Accommodating Complexity: WAC Program Evaluation in the Age of Accountability

WILLIAM CONDON
Washington State University

*A*ssessment. Accountability. These two closely related words are sufficient in and of themselves to chill the blood and roll the eyes of those who manage writing programs in general and writing-across-the-curriculum (WAC) programs in particular. Assessing writing is a complicated task made increasingly complex by teachers' desire for assessments that support instruction and, on the other end of the spectrum, the public's demand for *proof* of effective instruction. Once upon a time, the evaluation of writing seemed deceptively simple. Either a writing teacher in a single classroom applied what looked like his or her subjective judgment to each student's written products, or large numbers of students sat for multiple-choice-question (MCQ) tests that, the psychometricians assured us, measured verbal abilities indirectly, including the ability to write. Such assertions were always suspect since even the most basic common sense tells us that in order to assess writing ability, we must look at direct measures—at writing—rather than at bubbles filled in on an answer sheet. And so, over the years, writing teachers have led the way in establishing direct tests of writing (White; Morris) and, following the same impetus, portfolio-based writing assessment (Belanoff and Elbow; Belanoff and Dickson).

Each of these developments has accommodated the complexity involved in assessing writing ability, and each cycle of reform has produced a more complex, less positivist methodology for writing assessment—increasing the validity of the instrument while

at the same time satisfying the psychometricians' criteria of reliability (Hamp-Lyons and Condon, *Assessing*; LeMahieu, Gitomer, and Eresh). Implicit in that evolution is the sense that writing itself seems more consistent with chaos theory than with the epistemology expressed in positivism. The idea of the "butterfly effect" perhaps best captures the impact of a given class or learning experience on a given student. Changes in writing ability depend, we know, on a dizzying array of factors—among others, the student's readiness, openness, and willingness; the teacher's careful planning, theoretical and pedagogical knowledge, good timing, and even showmanship; and careful design of and timing in the curriculum. Even then, even when these complicated factors come together in what we've come to call a "teachable moment," the effects may take years to manifest—at which point, they are hard to connect with any single class, teacher, or learning experience. In sum, the more we learn about the enterprise of writing and about the enterprise of *teaching* writing, the more complicated the task of teaching writing seems. And as difficult as the teaching of writing is, assessing writing involves yet another layer of difficulty.

How much more complicated, then, is the enterprise of assessing writing across the curriculum? How much more complex is the activity of evaluating WAC programs? I imagine here a set of nesting eggs, one inside the other. The expanding layers represent the stakeholders in writing—students as the center egg, then faculty, administrators, parents, politicians, the public at large. Each has a different set of questions. Each wants some return on investment. Each larger egg involves more people and therefore carries a broader context and an expanding set of stakes. Each larger egg represents one more level of difficulty above the difficulty of "simply" assessing writing ability. Thus, each larger egg comes with a worrisome combination of greater complexity and higher stakes.

In the face of these higher stakes, we are also hampered by the failure of traditional measurement tools and the emergent, experimental nature of newer and better tools. We may have come a long way since the development of direct tests of writing in the late 1960s and early 1970s, but the newer tools that have proven effective for measuring writing ability are still extremely limited

in use and scope. Even the oldest of the improved methods, direct tests of writing—typically a timed writing holistically scored—are still only second to MCQ tests in frequency of use. Direct tests, of course, provide greater validity because they actually focus on a sample of what they purport to measure—writing. But their limitation to only one sample, collected under only one set of writing conditions—and that set itself the most constrained and unrealistic of all conditions under which people write—means that direct tests as well are able to answer only fairly simple, straightforward questions about a student's writing ability (such as whether that student is ready for the standard course in first-year composition or needs more practice first). Often, as Edward White points out, such questions depend on the most basic of writing skills—the ability to write consistently in complete sentences, or to use sentence-level punctuation correctly, or to arrange a short essay effectively into paragraphs (10–16).

Roberta Camp argues convincingly that we need more robust kinds of writing assessments, assessments that can answer more complicated and sophisticated questions about students' writing competencies ("Changing," "New Views"). Portfolios have begun to provide such assessments, but employing portfolios to provide reliable judgments involves levels of logistical and intellectual complication that sometimes stagger the teachers and administrators involved in the effort (Hamp-Lyons and Condon, "Questioning," *Assessing*). The experience can be so daunting that the group who developed Washington State's University Writing Portfolio assessment program compared their experience to "shooting Niagara" (Haswell, Wyche-Smith, and Johnson-Shull). Even strong advocates of the portfolio method have compared the experience to jumping off a cliff (Condon and Hamp-Lyons). Writing assessment, then, has become a much more complicated affair than it once was; in the attempt to measure more, and to make measures as fair and as accurate as possible, assessments themselves have become progressively more difficult to develop and to manage.

Emergent tools, greater complexity, higher risk. Each expansion of the audience outward raises the stakes, demands an accounting, affects the budget. Each audience for the evaluation comes to it from a different vantage point and looks for the evalu-

ation to meet different needs. Inside the program, so to speak, students and faculty have needs closely related to instruction. As we move up the hierarchy, further away from the classroom, evaluation gradually but inexorably turns into accountability—into the ability to document a program's effectiveness, to lay out the benefits it offers to different stakeholders, and to justify a program's existence or continued growth.

Internal audiences—students and faculty—have perhaps the most direct need for information about performance. Students want to know how well they are doing, of course, and they want to know at times and in ways that often do not fit within the traditional course and grading agenda. What does a particular grade in a composition course or a writing intensive course mean in terms of overall writing ability? How does it predict the usefulness of that level of writing ability as students approach writing assignments in other courses—whether WAC or "regular"? Should they feel satisfied with their current level of ability, or should they devote significant amounts of time and tuition money to further development? What will the curriculum demand of their writing, and how well prepared are they to meet those demands? How will the lessons they learn about writing in their chosen fields help them after graduation? These are just a few of the easier questions students bring to this assessment arena. Some of these questions are shared by those who teach the students, both in writing courses and in other courses in the curriculum. Teachers—WAC faculty in particular—need to know what they can reasonably expect students to be able to do with and in writing, and they need to match those expectations with the level of expectations that are implicit in the teachers' own course objectives, objectives which, in turn, are determined by their location within the curriculum. Teachers need to know how to build more effective assignments—knowledge that involves both information about the writing students will do after taking a particular course (in careers or in subsequent courses) and information about the writing students have done to that point in the university's curriculum.

Audiences external to the WAC program want to know how well it works, but they want that information for varying reasons. University administrators, as Haswell and McLeod have

pointed out, need information that can educate them about WAC in general and that can demonstrate the program's effectiveness, its impact on students, on faculty, and on learning in specific and in general. Beyond the academy, parents, legislators, employers, and the public in general want *results*—graduates who can write effectively as they enter their careers (our students, too, as they become our alumnae/i, share this need). At this level, evaluation works in the service of accountability, and as the scope for evaluation moves outward, evaluation becomes more and more involved with the overall accountability of the institution. Thus, WAC data can—and should—figure prominently in university accreditation; they should provide administrators with evidence that legislators can understand, evidence that documents the institution's efforts to provide more effective, more responsive learning opportunities for its students.

Each of these levels, each of these audiences, has complex needs that go far beyond the information we can gather by merely assessing students' writing. To date, WAC programs have done a poor job of addressing most of these audiences. Understandably, since WAC has been primarily a faculty development movement, program evaluation has focused on the effectiveness of those efforts. Even here, though, the results have been mixed. Fulwiler and Young admit that their early efforts at WAC evaluation led to the realization that they needed better assessment tools (2). As time passes, our efforts in this area are producing more useful results (Walvoord et al.). Still, the literature about WAC is only beginning to address questions that extend beyond the effectiveness of faculty seminars. In the latest—and to my mind the best—collection of essays about evaluating WAC, Kathleen Blake Yancey and Brian Huot's *Assessing Writing Across the Curriculum,* only two selections (Beason and Darrow; Kinkead) address student outcomes from WAC, and only two others (Morgan; Haswell and McLeod) address the administrative audience for WAC evaluation. The rest address the "same-old, same-old" issues that surround faculty development *qua* curriculum reform. We have to do more. We must do better.

As difficult as the problems are, as complex as they have become, the solutions involve, in effect, treating that complexity as an advantage. As long as we fall for the positivists' notion that

the way to measure a complex construct is to reduce it to its simplest components and then measure each of those independently of the others, we will be unable to measure a construct as complicated as writing—and seriously at sea trying to measure the even more complex effects of a WAC program. If we make the complexity of the task clear to all, however, and if we resist the urge to oversimplify, we can open up space to explore new methods of assessment and evaluation, methods that promise to contribute to a fuller understanding of what happens throughout our WAC programs.

The very fact that we can frame the issue in these terms is an indication of how markedly assessment has changed in the last decade. For many years, Edward White's maxim, "Assess yourself, or others will do it for (to) you" represented not only good advice, but also current practice. Assessments were enforced from outside the instructional context; and those assessments were generally hostile to instruction—reducing learning to a set of questionable skills, ignoring local curricular goals and objectives in favor of some putative national norm, taking major amounts of time away from instruction so that students could learn effective test-taking strategies, and so forth (Smith and Rottenberg). In the face of such a threat, White's advice made sense. But the necessity of such advice is part of the reason we tend to dread assessment. It was the devil we knew, and we used it to fend off the devil we didn't know.

Changing the Paradigm for Assessment

Today, however, positivist models of assessment are giving way to constructivist models, local assessments based in inquiry and collaborative investigation rather than outside assessments delegated to national testing companies or experts in psychometrics. This newer model, pioneered by Guba and Lincoln, engages all the assessment stakeholders in designing the evaluation, carrying out the methodology, examining the results, and formulating responses to the findings. The constructivist model mirrors the research process that is already a fact of life for college faculty. As a result, assessment seems far less threatening and myste-

rious than it was in the days when psychometricians controlled the processes. Today, we have better reasons to assess, and we have tools for evaluation that are far more familiar to us, tools we can control. Thus, we can respond to the task of assessment without feeling defensive in the way White's maxim implies; instead, we can welcome assessment as a process that helps us achieve goals that are important to us.

The first step in a constructivist evaluation is to involve stakeholders in setting goals and objectives for the evaluation that are as close as possible to the goals and objectives for the activity being evaluated. If a WAC program seeks to increase the amount of writing students do as they move through the curriculum, then a constructivist evaluation would seek to discover whether, in fact, students write more than they did before the implementation of WAC. If the program objectives involve helping students become better writers, then the constructivist evaluation entails collaboratively defining "better" and determining the best ways of discovering whether students are better writers as a result of the program's efforts. (For an excellent example of how to document improvement in student writing, see Haswell, "Documenting.") Any and all measures in this process come out of the local context for the evaluation, and data collecting is designed to be as nonintrusive as possible. So, rather than require students in a course to sit for a timed writing that at best is only tenuously related to their curriculum, data collection would entail looking at the products of their class work—at the instructional outcomes they would have produced anyway. An added benefit, of course, is that these outcomes flow directly from the instructional objectives the teacher sets in designing the course in the first place, completing the constructivist cycle in such a way that the feedback from one iteration of the evaluation acts more as feed-*forward*, since its most immediate use is in improving instruction in the next iteration. This emphasis on engaging assessment with instruction in order to improve instruction first and *then* supply data for accountability to audiences outside the classroom fits well with learners' and teachers' needs—in addition, the priorities inherent in the process are more consistent with an educational process.

The proof of that last assertion lies in the literature that has grown out of the constructivist paradigm. Thomas Angelo and Patricia Cross's *Classroom Assessment Techniques*, for example, contains example after example of evaluations that grow out of and in turn support improvement in classroom instruction. Similarly, Banta's *Making a Difference: Outcomes of a Decade of Assessment in Higher Education* provides examples of program evaluations that examine instructional outcomes in order both to improve instruction and to provide data that are useful in establishing a program's accountability. Indeed, if programs are to survive the sheer weight of the demands for assessment and accountability, then programs must develop means of evaluation that focus first on improving learning and then, by extension, on being accountable to administrators, parents, the public, and the legislature. Washington State University's experience provides two useful examples of this kind of evaluation.[1]

Each year, incoming first-year students sit for the Writing Placement Exam (WPE), which determines whether a given student needs extra assistance in English 101, WSU's first-year composition course. Those who need help—about 14 percent of the incoming class—add a weekly small-group writing tutorial, English 102, to their English 101 enrollment. Later, as rising juniors, these students complete the University Writing Portfolio, a midcareer assessment of their writing. At that point, these students, whose WPE's placed them in the bottom third of entering students with regard to writing ability, perform almost identically to their classmates whose WPE's had indicated they did not need the extra assistance provided in English 102:

◆ Of 2,130 students who placed into English 101, 192 (9 percent) received a "Needs Work" rating on the Junior Writing Portfolio.

◆ Of 356 students who placed into English 101 + 102, 39 (11 percent) received a "Needs Work" rating on the Junior Writing Portfolio.

As these percentages demonstrate, the difference at the junior level is insignificant. Only 9 percent of the students who enter as competent writers evince a need for additional assistance as they

enter their upper-division course (two of which will be Writing in the Major, or WID, courses). By comparison, only 11 percent of those who had been weaker writers at entry still occupy that niche. Does the peer-facilitated small-group tutorial help students improve their writing abilities? The figures indicate that the program works—feedback that was important to those involved in the instruction, but that proved equally impressive when cited to the provost, the Board of Regents, and the Higher Education Coordinating Board.

A second example also derives from the University Writing Portfolio. Ongoing assessment—in this case the biennial portfolio study for 1995–97—revealed a problem: among WSU's transfer students, more than 37 percent of non-native speakers (NNS) of English received a "Needs Work" rating on the portfolio—a rate more than three times higher than for the student body as a whole, which was 11 percent (Bonnema, Haswell, and Norris). Although we might expect students whose native language is not English to have a somewhat higher "Needs Work" rate than native speakers of English, 37 percent seemed far too high. Clearly, these students' needs were not being met. In examining the reasons for the high rate, we discovered that many of these students made poor selections of writing to include in their portfolios. Therefore, we changed how we work with these students so that they provide longer, more complex samples of their class work. In addition, the English department completely reformed its ESL offerings, building the portfolio process as a classroom assessment tool into English 105 (the equivalent of English 101 for non-native speakers) and changing that course so that it more completely parallels 101. Thus, more of our NNS transfer students take English 105 at WSU, rather than taking 101 elsewhere and transferring the credit. One result of these changes is that during the 1997–99 biennium, the rate at which NNS transfer students received a "Needs Work" rating dropped to 27 percent. We suspect this figure is still too high, but it represents good progress—and fast progress—in both formally and informally accommodating the instructional needs of these students.

These two examples lead to the next important reason that WAC programs should perform their own evaluations: we do it better. Indeed, in what Kathleen Yancey has called a "third wave"

of assessment (491), all good assessment, like politics, is local. The constructivist paradigm takes advantage of access to local contexts—to curriculum, faculty, administrators, students, institutional values, etc.—in order to increase the evaluation's usefulness by increasing its relevance to the local context. Positivist methodologies tend to distance evaluation from the local context not only by employing outside experts to perform the evaluation, but also by using standard methodologies rather than developing methods that fit the context of the program being evaluated. The results are often disastrous, as even a casual perusal of Stephen Jay Gould's *The Mismeasure of Man* reveals. As I argued earlier, indirect tests of writing are prima facie invalid—lacking construct validity in particular—because they do not involve looking at even a small sample of the construct being assessed (writing). In such cases, mismeasurement is a foregone conclusion. Worse, since positivist models most often culminate in statistics, they report numerical measurements, which are too often subject to misuse. Both Educational Testing Service (ETS) and ACT administrators assert forcefully, for example, that the SAT and ACT should not be used for purposes of writing placement, yet these tests are routinely used for that purpose. Finally, indirect tests lack universal fairness. Again, the statistics and reports from the agencies themselves—from ACT and ETS—indicate unintended yet marked differences in performance by race, ethnicity, and culture. Put simply, these tests discriminate in favor of white, middle- and upper-class, urban and suburban students. By contrast, White and Thomas found that direct tests of writing resulted in fairer outcomes (186–87), and newer methods such as performance assessments clearly provide fairer opportunities for students to establish their competencies (Hamp-Lyons and Condon, *Assessing*). I will not argue that local assessments are free from problems, only that assessments designed locally to address local initiatives and contexts are more likely to portray those contexts accurately and treat the stakeholders fairly than are large-scale state, regional, or national assessments which are much more likely, of necessity, to use positivist methodologies.

Finally, by assessing locally we can develop strong ties with other units within our institutions whose missions affect WAC

or whose philosophies are similar. The statistics on the success of WSU's English 102 program inspired the Student Advising and Learning Center (SALC) as that unit designed its Freshman Seminar Program—small, peer-facilitated courses that support students' learning and promote coherence among the students' classroom experiences. This program has proven highly effective (Henscheid), and the collaborations between SALC and the Campus Writing Programs continue to provide exciting opportunities for teaching and learning. Similarly, the Writing Programs' promotion of active learning, alternative assessments, and critical thinking has resulted in a natural partnership with WSU's Center for Teaching, Learning, and Technology (CTLT). CTLT, in turn, provides vital expertise as Writing Programs faculty search for better ways to measure the effectiveness of various programs. Most recently, students in a graduate seminar on writing assessment helped devise a rubric for measuring critical thinking. CTLT then "adopted" the rubric, conducting vital work in validating the rubric and giving it wider trials by encouraging faculty in general education courses to use it as a measure of instructional effectiveness. As a result, the director of general education, the senior fellow of CTLT (a faculty member), and the director of Campus Writing Programs just received a grant to do further work on the rubric—a grant that came from Washington's Higher Education Coordinating Board. Collaborations that grow out of evaluation activities lead to useful and beneficial relationships that reflect credit on all the partners to the collaboration. No positivist evaluation that I know of provides this kind of payoff.

In the end, what do these new methods mean for evaluating WAC programs? Clearly, no single form of assessment will give us all the information we seek. Just as clearly, collecting data will involve moving beyond traditional forms of writing assessment and research. The benefits of moving into new methods for evaluation, however, are substantial. Constructivist methods engage as wide a range of stakeholders as possible in order to frame questions, set goals, and devise methodologies—providing a context within which vital collaborative relationships are established and nourished. In addition, these evaluations yield richer sets of information and outcomes, so that improving WAC programs becomes easier, if only because the arguments for improvement

are based on rich, convincing data. Finally, the constructivist paradigm recognizes that evaluation and improvement constitute a continual cycle—the bad news is that assessment becomes a constant, continuing activity, but the good news is that improvement also becomes a constant, continuing activity. By paying attention to the stakes and the stakeholders, by using multiple methodologies, by exploiting the relationships between WAC and other university initiatives, and by breaking down larger questions about effectiveness into smaller, more easily addressed questions, we can make significant progress toward accommodating the level of complexity that WAC programs inevitably face, not just in evaluation but in the very act of addressing their missions.

Implications

Accommodating complexity begins when we extend current efforts at WAC evaluation to include a wider range of stakeholders. While WAC has focused on faculty practice—on reforming pedagogy—its primary effect is on the students whose learning is affected by that pedagogy. We need to know more about the effects of WAC courses on students' writing. We can measure some of these effects by assessing the writing students do, of course. Measuring writing competencies at entry and at several points along students' college careers allows us to make some statements about the impact a WAC curriculum has on the quality of students' writing. Two examples suffice to demonstrate ways in which such assessments might prove useful as WAC evaluations.

The first example is drawn from Washington State University, where (as mentioned earlier) entering first-year students write two timed essays in a single two-hour sitting in order to help faculty place the students into the appropriate first-year composition course. Then, at the junior level, students sit for another timed writing, identical in format to the Writing Placement Exam, and this writing, along with three essays written in other classes, makes up a University Writing Portfolio, which serves as a qualifying exam for WSU's Writing in the Major courses. Thus, students entering their major concentrations receive feedback on their writing, and those who need additional assistance with writ-

ing are identified and guided into an appropriate level of assistance. One way in which these assessment experiences have proven useful is the comparisons we can make between the entry-level Writing Placement Exam and the junior-level Portfolio Timed Writing: in a preliminary study of students who, by chance, wrote on similar topics at the two levels, Rich Haswell discovered several areas in which students had gained ground in writing as they moved through WSU's writing-rich general education program. This "value-added" form of assessment allows us to document the fact that widespread learning is occurring in the curriculum, that students are becoming better writers in general and along specific dimensions such as organization, focus, use of support, style, mechanics, and so forth. The study also allows Haswell to argue that students become more *efficient* writers, since juniors wrote longer sentences, longer paragraphs, and more words on the same task they performed as first-year students, even though the juniors had half an hour less time to write than they did as first-year students (Haswell, "Preliminary Results"). This evidence proved extremely useful to the institution as it prepared for its ten-year accreditation process, and the final study has become part of WSU's reports to Washington's Higher Education Coordinating Board (HECB), the body that mediates between the universities and the legislature. Thus, a report which assures inside stakeholders that WSU's WAC program is having positive effects also serves to help WSU's administration argue in several critical venues that the institution is doing its job vis-à-vis writing instruction. (A fuller version of this study can be found in Haswell, "Documenting.")

The second example speaks more to external audiences and is happening as I write. In response to the legislature's call for "performance measures" that can be used to evaluate the effectiveness of Washington's six four-year universities, the six provosts asked the universities' assessment officers and writing programs staff to attempt to develop a writing assessment that could serve as an accountability measure—in other words, an assessment that the universities could ask the legislature to use in allocating state higher education funds. Obviously, if the plan works, this will be a high-stakes assessment for the institutions. They will be able to choose writing as a performance measure

and to stake some portion of their state funding on being able to improve their students' performance on the assessment. The plan, as drawn up by Gerald Gillmore, the University of Washington's assessment officer, calls for collecting, at random, ten papers from each of ten senior-level classes (from ten different departments) on each of the six university campuses. These six hundred papers would be evaluated each summer by a combination of (1) writing programs faculty, (2) faculty in the disciplines from which the papers were selected, and (3) community members who are in careers in those disciplines. A university's performance would be measured by the proportion of students achieving a score of "Acceptable" in six categories of performance: Content, Organization, Reasoning, Rhetoric, Conventions, and Disciplinarity. In the pilot study conducted during the summer of 1998, an interdisciplinary team of writing specialists, assessment specialists, faculty from several other disciplines, and community members from those same disciplines was able to derive a scoring rubric and rate sample essays (see Gillmore). A second pilot in the summer of 1999 engaged a similar group of raters who evaluated approximately sixty essays and, using the rubric, achieved reliability of .79 on their overall scores.

This project needs further development before it can fulfill its promise as a fair, nonreductive, rigorous assessment of writing that can provide the major stakeholders (the universities and the legislature) with the information they need to make the decisions each must make. This feedback would be useful to the institutions as they focus their efforts on improvement: if their students do well in "Content" and "Conventions" but poorly in "Reasoning" and "Rhetoric," then they can focus their efforts on the latter two categories. If students perform well, for example, in all the categories except "Disciplinarity," then the university might enact some kind of WID program, along with examining the other ways in which departments prepare their majors to enter the discourse community of a given field. In turn, the legislature would tie some portion of the university's funding to the institution's ability to increase the proportion of students receiving scores of "Acceptable" in specific categories or across all the categories.

In both these examples, common assessment tools (performance assessments and direct tests of writing that use timed writings) are focused on particular questions that institutions must ask about WAC performance. From the far too general question, "Is WAC working?" we draw less often included questions: "Are students making progress as writers?" and "Are students able to perform satisfactorily as writers by the time they graduate?" These questions, together with others we might ask, begin to paint a more complete picture of students' progress and competencies as writers moving through the curriculum. In turn, that picture contributes to an overall understanding of how well the WAC program is working.

Clearly, evaluations of this sort cannot rely principally on assessments that are separated from the curriculum. For one thing, such assessments involve students in tasks in which they have no real stake. Even if we could require all students to complete a writing assessment task—say, a timed writing—what assurance can we have that students are motivated to do their best? Such tasks are typically barrier tests, requiring only a competent performance in order for the student to pass. If we are to gain an accurate picture of students' writing abilities, WAC measures need to come from high-stakes performances, from products of the students' degree work. Unless we collect samples of such high-stakes work, we cannot be certain that our portrait of abilities will be accurate or that it will serve us as we look to improve our performance as a whole.

Collecting actual classroom performances has another advantage, one so valuable that it is reason enough for collecting such samples. The opportunity to involve assessment with instruction provides many chances to examine students' progress, provide assistance to those who demonstrate a need for assistance, and target faculty development and curriculum reform wisely. In order to provide such rich and varied feedback and to perform the kinds of program evaluation that such data allow, we need to tap into instruction. To date, WAC evaluators' best and most prominent efforts have focused on what happens to faculty and to the courses and materials they design (Walvoord and McCarthy; Walvoord et al.). As important as that kind of

evaluation is, its limits in today's context are clear: such evaluations stop far too short of outcomes, of the effects that a WAC curriculum has on students' development as writers. If we are to examine this crucial area, then we need to use performance assessments. We need to collect students' performances on the assignments they fulfill within the WAC curriculum, and we need to examine those performances in ways that help us identify the outcomes of our WAC programs.

To some extent, this kind of assessment is already being done at a handful of small liberal arts colleges around the country, the most prominent example of which is Alverno College in Milwaukee. Students at these schools keep what is being called a "developmental portfolio," a record of their progress toward the objectives the school sets in its curriculum. In smaller schools, the logistics of such assessments are less daunting than they are at large schools, and the lines of communication needed among faculty and students are easier to build and maintain. Yet this kind of performance assessment, in some form or another, is the only tool we have that allows us to collect the data we really need to examine. In the past—and still today in large-scale assessments such as ACT, SAT, NAEP, CAT, etc.—testing and evaluation have been separate from the curriculum. Tests and other tasks were set without regard to what happened in a particular school or a particular classroom. And, as often happened, if a school set curricular goals that were significantly different from the goals assumed by the test makers, that school's students would not perform well on the test. That is the stranglehold that large-scale assessment has on our K–12 school system. If we are to avoid a similar stranglehold on higher education, then we have to find ways to evaluate students' work as they try to achieve the goals our institutions set for them. We have to find ways, no matter the scale of our operations, to collect information on how well our curriculum is serving our students. Robust performance assessments, as the smaller schools have shown, can provide to outside stakeholders rich, credible, convincing data on performance; and it can provide, to internal stakeholders, evaluations of curriculum and pedagogy that can focus efforts to enhance and extend learning.

Performance assessments, however, are invasive assessments. They get into the curriculum and into the classroom in ways that standardized testing or even direct tests of writing do not. Unlike previous methods and tools for assessment, performance assessment looks directly at the responses students make to the tasks they are assigned in a class or set of classes. Therefore, these actual tasks are exposed to view, along with the teaching methods and practices that frame those tasks. Many faculty members find this prospect worrisome, for obvious reasons. Once the lens of an evaluation is focused on the classroom, how can the faculty member avoid losing control of the assessment and perhaps being unjustly taken to task by its results? The invasive quality of these assessments necessitates an evaluation mindset on the part of the institution that makes every effort to include as many stakeholders as possible. If we listen to external stakeholders, we must look at outcomes; we must be able to show and explain the effects our curricula are having on the learners in our classrooms. As important as these external audiences are, however, we cannot focus on them alone. We must also see to the needs of internal stakeholders—primarily faculty and students—if the evaluation of anything as complex as a WAC program is to be effective. Evaluation cannot be something done *to* faculty and students; it must be something in which they participate—knowingly, at least, if not always willingly.

One way to recruit faculty and students as willing participants is to involve WAC with larger learning outcomes. On an institutional level, this can mean integrating WAC with assessment programs, for example, or developing WAC within the context of a writing-rich general education program. Both these strategies have benefited WAC at WSU, where the University Writing Portfolio, a junior-level assessment, provides data that help evaluate the university's newly revised general education curriculum. The portfolio also acts as a qualifying exam for the two upper-division Writing in the Major (M) courses that each student must take. The position of the portfolio serves WAC in two significant ways: (1) because students must have writing to incorporate into a portfolio, assessment provides an incentive to create and maintain a writing-rich general education curriculum; and (2) because the portfolio identifies students who need fur-

ther assistance with writing—and because it requires that they get that extra assistance—the portfolio acts as a strong support for the upper-division writing-in-the-disciplines curriculum. Portfolio raters are drawn from faculty who teach M courses since these faculty are the ones setting expectations for students' writing at that level. Thus, these programs serve larger agendas—curriculum reform, faculty development, definition of standards, etc.—and so WAC can participate in the evaluations of these other programs. When the Writing Assessment Office carries out its biennial self-study, it provides information, for example, about students' writing experiences in general education and M courses, about the number of faculty who are assigning writing, and about particular populations such as non-native speakers of English or students in different programs or departments. This information in turn acts as one component of WAC evaluation. Similarly, when general education undergoes its regular evaluations, it yields data about WAC at the lower-division level. In other words, much of the process of evaluating WAC at WSU takes place in ongoing evaluations of programs with which WAC is imbricated. To an extent, then, WAC evaluation involves refocusing and reinterpreting portions of these other programs' data. And, in true symbiotic fashion, the data generated in evaluations focused directly on WAC are useful to these other programs as well.

WAC can also serve individual faculty agendas. For example, results from surveys of WSU faculty who have taken an online version of Angelo and Cross's teaching goals inventory and their students who have taken a corresponding learning goals inventory indicate that faculty and students set a high priority on improving higher-order thinking skills. We know, too, that writing assignments promote higher-order thinking skills (see the preliminary report of the National Center for Higher Education Management Systems). Therefore, the results from the teaching goals and learning goals inventories allow us to join our WAC efforts to a goal that both faculty and students have identified as one of their most important. In this simple example, the WAC program can be framed as serving objectives that are important to both faculty and students. By turning the lens of performance assessment on course curricula and writing assignments (which we might describe as faculty outcomes in a WAC course), we can

help faculty improve their course performances by helping them develop assignments that more closely address course goals and objectives (or, in an earlier step, by prompting faculty to develop explicit course goals and objectives).

In a related example, many schools—WSU among them—are busily developing a variety of online learning environments that can serve as extensions of onsite courses and as the environment for distance education. To the extent that these environments allow interaction between students and between students and teachers, they also create a context within which WAC can flourish (see Reiss and Young, Chapter 3, this volume). Online learning environments that foster interaction between students and faculty are, as I have argued elsewhere, *written* classrooms (Condon). Because the primary means for interaction is writing, these environments incorporate writing into the learning experience more thoroughly than any purely onsite classroom can hope to do. As institutions across the country move online, WAC programs need to engage in that effort and participate in the kinds of evaluation that will of necessity accompany these new initiatives (Bober). WAC programs can save money, effort, and time by incorporating WAC into these new environments so that when the environments are evaluated, so is WAC—at least to the extent of its presence in the online environment.

The efforts mentioned in this essay by no means exhaust the possibilities for WAC evaluation. They do, however, begin to lay out an agenda that can lead us productively beyond the current state of the art, which focuses almost solely on faculty development. If we are to demonstrate the ways in which WAC serves its many stakeholders, we must move evaluation beyond current efforts, and in ways already being suggested by Gail Hughes (170–73) and Kathleen Yancey and Brian Huot (7–15). That is, we must begin to employ multiple measures, some quantitative and some qualitative; and we must engage more of WAC's stakeholders both as planners of the evaluation and as audiences for it. Basically, I want to suggest three major directions for WAC evaluation:

1. Using Guba and Lincoln's *Fourth Generation Evaluation* as a touchstone, we need to ask our stakeholders—both within

and outside the university—what outcomes we ought to evaluate when we look at our WAC programs and what interests such an evaluation needs to serve. We also need to involve those stakeholders—students, faculty, administrators, parents, higher education coordinating boards, the public in general— in designing the methods and the specific strategies and tactics we will use to evaluate our programs. This process helps ensure that our WAC curricula serve students' interests, that WAC supports the teaching and learning goals that faculty and students hold in our institution, and that the outcomes of college writing, broadly conceived, prepare students to embark on their careers ready to produce the kinds of writing that will help them perform at a high level.

2. We need to tie evaluation to actual performances—to the concrete outcomes of our WAC programs. In other words, we need to examine WAC course syllabi and assignments, and students' performances in those courses and on those tasks. Separating evaluation from the classroom context results in poorer data and less direct—and hence less useful— evaluations. If faculty are to invest in this kind of intrusive evaluation, the results need to serve the faculty's agendas— results need to figure into faculty development in positive ways, helping faculty satisfy the institution's demand for evidence of strong teaching performance. In addition, such an evaluation needs to serve other agendas that faculty identify as important: promoting higher-order thinking, for example, or maintaining high standards.

3. Evaluating WAC must be a continual effort. In part, WAC evaluation must be continual for reasons of self-preservation. WAC's very complexity demands complex forms of assessment. Continual evaluation allows us to spread the various evaluations out in ways that make them manageable. In addition, continual evaluation allows WAC directors to provide frequent "mini-reports" to stakeholders about the effectiveness of some aspects of the program and the need for reform in others. Thus, continual evaluation creates a context in which WAC evaluation will be perceived as re-

sponsive and responsible; continual evaluation also keeps the WAC program visible to central administrators, who receive the reports the evaluation generates. All these outcomes help the WAC program be a "good scout" within the institution.

The overall purpose of this kind of evaluation, of course, is to maintain the WAC program's health and effectiveness. Evaluation involves far more than protecting the program or providing statistics for their own sake. WAC evaluation ultimately must focus on improving the program, an objective that demands measures that will reveal weakness as well as strength. Over time, the weaknesses can be eliminated and the evaluations can document improvement—hence the need for long-range evaluations. And to the extent that the program is strong and effective—or to the extent that it can demonstrate improvement—the information gathered in these assessments feeds directly into accountability. Thus, WAC evaluation serves multiple needs, helping the program thrive, helping the institution evolve, and helping explain one way in which the university's curriculum serves important aspects of the public's agenda for higher education.

Note

1.   For a complete discussion of writing assessment and instruction at Washington State University, see *Beyond Outcomes: Assessment and Instruction within a University Writing Program*, edited by Richard Haswell (Westport, CT: Ablex, 2001).

Works Cited

Angelo, Thomas A., and Patricia Cross. *Classroom Assessment Techniques*. 2nd ed. San Francisco: Jossey-Bass, 1993.

Banta, Trudy W. *Making a Difference: Outcomes of a Decade of Assessment in Higher Education*. San Francisco: Jossey-Bass, 1993.

Beason, Larry, and Laurel Darrow. "Listening as Assessment." *Assessing Writing Across the Curriculum: Diverse Approaches and Prac-*

tices. Ed. Kathleen Blake Yancey and Brian Huot. Greenwich, CT: Ablex, 1997.

Belanoff, Pat, and Marcia Dickson, eds. *Portfolios: Process and Product.* Portsmouth, NH: Boynton/Cook, 1991.

Belanoff, Patricia, and Peter Elbow. "Using Portfolios to Increase Collaboration and Community in a Writing Program." *WPA: Writing Program Administration* 9 (1986): 27–39.

Bober, Marcie. "Online Course Delivery: Is Meaningful Evaluation Possible?" *Distance Education Report* 2.11 (1998): 1–3.

Bonnema, Doug, Richard Haswell, and Joel Norris. *The Washington State University Writing Portfolio: Second Findings, June 1995–May 1997.* Pullman, WA: WSU Office of Writing Assessment, 1997.

Camp, Roberta. "Changing the Model for the Direct Assessment of Writing." *Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations.* Ed. Michael M. Williamson and Brian A. Huot. Cresskill, NJ: Hampton, 1993.

———. "New Views of Measurement and New Models for Writing Assessment." *Assessment of Writing: Politics, Policies, Practices.* Ed. Edward M. White, William Lutz, and Sandra Kamusikiri. New York: MLA, 1996.

Condon, William. "Virtual Space, Real Participation: Dimensions and Dynamics of a Virtual Classroom." *The Online Writing Classroom.* Ed. Susanmarie Harrington, Rebecca Rickly, and Michael Day. Cresskill, NJ: Hampton, 2000.

Condon, William, and Liz Hamp-Lyons. "Introducing a Portfolio-Based Writing Assessment: Progress through Problems." *Portfolios: Process and Product.* Ed. Patricia Belanoff and Marcia Dickson. Portsmouth, NH: Boynton/Cook, 1991. 231–47.

Fulwiler, Toby, and Art Young. "Preface—The WAC Archives Revisited." *Assessing Writing Across the Curriculum: Diverse Approaches and Practices.* Ed. Kathleen Blake Yancey and Brian A. Huot. Greenwich, CT: Ablex, 1997.

Gillmore, Gerald M. "UW OEA: Research and Assessment." University of Washington, Office of Educational Assessment. 9 Nov. 1997 <http://www.washington.edu/oea/asessen.htm>.

Gould, Stephen Jay. *The Mismeasure of Man.* Rev. ed. New York: Norton, 1996.

Guba, Egon G., and Yvonna S. Lincoln. *Fourth Generation Evaluation.* Newbury Park, CA: Sage, 1989.

Hamp-Lyons, Liz, and William Condon. *Assessing the Portfolio: Principles for Practice, Theory, and Research.* Cresskill, NJ: Hampton, 1999.

————. "Questioning Assumptions about Portfolio-Based Assessment." *College Composition and Communication* 44 (1993): 176–90.

Haswell, Richard. "Documenting Improvement in College Writing: A Longitudinal Approach." *Written Communication* 17 (2000): 307–52.

————. "Preliminary Results from a Value-Added Study of WSU Undergraduate Writing." Washington State University Writing Assessment Office Internal Report. 21 June 1996.

Haswell, Richard, and Susan McLeod. "WAC Assessment and Internal Audiences: A Dialogue." *Assessing Writing Across the Curriculum: Diverse Approaches and Practices.* Ed. Kathleen Blake Yancey and Brian A. Huot. Greenwich, CT: Ablex, 1997. 217–36.

Haswell, Richard, Susan Wyche-Smith, and Lisa Johnson-Shull. "Shooting Niagara: Making Assessment Serve Instruction at a State University." *WPA: Writing Program Administration* 18.1 (1995): 44–53.

Henscheid, Jean. "Preparing Freshmen to Excel." *Bridges to Student Success: Exemplary Programs 1998.* NASPA/Student Affairs Administrators in Higher Education, 1998. 34–41.

Hughes, Gail M. "The Need for Clear Purposes and New Approaches to the Evaluation of Writing-Across-the-Curriculum Programs." *Assessment of Writing: Politics, Policies, Practices.* Ed. Edward M. White, William Lutz, and Sandra Kamusikiri. New York: MLA, 1996.

Kinkead, Joyce. "Documenting Excellence in Teaching and Learning in WAC Programs." *Assessing Writing Across the Curriculum: Diverse Approaches and Practices.* Ed. Kathleen Blake Yancey and Brian A. Huot. Greenwich, CT: Ablex, 1997.

LeMahieu, Paul G., Drew H. Gitomer, and JoAnne T. Eresh. *Portfolios beyond the Classroom: Data Quality and Qualities.* Princeton, NJ: Educational Testing Service, 1995.

Morgan, Meg. "The Crazy Quilt of Writing Across the Curriculum: Achieving WAC Program Assessment." *Assessing Writing Across*

the Curriculum: Diverse Approaches and Practices. Ed. Kathleen Blake Yancey and Brian A. Huot. Greenwich, CT: Ablex, 1997.

Morris, Barbra S. "The English Composition Board at the University of Michigan." *Literacy for Life: The Demand for Reading and Writing.* Ed. Richard W. Bailey and Robin Melanie Fosheim. New York: MLA, 1983. 265–68.

National Center for Higher Education Management Systems. *A Preliminary Study of the Feasibility and Utility for National Policy of Instructional "Good Practice" Indicators in Undergraduate Education.* Washington, DC: National Center for Education Statistics, U.S. Department of Education, Office of Educational Research and Improvement, 1994. 94–437.

Smith, Mary L. and Cart Rottenberg. "Unintended Consequences of External Testing in Elementary Schools." *Educational Measurement: Issues and Practice* 10 (1991): 7–11.

Walvoord, Barbara E., and Lucille P. McCarthy. *Thinking and Writing in College: A Naturalistic Study of Students in Four Disciplines.* Urbana, IL: NCTE, 1990.

Walvoord, Barbara E., et al. *In the Long Run: A Study of Faculty in Three Writing-Across-the-Curriculum Programs.* Urbana, IL: NCTE, 1997.

White, Edward M. *Teaching and Assessing Writing: Recent Advances in Understanding, Evaluating, and Improving Student Performance.* Rev. ed. San Francisco: Jossey-Bass, 1994.

White, Edward M., and Leon L. Thomas. "Racial Minorities and Writing Skills Assessment in the California State University and Colleges." *College English* 43 (1981): 267–83.

Yancey, Kathleen Blake. "Looking Back as We Look Forward: Historicizing Writing Assessment." *College Composition and Communication* 50 (1999): 483–503.

Yancey, Kathleen Blake, and Brian A. Huot, eds. *Assessing Writing Across the Curriculum: Diverse Approaches and Practices.* Greenwich, CT: Ablex, 1997.