WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

Assess Locally, Validate Globally: Heuristics for Validating Local Writing Assessments

Chris W. Gallagher

Abstract

Drawing on recent assessment scholarship in rhetoric and composition, this article describes an assessment quality review heuristic that allows writing program administrators to develop validation arguments framed within the principles and terms that the field and individual programs value. Such heuristics may be used to evaluate local assessments and to make a case for them beyond programs. Because they are grounded in our disciplinary knowledge, values, and commitments, these heuristics can serve as a vehicle for exerting leadership for writing assessment at a time when standardized tests are gaining traction in higher education.

Though compositionists specializing in writing assessment find plenty on which to disagree, the "basic principles" proposed by O'Neill, Moore, and Huot in their *Guide to College Writing Assessment*—that writing assessment be site-based, locally controlled, context-sensitive, rhetorically based, accessible, and theoretically consistent (57)[1]—are widely echoed in our assessment scholarship. Their contextualist approach is shared, for instance, by Bob Broad, who advocates "organic" and "locally grown" writing assessments (see his *What We Really Value and Organic Writing Assessment*). Similarly, Haswell et al. use the term "homegrown" (40) to describe the assessments they value, noting that the moral of the story they tell about Washington State's assessment program "is that writing teachers should be leery of assessment tools made by others, that they should, and can make their own." The authors note that their "adventure into assessment was *sui generis*, as [they] believe all such ventures elsewhere might be" (14). Even Patricia Lynne, who is critical of much of what she considers to be mainstream assessment scholarship in the field, advocates a "contextualist para-

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

digm" that attends closely to local environments and "the relationships among those involved in the assessment" (117).[2]

This insistence on local, "organic" assessments, while admirable, creates certain challenges for WPAs. First, there are conceptual difficulties. "Local" and "organic" are not the same thing. An assessment may be local without having arisen naturally from a local ecology (assuming we can conceive assessment in these terms in the first place). Further, such an assessment might serve the needs of teachers and students within a program quite well, which suggests that adapting assessment methods or strategies originating elsewhere might be perfectly appropriate and effective. More troubling, local assessments might *not* serve local needs and interests well. They may be poorly constructed, unaligned (or misaligned) with curriculum and instruction, and even blatantly unfair and discriminatory. Finally, to the extent that they *are* local, assessments are likely to be difficult to make legible to those who do not share the local values, beliefs, assumptions, conventions, and discourses. For instance, it may be difficult for those outside the program Bob Broad examines in *What We Really Value* to make sense of the long and complex maps and lists of criteria such as "Significance/ Development/Heart," "Focus/Pace/Concise," and "Tight/Subtle/ Minimalist/Show Don't Tell" (34-37).

This accessibility challenge—what O'Neill et al. describe as making the assessment program "transparent to those affected by it as well as others invested in its results"—often turns out to be more complicated than "communicat[ing] in language that is accessible to the constituencies" (57). What we must communicate about our local assessments, after all, is not just information, but an *argument*. Even the most conventional criteria for establishing assessment quality—validity and reliability—are now understood in writing assessment and educational measurement scholarship to consist of arguments supported by theoretical rationales and evidence (Parkes, Llosa, Huot). Jay Parkes, for example, suggests that a reliability argument "articulates the reliability values most relevant to the particular measurement situation and then the most appropriate evidence and theory to support an argument for the presence of that value" (2). Though not all compositionists value reliability as it is traditionally conceived—a point I will return to later—we should be able to appreciate Parkes' suggestion that the onus of an assessment argument is to present a case not only for our values but for the presence of those values in our assessments.[3]

Parkes' work challenges WPAs to develop methods of evaluating, documenting, and arguing for the quality of our local assessments. The trick is to develop methods that are rooted in local interpretations of disciplinary values and the local values of our programs while being intelligible and

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

persuasive to audiences beyond our programs. To that end, this article proposes a method, in the form of a heuristic, for validating[4] local assessments.

In this article I use "heuristic" to mean a method for invention and problem-solving—in this case a tool for evaluating and reporting on the assessment quality of local assessments. This validation heuristic is intended to be generic enough to suggest some typical moves and conventions but flexible enough to accommodate a wide range of local (program) values. It is also intended to suggest ways in which we can craft persuasive arguments, both within and beyond our programs, about what we value in assessment, writing, teaching, and learning. In the sections to follow, I will present a sample heuristic, describe its potential uses, and demonstrate how it reflects local and disciplinary values in an accessible and I hope persuasive fashion. I conclude with some thoughts about the implications of using such heuristics within the current educational and assessment policy context.

A Heuristic for Assessment Quality Review

This heuristic asks program participants to study an individual assessment—be it for placement, proficiency, program assessment, or faculty evaluation—within the context of their program and its values. By "assessment," I mean a procedure for gathering, interpreting, and evaluating student work for any programmatic purpose as well as any use of data generated from that procedure. The heuristic should be completed for each assessment *use*. It is not enough to declare a certain type of assessment—directed self-placement, for instance—"valid" and "reliable." We need to validate particular *decisions* we make based on assessment data, not just the test/instrument (see Broad; Huot; O'Neill and Huot).

The product of the heuristic is a highly contextualized but broadly intelligible profile of the assessment. This profile includes brief narrative descriptions and analyses as well as relevant appendices, including assessment artifacts (copies of the instruments, sample student work, meeting agendas and protocols, statistical results, technical reports, etc.). Here I present a sample heuristic. Please bear in mind, however, that I am not proposing *the* validation heuristic for all writing programs; this heuristic is the result of a particular negotiation of disciplinary and local values, and while it could be *adapted* by any writing program, it is not intended to be *adopted* wholesale. What I am proposing in this article is the development and use of heuristics *of this sort* by writing programs. (See the appendix for an example of a completed profile based on the featured heuristic.)

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

An Assessment Quality Review Heuristic

A. Briefly describe the writing program, including curricular and instructional goals, institutional constraints and opportunities (e.g. resources issues, labor conditions, professional development offerings), and student and teacher demographics. Append relevant documentation.

B. Briefly describe the assessment and its relationship, if any, to other assessments conducted in the program. If this assessment is part of an overall assessment plan, append the plan.

C. Answer the following questions about the assessment under review.

1. Meaningful

- What are the purposes of this assessment? What are its intended uses?
- How were these purposes arrived at? Who formulated them?
- Why and to whom are those purposes significant?
- How were these purposes made known to students and teachers?
- How does the content of the assessment match its purpose?

2. Appropriate

- How is the assessment suitable for this context, these participants, and its intended purposes and uses?
- How does the assessment reflect the values, beliefs, and aspirations of the participants and their immediate communities?

3. Useful

- How does the assessment help students learn and help teachers teach?
- How does the assessment provide information that may be used to improve teaching and learning, curriculum, professional development, program policies, accountability, etc.?
- Who will use the information generated from this assessment and for what purposes?

4. Fair

- How does the assessment ensure that all students are able to do and demonstrate their best work?
- How does the assessment contribute to the creation or maintenance of appropriate working conditions for teachers and students? How

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

does it ensure adequate compensation and/or recognition for the labor required to produce it?

5. Trustworthy

- How are the assessment results arrived at and by whom?
- How does the assessment ensure that these results represent the best professional judgment of educators?
- How does the assessment ensure that the results derive from a process that honors articulated differences even as it seeks common ground for decisions?

6. Just:

- What are the intended and unintended consequences of this assessment—for students, teachers, administrators, the program, the institution, etc.?
- How does the assessment ensure that these consequences are in the best interest of participants, especially students and teachers?

D. In light of this review, what changes, if any, do you plan to make to this assessment?

Uses of Validation Heuristics

While I can imagine that Parts A, B, and D of this heuristic could be useful, both as an inventorying exercise and a reporting mechanism, for any writing program, Part C—the crux of the assessment quality review—is flexible and should be crafted by each writing program. In the next section, I will show how Part C of the heuristic I present is a particular working-out of disciplinary and local values. Here, I want to explore potential uses of such heuristics.

The primary purpose of validation heuristics is to guide program participants through a contextual evaluation of the effects of assessment on their teaching and learning. At a minimum, to answer the questions thoroughly, program administrators would need to examine and account for teachers' and students' experiences with the local assessment. More ideally, groups of administrators, teachers, and possibly students would collaboratively evaluate their assessments and complete their heuristic. This heuristic, then, would enable these stakeholders to explore together the extent to which the assessments reflect both their local values and disciplinary values—thereby making assessment design, validation, and planning visible

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

and collaborative features of the writing program. (Such heuristics could easily be integrated into self-studies conducted for QEP, accreditation, and ongoing programmatic assessments.) Because heuristics are adaptable to local values and interpretations of disciplinary values, they honor the diversity of writing programs and accommodate changes in local and disciplinary values over time.

At the same time, heuristics may be used to inform and persuade a range of audiences outside writing programs, including colleagues, upper administrators, accreditors, parents, and the local and regional community. They are a method for formative review, but also for summative documentation. (As the example in the appendix shows, heuristics may help *in*validate an assessment, and the arguments and profiles that result might be used to persuade audiences that the assessment should be discontinued.) The documentation is intended to be accessible to the program's audiences, but framed in the program's terms, based on its disciplinary and programmatic values. It is both an argument for those values and an argument about the extent to which those values are present in particular assessments.

Depending on a program's circumstances—and most particularly the expectations of its various audiences—validation heuristics could produce a single document (a profile) meeting both formative and summative purposes, or they could produce multiple documents, each tailored to its specific audience. In addition, WPAs might choose to feature only certain parts of the profile for particular purposes: to focus instructors' attention on assessment purposes, for instance, or to make a case to administrators for the reliability of the assessment.

Validation heuristics also provide an opportunity to collaborate across programs on the basis of shared values. Because validation arguments are grounded in part in disciplinary values, it is likely that there will be significant overlap in institutions' heuristics. Thus, groups of institutions may decide to use the same heuristic. Nor is it difficult to imagine that a single heuristic or a small set of heuristics could be used for professional peer review as part of the CWPA's Consultant-Evaluator Service or CCCC's Writing Program Certificate of Excellence program.

To be sure, some compositionists will resist these standardized uses of heuristics, agreeing with David Reinheimer that "[j]ust as local assessment grows out of the local curriculum, the local population, and other local parameters, so too should a local validity argument grow out of the local assessment context, rather than being imposed on a context for which the argument is not appropriate" (171). I agree that validity arguments should never be foisted upon programs; that is not what I am arguing for here. But I would argue that even local validation—Reinheimer's primary concern—

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

is made more meaningful and defensible when conducted in the context of local interpretations of disciplinary values. This does not mean that local values are sacrificed to disciplinary values; indeed, in order to be consonant with composition and rhetoric's disciplinary values, any such heuristics would need to be rooted in the local, in particularity. This sounds like a contradiction, but it's merely a paradox: a *general* expectation to be responsive to the *local* context.

A final point: validation heuristics need not be applied to every assessment a teacher or even a program administers. Rather, programs would develop reasonable sampling plans and use technological platforms that make this work manageable. The point is not to join our K-12 colleagues under the mountain of paperwork they have been subjected to by the standards and assessment movement. The point is to evaluate the quality of our major assessments and to provide adequate and persuasive information to our various audiences about the quality of our local assessments.

Negotiating Local and Disciplinary Values

The sample heuristic presented earlier represents a particular negotiation of local and disciplinary values. While far from idiosyncratic, the criteria in Part C—meaningful, appropriate, useful, fair, trustworthy, and just—are nonetheless a result of deliberate choices among an array of options. They are meant (hypothetically, at least) to represent the values and goals of a particular writing program at a specific moment in time.

While I have composed this heuristic (and profile) for illustration purposes, an appropriate first step for writing program administrators and instructors—a step I am just now embarking on in the program I recently began administering—would be to review writing assessment literature and determine which assessment quality criteria best suit the values, needs, and beliefs of their program and its various audiences. Even this initial activity has great heuristic value, as program participants discover and clarify their ideas about writing, teaching, and assessment. Participants first need to decide *which* criteria are most important to them, and then they need to decide what those criteria *mean* and how they would be realized and documented in assessment practice. The key is to choose criteria that will be persuasive to program participants and audiences, since any validation is an argument for using the specific criteria identified; essentially, each heuristic says, "*These* are the important questions to ask."

Assessment quality criteria are different from the "basic principles" identified by O'Neill et al., the "principles for effective assessment" proffered in the *NCTE-WPA White Paper on Assessment*, or the "best assessment prac-

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

tices" proposed in the CCCC's "Writing Assessment: A Position Statement" (some of which are discussed below). These principles and best practices can and should guide our choice of criteria, but they do not operate at a sufficient level of specificity to allow us to make judgments and decisions about the quality of particular uses of assessments. In fact, when we turn our attention from these more general principles—on which our field has reached broad consensus—to specific criteria, we find vigorous and healthy debate. This is why Part C of the heuristic must be adaptable: it represents a local interpretation and negotiation of disciplinary values that reflect a range of professionally defensible options. In this way, the heuristic preserves local choice while tying programs to the discipline's ongoing work on writing assessment.

Perhaps the most pointed debate among assessment experts in the field revolves around the twin pillars of educational measurement: "validity" and "reliability." While scholars such as Brian Huot insist that sophisticated understandings of these terms can harmonize with our values, others—such Patricia Lynne—argue that they are inextricably linked to an objectivist paradigm that conflicts with our social constructionist disciplinary worldview. Writing administrators and instructors may, and do, take a range of defensible positions on this debate. The heuristic presented here reflects a particular stance on it: that Huot is right to suggest we cannot and should not ignore these terms, which are widely recognized within and beyond the discipline, but Lynne is right to suggest that (for that very reason) we cannot fully disassociate them from their objectivist "baggage" (3), and therefore we should not allow them to circumscribe our thinking about what we value in writing assessment. In other words, the heuristic frames both a validity and reliability argument within rhetoric and composition's contextualist framework without explicitly invoking these terms.

Briefly, the validity argument of the heuristic follows Huot and O'Neill's suggestion that "validity as it is currently understood is about validating decisions based on an assessment" (4). It proposes that we must examine both the conditions in which an assessment is conducted (and that it helps create) and the consequences of the assessment. For instance, *fairness* requires that conditions are in place for all students to do and to demonstrate their best work. This is traditionally called "opportunity to learn." But we are not concerned only with whether material has been taught to all students before testing, as this quality criterion is conventionally understood; as James Paul Gee has demonstrated, presenting the same material to all students in the same way and then assessing them in the same way on that material, without accounting for their varying embodied and textual experiences with the desired literacy practices and with others who engage

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

in those literacy practices, is patently unfair. To meet the standard of *fairness* necessary for validity here, then, the assessment must acknowledge, value, and make room for students' and teachers' linguistic and cultural differences. In terms of consequences, *justice* requires that we examine both the intended and unintended consequences for all parties: students, teachers, administrators, programs, institutions, and so on. We must examine the extent to which the implementation and the results of each assessment use are in the best interests of all participants. We are concerned with *effectiveness*: we evaluate the validity (or lack of validity) of an assessment based on the effects it brings about within its context of use.

The reliability argument of this heuristic is framed in similar terms. The standard for *trustworthiness*—the core of the reliability argument here—is collective professional judgment rendered in the context of a search for common ground through articulated differences. The point is not to *de*contextualize the reading practices of teacher-scorers in search of some Platonic "true score" ratified by consistency; it's precisely the opposite: to contextualize and support professional judgment through conversation. Moreover, in this model, that conversation seeks to use human differences—in interpretations, values, beliefs, experiences—to generate *multiple, usable meanings*. Here, following the lead of Huot, Huot and Neal, Broad, Lynn and others, the heuristic builds on Pamela Moss's "hermeneutic" approach to reliability, which involves

> holistic, integrative interpretations of collective performances that seek to understand the whole in light of the parts, that privilege readers who are most knowledgeable about the context in which the assessment occurs, and that ground those interpretations not only in the textual and contextual evidence available, but also in a rational debate among the community of interpreters. (86)

At the same time, this heuristic does not demand consensus, which composition and rhetorical theory long ago revealed as a problematic condition for community (see Harris, Myers, Trimbur). Rather, it calls for acknowledgement, documentation of, and collective inquiry into differences. In this way, it is located within the interpretive tradition described by Clifford Geertz in *Interpretation of Cultures*, in which "progress is marked less by a perfection of consensus than by a refinement of debate. What gets better is the precision with which we vex each other" (29).[5]

It is important to note that the validity and reliability arguments presented by this heuristic are negotiations, rather than applications, of disciplinary approaches to these concepts. For instance, the validity argument is

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

consistent with the conception of validity in the NCTE-WPA *White Paper on Writing Assessment in Colleges and Universities,* which in turn relies on the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education's definition of validity in its 1999 *Standards for Educational and Psychological Testing*: "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (American 9). But it is inconsistent with the more traditional but still widely held notion that validity represents the extent to which a test measures what it purports to measure (see Huot, O'Neill and Moore, 503-509, on the persistence of this notion in our field). At the same time, it complicates the white paper's construction of reliability as the provision of "consistent results, no matter who conducts the assessment." Instead, the heuristic is more in line with the notion of reliability advanced by CCCC's "Writing Assessment: A Position Statement." While this statement contains no explicit mention of reliability, it insists that "[b]est assessment practice is direct assessment by human readers," noting that while machine scoring "may promise consistency, [it] distorts[s] the very nature of writing as a complex and context-rich interaction between people." The statement also indicates that best assessment practice gathers "multiple perspectives on a performance," including those of peers, instructors, and students themselves. Like this statement, the heuristic presented here frames its reliability argument in terms of richly contextualized, collaborative judgments rendered by diverse human interpreters with a rich understanding of the context.

The heuristic also follows the lead of both of these professional documents in valuing assessments that are "appropriate" and "fair." But while it basically adopts the NCTE-WPA's understanding of "appropriate" assessment as "fit[ting] the contexts and decisions that will be made based on it," it goes beyond that statement's definition of "fair" as "provid[ing] an equal opportunity for students to understand the expectations, roles, and purpose of the assessment [and] guard[ing] against any disproportionate social effects on any language minority group" to include affirmation of cultural and linguistic difference as well as attention to labor practices.

Similarly, the heuristic is influenced by Patricia Lynne's work. Its emphasis on *purpose* is consistent with her understanding of "meaningful" assessment as "conducted for specific and articulated reasons" (15). And though the heuristic does not explicitly use "ethical" as a criterion—mainly because that notoriously vague and slippery term might be a rhetorical liability in some local contexts—it does seeks to address "the broad political and social issues surrounding" the assessment (118) and incorporates her thinking into the criteria of "fair" and "just."

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

The point is, this heuristic—indeed *any* such heuristic—reflects an *interpretation* and a *negotiation* of disciplinary values *and* an argument for that interpretation and negotiation. Some writing programs will choose to follow Lynne's lead and reject "validity" and "reliability" as key criteria, others will be committed to or will feel compelled to use them, and still others will invoke them without using them explicitly, as the sample heuristic does. This decision will depend both on local considerations and on how convincing program participants find Lynne's argument that these terms cannot operate coherently outside of an "objectivist paradigm." In any case, because they are locally adapted, validation heuristics are flexible enough to accommodate a variety of programmatic assessment values and approaches.

Conclusion

Unfortunately, our disciplinary presumption in favor of the local is overmatched by a more general presumption against it: "local" equates in many people's minds to chaos, a free-for-all. As Broad notes, our "home-grown, do-it-yourself" paradigm competes with attractively packaged "fast-food-style" offers by commercial testing corporations "to make assessment faster and simpler by cutting it off from the rest of our work as educators" (2). Standardized testing continues to make in-roads in higher education, and upper administrators, policymakers, and the general public continue to imagine faculty and students as targets of assessment rather than generators of it. In fact, standardized tests are presumed to be more valid and reliable precisely *because* they are created outside the local context. In the wake of the Spellings Commission's attack on the "complacency" and "recalcitrance" of the professoriate, higher education policy leaders have adopted accountability systems (such as the Voluntary System of Accountability) that rely on standardized tests (Schneider in Association), assessment vendors have piled on by calling into question faculty's assessment capacity (see, e.g., Hersch; Dwyer, Millett, and Payne), and both nonprofit and for-profit assessment companies have redoubled marketing of their off-the-shelf wares (see Hutchings). In this context, local assessment is a hard sell. We need to demonstrate that our assessments—whether directed self-placement, dynamic criteria mapping, eportfolios, learning records, etc.—are of high quality. We need to show that there are multiple ways to achieve high assessment quality. And we need to do so on our terms, framed in arguments congruent with our local *and* disciplinary values.

The advantage of using validation heuristics for this purpose is that they provide a useful method for internal, formative review while offering an accessible and persuasive reporting format. These heuristics may be used

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

by individual programs to educate and edify their various audiences. For example, I used the heuristic here as an invention activity to generate arguments to university administrators that we ought to move away from the existing placement program and toward a version of directed self-placement (see the appendix). Though the heuristic reveals the *lack* of quality of the assessment we were using—a possibility we all must prepare for[6]—it gave our program an opportunity to articulate our values and bring our practices more in line with them. In addition, administrators were quite pleased to learn that the Writing Program had conducted such a thoughtful, careful review of its own assessment. So another advantage of using such heuristics is that they demonstrate an incontrovertible professional commitment. This suggests that we needn't approach assessment validation as a PR endeavor; instead, we can and should undertake serious reviews that reveal both the strengths and the weaknesses of our assessments. The purpose of using validation heuristics is to inquire into the quality of our assessments in order to improve them.

As I have suggested, such heuristics may be used by groups of programs that desire a comparative look. But even then, the heuristics don't lend themselves to ranking or standardization because they rest on the notion that local assessments can be shown to be of similar quality despite—actually, because of—their differences. This kind of inter-institutional evaluation is criterion-referenced, not norm-referenced. It bypasses the kinds of insidious rankings of radically dissimilar programs and institutions that issue from standardized tests in favor of contextual portraits that include data on student learning as well as information about the quality of the assessments used to generate those data. Such a move should gladden the heart of even the cold consumer who rules accountability's cold calculus.

Notes

1. This list is based on Brian Huot's earlier formulation in *(Re)Articulating Writing Assessment for Teaching and Learning*; the authors added the final principle to Huot's original list.

2. In addition to the scholars cited here, see Barlow, Liparulo, and Reynolds; Reinheimer.

3. Parkes identifies a set of "components" of a reliability argument, but they are too confined to a technical understanding of reliability to be borrowed wholesale by WPAs. They are: "1. A determination of the social and scientific values… most relevant to the scenario at hand. 2. Clear statements of the purpose and the context of the assessment. 3. The definition of a replication in the particular

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

assessment. 4. A determination of the tolerance or level of reliability needed. 5. The evidence. 6. The Judgment: Putting it all together" (6).

4. In this article, I use the term "validation" advisedly. It is a term of controversy in our field because it implies that "validity" is our master term. While some of us (most notably Huot) accept this premise, others (most notably Lynne) do not, as I discuss later in this article. When I use "validation," then, I mean, broadly, inquiry into and reporting on assessment quality.

5. Thanks to Matt Noonan for directing me to this quotation.

6. Huot, O'Neill, and Moore suggest, following Michael Kane, that most validation research generated by testing companies has "a confirmationist bias" (Kane qtd. in Huot et al. 507). WPAs have the opportunity to conduct more rigorous, thorough, and honest validation research on their own assessments. We might also gain insight, and rhetorical advantage, by applying our validation heuristics to standardized tests

Works Cited

American Psychological Association, American Educational Research Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association, 1999. Print.

Association of American Colleges and Universities. "Association of American Colleges and Universities Calls for Campuses to Articulate and Assess Broad Set of College Learning Outcomes at the Highest Levels of Achievement." Press Release. 7 January 2009. AAC&U. Web. 15 July 2009.

Barlow, Libby, Steven P. Liparulo, and Dudley W. Reynolds. "Keeping Assessment Local: The Case for Accountability through Local Assessment." *Assessing Writing* 12.1 (2007): 44-59. Print.

Broad, Bob. *What We Really Value: Beyond Rubrics in Teaching and Assessing Writing*. Logan: Utah State UP, 2003. Print.

Broad, Bob, Linda Adler-Kassner, Barry Alford, Jane Detweiler, Heidi Estrem, Susanmarie Harrington, Maureen McBride, Eric Stalions, and Scott Weedon. *Organic Writing Assessment*. Logan: Utah State UP, 2009. Print.

CCCC Committee on Assessment. "Writing Assessment: A Position Statement." November 2006 (Revised March 2009). Conference on College Composition and Communication. Web. 8 Feb. 2010.

Dwyer, Carol S., Catherine M. Millett, & David G. Payne. *A Culture of Evidence: Postsecondary Assessment and Learning Outcomes*. June 2006. Educational Testing Service. Web. 8 Feb. 2010.

Gee, James Paul. "Opportunity to Learn: A Language-Based Perspective on Assessment." *Assessment in Education* 10.1 (March 2003): 27-46. Print.

Geertz, Clifford. *The Interpretation of Cultures*. New York: Basic Books, 1973. Print.

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

Harris, Joseph. "The Idea of Community in the Study of Writing." *College Composition and Communication* 40.1 (1989): 11-22. Print.

Haswell, Richard H., ed. *Beyond Outcomes: Assessment and Instruction Within a University Writing Program*. Westport: Ablex, 2001. Print.

Hersh, Richard H. "What Does College Teach?" *theatlantic.com*. November 2005. *The Atlantic*. Web. 8 Feb. 2010.

Huot, Brian. *(Re)Articulating Writing Assessment for Teaching and Learning.* Logan: Utah State UP, 2002. Print.

Huot, Brian and Michael Neal. "Writing Assessment: A Techno-History." *Handbook of Writing Research*. Ed. Charles A. MacArthur, Steve Graham, and Jill Fitzgerald. New York: Guilford, 2006. 417-32. Print.

Huot, Brian and Peggy O'Neill, eds. *Assessing Writing: A Critical Sourcebook*. Boston: Bedford/St. Martin's, 2009. Print.

Huot, Brian, Peggy O'Neill, and Cindy Moore. "A Usable Past for Writing Assessment." *College English* 72.5 (2010): 495-517. Print.

Hutchings, Pat. "The New Guys in Assessment Town." *changemag.org*. May/June 2009. *Change*. Web. 8 Feb. 2010.

Llosa, Lorena. "Building and Supporting a Validity Argument for a Standards-Based Classroom Assessment of English Proficiency Based on Teacher Judgments." *Educational Measurement: Theory and Practice* 27.3 (Fall 2008): 32-42. Print.

Lynne, Patricia. *Coming to Terms: A Theory of Writing Assessment*. Logan: Utah State UP, 2004. Print.

Moss, Pamela A. "Can There Be Validity Without Reliability?" *Assessing Writing: A Critical Sourcebook*. Ed. Brian Huot and Peggy O'Neill. Boston: Bedford/St. Martin's, 2009.81-96. Rpt. *Educational Researcher* 23.4 (1994): 5-12. Print.

Myers, Greg. ""Reality, Consensus, and Reform in the Rhetoric of Composition Teaching." *College English* 48 (1986): 154-73. Print.

NCTE-WPA. *White Paper on Writing Assessment in Colleges and Universities*. 2008. Web. 8 Feb. 2010.

O'Neill, Peggy, Cindy Moore, and Brian Huot. *A Guide to College Writing Assessment*. Logan: Utah State UP, 2009. Print.

Parkes, Jay. "Reliability as Argument." *Educational Measurement: Theory and Practice* 26.4 (Winter 2007): 2-10. Print.

Reinheimer, David A. "Validating Placement: Local Means, Multiple Measures." *Assessing Writing* 12.3 (2007): 170-79. Print.

Trimbur, John. "Consensus and Difference in Collaborative Learning." *College English* 51.6 (Oct. 1989): 602–16. Print.

Appendix: Sample Profile

Note: This sample is intended to be illustrative. It was not composed in the ideal (i.e., collaborative) manner described in this essay, and so it must be considered hypothetical. Moreover, as one reviewer of this article rightly noted, it does not include the kind of empirical data one would ideally

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

include in a serious validation activity. However incomplete, this profile demonstrates how validation heuristics may be used to evaluate extant assessments against a program's theoretical rationales, invent arguments that support or challenge the current assessment, and lay the conceptual groundwork for new assessments. Finally, this sample is not meant to function as a criticism of anyone involved in the profiled assessment, which was created in response to an institutional crisis. Indeed, the assessment suffers under the scrutiny of the writing program's strong, well-conceived theoretical rationales.

A. Briefly describe the writing program, including curricular and instructional goals, institutional constraints and opportunities (e.g. resources issues, labor conditions, professional development offerings), and student and teacher demographics. Feel free to append relevant documentation.

Our First-Year Writing Program (FYWP) consists of four courses: Introduction to College Writing, College Writing, and versions of these two courses for English Language Learners (ELLs). The two introductory courses (one for native speakers and one for ELLs) are credit-bearing, but do not fulfill the university's first-year writing requirement. Students placed in these courses subsequently must take and pass either College Writing or the ELL version of College Writing to fulfill the requirement. We run approximately 150 sections of these courses per year staffed by a diverse corps of instructors, including TAs, part-time lecturers, and full-time lecturers. Their respective pay rates are appended [*not included in this sample*]. Instructors receive some initial and ongoing training (workshops and meetings), but the latter offerings are limited. The FYWP, located in an English Department, has no independent budget. All first-year writing courses use Bartholomae and Petrosky's *Ways of Reading*, though instructors usually supplement the textbook with other texts in various genres and media. Our focus is on critical reading and writing of academic and multimedia essays. Students are asked to locate their own ideas in the context of scholarly and public conversations on complex topics. Classes are conducted in a collaborative workshop setting, and students are asked to compose multiple drafts of each essay. Their work culminates in reflective portfolios that include polished essays totaling 6000-7000 words. See attached Learning Goals [*not included in this sample]*.

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

B. Briefly describe the assessment and its relationship, if any, to other assessments conducted in the program. If this assessment is part of an overall assessment plan, append the plan.

The assessment under review is a diagnostic exam conducted for placement in the four courses described above. Almost all students are initially placed in College Writing. On the first day of class, all students in required first-year writing courses receive a prompt, typically involving a short passage from the Introduction to *Ways of Reading* and a question that asks them to interpret the passage and apply it to their own experience as readers and writers. (See attached sample prompt [*not included in this sample*].) The prompt calls for a detailed, coherent essay composed in class and passed in to the instructor. Students are given the class period to compose their essays, though many of them do not use the entire allotted time. After class, the instructors read the essays and identify any students whom they believe would be better placed in another course: either an introductory course or one of the ELL courses. These essays are shared with a Writing Program (WP) administrator, who conducts a second read and either confirms or rejects the instructor's placement recommendation. The English Department staff reassigns students who are placed out of College Writing. Every effort is made to effect these placements during the first week of classes. There is no official appeal mechanism. This placement assessment is not, at this time, tied to an overall programmatic assessment plan.

C. Answer the following questions about the assessment under review.

1. Meaningful:

- What are the purposes of this assessment? What are its intended uses?
- How were these purposes arrived at? Who formulated them?
- Why and to whom are those purposes significant?
- How were these purposes made known to students and teachers?
- How does the content of the assessment match its purpose?

The purpose of this diagnostic exam is to place students into the writing course in which they are most likely to succeed. It was implemented several years ago as a stopgap after a summer placement exam was defunded. (University administrators determined that the summer placement mechanism, in which WP instructors read and made placement decisions before the semester began, placed too few students out of College Writing to justify the cost, which mainly consisted of hourly pay for readers.) WP adminis-

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

trators feel they did not have much time to plan or pilot this exam, and as a result, it has always had an *ad hoc* dimension to it. They have felt uneasy about the exam on several counts (holding it on the first day of classes, its impromptu nature, the use of one writing sample to make an important placement decision, etc.), but they have not had the time or the resources to design and implement a viable alternative.

Placement is significant to all parties involved. Students who are placed into a lower-level course must take (and pay for) two courses to fulfill their requirement. On the other hand, students who are unprepared for the rigors of College Writing may benefit from an extra semester of writing experience and instruction and may well fail College Writing if they are not placed out of this course. Placement out of College Writing also impacts students' schedules, and this happens after the semester begins. Similarly, the Writing Program must adjust to last-minute shifts in enrollment. WP administrators and instructors also have an interest in providing students with the most appropriate instruction. Students' advisors must work to accommodate changes in students' schedules. The university as a whole is impacted by these shifts in enrollment (and the attendant tuition implications), and its leaders too are invested in providing students with the most appropriate instruction.

Students are given no advance warning that the diagnostic exam will take place on the first day of classes, though information about the exam is provided on the WP website. The purpose of the placement procedure is described to them by their instructor before the exam begins. Instructors are provided the prompt and procedure, including criteria for evaluating the essays, before the semester begins. No formal on-the-spot norming is offered; instead, this assessment is influenced by William Smith's "expert reader" model, in which teachers are essentially "trained" by their experience teaching the course into/out of which students are being placed. The exam is frequently discussed at workshops and meetings, however.

The content of the exam—a short passage from *Ways of Reading* and a question about students' own reading and writing experiences—is meaningful in the context of courses in the program, as they use that textbook. The WP administrators have worked hard to design prompts that would encourage students to write about their reading and writing experiences in the context of the goals of the writing course. Instructors find that the diagnostic essays provide them information about their students and help prompt discussion about the purposes and goals of the course.

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

2. Appropriate

- How is the assessment suitable for this context, these participants, and its intended purposes and uses?
- How does the assessment reflect the values, beliefs, and aspirations of the participants and their immediate communities?

While the *content* of the exam may be meaningful in the context of the FYWP, the testing *conditions* are not appropriate to this context. Students are asked to write impromptu essays on topics they have not chosen in a timed environment without the benefit of research, sustained reflection, or extensive drafting and revision. The FYWP values research, reflection, and writing process. It also values supportive communities of writers, which is difficult to establish after the first-day exam puts teachers and students into an immediate evaluative relationship.

3. Useful

- How does the assessment help students learn and help teachers teach?
- How does the assessment provide information that may be used to improve teaching and learning, curriculum, professional development, program policies, accountability, etc.?
- Who will use the information generated from this assessment and for what purposes?

It is not clear that the exam has learning benefits for students; indeed, it may send the wrong message to students about what their writing course will entail and what is valuable about writing. Again, the information provided by the essays may be of some pedagogical use to instructors early in the courses. WP administrators also have used this procedure to gather information about, and to prompt discussion on, how teachers read student writing.

However, the information provided by this assessment has not been used systematically by the program or the university. It would be possible to run correlations between placements and indicators of students' preparation (high school GPA, SAT/ACT or TOEFL scores, etc.) or the grades they eventually receive in their writing courses. However, no such analyses have been undertaken. It would be possible to determine if any groups of students fared particularly poorly on the diagnostic. This analysis has not been conducted, either. It would be possible to examine placement trends over time in order to plan for course enrollments. This has been done in a loose way, but not formally. These analyses have not been conducted in a serious

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

way because institutional data have proven difficult to obtain by busy WP administrators and because those administrators have had no alternative at hand should the analyses prove the diagnostic to be problematic.

4. Fair

- How does the assessment ensure that all students are able to do and demonstrate their best work?
- How does the assessment ensure that cultural and linguistic differences among teachers and students are recognized, valued, and given voice in the assessment and in the teaching practices it encourages?
- How does the assessment contribute to the creation or maintenance of appropriate working conditions for teachers and students? How does it ensure adequate compensation and/or recognition for the labor required to produce it?

The diagnostic does not ensure that all students are able to do and to demonstrate their best work; it provides a limited snapshot of their ability to perform a particular kind of writing—short essays on assigned prompts—under specific constraints. Writing assessment experts agree that high-stakes decisions—for some students, taking an extra class is high stakes—should not be made on the basis of a single piece of writing (see for example CCCC's "Writing Assessment: A Position Statement").

Students whose first or strongest language is not English and students with disabilities may be put at a particular disadvantage by the testing conditions. Again, we do not know if the instrument demonstrates a consistent bias against particular groups of students. We do know that some instructors are quick to identify variations from standard edited American English as errors and that some consistently nominate ELLs for placement out of College Writing even when the essays are strong apart from typical minor errors we encounter in ELL writing.

Neither instructors nor students were involved in the design of the diagnostic, which was formulated quickly by WP administrators in response to a funding decision by university administrators. No formal surveys have been conducted, and no recommendations have been solicited, from either students or instructors, but judging by anecdotal evidence, it does not seem to WP administrators that the exam enjoys wide support.

The summer placement procedure paid WP instructors for reading essays and making placement decisions. Full-time and part-time instructors, always financially strapped, counted on this summer pay. The first-day diagnostic reassigns this work as a teaching duty conducted during the

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

first week of classes. For instructors who are already not highly paid, this extra labor, in addition to provoking general anxiety, erodes staff morale. Though a good deal of labor goes into the assessment—both by the WP administrators who design it and the instructors who implement it—it costs the university nothing.

5. Trustworthy

- How are the assessment results arrived at and by whom?
- How does the assessment ensure that these results represent the best professional judgment of educators?
- How does the assessment ensure that the results derive from a process that honors articulated differences even as it seeks common ground for decisions?

Under the "expert reader" model, most student essays receive one reading. This seems defensible, given that Smith's studies found very high reliability among teachers of the same course (90-96%), even without any on-the-spot formal training or norming (198). However, the WP has never sampled essays, as Smith did, to determine agreement rates. Nor has it systematically gathered data on agreement between instructors and the administrators who conduct second reads.

Instructors are provided with a set of criteria (formulated by WP administrators) for evaluating the essays and a sheet on which they record the rationale for their judgment. In practice, these sheets are little used and decisions are generally arrived at through conversation between instructors and program administrators. These conversations tend to be thoughtful and text-based (i.e., grounded in a close reading of the student's essay), and they are aimed at rendering shared professional judgment about the student's likelihood of success. The WP administrator assumes the role of final arbiter. Occasionally, the administrator is convinced by an instructor's reading or vice versa. Interpretive differences are not quashed, but nor are they generally given full voice, given the vagaries of the structural hierarchy.

6. Just:

- What are the intended and unintended consequences of this assessment—for students, teachers, administrators, the program, the institution, etc.?
- How does the assessment ensure that these consequences are in the best interest of participants, especially students and teachers?

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

The primary intended consequence for this assessment is that students are placed in the courses in which they are most likely to receive the most appropriate instruction and to succeed. Secondarily, instructors receive useful information about students and a place to start class discussions early in the semester. It is not clear that the assessment succeeds in ensuring the primary intended consequence; no data have been systematically collected to demonstrate that students are properly placed. Anecdotal evidence supports the achievement of the secondary intended consequence; that is, instructors regularly describe using the diagnostic essays for pedagogical purposes in the first couple weeks of class.

The unintended consequences are legion, including the sacrificing of instructional time—on the first day of classes—to testing; the immediate establishment of an evaluative relationship between instructors and students; significant anxiety for students and instructors alike; last-minute disruptions in students' schedules; the sending of a false message about what the program values; and additional labor for poorly paid instructors and program administrators during the first week of classes. All of these unintended consequences are inherent in the design of the assessment.

D. In light of this review, what changes, if any, do you plan to make to this assessment?

Though WP administrators, instructors, and students generally have done their best to make this assessment meaningful, we view it as falling short on all the quality criteria that matter to our program. For this reason, we are abandoning it. After conducting research on various placement assessments in use around the country, the Writing Program Committee, which consists of representatives of the various groups of instructors in the program, believes we ought to consider re-instituting the summer placement procedure. In the meantime—that is, next year, while we pursue possible resource avenues—we will pilot a procedure we are calling "guided self-placement." This procedure is a modified version of "directed self-placement," a mechanism currently in use at a wide range of institutions (for information on DSP, see http://faculty.gvsu.edu/royerd/dsp/). In our version, students familiarize themselves with our offerings, conduct a self-assessment, and, in consultation with the advisors, place themselves into one of the four first-year writing courses. In advance of the first day of class, they write a short essay in which they describe their prior experiences with writing and reading and explain how these experiences led to their current level of confidence and competence in writing. These essays will be col-

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

lected and reviewed by instructors who, in consultation with WP administrators, will make placement recommendations. These recommendations are not binding, but we will encourage students and their advisors to take this professional advice seriously. See the attached "guided self-placement" flyer for students and the Frequently Asked Questions sheet for advisors [*not included in this sample*].

Though guided self-placement does not overcome all the weaknesses of our diagnostic exam, and while we will continue to explore other options as we pilot this one, we view it as more consonant both with our programmatic values and the values of our discipline, rhetoric and composition. It is a vast improvement over the diagnostic exam. It doesn't require students to take a high-stakes test on the first day of class. It doesn't require teachers to sacrifice their crucial first day of instruction. It doesn't require students to write under testing conditions that don't correspond to the conditions under which they produce writing in our courses. It doesn't involve the English Department staff in changing students' schedules. Instead, it asks students to take responsibility for their education while providing them guidance; allows teachers and students to form a teaching-learning relationship from the first day of class; asks students to produce a piece of writing under more natural conditions; continues to use the expertise of teacher-readers; and involves students, teachers, and advisors working together in the best interest of students.

As we implement the guided self-placement procedure, we will gather and analyze various kinds of data: feedback from instructors and students (via surveys and interviews); percentages of students accepting vs. rejecting placement recommendations; agreement rates between readers (we will select a sample for second reads as well as document the agreement rates between instructors and administrators); correlations between placement recommendations and writing course grades; correlations between our placement recommendations and various indicators of student preparedness (high school GPA, SAT/ACT/TOEFEL scores, etc.); analyses designed to reveal differential impact of certain groups of students; trends in placement recommendations and acceptance/rejection of these recommendations (assuming we continue to use the procedure over a period of some years); and more. We will also build a database of students' essays about writing, as we anticipate they will provide a treasure trove of valuable insight into students' writing and reading experiences.

Works Cited

Bartholomae, David, and Anthony Petrosky. *Ways of Reading.* 8th Ed. Boston: Bedford /St. Martins, 2008. Print.

WPA: Writing Program Administration, Volume 34, Number 1, Fall 2010
© Council of Writing Program Administrators

CCCC Committee on Assessment. "Writing Assessment: A Position Statement." November 2006 (Revised March 2009). Conference on College Composition and Communication. Web. 8 Feb. 2010.

Smith, William. "Assessing the Reliability and Adequacy of using Holistic Scoring of Essays as a College Composition Placement Technique." *Validating Holistic Scoring for Writing Assessment*. Ed. Michael M. Williamson and Brian A. Huot. Creskill: Hampton, 1993. 142-205. Print.