

6

THEORIZING WRITING ASSESSMENT

In “Toward a New Theory of Writing Assessment,” Brian Huot contends that “it is premature to attempt any full-blown discussion of the criteria for newer conceptions of writing assessment” (1996b, 561), a position he reiterates in 2002. I am not sure this was true, even in 1996, and I am even more convinced that it is not true now. Huot, like many compositionists, chooses to support contemporary validity theory on the belief that contemporary educational measurement theory will work for writing assessment. And it is possible. It is certainly the safer and surer route. If we do what the establishment is saying is a “good thing,” we cannot go wrong—at least not very far.

But safe is not necessarily best or most productive. Pressures for assessment are rising as a result of legislation such as the “No Child Left Behind” Act, and traditional theory is supporting assessments that most of us find at least questionable. For example, local elementary schools in my county are switching to a numeric “grading” system where students receive 4, 3, 2, or 1 based on “whether students have met state standards and are ready to go to the next grade” (Hui 2003). This system bears a striking and chilling resemblance to the 4–point holistic system used to evaluate portfolios as an exit from first-year composition at some universities. It is a system being used to cut out the credit students get for effort that appears in letter grades, and to better reflect “the scale used on the state’s end-of-grade tests” (Hui 2003).

I think we are finding, like those educational measurement scholars working in complex performance assessment, that existing theoretical structures in educational assessment do not provide the kind of support, justification, and direction that we need as we evaluate the work our students do. As I have tried to show thus far, existing approaches tend to rely too heavily on existing terminology that carries far too much baggage. In educational assessment theory, validity is still trailing reliability in its wake, in spite of well-supported calls for hermeneutics. In composition studies, alternative approaches so far end up largely ignored, or contorted to conform to the norms of validity and reliability. I believe we need a new set of principles, not simply a re-configuration or a re-articulation of the old ones.

I would further argue that Huot's work, in part, makes possible just such a discussion by contributing both a preliminary set of principles and, more importantly, a methodology. His methodology—extrapolating principles from existing practices—persuasively suggests that the tools needed for theoretical work in writing assessment already exist within composition studies. Yet, contrary to what his methodology would suggest, he chooses to revert to validity for theoretical support for his efforts. This chapter continues the work Huot's essay begins, but it begins at this point of contradiction. Instead of adopting a theoretical vocabulary, I want us to begin the work of developing such a vocabulary for writing assessment that is grounded in and responsive to theoretical principles already at work within composition studies.

Developing an alternative theoretical vocabulary has precedents, even within the limited review of contemporary practices I offer in this text. Guba and Lincoln's use of "fairness" and "authenticity," and Smith's use of "adequacy" have contributed significantly to my thinking about the lexicon of writing assessment. While these criteria have merit, they are not ultimately satisfying; the former pair of terms tends to shape assessment as a research project, and the latter term tends to function as a sort of minimalist validity rather than as a step away from educational measurement theory. I find "hermeneutics" somewhat helpful, but because it has been framed as an alternative to reliability rather than a reconceptualization I find it too limiting. Adopting criteria has the advantage of building on what comes before, but developing criteria has the advantage of fresh perspective. What if, instead of looking at what historically has been required of assessment practices, we look at what we want when we evaluate writing?

Ultimately, I think composition studies would be better served by criteria that arise from the values implicit in theories about composing and learning to compose. In chapter two, I argue that those working in literacy instruction and scholarship are more interested in the variations among those who can read and write and the material conditions that influence and result from these differences; I also argue that these variations and conditions are often the result of power differentials. An assessment designed to describe and explain these variations would need to take into account a broad understanding of the context in which the literate ability is learned and tested, and it would need to be carried out in the interests of those affected by the assessment with an awareness of how power shapes decisions. The names I have chosen to represent these two qualities are "meaningful" and "ethical."

I offer these terms in place of “reliability” and “validity.” While there is a rough correlation between “meaningful” and “valid” on the one hand and “ethical” and “reliable” on the other, the terms I am proposing have the advantage of looking outward and inward at the same time. Instead of focusing primarily or even exclusively on assessment methodology—as “validity” and “reliability” do, at least in the traditional configuration—the criteria of “meaningfulness” and “ethics” highlight the context of assessment and the relationships among those involved in the assessment. Such an outward turn would counter the myopic tendencies of educational measurement theory that, I would argue, have led to an unhealthy obsession with technicalities. “Meaningfulness” and “ethics” provide a broader view and would, at least potentially, address the primary assessment concerns of compositionists: that, for example, assessment be substantive, purposeful, responsive, and fair. These terms, I believe, provide a lexicon for theorizing writing assessment dependent not on inappropriate objectivist epistemologies, but on principles accepted within the composition community.

In this chapter, I explore the merits of these terms as principles around which a theory of writing assessment might take shape. “Meaningfulness,” as I am developing the term, describes the significance of the assessment, its purpose and substance, and draws attention to the object of assessment. In addition to the more obvious object-oriented questions—“what is the assessment about?” and “what does it mean?”—“meaningfulness” raises a series of related questions, including “meaningful to whom?” and “for what purpose?” The answers to queries of this sort tie evaluation explicitly to the situation in which literacy instruction takes place, to the parties involved and the location in which it is conducted, that is, to assessment in context. “Meaningfulness,” therefore, structures the relationships among the object(s) of writing assessment, the purposes of that assessment and the circumstances in which the assessment takes place.³⁹

While “meaningfulness” highlights the object of assessment, “ethics” focuses on the effects of evaluation practices, the means of assessment, and the uses of the results—all of which draw attention to the participants in and processes of assessment projects. The inclusion of “ethical” here comes from the initial impulse that led to objective testing: the desire to be fair. Whether or not reliability in the statistical sense is moot, as Huot argues, the impulse to be fair remains. The concept of “ethics” potentially provides a set of conceptual tools for organizing, analyzing, and describing the relationships among participants in assessment situations,

which would incorporate ideals of fairness. For contemporary writing assessment, what is needed is a formulation of “ethics” that depends not on abstractions and absolutes such as reproducibility, which would replicate the aims of objectivity, but rather on social or community values. The principles of communicative ethics in the work of Jürgen Habermas and Seyla Benhabib in critical theory, with some modification due to the particular situation of assessment, sufficiently parallel current thinking in composition studies to provide a basis for organizing the relationships among participants in writing assessment situations.

Figure 2 summarizes the objectives of meaningfulness and ethics as principles for writing assessment. Each term is discussed in more detail later in this chapter.

FIGURE 2

Objectives of Meaningfulness and Ethics

Meaningful Assessment

- Defines the significance of the assessment
- Structures the relationships among the object(s) of writing assessment, the purposes of that assessment and the circumstances in which the assessment takes place.

Ethical Assessment

- Addresses the broad political and social issues surrounding assessment
- Organizes and provides principles for understanding the conduct of the participants and the procedures for evaluation

These terms begin the work of theorizing writing assessment in composition's terms. Theories provide explanatory schemes: systematic statements of principles that constitute a perspective, not unlike a Burkean terministic screen, more or less insightful depending on the context. Composition studies has labored for the better part of a century with an explanatory scheme for writing assessment that has little to do with what we know about literate ability and that does not or cannot acknowledge the value of commonplace practices in the discipline. The vocabulary I offer here draws from theories and practices valued within composition studies. The terms “meaningful” and “ethical” suggest one possible way to structure and analyze writing assessment that coincides with what compositionists already know about writing and learning to write.

PRINCIPLES IN PLACE

“What compositionists already know about writing and learning to write” covers an extensive and disputed territory, and I do not pretend that the principles I am exploring here could thoroughly address such an expanse. Within the discipline, however, there exist a limited number of principles about which, I would argue, compositionists agree. One such, as I discuss in chapter two, constitutes the heart of the contextual paradigm of literacy: context is integral to the meaning and value of any literate act. This contextual paradigm is arguably a subset of the broader paradigm of social constructionism—an epistemological position influencing disciplines throughout the academy and one generally accepted by compositionists.⁴⁰ Social constructionist thought is based on the notion that knowledge is constructed in social settings, by persons in context. Although this epistemology is hardly uncontested, scholars in composition studies generally accept its central tenet. Because of this widespread endorsement, it serves as a reasonable starting point for developing theoretical principles for writing assessment.

For those unfamiliar with the concept, here is a quick summary. Social constructionist thought was introduced to composition studies primarily through the work of Kenneth Bruffee in the mid-1980s. In a series of articles and interchanges in *College English*, Bruffee argues that

A social constructionist position in any discipline assumes that entities we normally call reality, knowledge, thought, facts, texts, selves, and so on are constructs generated by communities of like-minded peers. Social construction understands reality, knowledge, thought, facts, texts, selves, and so on as community-generated and community-maintained linguistic entities—or, more broadly speaking, symbolic entities—that define or “constitute” the communities that generate them. (1986, 774)

Social constructionism is first and foremost an epistemological position, offering “assumptions about the very nature of the known, the knower, and the discourse community involved in considering the known” (Berlin 1987, 3). Social constructionism stands in primary contrast to objectivist and subjectivist epistemologies. Where social constructionism foregrounds the ways in which understanding is shaped and assembled by those doing the understanding, an objectivist or positivist epistemology finds knowledge and reality to be outside of the knower, independent and immutable. The objective knower’s job is to find—not make—knowledge, and Truth is the ultimate goal. The subjective knower finds Truth inside herself or through introspection, and its correlation with an external reality is largely immaterial.

This idea that knowledge is constructed and not discovered—either externally or internally—has had a significant impact on writing instruction. Early work on the consequences of social constructionism was undertaken by James A. Berlin, most notably in his 1987 monograph *Rhetoric and Reality*. Berlin argues that an objectivist epistemology results in a “current-traditional rhetoric” where Truth exists prior to language and thus language is at best a transparent medium and at worst a distortion. Writing in a current-traditional model, then, is a matter of proper precision, clarity, and form so as to reveal Truth. This paradigm, he points out, has dominated the twentieth century. A subjectivist epistemology is no less troubling for composition studies. If all knowledge is gained internally, writing cannot be taught. Writing in this model becomes largely a matter of empowering the authentic voice within each student, the success of which only the writer herself can judge. The result is “writerly” prose, prose that may be meaningful to the writer, but which ignores the needs and wishes of the audience.

A social constructionist epistemology, which results in what Berlin calls “transactional rhetoric,” “sees truth as arising out of the interaction of the elements of the rhetorical situation” (1987, 15). Berlin favors the “epistemic” variety of transactional rhetoric, which argues that truth is constructed in the interaction among writer, reader, language, and material reality. Of these elements, language is paramount; the other three elements are considered “verbal constructs” (1987, 16). Writing in this paradigm is a means of fostering this interaction in the construction of situated and contingent meaning. While I am necessarily oversimplifying Berlin’s work, which itself has been accused of oversimplification, his descriptions of the differences among these three epistemological positions provide a reasonable overview—or construction—of the primary paradigms affecting writing instruction in the twentieth century.

Social constructionism has been the dominant paradigm for more than a decade (and nearly two), in composition scholarship if not in practice.⁴¹ That is, scholars more readily accept the notion that knowledge—and by extension, writing—is socially constructed, than that knowledge resides in material reality apart from human and linguistic perception or that knowledge is the property of the autonomous individual. Instead of challenging the idea of social construction, scholars argue about what the construction looks like, how community works, and the consequences of such an epistemology. Kurt Spellmeyer for example, challenges Bruffee’s version of social constructionism on a number of grounds, from its avoidance of the problem of dissent to its inherently contradictory assertion that community-developed knowledge is disinterested (1993, 155–92).

Spellmeyer does not, however, reject the basic premise that knowledge is socially constructed.

In my own development of theoretical principles for writing assessment, I rely on this same premise at the same level of generalization. While there are a number of complications and directions for further research suggested by the complexities of social constructionism in composition, some of which I briefly discuss in my conclusion, this point of disciplinary agreement provides an appropriate starting location for establishing principles for writing assessment. Social constructionist thought provides a bridge between the concepts of “meaningfulness” and “ethics” on the one hand and the discipline of composition studies on the other. While neither term in and of itself bears a distinctive mark that would tie it to composition studies, both suggest a relational or collective value structure indicative of the same social constructionist impetus that gave rise to the contextual paradigm of literacy. “Meaningfulness” and “ethics,” that is, bespeak social values: the terms suggest neither objectivity nor subjectivity, but social reality constructed by persons in concert and in specific settings.

And certainly, the application of social constructionism to writing assessment is not without precedent. Social constructionist thought already serves as the epistemological base for the development of the more progressive guidelines for assessment. The authors of the NCTE and IRA *Standards*, for example, draw attention to the number of educators, researchers, parents, and legislators who participated in the development of the final document. While the authors never claim full consensus, they state that in “reflect[ing] the many different voices, interests, and concerns of these diverse contributors” the *Standards* “captures the essential goals of English language arts instruction at the turn of the century in the United States of America” (NCTE and IRA viii). The process of integrating the positions of numerous participants derives from social constructionist principles about the ways in which knowledge develops via collaboration and exchange, a process that makes little sense in an objectivist or subjectivist epistemology. Similarly, both the CCCC Committee on Assessment (1995) and Brian Huot (2002) argue that evaluation occurs in a social context, that reading and writing are social acts, and that meaning is made through social engagement. Both Haswell (1998; 2001b; 2001a) and Broad (2000; 2003) take the social nature of reading and writing as a basis for their work; they almost take it for granted.

Beyond precedent and beyond correlation with existing paradigmatic influence, there is at least one more compelling reason to actively adopt social constructionism as an epistemological base for principles of

writing assessment: social constructionism promotes radical contingency. If knowledge is dependent on community and context, as it is in social constructionism, whenever either of these elements changes, knowledge must be reconstituted. Changes occur inevitably as faculty, curricula, and programmatic goals evolve; moreover, the specific contexts of differing institutions also require assessment methodologies to differ. Contemporary writing assessment practice, however, has been founded on the criteria of validity and reliability, which tend to either erase or ignore differences in context. In a social constructionist paradigm, assessment principles and procedures—including those I am advocating here—could not be simply codified once and for all; they would be entirely contingent on context and community. Moreover, should social constructionism lose viability, the entire basis for writing assessment would require reconsideration, a rethinking that an objectivist paradigm has not been able to accommodate. The radical contingency of social constructionism encourages revisionary practice such that all assessment procedures should be continually under review to ensure that they are productive, necessary, and appropriate, that they are meaningful and ethical.

MEANINGFUL ASSESSMENT

To say that an assessment is “meaningful” is to claim that it has significance. In an objectivist paradigm, the significance of an assessment would be tied to (ostensibly) universal ideals of writing ability. In a contextual constructivist paradigm, however, significance would be dependent on the scene in which the assessment takes place. That is, an assessment would be meaningful in relation to the needs and values of those within the assessment context. Those needs would, at least ideally, direct the “purposes” of the assessment and the values would shape the object or “substance.”

Thus, the term “meaningful” as I am developing it begins with notions of purpose and substance. The idea of purpose as an element of meaningfulness draws attention to what evaluators expect an assessment to accomplish and for whom. Substance serves as a partner term, focusing on the content or subject matter of an assessment. These qualities shift the evaluator’s gaze away from the means of testing—which has historically been the focus—and toward the reasons for and the object(s) of assessment which, I would argue, constitute a more appropriate center of attention for assessment. In relying on these terms, I am implicitly arguing that before any assessment takes place, those developing the procedure

need to understand as thoroughly as possible why they are assessing and what information they are attempting to gather, or more accurately, to construct. Without this understanding, assessment practices easily become meaningless.

Of the two, purpose seems the more fundamental issue. Because it focuses attention on the intention of the assessment, on why the assessment is productive or necessary or appropriate, purpose generates questions about why evaluators want to know the information that the assessment will ostensibly produce. Since assessments provide information about the subject at hand (in both the person and subject matter senses of the word), the reasons for assessment should first be grounded in the need for information about the subject: which course, for example, the student should be placed in; whether the student is ready for second semester composition; if the student has shown progress and the nature of that growth. Without a purpose, there is no reason to assess in the first place, so questions about the substance—what to assess—never arise.

There are also secondary purposes for assessment practices, such as faculty development and curriculum reform, which should also be explicit and integrated into any given assessment project. The distinction I am making between primary and secondary purposes has to do with the difference between necessary information and positive side effects. While both are important, the need for information is the direct purpose for assessment and should take precedence over anticipated or desired benefits. One way to think about this difference is to consider the effect of the assessment on various participants, and students stand a greater chance of being harmed by any particular assessment procedure than do faculty. For example, in the large-scale proficiency assessment at the University of Wisconsin-Milwaukee described by Mary Louise Buley-Meissner and Don Perkins (1991), students who fail the proficiency exam—a number which includes a high percentage of minority and ESL students—cannot achieve junior standing and cannot graduate.⁴² Those who fail suffer immediate negative consequences, and thus the assessment holds the potential to materially harm students. Because they can be immediately and directly harmed, assessment purposes that address their needs, such as proficiency requirements or placement, should be paramount. Purposes that center on faculty, such as faculty development, would be supplementary.⁴³ The purposes affecting those with more at stake in the results of any specific assessment should take precedence over the purposes affecting those with less.

FIGURE 3

*Questions of Purpose Raised by the Principle of Meaningfulness***Purpose: Why the assessment is productive or necessary or appropriate***Primary Purpose*

- What reasons are there for assessing? Why evaluate?
- What effects should the assessment to have on students, at least ideally?

Secondary Purpose

- What effects should the assessment to have on the curriculum, at least ideally?
- What effects should the assessment to have on the faculty, at least ideally?

Purpose—the reasons for assessment—may be logically prior, but substance is likely the more complex and difficult issue.⁴⁴ When evaluators develop the substance of an assessment, they are determining what they want to know and what they need to do to satisfy their defined purpose(s). The former, I would argue, should be a function of disciplinary knowledge; that is, the subject matter of an assessment should draw on an established body of knowledge, particularly in the case of large-scale assessments, which draw together multiple local contexts such as classrooms. Substance, then, would include the subject matter to be assessed and the assessment format(s) that will best yield the necessary information. These substantive elements focus explicitly on the relationships between the information to be constructed (the results desired and produced) and the subject producing the information (the student). These elements constitute primary substance.

It is also possible to think about a category of secondary substance that would focus on what the evaluators need to do in order to achieve their secondary purposes. If, for example, curriculum reform is a secondary purpose, the evaluation designers should ask how the assessment could contribute to the desired changes. For example, in a move from a modes-based first-year curriculum to a rhetorically based one, a change in the subject matter of the test from forms to rhetorical principles will encourage a concomitant change in course design. While some of the concerns about what an assessment does or should do are ethical—a subject I will consider in the next section—some are substantive. That is, I am suggesting that the substance of the procedure is an appropriate concern for those developing assessments.

FIGURE 4

*Questions of Substance Raised by the Principle of Meaningfulness***Substance: What assessors want to know and what they need to do to satisfy their defined purpose(s)***Primary Substance*

- What is the subject matter students should know or what abilities should they be able to demonstrate?
- What disciplinary knowledge supports this choice of subject matter and/or abilities?

Secondary Substance

- What secondary effects or changes are desired through assessment?
- What assessment processes will achieve these effects or changes?

These questions of purpose and substance, both primary and secondary, are not independent of one another. I have already suggested ways in which primary and secondary substance might address primary and secondary purposes, respectively. In addition, I would argue that for meaningful assessment, purpose and substance, taken together, should exhibit consistency. Take, for example, a portfolio assessment designed both to evaluate students for university requirements of proficiency and to evaluate the writing program's effectiveness. While the same object of assessment—student portfolios—can be used for both, the assessments themselves have different purposes—determining student achievement and determining writing program achievement—and, consequently, should have different substances. In the case of establishing student achievement, the substance of the assessment should focus on the work of individual students and thus would include judgments about students' abilities to work with concepts like rhetorical awareness, academic conventions, and critical thinking. In the case of program evaluation, the substance should focus on the program as a whole and thus would include determinations about the appropriateness and effectiveness of the criteria for judging student abilities in relation to the university's goals for student writing ability. Students might only produce one portfolio, but the evaluation of the individual student's work and the evaluation of the program using that portfolio should generate separate assessments. One assessment procedure would be hard pressed to *simultaneously* address both sets of purposes; the teachers who might read the portfolios, for

example, would have difficulty, *in the same assessment*, judging the legitimacy of the criteria they are using. The two purposes, that is, require inconsistent, though not necessarily conflicting, substances and should thus require different assessment designs.

Continuity, however, cannot ensure meaningfulness. For example, the use of multiple-choice tests for placement purposes can easily be described as consistent. Their purpose is to place students according to their needs, and the correlation between grammatical knowledge and writing ability has been well documented. While there are arguments that counter the existence and value of the correlation, such assessments exhibit an internal consistency. Internal consistency among purposes and substances, without reference to some relevant body of knowledge independent of the assessment, is insufficient by itself to establish a meaningful assessment. I see this as a primary weakness in educational measurement theory. The attempt to develop universal principles for assessment ignores the particular demands and values that adhere to particular subject matters. For some disciplines under some conditions, the principles of educational measurement might adequately reflect the demands and values of the discipline (e.g., objectivity and universality), or alternately, the discipline might hold enough institutional power to ensure that its demands and values are addressed whether or not a correlation exists with educational measurement theory. Historically, however, neither has been the case for composition studies at any level of schooling.

My point here is that in large-scale writing assessment, and indeed in any assessment that claims broad significance, meaningfulness should be a paradigmatic function. In general, the “meaning” of any concept—and even whether or not something is considered “meaningful”—can be determined only with reference to some set of assumptions about what is valuable, knowable, reasonable, and so on. Where the application of meaning is broad, as in large-scale assessment, those assumptions would need the status of paradigm in order to carry weight beyond the local context. Because large-scale assessments are widely applicable to a large number of people whose immediate contexts differ somewhat (such as the individual classes in a first-year writing program), the current dominant paradigm relevant to the object of assessment should determine the meaningfulness of the assessment. Moreover, because paradigmatic shifts in disciplinary knowledge require substantial agreement among the members of disciplinary community, they do not occur often. Paradigmatic knowledge, then, provides a reasonably stable, but not immutable, reference point outside

any given assessment, as a sort of reality check in the form of primary substance. In addition, paradigmatic knowledge is sufficiently flexible to allow a wide range of assessment practices. Instead of prescribing a specific agenda, such knowledge provides a conceptual framework that guides the work in a particular field, including the assessment of students.

The need for paradigmatic correlation between scholarship and assessment that I am advocating here helps explain the troubled reception of the NCTE and IRA *Standards for the English Language Arts*, even beyond the problem of defining a “standard.” As I argue in the last chapter the two professional groups relied on the contextual paradigm of contemporary literacy theory to develop their standards. Critics of the *Standards*, however, relied on a technocratic notion of literacy, as indicated, for example, by their calls for benchmarks that would enumerate skills and consequently atomize literate ability into teachable and testable units. In relying on an older paradigm of literacy, unsupportable in light of current disciplinary knowledge, critics called for standards and implicitly for assessments that would ultimately be meaningless for evaluating literate ability according to those with expertise in the field of writing instruction. In order to develop meaningful assessments, those designing procedures should be familiar with the relevant current literature and responsible for applying that scholarship.

In the case of writing assessment, the object is written evidence of literacy, and as I argue in chapter two, the body of contemporary literacy scholarship evinces a contextual paradigm. For a large-scale writing assessment procedure to be meaningful, as I am developing the term, the primary substance—what evaluators want to know—would draw on that scholarship. For example, contextual literacy attaches meaning in writing to the location and purpose of that writing, so contextual assessment would involve evaluating writing for its ability to respond to rhetorical situations. The paradigm also claims an integrated view of writing and consequently would encourage the assessment of whole writing tasks which treat literate ability as a situated act rather than as discrete skills or pieces of information, as would be preferred in a technocratic paradigm. The CCCC Committee on Assessment and Brian Huot suggest several other assessment “content” issues in line with a contextual view of literacy—although neither explicitly characterizes them as such—including the need for a variety of texts that address a variety of contexts (CCCC Committee on Assessment 1995, 432) and the reliance on situated judgments about the success of a piece or pieces of writing without reference to abstractions such as “writing quality” (Huot 2002, 102).

In addition to shaping the content—or primary substance—of assessment, the specific context in which the assessment occurs organizes the meaningfulness of the particular assessment practice, serving as a sort of container that both shapes its form and limits its scope. Context influences the form of any particular writing assessment in some obvious ways. The requirements of the program and the curriculum, for example, should determine both purpose and substance, so that an assessment aimed at determining proficiency with conventions of academic writing should look different than one aimed at determining facility with a range of rhetorical situations, both academic and non-academic. So, too, should assessments differ depending on the experience and qualifications of the assessors. A program in which first-year composition is taught entirely by tenured and tenure-track English department faculty, regardless of scholarly expertise, for example, is likely to result in a different set of evaluative criteria than one in which the faculty consists entirely of adjuncts and teaching assistants.⁴⁵ When I say that context shapes assessment, I am referring to the ways in which the pedagogical and practical demands of the specific situation influence the form and the meaning of the assessment. This connection between literacy instruction and assessment means that the instructional needs and goals of the context define the substance of the assessment and consequently its meaningfulness. Moreover, as the features of this context change, assessments should be reconsidered in light of those changes.

Just as the context shapes assessment, it also contains it. Contexts, that is, are boundaries as well as locations. This would explain why assessments do not migrate well from context to context. No two contexts are identical; what would be meaningful in one would not necessarily remain meaningful in another. This is not to say that each assessment must reinvent the wheel, but rather that no assessment should be indiscriminately grafted onto a different context. Models from other contexts can serve as guides, but they should be considered in light of the requirements specific to the current context. Nor can a single assessment procedure successfully serve multiple purposes. As purpose changes, so does the substance of the assessment, making it a different evaluation that is subject to the influence and limitations of its own context.

If context shapes and limits assessment as I am suggesting, then control of any assessment procedure should remain within that context as well in order for the results to be meaningful. Only those within the specific literacy instruction context can understand the situation sufficiently to ensure the continuity among purpose, substance, and instruction and

assessment. This means that specific assessment procedures should not be mandated from without, although general requirements for assessment certainly can be. The authority over assessment, however, is not without obligation. Those in charge of the local scene have a responsibility not only to apply principles from the relevant scholarship, but also to consult those outside the immediate context who have a stake in the assessment both to gather information about their needs and expectations and to communicate their principles and criteria. For example, writing program directors have a responsibility to explain their criteria for assessing exit portfolios to the relevant institutional committees and administrators when that assessment serves a university-wide requirement for writing proficiency. Control does not mean autonomy.

The need to address the concerns of extra-contextual stakeholders points to one final issue in making assessment meaningful. As I mentioned earlier, the idea that assessment must be meaningful begs the question “meaningful to whom?”—the answer to which influences both purpose and substance. For example, the reconsiderations of assessment of the last decade I have been discussing all argue—either explicitly or implicitly—that writing assessments must be meaningful first to students and then to teachers. The CCCC Committee is particularly adamant that writing assessments at the post-secondary level serve pedagogical ends first. Huot also places pedagogical purposes before those further from the classroom, and this claim is implicit as well throughout the NCTE and IRA *Standards*. Meaning, however, can vary according to the stakeholders’ relationship to the assessment, so that what teachers and students find meaningful will not necessarily translate to administrators or to funding agencies. This variation has real consequences for design and implementation of assessment practices and for use of results. The intended meaning of an assessment—pedagogical or otherwise—should be understood as thoroughly as possible prior to the assessment itself. Moreover, its meaningfulness within the immediate assessment context needs to be accessible to those outside of that situation, too.

That is, in order for an assessment to be “meaningful” as I am describing it here, its purposes should be clear and its substance defined so that those involved in the assessment—especially those who could be harmed by it and those who need to use it—fully understand it. Students and deans alike need to be able to understand not only the assessment procedure but also the reasons an assessment is necessary and what it consists of. Moreover, all parties, but most particularly those making the assessment judgments, need to understand how the assessment stands

in relation to what is known and accepted about writing and learning to write. Paradigmatic knowledge balances meaningfulness within a particular context with the predominant thinking of the larger community of experts. The balance, however, is not stable. Paradigmatic knowledge changes; contexts change far more rapidly. Both conditions reinforce the contingency of meaningfulness. Without reference to disciplinary knowledge and without consideration of the particular context—without references to contingencies—assessments lose their meaning. Those responsible for assessments should continually reevaluate their procedures, particularly in terms of the goals and objects of assessment—its purposes and substance—to ensure that their assessments remain meaningful.

ETHICAL ASSESSMENT

“Meaningfulness” as an organizing term provides direction for developing purposeful and substantive assessments, but it only indirectly suggests how to conduct them, either practically or ethically. Practical concerns, such as funding, affect the conduct of assessment primarily at the level of local administration and are perhaps best addressed through policy statements such as the CCCC Committee’s (1995) or by procedural guides such as White’s *Teaching and Assessing Writing* (1994c). Ethical issues, however, are matters for theoretical consideration where a conception of “ethics” can suggest systematic criteria for addressing the broad political and social issues surrounding assessment. As I see it, an ethics of assessment ideally would focus on the relationships among those affected by the assessment: students, teachers, administrators, legislators, parents and so on. Ethical principles would then provide a way of organizing and understanding the conduct of the participants, not only in terms of their behavior or the procedures for evaluation, but also in terms of the broader social and political implications for writing assessment.

The concern for ethics has most frequently appeared in the narrowed guise of “fairness.” White, in particular, treats the desire to be fair as a paramount concern in writing assessment: “Reliability is a simple way of talking about fairness to test takers, and if we are not interested in fairness, we have no business giving tests or using test results” (1993, 93).⁴⁶ Historically, “fairness” has indicated a desire on the part of teachers and test administrators to be just and impartial in their treatment of students. “Reliability” has provided a set procedural criteria for ensuring that all students would be treated equally, and thus has been roughly equated with “fairness” in the literature on writing assessment.⁴⁷ Arguably, most, if not all teachers and administrators want to be fair in their assessments: it

would be difficult to find any educator who would say otherwise. Fairness was, after all, one of the impulses that led Horace Mann to praise written examinations as “impartial” and “just” and to implement them in the first place (1845, 330–31).

However, “fairness” seems too limited to provide a sufficiently articulated set of principles for conducting writing assessment projects. While the ideal of fairness does succeed in drawing attention to the effect of assessment on students and on the relationship between teachers and students, this is not the sum-total of the social and political relationships at work in evaluative situations. Limiting the relationships to these two parties would ignore a great deal of the context, particularly for large-scale assessments. Perhaps more importantly, “fairness” implies a decontextualized even-handedness indicative of an objective epistemology. To be fair is to be “unbiased” and “impartial.” Instead, I am suggesting the term “ethics,” which would encompass the motivation behind the use of “fair”—assessment that treats all individuals as equitably and justly as possible—but which would add an emphasis on the responsibilities and accountabilities among all parties involved. The project of assessment generally is to establish norms by some means and to evaluate some object in terms of those norms. While “meaningfulness” governs the properties of the object and the goals of the project itself, “ethics” governs the relationships among participants, including the process of establishing and revising such norms and the responsibilities entailed in any assessment project. Thus, ethics draws attention to assessment in practice.

Under the rubric of “meaningful” and “ethical” assessment, once the purpose and object of assessment have been determined, those developing an assessment would turn their attention to the evaluative procedure itself, and early on to the development of criteria for evaluation.⁴⁸ These criteria, as should be clear from my discussion of meaningfulness, would derive from the dominant paradigm and should adhere to the purpose and substance of the assessment. The process for developing them should also derive from that paradigm, and specifically would reflect a relevant epistemology. That is, the manner in which criteria for assessment are determined within any given paradigm should reflect corresponding theories about how knowledge is made within that paradigm. In a positivist paradigm, for example, norms are established through reference to “facts” and other empirical evidence treated as if they were objective. This is why “reliability” could function as an ethical proxy in positivist assessments. The best way to be “fair” in such situations would be to treat everyone the same, and the best way to do that would be to refer all decisions

to objective reality. If the reality were indeed objective, the results would be reproducible, and reproducible results would ensure that everyone was treated objectively. Insofar as ethics would focus on consistent and principled conduct, this principle would require theoretical consistency between the paradigm governing the object of assessment and the epistemology governing the procedure by which the process of assessment occurs. The social constructionist paradigm at work in composition scholarship would encourage the development of norms and procedures for assessment by those in the community and context at hand which are relevant to their specific situation. I look at some of these models in the next chapter.

Scholarship in critical theory during the last decade or so suggests an ethical structure which relies on a social constructionist epistemology and which is suitable to a contextual paradigm of large-scale writing assessment. Communicative ethics, as associated primarily with the work of Jürgen Habermas and Seyla Benhabib, focuses on the process by which ethical and/or moral norms are established. According to Habermas, communicative ethics—also called “discourse ethics” or “communicative action”—focuses on how moral norms attain validity through a dialogic process of “moral argumentation” (1990b, 197).⁴⁹ Benhabib describes communicative ethics in terms of the central question it asks: “Instead of asking what an individual moral agent could or would will, without self-contradiction, to be a universal maxim for all, one asks: what norms or institutions would the members of an ideal or real communication community agree to as representing their common interests after engaging in a special kind of argumentation or conversation?” (1990, 330). Communicative ethics is premised on the notion that communities can (and should) reach agreement about ethical judgments through discursive or, I would suggest, rhetorical means. In the context of an “ideal speech community” where all parties have equal authority and opportunity to speak—a problematic notion which presents some of the same difficulties as Guba and Lincoln’s work and which I will return to in a moment—community members develop their normative ethical principles through motivated argumentation. Within such a structure, Habermas claims, “one actor seeks *rationally* to *motivate* another” rather than “to *influence* the behavior” through threats or promises (1990a, 58).⁵⁰ Theoretically, the process of argumentation in communicative ethics provides each person with the opportunity to participate fully; moreover, such participation necessarily entails reflection not only on the individual’s own position(s), but also on the arguments of other participants. Thus, communicative

ethics is an intersubjective rational engagement via language in which all parties understand all other parties to have the same participatory obligations, opportunities and rights (Benhabib 1986, 333).⁵¹

Because “the theory of communicative ethics is primarily concerned with norms of public-institutional life, or with *institutional justice*” (Benhabib 1986, 283), it is particularly appropriate for a theory of writing assessment. Communicative ethics focuses on the process of justifying norms, such as standards and criteria for evaluation, in structured, public discourse communities, which would include universities, departments, districts, and classrooms. Other theories ground ethical decisions in the thought-experiments of autonomous individuals or in empirical data, and consequently move ideals of ethical behavior outside of intersubjective rhetorical consideration and community action and toward universal and formal principals that stand outside any given communicative act. Communicative ethics, however, makes the development of ethical norms part of a dialogic process and thus offers the opportunity for participants to shape the principles at work in a discourse community. Benhabib suggests that communicative ethics can be understood as “a procedural ethics”: “In fact the very thrust of the theory of discourse is to show that the idea of truth entails that of rational consensus, but that such rational consensus can only be explicated *procedurally* by defining the strategies and modes of argumentation through which it can be arrived at” (1986, 288). Communicative ethics, that is, treats norms in process and focuses on the means by which agreement is reached. Instead of dispensing personal or empirical foundations that leave little room or reason for discussion, communicative ethics actively incorporates the process of establishing of intersubjective norms into the overall process of developing assessment projects.

Communicative ethics is not without its critics and its conceptual difficulties. Most criticisms seem to have been leveled at two areas. The first is that communicative ethics tangles moral judgment with rational argumentation, and thus ignores independent will. The criticism is not particularly relevant to my use of communicative ethics. Writing assessment, as I am conceiving it, already accepts the slippage between rationality and morality that results in values and judgments following from those values, so whether or not there is confusion on this front in communicative ethics is not particularly a sticking point here.

The second criticism, however, is. In communicative ethics proper, the notion of the ideal speech community is both key and problematic. For Habermas, the ideal speech situation necessitates the possibility of

full, equal and unfettered participation by all concerned (1990a, 89), a description not unlike that offered by Guba and Lincoln in their calls for all stakeholders to participate as equals in evaluation (1989, 40–41, 51–57, 267).⁵² The problems with this ideal leap off the page. The certainty that consensus should be the goal of such dialogue and the pretense that power differentials could evaporate—two primary difficulties—are already familiar to those in composition studies through discussions in the mid-1980s about the value of consensus and through more recent discussions about the training of evaluators to reach consensus in holistic scoring. Benhabib—who favors Habermas’s position—reviews the criticism from within critical theory and argues that the ideal speech situation is actually an illustration of communicative ethics and not a precondition for it (1986, 289). That is, it “describes a set of rules which participants in a discourse would have to follow . . . and a set of relations . . . which would have to obtain between them, if we were to say of the agreement they reach that it was rationally motivated, dependent on the force of the better argument alone” (1986, 285).⁵³ Habermas and Benhabib are theorizing about the dialogic construction of ethical decisions at the level of the general populous. And when Guba and Lincoln apply their similar structure to qualitative research, they bring all stakeholders to the table under theoretical conditions that substantially resemble Habermas’s ideal speech community.

In both cases, the dialogic situation under consideration aims at a decision that differs fundamentally from that made in writing assessment situations. Neither moral dilemmas nor research designs require that the end result be a definitive answer delivered in a limited time frame. Guba and Lincoln acknowledge this in the idea that fourth generation evaluation is only “paused” and not concluded (1989, 74). Habermas, instead, differentiates between the justification of general ethical principles and the application—or “contextualization”—of such principles and reaffirms that his focus remains on general principles (1990b, 206–07). The assessment of writing is a case of the latter, contextualized situation, and because of the nature of the decision to be made, the situation calls for an explicit and expeditious conclusion rather than a pause.

Habermas and Benhabib both argue for the universalism of ethical principles, which Benhabib defines as “the principle that all human beings, by virtue of their humanity, are entitled to moral respect from others, and that such universal moral respect minimally entails the entitlement of individuals to basic human, civil, and political rights” (1994, 173). My guess is that most of us in composition studies would accept such

a universal condition—in spite of its acontextuality—but note that this is a moral principle. There is a division between the search for morality taken on by Habermas and Benhabib and sound principles for developing assessment procedures that I am taking on here. Universalizing the meaning of writing assessment has had damaging consequences for writing instruction; universal ideals of ethical conduct, I would argue, do not carry the same repercussions: people do not fail out of being human.

These differences of condition prompt me to argue that the situation of writing assessment is not an *ideal* speech situation, but rather a *limited* speech situation. That is, because of the nature of the decision to be made and the time constraints of assessment—if only in the form of the date grades must be turned in—the situation of writing assessment is sufficiently restricted in scope and purpose to warrant limitations on the conditions of communicative ethics proper. Specifically, I would argue that participation be grounded in expertise and accountability. Those who have expertise regarding the specific decision and who are responsible for performing the assessment itself are the ones who should hold authority in the decision-making process and in the evaluation itself. This means that at the level of classroom assessment, teachers should be the primary participants, while at the level of program assessment, writing program administrators, faculty and other officials should take part.

This does not exclude other participants, but rather relegates them to a secondary role. Those who are not directly accountable for assessment results but who are affected by those results should participate in assessment design where appropriate—a condition determined by contextual constraints—and should be entitled to explanations about the actual assessment process, but they should not have undue influence. One way to think about the distinction I am suggesting is that those to whom evaluators are accountable (including deans and students) are limited participants, while those who are accountable are full participants. Limited participants need to know that the assessment is meaningful, that is, what purpose(s) it serves and what substance(s) it addresses. Full participants also need to know *how* to perform the actual assessment, and consequently need to participate in the day-to-day dialogue.

In limiting the ideal speech situation, I am, admittedly, circumventing the very practical problem of inclusiveness. However, it would be neither possible nor productive to fully include all stakeholders in every assessment decision from the conceptual design through the final evaluations. Right now, as a result of our national push for testing, we are seeing plenty of inappropriate decisions made by those without appropriate expertise.

If nothing else, these decisions and their effects provide us with good reasons to limit authority in assessment situations. I have argued that context serves as a boundary on meaningfulness; I would now add that context also serves as an ethical boundary. Not everyone who has a stake in an assessment will be held equally accountable for the norms and the decision-making process; so, I would argue, not everyone should have the same authority. Procedural questions about whether or not everyone has equal opportunity to speak and act in ethical dialogue are not as significant in writing assessment situations as are the appropriateness and value of the decisions. The principles of communicative ethics are applicable to writing assessment if we add the caveat that speech communities are grounded in specific contexts. Norms for conducting writing assessment arise from dialogue within a situated community. Consequently, as that community and/or its context change, norms are subject to change. The process is thus contingent, always subject to reconsideration and revision.

There are, however, more compelling reasons to apply the principles of communicative ethics in developing ethical writing assessment. For example, its practice reflects a social constructionist epistemology, structuring value judgments without resorting to either objectivity or subjectivity. Instead, ethical conduct is determined dialogically by the community in concert as part of the process; we see precursors of this work in Broad's study of City University (2003). Moreover, intersubjectivity, the condition of dialogue, functions reflexively: in the process of argument, the parties, at least ideally, continually reconsider their own positions in the light of the others' arguments. Thus, the self-reflective nature of communicative ethics parallels the emphasis on self-reflection in contemporary composition pedagogy.⁵⁴

The focus on and function of dialogue in determining ethical conduct restructures the ideal of "fairness" in writing assessment. Under positivism, fairness is ostensibly ensured through reproducibility. I would like here to examine Brian Huot's argument about this equation a bit. Writing is necessarily a varied act and consequently so is writing assessment, which is at least in part, an act of reading. Because writing is a contextually dependent interchange among writer, reader, language and subject matter, changes in any one of these elements effect the entire structure of reading. This is why judgments about writing ability change from time to time. Fluctuations in an evaluator's perceptions about the relative importance of certain rhetorical features, for example, can easily influence her judgments of a particular essay, even if she has evaluated it before. The restrictions necessary to establish reproducible results in writing assessment are

an attempt to prevent such changes, and they narrow the range of writing and reading so much so that they do not result in a fair product, that is, in a legitimate representation of student work or the reading of it.

In communicative ethics, the ideals of “fairness” would be based, not on reproducibility, but rather on full consideration—that is, on as complete an analysis as necessary to reach a decision, with the understanding that there are restrictions on time and resources. The principles of communicative ethics—with the caveat of a limited speech community—as applied to writing assessment provide that the process for assessment be dialogic, that the goal be consensus about the assessment decision, and that rational argumentation be the means by which such consensus is reached. The conditions of dialogue and rational argumentation promote assessment decisions based on extensive attention to and reflection on not only the principles of meaningfulness—purpose and substance—as determined by the community, but the relationship between those principles and the actual piece(s) of writing to be assessed. This full consideration can take into account the variety of writing—including multiple pieces of writing or multiple drafts—and the distinctiveness of particular contexts without being unfair because *all* assessments would receive the same full consideration, even though the specifics considered might vary from assessment to assessment.

An ethical assessment, as I am developing the term, would occur first and foremost in dialogue. Evaluators would ideally develop criteria through argument and discussion about the needs of the program and the purposes of the specific assessment. Limited participants, such as deans and legislators, would be invited to join or to forward their concerns and priorities, but the final decisions would rest with the full participants, particularly teachers and program administrators. The application of these criteria would also occur in dialogue. Evaluators would discuss each student’s work as thoroughly as is necessary to reach a community assessment decision. The procedures I am describing here stand in stark contrast to holistic scoring, where evaluators read quickly and independently, without discussion, except in “norming” situations where individual readers who deviate from the norm are expected to modify their decisions.

Here I am positing a dialogue that is in many ways as idealistic as Habermas’s inclusive ideal speech community. Like Benhabib, however, I would argue that this is an illustration of what ideal ethical assessment might look like. I am arguing for a limited speech community, not because the ideal version is too utopian, but rather because I find the inclusiveness of the ideal version both unnecessary and counterproductive to the

context of writing assessment. In addition, such inclusiveness strips compositionists of their expertise, a consequence I have underscored before and to which I will return shortly.

A limited speech community, however, would require increased attention to accountability. As I have argued, those parties who assess other parties have a responsibility for conveying the meaningfulness of the assessment. This includes articulating the procedural principles, particularly when all parties do not have equal access to the dialogue. Students, for example, should have access to information about the criteria for assessment and the substance of the dialogue surrounding the evaluation of their writing. Similarly, administrators should be informed about the particulars of purpose and substance most applicable to the assessment process as a whole. In principle, heightened accountability means that those privy to specific dialogues should be able to explain the salient and relevant aspects of those dialogues to those who are affected but who are not full participants.

FIGURE 5

Questions Raised by the Principle of Ethics

Community: Who is involved in the decision-making?

- Who has a stake in this assessment?
- How accountable is each party for the assessment decisions?
- Based on their accountability, what role should each play in the design of the assessment?
- What can each party contribute to the assessment? What is their area of expertise?
- Based on their potential contributions, what role should each play in the design of the assessment?

Process: What procedures will the assessment requirement?

- What is the decision to be made?
- What procedures will result in the fullest consideration possible for all concerned?
- What limitations and restrictions inhibit full consideration?

Ethics governs the relations among participants as they apply the principles of meaningfulness in concrete assessment situations. It is, thus, a function of assessment in practice. Ethical assessment, through the ideal of dialogic argumentation in a limited assessment community, I have argued, gives students full consideration and remains accountable not just to those students, but to all stakeholders. Like the qualities of

meaningfulness, however, the procedures for ethical assessment are subject to revision as the context changes. This is one of the functions of accountability, which includes not only the notion that an evaluator should be able to explain herself, but also that her decision should be a responsible one in her own estimation. Ethics in writing assessment, then, refers as well to purpose and substance, linking these qualities of meaningfulness to specific persons in specific contexts with specific relationships to one another.

THE IMPORTANCE OF TERMS

In the vocabulary I am developing here, meaningful writing assessment would be purposeful and substantive, while ethical assessment would be discursive and would work toward full consideration of the object of assessment. It would be easy to argue that there is little innovative or radical about what I am proposing here. Neither term is particularly provocative when understood within the framework of social constructionist theory. Like Huot's work, my own extrapolates principles, but I am drawing more from theoretical work in literacy and composition studies than from contemporary practice in writing assessment. Furthermore, arguably, the basis for some of the general principles I have outlined here can be found as well in literature in educational testing, particularly Samuel Messick's work on validity, as Huot's arguments suggest.

However, if the imposition of "validity" and "reliability" has taught those in composition nothing else, it has taught us that power and paradigm adhere at the level of lexicon. Composition studies and writing instruction at all levels of schooling have suffered as a result of the deployment of these educational measurement criteria in writing assessment contexts. The changes, for example, in the definitions of and the qualifiers added to "validity" have not translated to actual practice, despite being the "norm" for well over a decade, nor has hermeneutics been accepted as an alternative to reliability, despite work on this principle in assessment circles since at least the early 1990s. Moreover, it is still the norm for educational measurement experts and not composition professionals to define the majority of writing assessment procedures, in spite of exceptions such as Washington State. Yet "validity" and "reliability" provide little meaningful support and have regularly proven damaging for those in composition studies charged with designing large-scale writing assessments.

I believe these principles, or something like them, can find a home in post-secondary composition studies, and from there—with sufficient

research, solid results, and strong arguments—can work their way into other forums, as I suggest in chapter five. My hope is based in part on the fact that the theory I present here does not produce readily quantifiable results, a key difference between the dominant and the alternative I offer here. Within the field of educational measurement, as Moss and Delandshere and Petrosky point out, even those who advocate more complex validity determinations still resort to numbers to justify their work. But writing does not easily or effectively reduce to numbers. When we force that reduction, we end up with information useful by legislatures and governing boards for quick comparisons, budget decisions, and public influence—none of which compositionists would consider a primary purpose of writing assessment. As Elbow (1993) has argued—persuasively, if not successfully—quantification does not help provide any truly useful information to the writer, the teacher, or even writing programs. I agree and, I believe, so would many compositionists. But we have not had a strong theoretical alternative to psychometrics that would result in qualitative judgments. The problem with using validity, even in its more complex and radical forms and even without reliability, is that its justification has been attached to quantification, and quantification does not provide what students, writing instructors, and writing programs need. It does not provide what legislatures and governing bodies need, either; they just don't know it . . . yet.

My development and presentation of “meaningfulness” and “ethics” is not, however, the end of the theoretical yellow brick road. We need the research. “Meaningfulness,” in particular, is a seriously underdeveloped concept in composition studies, as I have argued. The majority of published work on writing assessment focuses on *how* we assess, not *what* we assess or *why* we (should) do so.⁵⁵ Some projects have begun this work, including the NCTE and IRA *Standards*, the “Outcomes” developed by the Council of Writing Program Administrators (1999), and most recently Bob Broad’s study (2003). However, a need for research remains, particularly into the substance of writing assessment. Nor are ethical considerations settled: few writing assessment projects exhibit the principles of full consideration and communicative engagement I have outlined, and interactions among evaluators need further study. There is a lot of work to be done.

What I offer here is a starting point, a set of terms that draw from and are compatible with the paradigmatic principles at work in literacy and composition studies scholarship. As such, they are steeped in contextuality: context in this model is a primary defining quality for developing the

particulars of any given writing assessment situations. Moreover, these terms are dependent at the level of principle on their context. That is, because they are contingent on the contextual paradigm of literacy, they are subject to revision according to changes at the level of paradigm. Should the paradigm change entirely, as it has in the past, the entire conceptual project defined here is up for reconsideration. The kind of reconsideration I am suggesting is what has *not* happened even as the literacy paradigm has shifted from a technocratic orientation to a contextual one. Consequently, I am arguing for the contingency of these terms. They are only valuable as long as literacy scholarship espouses the primacy of context and as long as composition studies accepts the principles of social constructionism.

Composition studies needs theoretical work in writing assessment that, in turn, requires a vocabulary that embraces the values of literacy scholarship and instruction. To date, assessment theory has been grafted onto writing situations as if the terms did not matter—although compositionists certainly know otherwise. The terms I offer here—meaningfulness and ethics—continue the project of developing theories of writing assessment founded not in educational measurement but in composition studies.