

INTRODUCTION

Choosing Our Terms

On September 18, 1989, I returned my first set of graded essays. There were six Ds, eleven Cs, five Bs, no As, no Fs, and one missing paper—I still have the gradebook. The Ds weighed most heavily on my mind. In my carefully scripted percentages, this first essay was worth 5% of the total grade for the course; those students with Ds had suddenly seriously damaged their chances for an A, yet they had barely begun the course. One of them—I can still see his face—met with me for about an hour after class that day, trying to understand how he could have gotten a D. “I’m not asking you to change the grade,” I remember him saying several times, but he did want to understand how he had done so poorly. I remember painstakingly going over his paper with him, and although I do not recall the particulars of the discussion, I do remember feeling that my explanations were entirely inadequate. In class we had discussed “focus,” “coherence,” “content,” “organization,” “style,” and “mechanics,” but I am certain that these concepts floated over the class’s heads as abstractions; I am not certain that I had a clear handle on them myself at the time. Worse—or at least I thought so at the time—my plan for the course, which followed departmental guidelines, did not allow for revision after I graded a paper. The grades stood. Looking at the gradebook today, I am not surprised to see that the final grades for the class were all Bs and Cs; no one failed, but no one made an A either. Without the option to revise their essays, students would have had to exhibit a fairly high proficiency in writing prior to entering the class—a condition that would have placed them in second semester composition rather than my first semester class, or even exempted them from the composition requirement entirely. Add to that my relatively vague understanding and explanations of the criteria for evaluation, and I do not see how any of my students could have excelled.

Sometime during that same semester, I heard about portfolios as a way of deferring grading until students had had a chance to revise. While I used peer workshops in my class so that students received feedback on rough drafts prior to handing in their essays, I had never been introduced to a procedure that allowed students to use *teacher* commentary for

revision. Portfolios, I felt, would solve the worse of my two pedagogical dilemmas—and at the time I was certain that the immutability of early semester grades was a bigger problem than a vague understanding of the criteria. I wanted to use portfolios during the second semester but was not allowed to; I was a TA and my mentor used a percentage-based system. In fact, only one mentor in the department used portfolios. Because I was not in her group and no one else knew enough about portfolios to use them effectively, I could not. However, I was not entirely deterred. Using the portfolio idea as a guide, I got around the restriction by allowing anyone with a D or F to revise for a higher grade. The result, however, was less than stellar. Of the forty-two students who completed the work for the course, only one student managed to get an A- and one got a D+; the rest, again, were Bs and Cs. Revision was certainly a good idea; I remember believing that my students learned more about writing, and quite a few of them took advantage of the revision policy to change what might have been Ds to higher grades. Still, I felt their grades did not reflect the quality of their work for the semester.

Throughout my TA training I was reminded by my mentor to emphasize that students *earn* their grades, we do not *give* them. Dutifully I passed this perspective on to my students, making them claim responsibility for their own performances. I could not help but feel, however, that part of the responsibility was mine, particularly when I had trouble articulating exactly what had *earned* them a particular grade. While I could compare papers and see that one was better than another and mark errors in grammar and mechanics, for more important issues like focus, organization and development, there was little to guide me. My mentor reviewed my grades and agreed most of the time, so my judgments must have been “right.” In our conversations about the papers, we used the same terms, and for brief moments I even knew what a strong controlling idea and good organization looked like. That certainty, however, was fleeting. Too often, particularly when I was using my comments to show students how they *earned* those Cs, I was even more certain that my grading was arbitrary. Since my mentor and I were in agreement, however, I decided that either we were both arbitrary in the same way or that I just knew, perhaps instinctively, how to grade. At the time, the latter interpretation seemed more likely, or perhaps more appealing.

In the years since that first set of essays, I have often revisited and occasionally obsessed over the issue of assessment in both my pedagogical and my scholarly work. I have developed courses in which students grade each other and worked on committees to change assessment policies and procedures.

I have designed and implemented portfolio assessment pilots. I have driven at least two of my graduate professors to distraction over “the problem of grades” as they were trying to teach me about process pedagogy. I have presented conference papers on the subject more than half a dozen times. I have written this study. Looking back at my first experience in evaluating student writing, I realize that my judgments were neither arbitrary nor instinctive. Instead, my mentor and I valued similar qualities in writing, probably because of similar training, and even though we had difficulty articulating the nature of those qualities, we knew them when we saw them.

Focus, organization, and the like are no longer particularly elusive to me, although they are certainly abstractions. My dissatisfaction with the grading, I now realize, had less to do with my inability to articulate their meanings to my students than my sense that when I emphasized these elements, I was not evaluating what really mattered in good writing. But I had no way to step back from looking at those particulars; I had no way to evaluate my evaluations. I know now that “validity” and “reliability” comprised the dominant theoretical framework for evaluating assessments—and still do—but these concepts were not available to me at the time. My mentors never communicated them to me, and I was instead left with a growing dissatisfaction about my assessments.

I know these terms now, but I am still not satisfied. The current lexicon used to explicate assessment practices comes from educational measurement theory and carries the baggage of an objectivist paradigm that is ultimately incompatible with what compositionists know about writing and learning to write. Educational measurement principles, most often signified by these terms “validity” and “reliability,” tend to characterize writing assessment as a technical activity with objective outcomes. This approach ignores the complexities of written literacy valued by compositionists, including the influence and importance of communicative context, the collaborative and conversational aspects of the process of writing, and the social construction of meaning. The technical and objectively oriented principles of educational measurement theory do not provide an adequate method for evaluating such aspects of written literacy. This text represents my attempt to develop a theoretical vocabulary for writing assessment that grows out of principles accepted within composition studies. We need new terms.

Historically, compositionists have been inclined to accept the authority of educational measurement principles to define assessment procedures, particularly in large-scale situations where the sheer numbers of students prevent the carefully individualized responses that classroom teachers

favor. The discipline of psychometrics—the branch of psychology devoted to measuring mental ability—and formal work in the assessment of student writing arose almost simultaneously during the first two decades of the twentieth century. To most educators of the time, including those teaching writing, educational measurement theories seemed better equipped to handle large student populations, and their procedures resulted in statistically satisfying outcomes, an important quality during the first half of the twentieth century, the height of positivism in the United States.

More recently, however, compositionists have begun to acknowledge the mismatch between composition theory and assessment practices based on educational measurement principles, but they have yet to seriously challenge those principles directly. For instance, they question the value of practices that require extensive reader training to produce “reliable” results, even for experienced teachers and evaluators. They only occasionally challenge the ideal of “reliability” itself, however. Brian Huot’s recent *(Re)Articulating Writing Assessment*, for example, accepts the value of educational measurement theory and argues for compositionists to embrace cutting edge validity theory, even as he describes the “differing and often conflicting epistemological and theoretical positions” among the scholars in different fields involved in writing assessment research (2002, 23). My problem with this approach is that it does not appear to work. Huot acknowledges that the “cutting edge” theory has not trickled down to the practice of developing assessment instruments, despite the fact that it has been around for some time (2002, 90). Instead, we have a set of assessment principles from educational measurement theory that are amazingly persistent in their attachment to older positivist notions.

Educational measurement theory defines large-scale assessment as a technical activity. Consequently, each aspect of an assessment situation is treated as a variable more or less within the control of the assessment designer or administrator. Composition theory, however, treats writing as a complex of activities and influences, most of which cannot be cleanly isolated for analysis or evaluation. Until the mid-1960s, when composition began to develop professional status, writing instructors were generally content to leave large-scale assessment to psychometricians. In the decades since, however, composition scholars have come to reject objectivist approaches to writing. The social constructionist principles which have supplanted objectivist approaches among composition scholars—in theory, if not always in practice—generally contradict the notions of a knowable and universal reality independent of the observer, foundational

concepts in objectivism. But although theories of composing and philosophies of knowledge-making in composition have changed, objectivist-oriented psychometrics still exerts considerable influence on writing assessment practices at all levels of schooling.

My study characterizes the incompatible approaches to writing assessment generated by educational measurement principles and composition theory in terms of a paradigmatic clash. Most of us in composition studies are familiar with the term “paradigm,” but it is worth revisiting Thomas S. Kuhn’s thinking in *The Structure of Scientific Revolutions* (1996) for the relationship between a paradigm and a discipline. Kuhn argues that “advances” in science occur not through the weight of accumulated knowledge but through upheavals in the beliefs, values, and models of the members of a particular scientific community. While in the original text Kuhn provides multiple, and occasionally contradictory, definitions of “paradigm,” he devotes his 1969 “Postscript” to a more careful explication of the term. A “paradigm,” according to Kuhn, “stands for the entire constellation of beliefs, values, techniques, and so on shared by the members of a given community” (1996, 175). This shared “constellation” solidifies related sets of assumptions into foundational concepts that no longer require explicit justification within the community (1996, 18–20). Paradigms are dependent on community affiliations, and for Kuhn, specifically on disciplinary affiliations; the shared assumptions at work in paradigms are the result not of accident or proximity, but of training. Those who share the same disciplinary preparation are inclined, for example, to accept similar premises, to use similar models, and to define problems in similar ways.

“Paradigm,” then, is a descriptive term, a shorthand indicating a set of common models, values, commitments, and symbolic exchanges that unite disciplinary communities. Within any given community, paradigms do two related primary types of work. First, they provide foundational concepts and legitimate disciplinary scholarship by delineating what counts as knowledge and knowledge-making within any given field. For example, cognitive research using think-aloud protocols barely exists any more, not because we learned all we need to know about the cognitive aspects of writing, but because this approach was only valuable as long as objectivist principles determined the legitimacy of knowledge in the field. As social constructionist epistemologies gained acceptance, such objectivist approaches lost favor among researchers as a means for contributing knowledge to the field.

In addition to providing a legitimating context, paradigms also provide a knowledge base, a common language. They enable scholars to

assume that members of their community will share certain assumptions and models, and that they will understand references to those elements without substantial explanation. In composition studies, for example, concepts such as cognitive theory, process pedagogy, and social constructionism, as well as names such as Mina Shaughnessy, Shirley Brice Heath, and James A. Berlin require little explanation among scholars, except where such elements are specifically the object of study. Familiarity with concepts and scholars such as these are effectively prerequisites for membership and constitute much of the graduate education in the field. The existence of a paradigmatic structure, however, does not presuppose thorough agreement among the community members who speak the lingo. Members of the same paradigmatic community may disagree vehemently about the relative merits of Berlin's work or the particulars of social constructionism. What they share is a set of general disciplinary concepts; they speak the same language, work toward similar goals, and accept many of the same premises.

My study characterizes the tension surrounding current assessment practices as a clash between the objectivist paradigm dominant in educational measurement theory and the social constructionist paradigm of composition studies. While certainly an oversimplification, this paradigmatic delineation allows me to explore large-scale writing assessment as a site contested between two communities with differing values, goals, and lexicons. As composition studies has gained disciplinary status in the academy, its members have, at various times, attempted to reconfigure writing assessment practices so that they more nearly correspond with composition pedagogy and theory; the development of impromptu essay tests and portfolios are the best known examples. These methods have met with moderate success, but the number of institutions and programs still using objectivist instruments such as multiple-choice tests, particularly for placement purposes, is substantial.

Attempts to transform assessment practices within composition have been seriously limited by the principles of educational measurement in force for large-scale writing assessment situations, principles that have historically been held in place by political and economic factors. Even while accepting a social constructionist paradigm, compositionists working with large-scale writing assessment have continued to evaluate and theorize their practices according to the criteria of "validity" and "reliability"—terms developed within the objectivist paradigm of educational measurement theory and still carrying positivist baggage. In relying on these terms—whether actively or reluctantly—composition professionals

turn over decisions about the value of their existing procedures and the shape of plans for the future to scholars working in a different and incompatible paradigm.

The active support is probably the most damaging. When prominent assessment scholars such as Brian Huot (2002) and Edward M. White (1996c) argue that psychometric principles are appropriate for writing assessment and that compositionists need to understand the terms of educational measurement theory in order to justify their practices to those outside of composition, I feel they put our practices and our principles at risk. In this study, I argue that such a move not only authorizes supervision of composition's procedures by the educational measurement community, but also limits our ability to theorize our own practices and thus to understand how they work and to imagine what they might become. The analysis of writing assessment as a site of paradigmatic contention highlights the fact that composition studies needs to develop its own terms for theorizing assessment.

I am using "theory" here to mean a systematic statement of principles that provides an explanatory scheme or model. Theories serve two primary functions: they interpret existing practice and they propose modifications to existing practice. The interpretive variety provides models that explain what is already occurring, while the propositional variety performs a heuristic function, suggesting alternative programs for what is already occurring. In *Beautiful Theories*, Elizabeth Bruss argues that there is a real, though not absolute, distinction between explanatory and programmatic theories, and that the "real danger" lies "in confusing one kind of theory with the other: allowing procedural recommendations to mask themselves as explanations or failing to provide sufficient reasons for their use" (1982, 40). Writing assessment within composition studies seems to suffer from confusion in both directions. Educational measurement theory as embodied in the terms "validity" and "reliability" is descriptive: it provides a model for explaining how accurate, appropriate and reproducible the results of any given test are. In the translation to composition studies, however, this theory has become both programmatic *and* explanatory. To meet the criteria, numerous programs have adopted assessment procedures that produce valid and reliable results—the programmatic use of theory—while at the same time, these criteria have been used to explain the assessment procedures they helped create—the explanatory use of theory. In this way, the criteria have predetermined the theoretical implications for and results of writing assessment. Of course we have developed valid and reliable measures: the entire design

of instruments such as holistic scoring was based on the need for valid and reliable results.

More important than the confusion of functions, however, is the fact that, in my view, the criteria of “validity” and “reliability” have little theoretical significance within composition studies, except as has been provided by and interpreted from educational measurement theory. Definitions appearing in composition studies’ literature usually simplify the meanings of these terms—as in White’s work—or, in more conscientious (and much rarer) moments, refer to current debates in educational measurement theory, as in Huot’s. Both appropriations, however, belie the technical rigor with which these criteria are applied in educational measurement settings and the objective paradigm that supports their use. The paradigmatic analysis I develop here suggests that these adopted criteria are not particularly relevant to writing assessment, considering the theoretical distance between the disciplines of educational measurement and composition studies.

I am not saying that educational measurement theory is somehow “bad” or “wrong.” I am, however, arguing that there is an inherent problem in the forms that educational measurement theory takes and the pressures that it has imposed on composition studies and on writing programs. I agree with Huot that there is interesting thinking going on in contemporary psychometrics, and that we should pay attention to the work of scholars such as Samuel Messick and Pamela Moss. But their interesting work does not regularly appear in high stakes, large-scale assessments at any level, even at the university level, where we have more say in our programs and assessments than K-12 educators do in theirs. The principles of “validity” and “reliability” have cemented practice in ways that militate against the use of cutting-edge reconfigurations and re-definitions. Educational administrators do not apply them; policy makers do not accept them. In writing programs at the university level, compositionists must take our share of the blame; historically, we have accepted the principles without sufficiently questioning them.

Educational measurement theory has traveled to composition studies in the form of its primary terms, “validity” and “reliability.” Composition studies has no similarly influential lexicon, no concepts for understanding the process of evaluating writing that arise from scholarship in the field or that are compatible with the principles of social constructionism generally accepted in the field. Although members of the composition community have proposed alternative models for assessment, these practices have almost always been evaluated—by compositionists and

non-compositionists alike—on the basis of their adherence to the qualities of “validity” and “reliability.” Without an alternative set of principles, writing assessment practices remain limited by these criteria. In addition to analyzing the paradigmatic clash between educational measurement theory and composition theory, then, this study begins the work of developing such a lexicon by proposing a pair of theoretical terms: “meaningfulness” and “ethics.” Although not particularly identifiable as “composition words,” these interrelated terms provide a preliminary assessment vocabulary steeped in social constructionist principles generally accepted within the field.

The value of these terms lies less in their particular definitions and applications than in their potential for opening up assessment as an area of theoretical inquiry for composition scholars. Composition studies, as a discipline, has yet to claim the authority to define principles for assessing our assessments. Moreover, scholars’ reliance on principles beholden to an objectivist paradigm tends to circumvent, ignore, or even negate the paradigmatic tenets at work within the discipline of composition studies. Explorations and vocabularies such as those offered in this study suggest ways of claiming this definitional authority that are both responsive to disciplinary principles, and accountable to those in the profession and to those affected by writing assessment.

A FOUCAULDIAN APPROACH

Power adheres at the level of lexicon. In a frequently cited passage, Kenneth Burke argues that “[e]ven if any given terminology is a *reflection* of reality, by its very nature as a terminology it must be a *selection* of reality; and to this extent it must function also as a *deflection* of reality” (1966, 45). This passage is often used to demonstrate how particular philosophies or perspectives shape our view of a particular world. Burke, however, is more specific than this interpretation would indicate. According to Burke, the power to configure reality resides specifically in the terms chosen to describe that reality. The specific words used, that is, direct our observations, making some observations more likely than others and making some nearly impossible.

In the practice of writing assessment, the terms “validity” and “reliability” currently delineate reality. In order to understand how these terms function and the “reality” they constitute, in chapter one, I analyze the discourse of writing assessment within composition studies for the moments when these terms exert influence and the shape that influence takes. By drawing attention to these terms, I focus on the assumptions about

writing, knowledge, and the process of assessment that incur as a result of their deployment. This analysis is rhetorical; it looks specifically at the ways in which these terms shape our horizon of discourse, at the ways in which we are persuaded to understand the landscape of writing assessment.

The first part of this study, specifically the first four chapters, is critical, or archaeological in the Foucauldian sense of the word. “Archaeology,” for Michel Foucault, “is the systematic description of a discourse-object,” a rhetorical analysis describing the methods and implications of the ways specific discourses constitute the objects about which they speak (1972, 140). For example, in *The Archaeology of Knowledge*, Foucault describes his attempts to understand particular concepts, such as “madness,” as unified objects around which particular and consistent types of statements can be made (1972, 32). As he examined statements, however, he found not a unified object, but multiple objects with the same signifier that were treated as if they all referred to the same signified. Archaeology erases such apparent unity so as to examine the specific occurrences of a discursive object. In *Archaeology*, Foucault suggests that to begin examining discursive objects we should look to the discursive locations in which those objects appear—their “surfaces of emergence” (1972, 41). In his examination of madness, for instance, he looks at how the term emerged in the nineteenth century discursive locations of the family, the penal system, and sexuality. These surfaces serve a normative function: limiting, defining, and determining the status of those objects which emerge within them (1972, 41).

In a strictly Foucauldian configuration of my study, the discursive object would be “writing assessment” or some similar term, and the analysis therein would demonstrate how this term has historically and actually referred to a number of different objects which are dealt with as if they are essentially the same thing. Part of the rhetorical work of my argument is, in fact, the construction of such a historical account of the succession of writing assessment practices and the paradigmatic shifts in literacy scholarship and pedagogy across the same time period, which I explore in chapter two. What is more interesting and important about writing assessment, however, is that the objects that have emerged in the composition studies’ literature bear the mark of a value system that is alien to the current paradigmatic structure of composition studies itself. That is, composition studies, as a “surface of emergence” for writing assessment, produces a paradigmatically inconsistent discursive object. Instead of focusing on the discursive object, then, the first four chapters here analyze the surface itself, examining the ways in which the

discursive space of composition studies has been shaped so as to produce objectively-oriented assessments in spite of its own social constructionist shape and trajectory.

According to Foucault, the emergence of a discursive object is not a quiet event, and surfaces of emergence are not peaceful places. As sites of struggle “where forces are risked in the chance of confrontations, where they emerge triumphant, where they can also be confiscated,” the locations in which discursive objects take shape may be better called “scenes” (1977, 93), which indicate locations for action rather than the more passive location of a “surface.”¹ The struggles of a particular scene and the more powerful combatants within that scene delimit how a particular discursive object emerges and what that object may denote and connote as a part of that context. The power structures of a given discourse determine what may be spoken and what may not; there are winners and losers, those with more authority and power and those with less. Foucault calls the losers “subjugated knowledges,” which are those knowledges discounted and hidden by systemizing structures and which cannot emerge because they do not fall into patterns appropriate for a particular discourse (1980, 82).

My study casts these scenic struggles in writing assessment in terms of a clash of two paradigms. This representation has the advantage of highlighting the ways in which the particular terms and conditions of each paradigm influence—or fail to influence—the shape of writing assessment in the context of composition studies. The absence of social constructionist principles for writing assessment suggests that the objectively-oriented paradigm of educational measurement theory has considerable power to determine the shape of writing assessment, even where writing assessment exists as an area of activity within the domain of composition studies. Knowledge about writing assessment that would take social constructionist principles into account, then, is subjugated, discounted, and hidden in the face of the need to meet the criteria of “validity” and “reliability.”

However conceptually productive, archaeology results in descriptions that are insufficient by themselves to change power structures. Although Foucault tends to favor description over prescription, he does offer “genealogy” as the result and employment of archaeology. Genealogy is the understanding, gained from archaeology, of the struggles which have subjugated certain knowledges and privileged others and, more importantly, the tactical deployment of such subjugated knowledges (1980, 83). Archaeology exposes the rules of struggle; genealogy interprets and deploys them to obtain a different advantage.

The archaeological rhetorical analysis of large-scale writing assessment here describes the struggle as a lexical problem: educational measurement theory has terms for evaluating assessment practices; composition studies has not. Given the paradigmatic inadequacy of educational measurement's terms for work in composition studies, a genealogical interpretation would suggest that composition studies develop its own vocabulary for writing assessment. This is the work of the sixth and seventh chapters of this study. It is not enough to criticize existing assessment practices; historically, this has tended to result in the refinement of practices, but not in their reconceptualization. "Validity" and "reliability" are theoretical terms: they explain the value of existing practices and outline directions for change based on the values at work in the discipline of educational measurement. Composition studies has no terms with which to perform similar work. The genealogical move this study makes, then, is to propose such a set of principles.

Applying the principles of a Foucauldian archaeology, this text constructs an account of the history, theory, and practice of writing assessment which centers on the need for an appropriate theoretical terminology. Its aim in large part is to estrange the familiar. "Validity" and "reliability" are a normal part of the terrain of writing assessment. These terms, however, are not indigenous to composition studies. Rather than explicate and thus normalize these terms as some recent work in composition studies has done, this study challenges their suitability for rhetorical deployment in writing assessment.

ORGANIZATION

My first chapter constructs a historical account of large-scale writing assessment practices in the United States from the advent of written examinations in the 1820s through the grassroots development of portfolio assessment in post-secondary composition programs in the late 1980s and early 1990s. The most significant events for the purposes of this study, however, took place not at the beginning or the end of the period, but in the middle. During the first decades of the twentieth century, the field of psychometrics arose; at the same historical moment, writing teachers and researchers—there were a few at this time—were searching for a way to make teachers' "marks" on student papers more uniform. Psychometrics provided the answer in the form of standardized tests. As members of a society generally receptive to scientific advancements and procedures, many writing teachers greeted the regularity and reproducibility of the

results with enthusiasm. When writing teachers adopted these tests, they also adopted the principles of educational measurement that supported their use, namely, “validity” and “reliability.” These terms have since directed the shape of large-scale writing assessment practices, even when teachers and institutions have favored the direct evaluation of students’ written work over the indirect assessment of writing by multiple-choice testing. Even portfolio assessment, which resists the more rigid applications of the psychometric concepts, is nonetheless subject to evaluation according to these criteria. This chapter argues that continued reliance on these terms perpetuates an intolerable incongruity between the ideals of literacy education and the practice of writing assessment.

Chapter two examines the ideals of literacy education over roughly the same period as the first chapter. Working with the premise that evidence of literate ability is the object under scrutiny in large-scale writing assessment, this chapter examines paradigmatic changes in literacy pedagogy and scholarship. Whereas such assessment adopted an objectivist paradigm whose influence has endured, literacy education has undergone a series of shifts and currently operates within a contextual paradigm grounded in social constructionist theory. The social constructionism of literacy and the objectivism of assessment, however, are incompatible. In spite of postmodern challenges, objectivism continues to dominate assessment practices. Consequently, the literacy conceptualized and theorized in scholarship bears little resemblance to the literacy ostensibly evaluated in large-scale writing assessment.

Chapter three analyzes early significant changes and challenges to conventional writing assessment practices. Reviewing recent work on the expert reader method of assessment in composition studies and influential work by Guba and Lincoln (1989) in social science research, this chapter focuses on the ways that the literature about these cutting-edge assessment procedures negotiates the demands of educational measurement theory on the one hand and composition theory on the other. In their justifications and analyses, the authors of the research in composition studies tend to rely too heavily on the principles of validity and reliability for justification, while Guba and Lincoln’s principles sacrifice pragmatics and become too unwieldy for large-scale assessment situations. Ultimately, these practices do not meet the need for strong changes in writing assessment.

Even the more theoretically-oriented ventures in composition studies—including the “Position Statement” published by the CCCC Committee on

Assessment (1995) and Huot's (*Re*)*Articulating Writing Assessment* (2002)—continue to accept the value of validity and/or reliability to define their projects. Chapter four analyzes contemporary theoretical work in both composition studies and in educational assessment for the alternatives to traditional psychometric principles they offer. Those who try to co-opt educational measurement theory for composition purposes are the more prevalent group, and although they ultimately fail to divorce themselves entirely from objectivist principles, they do succeed in foregrounding possibilities for the application of social constructionist principles to large-scale writing assessments settings. In this chapter, I also examine contemporary challenges to traditional educational measurement theory, particularly to reliability, from within educational assessment, as well as challenges in composition studies that do not rely on educational measurement principles at all. While these latter challenges have a lot of potential, the principles they offer are incomplete and in the final analysis unable to provide the kind of broad theoretical support needed for large-scale writing assessment.

Ultimately, I believe, we need a stronger challenge to conventional educational measurement theory, and chapter five makes that argument. It begins with an analysis of the conflict over the *Standards for the English Language Arts* (National Council of Teachers of English and International Reading Association 1996) as a particularly public example of the clash of paradigms. The *Standards* relies explicitly on notions of contextuality and community indicative of a social constructionist perspective, while critics charge that the standards are inexcusably vague and ultimately unassessable. The distance between these positions reflects the conceptual distance between the social constructivist ideals of composition studies, of which written literacy education is a part, and the objectivist ideals of assessment theory and practice. The negative reception of the *Standards* was due, in part, to the timing of its release; those of us in post-secondary composition studies had not yet developed the evidence to support constructivist principles for literacy assessment. In this chapter, I argue that post-secondary compositionists have the responsibility to develop both principles and evidence, and that we need to do so apart from educational measurement theory. Whenever we return to conventional assessment theory, our innovations are subsumed and inherently weakened. Separate development would give us the opportunity to assemble persuasive arguments in favor of alternatives before bringing them back to the broader assessment discussion.

My research in the first part of this book suggests the need for an alternative set of principles, and chapter six provides a preliminary

sketch of just such a lexicon. Developing the concepts of “meaningfulness” and “ethics” as principles for describing and evaluating assessment practices, this chapter begins the work of theorizing writing assessment in social constructionist terms. “Meaningfulness” draws attention specifically to the purposes for and substance of any given assessment practice. Meaningful assessment, then, should be conducted for specific and articulated reasons, and its content should be intelligible to those affected by the procedure. “Ethics” draws attention to assessment as it is practiced and specifically to the relationships among those involved in the process. This chapter proposes an ethical structure based on Jürgen Habermas’s “communicative ethics,” which advances the notion that communities reach ethical decisions on the basis of rational argumentation (1990a). Habermas relies on “ideal speech communities,” a condition too broad and utopian for most assessment situations; however, with the caveat of a *limited* speech situation, in which those accountable for the assessment occupy positions of authority, communicative ethics provides a reasonable preliminary structure for assessment procedures. Thus, “meaningfulness” and “ethics” offer an introductory lexicon for theorizing writing assessment in ways that are compatible with the social constructionist paradigm at work in composition studies.

My seventh chapter provides some examples of what contemporary writing assessment practices look like when examined through the lens of “meaningfulness” and “ethics.” I begin by looking at the published scholarship on assessments at the University of Cincinnati and Washington State University. Both cases demonstrate strongly ethical practices, but have comparatively little to say about meaningfulness, particularly in terms of what I call in chapter six “primary substance,” the content of our assessments. To examine the substance more fully, I turn to two other studies, one by Bob Broad (2003) and the other based on a pilot I ran at North Carolina State University. These analyses, taken as a whole, demonstrate the value of looking outside reliability and validity to understand and evaluate what happens when we assess writing.

Finally, the conclusion argues that the imposition of educational measurement principles on large-scale writing assessment situations is no longer warranted and is indeed detrimental to pedagogical and theoretical efforts in composition studies. Although there is a good-sized body of published work on assessment within composition studies, the majority of it takes a thoroughly practical approach, focusing on procedures rather than on principles. While there are legitimate reasons for the practical emphasis, such a pragmatic focus—along with the historical precedents

described in the early chapters of this study—has encouraged compositionists to uncharacteristically accept educational measurement's principles rather than to develop their own or to at least actively modify them. Terms such as those suggested in chapter six would not only provide principles that correspond more nearly with composition's social constructionist paradigm, they would also help establish composition's disciplinary authority in matters of evaluating evidence of written literacy.

This book reflects my own attempt to come to terms with writing assessment and particularly with compositionists' acquiescence to theoretical principles which have little to do with current thinking about writing and learning to write. Although my work began in a moment of dissatisfaction with my own ability to evaluate student writing, I have since found that my personal difficulties reflect composition studies' own historically conflicted relations with objectivism, politics, economics and disciplinary authority. The discomfort I have experienced could hardly be mine alone, but few scholars have analyzed the ways in which these relations have enabled educational measurement theorists to put words in our mouths, whether they mean to or not. That ventriloquist's trick ultimately fails to provide sufficient guidance for developing the kinds of assessments we want. In order to reconceive assessment, compositionists require a different theoretical lexicon. We need, that is, to come to our own terms.