

CONCLUSION

Coming To Terms

In “Theory and Practice,” Charles I. Schuster analyzes the connections between these two “sides” of scholarship and concludes “that theory is a form of practice and that practice is the operational dimension of theory” (1991, 43). They function in tension, he claims, each grounding the other in ways that make them effectively inseparable. Not surprisingly, he finds the most compelling work in composition studies “positioned on that ambivalent threshold shared by both theory and practice,” not readily categorizable as one or the other (1991, 46).

Schuster’s analysis helps explain my own dissatisfaction with composition studies’ scholarship on large-scale writing assessment. While there has been a great deal of work produced by compositionists on the subject, the majority of it falls squarely in the category of practice. Current literature abounds with pragmatic advice, procedural descriptions, and case study analyses of the process of developing new assessments. As a whole, this body of work is thoughtful and perceptive, aware of the complexities of assessing writing, familiar with other scholarship on the topic, and attentive to the material limitations within which writing programs operate. It is, nonetheless, overwhelmingly—almost exclusively—practical.

Not without good reason. Scholars, administrators, researchers and teachers tend to take up the literature on assessment to address very immediate and practical situations—the need to reconsider current practice because of the real or perceived failure of existing instruments, for example, or the need to respond to new assessment mandates. These are material needs centered in actual practices; they favor pragmatic, “real-world” solutions over theoretical analyses. The practical emphasis in writing assessment scholarship, then, constitutes an appropriate response to the needs of its audience.

These practical needs explain the practical emphasis in the literature, but not the way in which compositionists have limited our engagement with the theory of large-scale writing assessment. It is not that compositionists function as if we have no theory—in the final analysis, an impossible condition—but rather that we have done precious little theorizing of our own on the subject. Within our discipline, even scholarship on pedagogy,

our most practical of practicals, has a strong theoretical bent—consider the work on collaborative learning or critical pedagogy, for example. Not so with assessment. The majority of writing assessment scholarship published by compositionists in the last decade either gestures in the direction of educational measurement theory or simply never addresses the question of theory at all. Even *Assessment of Writing*, the volume published by the Modern Language Association—a veritable mecca of theory—maps writing assessment according to the pragmatic concerns of its subtitle: *Politics, Policies, Practices* (White, Lutz, and Kamusikiri 1996). By and large, scholars in composition seem to have been content to leave assessment theory to educational measurement specialists.

This does not mean, however, that compositionists have been pleased with the results. Time and again, as I have emphasized, compositionists writing about large-scale assessment have argued that reliability and validity are troublesome criteria, inadequate and too limited for the distinctive task of evaluating writing. Yet rarely do these scholars propose alternatives. If theory indeed provides systematic statements of principles that in turn provide explanatory schemes or models—as I use the term in this study—then composition scholars have operated and continue to do so, for the overwhelming most part, with an explanatory scheme for writing assessment that has little to do with what we know about writing and learning to write. That is, with a few notable exceptions, composition scholars act as if “validity” and “reliability” constitute the best—or perhaps the only—theoretical lexicon by which assessment practices can be described and judged.

This acceptance strikes me as both uncharacteristic and unproductive. Compositionists have been actively proclaiming the limitations of educational measurement theory for at least two decades yet continue to employ the principles of validity and reliability as justification for large-scale assessment practices. This persistent reliance on objectionable principles is atypical of compositionists. For a variety of reasons, composition scholars tend to appropriate theories from other disciplines, to import them, reshape them, even domesticate them; the relative merits of this practice are a perennial subject of debate. We have not even accepted postmodern theory from within our own departments without some serious adjustment, limitations, and caveats. Rarely do we simply grant authority to some other discipline’s principles without modification. In the case of educational measurement principles for assessment, however, we are doing just that; even those who would co-opt the principles are not really proposing changes to them—merely the use of the more applicable parts of the original.

Practical necessity explains some of the reliance on external principles; historical precedent provides additional justification, particularly for the authority these principles maintain. But, during the last quarter century, and particularly in the last decade, theories of composition and composition pedagogy have diverged sharply from those of assessment; the former have taken a social constructionist turn while the latter have stayed the objectivist course. The distance between social constructionist principles in composition and objectivist principles in assessment is now sufficiently great so as to have exacerbated disagreements about literacy assessment; the controversy surrounding the NCTE and IRA *Standards for the English Language Arts* provides one particularly public example.

In spite of this disjunction, educational measurement principles continue to frame compositionists' understanding of large-scale assessment, and when the educational measurement frame and compositionists' expectations for writing assessment are at odds, compositionists have generally had one of three responses. The first calls for a (re)turn to practices consistent with the rigors of educational measurement principles; this is Edward M. White's career-long argument. The second type of response, more frequent than the first, attempts to push the frame aside entirely by arguing directly against the principles and claiming we can do without them. Elbow (1994) and Broad (1994) both take this approach in an effort to claim authority for compositionists in writing assessment situations; both, however, acknowledge that practical constraints cause them to return to practices compatible with educational measurement principles.⁶¹ The third approach—far and away the most common—is to co-opt them. This is Huot's tack in *(Re)Articulating Writing Assessment* (2002), and it is the rhetorical strategy of choice in descriptions of program development where authors are interested both in claiming their own authority and legitimating their efforts according to established norms. The last two types of response assert—implicitly or explicitly—the distance between educational measurement theory and composition studies theory, but they simultaneously tend to accept the authority of educational measurement theory to explain, define and justify writing assessment. None of these positions, in my opinion, presents an alternative to the frame itself.

The absence of such an alternative is hardly a failing on the part of composition scholars. Part of the work of this study has been to demonstrate the ways in which educational measurement principles have been a normalized part of large-scale writing assessment discourse for most of the last century and how at one time they coincided with what

writing teachers knew about literacy. The divergence between composition theory and assessment theory is a fairly recent phenomenon; only within the past decade or so have scholars really begun to challenge assessment practices justified solely by educational measurement concepts. Given composition's pedagogical imperative, it is scarcely surprising that we focus our challenges at the level of practice.

This method of confrontation at one time provided compositionists with a strong challenge to objective testing. When the issue was whether writing would be evaluated according to direct or indirect measures, compositionists' practically motivated arguments about the effect of assessment on pedagogy were relatively influential. Now, however, the focus on assessment practices either directs our attention toward procedures and methods, or results in assertions about institutional authority, both of which divert attention from the theoretical principles that shape practice. The former implicitly claims that educational measurement theory is a suitable framework for explicating and evaluating writing assessment practices. The latter approach takes sides regarding which faction—composition studies or educational measurement—has the institutional power to define assessment procedures. This is the shape of the Elbow-White debate that has often focused on composition's authority relative to that of organizations such as the Educational Testing Service.

"Validity" and "reliability" function as a terministic screen that obscures the theoretical disjunction between educational measurement and composition. With these principles in place, compositionists tend to treat writing assessment theory as a *fait accompli*, particularly when practical concerns are predominant. These principles, however, derive from an objectivist epistemology, and they bring a corresponding objectivist pressure to bear. Throughout much of the twentieth century, this has not been a problem; literacy education and educational measurement have shared an objectivist orientation. This is no longer the case, and for those working in composition studies, the effect has been a discursive object—writing assessment—that emerges within the scene of composition studies but according to the objectivist paradigmatic tenets and constraints of educational measurement theory. In other words, educational measurement theory delineates the model for large-scale writing assessment, even though that theory is incongruous with contemporary theories of writing and learning to write.

This theoretical imposition would not be possible without a substantial power differential between composition studies and educational measurement. In the academy, educational measurement shares in the

prestige (and funding) attached to the relative objectivity of the social sciences, while composition studies falls into the “softer”—and poorer—humanities category. The lack of scientific stature carries over to the public domain in the United States, where scientific results are frequently equated with “facts” and “reality” while other research produces merely “opinion” or, more kindly put, “interpretation.” Founded on educational measurement principles, the Educational Testing Service has built an assessment empire by foregrounding the scientific nature of its operation, and in so doing, instruments such as the SAT and GRE have become the publicly accepted model for “real” or “serious” assessment. As a result, organizations like ETS have secured not only public backing, but also the funding with which to continue their efforts in ways that composition studies professionals have not. While several of the most recent tests and revisions include essay sections, their mainstay is still the multiple-choice format that produces “objective” results. Next to such scientific rigor, social constructionist theories of knowledge appear relativistic to say the least; there simply are no thoroughly constructionist principles for large-scale writing assessment. This power imbalance reinforces the notion, even among compositionists, that in high stakes, large-scale assessments, objectivity is paramount. Consequently, compositionists have no assessment lexicon compatible with the discipline’s social constructionist paradigm, even if only for their own use.

In large part, the problem is political. The dominance of educational measurement principles continues in part because of the public response to and desire for quantification, even if the results in numeric form are nearly meaningless—or certainly less meaningful than a more qualitative result would be. First-year numbers for the No Child Left Behind Act Adequate Yearly Progress are being released all over the country at this writing. Saying that my daughter’s high school met 15 or 71.4% out of 21 target goals does not tell me what goals were missed or why—or, more importantly, why those goals are important and what the school is doing to meet more goals. My other daughter received a 2.5 on her yearly writing test two years ago, a result that stays in her academic file, and that tells future teachers that she barely passed the test. That result does not tell her future teachers that she passed by the skin of her teeth because she does not like to elaborate on her ideas in writing, that she prefers drawing and working with clay but will respond well when given projects that allow her to couple her artistic interests with writing. It also does not explain that prompts that pull on her problem solving abilities and inventive ways of thinking will get much better

results than prompts that ask her to tell a story. I know this because I know my daughter, and as a writing teacher, I pay attention to her writing. My daughter's teacher from last year figured it out, but it took her about a quarter of the year. Just think how much further ahead teachers would be if the results they saw, if students' academic records were at least partly discursive, particularly in areas where the discursive provides so much more information.

"Meaningfulness" and "ethics" are not going to take the K-12 world by storm today. Public opinion and educational practices at that level are currently too wedded to quantifiable results, which is what "validity" and "reliability" produce; the fiasco over the *Standards* gives us a fairly clear sense of how discursive and qualitative results would be received at that level of schooling. But at the post-secondary level, we have more autonomy, and we can make decisions about our practices that are not quite so beholden to public opinion and quantification. Washington State is doing this. Their placements and mid-career portfolios result in a decision, not a number, and the form seems quite adequate to the task and to the needs of the students and the program. In other words, at the post-secondary level we do have the power to dictate our assessment practices, as Elbow (1991) argues. But not if we continue to rely on validity and reliability. Educational measurement principles put constraints on our assessment practices that limit the range of what we can do. WSU's program as a whole does not meet the full technical criteria of reliability; if it had to, I cannot see how the program could exist. Educational measurement principles require that we reduce our results and the value of the assessments to numeric form. Yet our most interesting and useful results often do not come in the form of numbers, however, and sometimes we do not need numbers at all.

The terms I am offering here—or other principles like them if these are found wanting—would give us the tools to develop different kinds of assessments that provide results that we find useful. We will still need to justify and legitimate those results. We will still need to be able to explain the value of assessments both to the students and to our programs. But if we can start with principles that help us think through what we really want in an assessment, about what any given assessment is for, we are more likely to develop truly satisfying assessments, the kind that provide valuable information to guide our programs and our students.

Large-scale writing assessment is conflicted at the level of theory. The two paradigms—objectivism and social constructionism—are effectively irreconcilable. More importantly, the clash of paradigms situates

large-scale writing assessment in an unproductive tension so that while the surface features of the assessment change, little of substance does. Portfolio assessment procedures, for example, look different from those for single-sitting impromptu, but ultimately, they are very similar; the training and calibrating demands of reliability via holistic rating procedures sees to that. “Validity” and “reliability” are the theoretical linchpins suspending writing assessment between objectivism and social constructivism.

I recommend pulling those pins because we need alternatives as we carry out the kind of “threshold” work that Schuster recommends. Terms such as “meaningfulness” and “ethics” outline the shape of assessment once its theoretical principles exhibit paradigmatic consistency with principles already prevalent in composition studies. While these specific terms and the principles they signify will likely be revised, they nonetheless demonstrate possibilities for composition scholars to theorize assessment according to our own understanding of what it means to write and to learn to write.

Underlying my development of these terms is the implicit argument that composition studies professionals should claim the authority to define the principles by which to describe, evaluate, and reimagine what evidence of literate ability—as well as assessment itself—looks like. As social constructionist principles have become dominant within the discipline, compositionists have reconceived the development of literacy, but we have only begun to discuss how indications of such development appear in students’ work. Moreover, as a discipline, we have yet to consider seriously and systematically the possibilities and consequences of social constructionist assessment. Theorizing writing assessment in this manner would encourage compositionists to explore previously un(der)researched aspects of evaluating writing. For example, as long as “reliability” functions as a theoretical principle for writing assessment, assessment procedures cannot abide dissent. Yet dissent persists, even in “communities of like-minded peers” (Bruffee 1986, 774). The ideal of “reliability” presupposes that dissent is counterproductive, and historically, both compositionists and educational measurement specialists have worked to eradicate disagreement in assessment situations. Consequently, we know very little about what dissent may contribute to our evaluative capabilities and to our students’ resultant understandings of literacy. We need research in order to find out what dissent might contribute and how we might better accommodate it or even encourage it, should we find out that it is valuable. Scholarship of this type, difficult to justify in

an objectivist paradigm, becomes possible as compositionists claim the authority to theorize writing assessment.

Such a claim suggests a redistribution of power—most particularly a shift in disciplinary authority in favor of composition studies—that likely would be both controversial and uncomfortable. This reallocation implies that writing professionals have expertise more appropriate for evaluating written literacy than do educational measurement professionals. On the one hand, this seems self-evident. On the other hand, compositionists have tended historically to turn over procedural decisions about assessment to the testing community. Huot points out that “English teachers’ justifiable distrust of writing assessment has given those without knowledge and appreciation of literacy and its teaching the power to assess our students. The ability to assess is the ability to determine and control what is valuable” (2002, 107). Relinquishing the procedural decisions, that is, also results in conceding the authority to define what is valuable about writing and assessing writing. Reclaiming assessment, then, also means reclaiming this definitional work, which, in turn, implicitly argues for the priority of discipline-specific theories.

Although I am calling, in part, for a clearer demarcation of disciplinary boundaries at the same moment as many are calling for border-crossings, I do not mean that composition studies should ignore work in other fields, including educational measurement. I am arguing, however, for the acceptance of our own expertise, and for appropriate and necessary disciplinary boundaries. My own familiarity with educational measurement and with methods for evaluating processes, for example, does not give me the expertise or authority sufficient to evaluate laboratory procedures in chemistry or biology in any meaningful way. There is a curriculum-specific knowledge that I cannot demonstrate without becoming a member of the relevant community. The theoretical work I am proposing here argues that composition studies entails such disciplinary expertise as well, and that the value of such expertise goes well beyond the role of expert consultant.

The power to theorize is the power to define, to influence, to organize, to limit. Currently, we are limited by educational measurement principles. At the level of theory—though not yet at the level of practice—composition studies no longer needs to accommodate these principles; we continue to do so at the risk of widening the rift between what we know about learning to write and what we encourage our students to believe through our assessments. “Meaningfulness” and “ethics”—or any other principles we choose to develop that are based on social constructionist tenets—can

provide compositionists with the theoretical tools to make assessment work as an integral and integrated part of our discipline.