announcing, "This is a closed-book operation!" We also suggest that information be made as public as possible. Walls in today's classrooms are often covered with papers full of notes, graphs, and pictures. What a writer needs most is words. Students will write more compelling and better organized papers when they can lift their eyes to the walls and find the words and phrases that jumpstart them. Those who write for a living depend on this approach, hence the need for *scratch paper* and room to spread it out.

Our high-tech colleagues are curious about how computers are used for writing in the classroom. They notice that students prepare a draft on paper, then use the computer for revision and publication, and ask, "Why don't they write on the computer from the beginning, like I do?" In fact, many of these colleagues rely on paper to get started. Pages can be spread out and scribbled and graphed on. It takes experience and practice with the computer screen to write exclusively on a computer.

Writing can be a lonely task. We have done a lot of writing together, and some of it has been lonely, but the joy of the experience comes from the collaboration, which takes many forms. Our point in mentioning *support and advice* is partly to encourage teachers to provide scaffolding, but the real message for assessment is to provide students with op-

portunities and counsel about how to work together in the construction of compositions.

Finally, the matter of *strategy*—a synopsis of the writing process can be captured by three two-part alliterative phases: (1) *develop* and *draft*, (2) *review* and *revise*, and (3) *polish* and *publish*. Not every writing assessment incorporates all of these elements, but best practices build on this design and lay out the specific assignment scenario for students.

We hope that, in ranging over this array of topics, you will consider the planning and management of a writing assessment as an organic exercise, where you begin with an overarching design but know that, much like an orchestra conductor, your role is not to tell the harpist when to pluck a particular string or the timpanist when to strike the drumhead with a particular force. We hope, in addition, to have sensitized you to the social aspects of writing.

So much for the soft side of assessment and evaluation. What about the bottom line? How should one judge how well the student did? What grade should one assign to the work? Grades have recently been complemented by a new concept, the *rubric*. Advice about appropriate rubrics for writing assessment can be found in a variety of sources (Arter, McTighe, & Guskey, 2001). The primary division is between *holistic* and *analytic*, or *trait*, strategies. In holistic scoring, which dominates large-scale assessment, the rater gives the composition a brief reading (a few minutes at most) and assigns it a single score. Raters undergo intense training for this task, during which they review *anchors*, prototypical papers in each of the score categories. To check consistency, benchmark papers are inserted during the scoring process, and raters are recalibrated as necessary. This process leads to reasonably high inter-rater reliability, which means that judges agree with one another, both overall and in judging individual students. The problem, of course, is knowing exactly what the ratings mean and what to do with the information. The strategy is poorly suited for classroom assessment.

Analytic or trait rubrics (Spandel, 2004) are designed around specific facets of a composition, mostly related to familiar writing features. The most popular system today is the "Six Traits" approach, which encompasses six features, with a plus-one facet added recently:

- *Ideas*: The composition includes a central focus or theme, which is elaborated with relevant details, anecdotes, and similar features.
- *Organization*: The order and layout of the paper are coherent, with a clear sense of direction in communicating the focus or theme.
- *Voice*: There is a sense that the writer is speaking directly to the

reader and communicating a sense of purpose and an awareness of audience.

- *Word choice*: Precision, appropriateness, and richness of vocabulary are present.
- *Sentence fluency*: One sees flow, connectedness, and variety in the construction of sentences; this aspect depends somewhat on grammatical conventions, but is not exactly the same as the next item.
- *Conventions*: Mechanical features, including spelling, grammar, punctuation, and paragraphing, are correct.
- *Presentation*: This new facet covers the appearance of the composition, including handwriting, effective use of layout, illustrations, and so on (partly reflecting the emerging use of computers for polishing a composition).

A wide range of resources is available to illustrate the use of the multitrait system for analytical assessment, including rubrics for each trait, along with examples of student writing that illustrate different levels of accomplishment for each facet. Best practices in writing assessment can clearly build on these features, which prepare the student for the practice of writing over the long run, through high school into college and on to the variety of professions where writing is either central (newspaper reporters, magazine and book editors) or essential (anyone who prepares memos or documents as part of his or her job).

A couple of the traits listed deserve special attention, and a few additional items are worth mentioning. Voice is both important and challenging in classroom writing. For serious writing in college and beyond and for secondary students, a clear sense of purpose and awareness of audience are critical requirements for any writing assignment. Unfortunately, most writing prompts address these two features weakly at best. The audience is either the teacher (implicitly) or an artificial entity ("write a letter to your parents"), and purpose is missing or made up.

All too often, the reality is that an assignment is just an assignment. Under these conditions, expecting students to infuse their composition with personal voice—with an authentic sense of purpose and audience—is unrealistic. An honest voice might lead the student to begin, "I'm writing this paper for Ms. Martin because I have to. I only need a B, so I'm not going to really do my best, but hope this is good enough." This matter can be handled at least partly in two ways. One is to call upon situations *within* the classroom that are as genuine and engaging as possible (some topics are more interesting and personally relevant than others). The other is to look for opportunities *outside* the classroom; with the ar-

rival of the Internet, despite limited access in many schools, students can engage others from around the world in authentic dialogues.

The second point that we think deserves mention is length. Many writing experts are conflicted about this feature; indeed, many think it is a mistake even to mention it. More is certainly not necessarily better, but it should be possible to offer students advice and feedback about "not enough" and, in those rare instances when it becomes a problem, "too much." Teachers routinely include length as part of an assignment (five paragraphs, two pages, and so on). If length is not included, it often matters nonetheless. The fifth grader who hands in three sentences when everyone else is filling a page is likely to receive a low grade, even if the sentences are well crafted and on topic. Dealing with length is a complex matter, and we will not attempt to resolve it here, other than to encourage attention it in assignments. In the elementary grades, "more" is probably a positive outcome, certainly for a first draft. In the later grades, when it becomes critical for students to learn to manage their time, the nature of the assignment may be important. Even here, it is worth remembering that the best predictor of scores on most college entrance writing exams is not the content or style, but rather the length of the essay.

The third point centers on attention to genre, the type of writing called for by the assignment (Schleppegrel, 2004). Distinguishing between narrative and informational writing, between stories and reports, is an important first step, but in our review of the available resources, students could benefit from greater clarity in the distinction and from further distinctions within the informational genre. For example, we have seen rubrics for narrative writing that emphasize topical focus, introduction, conclusion, and so on. This language is a mismatch to the narrative form, which builds on concepts like theme, setting, and resolution (Lukens, 2002). For informational texts, planning a simple descriptive piece around the five-paragraph essay is quite different from laying out a compare-and-contrast analysis, a process explanation, or a persuasive argument (Chambliss & Calfee, 1998).

Our fourth point centers on the content or substance of a composition, which is absent from writing assessment systems with which we are familiar. Content was discussed earlier in contrast to process; for practical purposes, you can think of process as everything that the multitrait models cover. One approach to content uses topic as the criterion. If the assignment is about earthquakes, did the student stick with earthquakes? From one perspective, this point is simple, but staying on topic can take many forms. The student can reproduce material from various sources, from paraphrasing to outright cutting and pasting. A student can develop the topic as a story, recounting his experiences during the San Francisco

Loma Prieta quake. Other students can demonstrate understanding by transforming resources and experiences into a genuine composition—the building of something new from a collection of basic elements. Best practices in writing assessment should distinguish among these activities. Summarizing and note taking are important skills for students to learn to the point of fluency and as a context for practicing conventions. Storytelling is an engaging activity and provides another opportunity for practicing skills. To be sure, relatively few storytelling jobs are available for college graduates, but education should be about more than jobs.

At a practical level, we think that text-based writing assessments should include a rubric that gauges students' capacity to *transform* the substance of the topic. The challenge is to handle this task at a classroom level. Ideally, the teacher has led students through a topic like earthquakes (or a general domain like plate tectonics), and individuals or small groups have conducted additional reading and research, exploring the domain along paths that may be new to the teacher. Teachers cannot be experts on everything; how does one both offer students the freedom to explore and judge the quality and accuracy of diverse reports?

In a project on reading and writing about science (Miller & Calfee, 2004), we and our colleagues spent considerable energy wrestling with such questions. On the one hand, it is important for the student to include fundamental concepts and relations in his or her composition, what we refer to as *schemata* (Anderson, Spiro, & Anderson, 1978), or sets of ideas and words connected in particular ways, which can serve as a template for evaluating presentations. For instance, volcanoes, an engaging topic across the grades, take two wildly different forms in classrooms: (1) the vinegar-and-soda version in which these ingredients, along with red dye, are poured into a clay model to generate an eruptive fizz; and (2) the plate-tectonics account, which suggests that the earth is cracked into great chunks by the roiling of magma, where volcanoes emerge as "blurps," like a kettle of thick pea soup. However a student decides to treat the topic of volcanoes, a composition that captures the scientific content must include the pea-soup model in some form. A favorite example comes from a small-group project by students at a school near Honolulu, who prepared a lengthy project report on the difference between the vinegar-and-soda exercise in the weekly news magazine and what they had learned from reading about the volcanic terrain on which they walked. Their description of the contrast between the two models provided transformational evidence of deep learning.

Pursuing this assessment route poses a number of challenges, not the least of which is the demand on the teacher to read individual compositions thoughtfully. Addressing this matter can be tough, but here are a few words of advice. The first builds on the potential for computer-

based text analysis to do some heavy lifting here. Programs like the Intelligent Essay Assessor (Ericsson & Haswell, 2006; Shermis & Daniels, 2003), while still in the prototype stage, provide students and teachers with a quick evaluation of the substance of a composition, including an analysis of the match to the content schemas. Which critical concepts and relations from the topic are found in the composition, and which are missing? If the idea seems far-fetched, remember that, not too long ago, calculating the readability of a particular passage required a fair amount of work.

The second point, one that applies to all facets of composition assessment, is the suggestion to give away grading—indeed, the entire writing system—to students. The teacher can accomplish this goal in a variety of ways, including cooperative learning. What could be more sensible than teaching students to collaborate on projects, including writing tasks, and to learn to critique their own work? The main advice here centers on teaching students about the concepts of genre, traits, and rubrics. This strategy brings at least two clear advantages. One is that students become independent learners in the fullest sense, responsible for handling all facets of communicating their mastery of a topic. The other is that the teacher no longer bears the entire responsibility for student learning; in particular, he or she does not have to read and review every piece of student writing in detail. Rather, his or her task is to monitor and discuss the students' reading of their work.

Two potential problems emerge from this advice. First, what if students do it wrong or cheat? Second, might it not be easier for the teacher to do the work than to spend the energy needed to teach students how to handle the task? The response to both questions is the same: Teaching students to become independent and responsible learners is difficult, but addressing this challenge is critical for reform of schooling in our country. These issues emerge with special clarity for content-area writing. When the conditions are right, writing reveals thinking with unusual clarity. The results show up partly as scores and grades, but more important are the portraits that students construct in demonstrating their understanding of a topic, revealing their capacity to "go beyond the information given" (Bruner, 1973).

MEETING THE CHALLENGE

Basic skills in reading and mathematics have taken center stage in recent years. Without questioning the need for attention to these fundamental areas, we believe that improving educational outcomes for students in the United States will require engagement with the challenges of helping

students become proficient writers in the content areas. The principle here is that writing reveals thinking, that the capacity to lay out one's understanding in a clear, organized, and compelling fashion is an essential communication tool.

Effective and efficient formative assessment is a key consideration for this task. Writing assignments provide only limited benefit unless accompanied by informed and informative feedback on both the process and content of the compositions. The techniques for planning and conducting such assessments are not mysterious; we know a great deal about best practices from both research and practical experience.

It is tempting to try to assemble these techniques as a set of blackline masters, along with simple formulas to guide teachers and students. In reality, these accomplishments depend on classroom teachers who possess appropriate knowledge and skill for developing and implementing the best writing assessments that we can imagine from current research and practice. They need opportunities for professional dialogue around these matters, much as physicians have opportunities for medical rounds (time when they can discuss cases), and they need the benefit of institutional support that recognizes the validity of assessments that are grounded in genuine performance activities.

ACKNOWLEDGMENT

REFERENCES

Anderson, R. C., Spiro, R. J., & Anderson, M. C. (1978). Schemata as scaffolding for the representation of information in connected discourse. *American Educational Research Journal, 15*, 433–440.

Arter, J. A., McTighe, J., & Guskey, T. R. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance.* Newbury Park, CA: Sage.

Black, P., Harris, C., Lee, C., Marshall, B., & William, D. (2003). *Assessment for learning: Putting it into practice.* Buckingham, UK: Open University Press.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning.* New York: McGraw-Hill.

Bruner, J. S. (1973) *Beyond the information given: Studies in the psychology of knowing.* Oxford, UK: Norton.

Calfee, R. C. (1997). Assessing development and learning over time. In J. Flood, S. B. Heath, & D. Lapp (Eds.), *Handbook for literacy educators: Research on*

teaching the communicative and visual arts (pp. 144–166). New York: Macmillan.

Calfee, R. C., & Hiebert, E. H. (1990). 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.

Calfee, R. C., & Miller, R. G. (2005). Breaking ground: Constructing authentic reading–writing assessments for middle and secondary students. In R. Indrisano & J. Paratore (Eds.), *Learning to write, writing to learn: Theory and research in practice* (pp. 203–219). Newark, DE: International Reading Association.

Chall, J. S. (1995). *Stages of reading development.* New York: Harcourt College Publishers.

Chambliss, M. J., & Calfee, R. C. (1998). *Textbooks for learning: Nurturing children's minds.* Oxford, UK: Blackwell.

Chappuis, S., Stiggins, R. J., Arter, J., & Chappuis, J. (2005). *Assessment for learning: An action guide for school leaders.* Portland, OR: Assessment Training Institute.

The College Board (2006). *The keys to effective writing.* Retrieved March 13, 2006, from *www.collegeboard.com*

Ericsson, P. F., & Haswell, R. (2006). *Machine scoring of student essays: Truth and consequences.* Logan: Utah State University Press.

Freedman, S. W. (1997). *Exchanging writing, exchanging cultures.* Cambridge, MA: Harvard University Press.

Gray, J. R. (2000). *Teachers at the center.* Berkeley, CA: National Writing Project.

Harp, B. (2006). *The handbook of literacy assessment and evaluation.* Norwood, MA: Christopher-Gordon.

Heath, S. B. (1983). *Ways with words.* New York: Cambridge University Press.

Lukens, R. J. (2002). *Critical handbook of children's literature.* Boston: Allyn & Bacon.

Miller, R. G., & Calfee, R. C. (2004). Making thinking visible: A method to encourage science writing in upper elementary grades. *Science and Children,* 42(3), 20–25.

Schleppegrel, M. (2004). *The language of schooling.* Mahwah, NJ: Erlbaum.

Schneps, M. H. (1989). *A private universe.* Cambridge, MA: Project STAR, Harvard University.

Shermis, M., & Daniels, K. E. (Eds.). (2003). *Automated essay scoring: A cross-disciplinary perspective.* Mahwah, NJ: Erlbaum.

Spandel, V. (2004). *Creating writers through 6-trait writing assessment and instruction.* Upper Saddle River, NJ: Pearson Education.

Stiggins, R. J. (2004). *Student-involved assessment for learning.* Upper Saddle River, NJ: Prentice Hall.

Urquhart, V., & McIver, M. (2005). *Teaching writing in the content areas.* Alexandria, VA: Association for Supervision and Curriculum Development.