

Composition Forum 35, Spring 2017

## Let's Disagree (to Agree): Queering the Rhetoric of Agreement in Writing Assessment



Paul Walker

**Abstract:** This article describes and theorizes a failed writing program assessment study to question the influence of “the rhetoric of agreement,” or reliability, on writing assessment practice and its prevalence in validating institutional mandated assessments. Offering the phrase “dwelling in disagreement” as a queer perspective, the article draws on expertise theory and notions of ambience and attunement in rhetorical scholarship to illustrate the complexity, unpredictability, and disorder of the teaching and assessment of writing. Adopting a queer sensibility approach, the article marginally disrupts “success” as assumed by order, efficiency, and results in writing assessments and explores how scholars might reimagine ideas, practices, and methods to differently understand a queer rhetoricity of assessment and learning.

But man seeks to worship what is established beyond dispute, so that all men would agree at once to worship it.

-- Fyodor Dostoevsky, *The Brothers Karamazov*

Absolute curiosity, and the love of comprehension for its own sake, are not passions we have much leisure to indulge: they require not only freedom from affairs but, what is more rare, freedom from prepossessions and from the hatred of all ideas that do not make for the habitual goal of our thought.

-- George Santayana, *The Sense of Beauty*

The very idea of individual *intuition* disconcerts the academic value we place upon collaborative deliberation. Teaching and evaluating writing, as well as the administration of it, requires numerous decisions, and generally we consult our local or disciplinary colleagues when another viewpoint could be helpful. But we also make many decisions on our “own,” even in high-stakes situations, especially if we have experienced similar situations before. By definition, such situational decisions improve our expertise, and expertise studies (Dreyfus and Dreyfus, Ericsson) have confirmed that experts develop intuitive abilities in making decisions that relate to those frequent and similar situations. In previous work, I have explored theoretical facets of intuitive expertise, or connoisseurship, as applied to teaching and evaluating student writing, drawing on ontological principles and positing that “teachers’ perceptions of student papers are influenced by the numerous swirling factors involved in ‘knowing’ students and themselves” (Walker), which I relate here to what scholars Thomas Rickert and Krista Ratcliffe respectively call *ambient rhetoric* and *rhetorical listening*. My recognition of the relationship between those “swirling factors” and an ecological, complex, nonsovereign subjectivity helped me develop a practical expertise-based assessment model (Osborne and Walker) which, to frame it in Rickert’s terms, collects rhetorical ambience without fully accounting for its elements. That work emphasizes how altering our perception and utilization of intuitive expertise in the *kairotic* “emplacement” of assessment enables a modest, generative trust in teachers as attuned evaluators of writing within a specific writing program. The expertise-based model developed for our first-year writing program produces numerical results that satisfy administrators and accreditors, but does so in a way that simultaneously challenges the contingency of assessment in our educational situation.<sup>{1}</sup> My intention is to extend this conversation regarding writing assessment practices, which, while usually well-intentioned and self-aware of the nuance and complexity of writing and its instruction, also undermine the valuing of difference and disagreement through reductive institutionalization. I am not naïve to the complex realities of our institutional and educational landscapes, but I find that the most satisfying theoretical and pedagogical pathways shared in the field of rhetoric and composition meander rather than bulldoze a clearly articulated and “straight” trail.

The pathway I follow in this essay therefore meanders—there are both obstacles to sidestep and scenic detours that distract us. While some of us instinctively seek a straighter route and its perceived efficiency, Jonathan Alexander and Jacqueline Rhodes remind us that Composition's inherent bulldozer-like qualities often push the scenic "excess" to the side of the "straightened" path because it is deemed inefficient or complicated and potentially "disrupts the containment" presumed necessary for Composition's standing (196). In such an "uncontained" sense, here I adopt a queer positionality from which perspective I consider the "disorienting excess" emerging from an assessment study relegated to the margins as a result of its failure to meet empirical measures of significance—statistical reliability standards—that orient and are "contained" by writing assessment scholarship. Rather than re-do my failed assessment with calibration adjustments until it met those standards, I wandered in another direction, and the study itself represents a marginal path that is neither abandoned for a clearer, well-trodden one nor one held in esteem as an epiphany of what I was doing wrong so that I could right myself. Instead, my failed study represents a non-centered space, an alterity, where I wasn't sure if I should remain; a scenic spot off the main path that doesn't necessarily lead anywhere, which admittedly endangers the path's value in providing orientation and guidance, but its ephemerality engages my mood and perception of rhetoric in unexpected ways. However, this unexpected novelty is not really a kind of "success," nor is it liberating. Frankly, it has made things more difficult and disorienting. Examining this difficulty and nonsuccess harkens to J. Jack Halberstam's alternative way of viewing failure; he refers to failure as an "art" known particularly by those constantly existing in and engaging with the margins of society. In other words, being "queer" means to *always* fail in society's definitions of success and legitimacy, and so in queerly refusing to view failure as something to avoid yet also an inevitability on the path to success, failure becomes instead a path to nowhere, a space wherein we cannot predict, and by doing so, generate alternative ways of sustaining. As Halberstam states, "failure allows us to escape the punishing norms that discipline behavior and manage human development with the goal of delivering us from unruly childhoods to orderly and predictable adulthood" (3).

Therefore, it is within this queer space of failure that I combine ideas of intuitive expertise, assessment, and ambience in order to theorize how we might attune ourselves to difference in writing assessment situations. Specifically, I want to suggest a rhetorical attunement to difference wherein we listen generously yet accept the lack of mutual understanding that radical alterity always produces. In other words, the complexity of writing assessment or writing program assessment *should* result in difference and disagreement, but I propose viewing such alterity as neither a failure nor a temporary obstacle on the way to calibrated agreement and similarity. Hence, I find in queer studies a dynamicity recognizing the ambience and fluidity of human interaction and hope such an idea resonates in meaningful ways for assessment scholarship, which has a rich mix of dissent and attunement to difference despite widespread institutional mandates and structural impediments (e.g. Gallagher; Inoue; Lynne; Wilson). We know that in education policy, ambient factors—in addition to intrinsic learning and intellectual worth—are increasingly dismissed as antithetical to fixed notions of job-aimed education and mainstream narratives of heteronormative, masculine, and capitalistic success. Not only is this evident by random polling and populist political pandering, but also by corporate-based reform and the redefining of university education as exclusively an upward, assertive, merit-based pathway to a "worthwhile," "successful" career rather than valuable in and of itself. Teachers of writing acknowledge writing as an ecological activity comprising at a minimum, reading, thinking, interpreting, conversing, drafting, and revising. Yet in a quantophrenic educational culture moving rapidly toward overemphasis on STEM fields and the certainties presumed by that path, we should recognize that "assessments codify particular value systems" (Scott and Brannon 277), and be aware that in failing, purposefully, to codify value systems that reduce complexity, we will also fail to change the system. But rather than bemoan our lack of success, trying on a queer perspective can help us re-think how we perform our disciplinary assessments as expressions of "success."

## The Rhetoric of Agreement

As a wily guide along this meandering path, I utilize the phrase *rhetoric of agreement* as representing the growth of social-science parlance encompassing *reliability*, which in the assessment of student writing often—but not always—acts as a prime warrant in validating (defining as "successful") placement, programmatic evaluation, and curricular decisions. Problematizing *reliability* does not paint all uses of it as flawed, nor does it indicate a lack of understanding of its relatively small role in the ways writing assessment scholars have meaningfully framed and reframed *validity*. But the provocative focus here allows identification of potential underpinnings of assessment that are less extensively discussed or questioned in many institutional practices. Therefore, unlike Wayne Booth's "rhetoric of assent," which simultaneously seeks answers and withholds doubt in order to balance between the modernist dogmas of scientism and irrationalism, *the rhetoric of agreement* is socialized trust in a group of subjects nodding their heads together—an adult version of the childhood notion that two is *always* better than one. Again, I am in no way suggesting agreement, confirmation, or reliability are wrong intrinsically; rather, I propose that agreement as the subsumption of difference can, through institutional mandates and what D. Diane Davis calls the "rhetoric of totality" (12), marginalize queerness by reifying masculinized and capitalistic traditions, including the persistent upward trajectory of merit or value-added results and the assumption that answers to difficult questions about learning and performance and

identity are *waiting* to be found by *acting* subjects. Writing assessments based in or reliant partly on an agreement paradigm inherently distrust individual evaluations like many people distrust fluid/queer identity, treating their rough and messy difference and disagreement as obstacles to order, rules, boundaries, explanations, efficiency, and solutions—all aspects of composition's carefully composed “harmony” and as such impervious to nonsense from the margins. Citing Robert McRuer, Alexander and Rhodes state: “Composition theory may not be able to ‘work against the simplistic formulation of that which is proper, orderly, and harmonious.’ To do so would be to engage in work that is *not* composition. Such work is *impossible* for composition” (196).

For example, a calibrated rubric operates as a technological agreement construct, designed to enable or enhance human individual ability to “orderly” and reliably assess writing. All technologies and media “extend” human capability but, as Marshall McLuhan said, they also have a “massaging” influence always already acting upon our perception of the content actually delivered. In this sense, the possibility exists that the more we assess and are assessed through formal mediation, the more the results, or “content,” blind us to the way the technological construct affects us, perhaps causing a distrust of our own and others’ different ways of arriving at decisions. Some might ask why such distrust is a problem; I answer that skepticism is different than distrust, and recent trends in education, in K-12 especially, illustrate the problems that occur when teachers are not trusted as professionals. Pushing testing and accountability, rather than participating in a healthy discussion of systematic factors in low student performance with genuine skepticism of common practices, seeks to root out “bad” teachers, who, by and large, are invented to justify the policy that no individual teacher is trustworthy to develop, deliver, and assess a curriculum that is “better off” driven by Big Data, analytics, legislative wisdom, or free-market capitalism, paralleling historical rationalizations for civilizations being “better off” if white men are in charge.

As stated, I acknowledge the numerous efforts to nuance validity and reliability by rhetoric and composition scholars in the last decade or two, as well as the breadth and depth of earnest *praxis* to make writing assessments responsible, local, ethical, and meaningful (Gallagher “Local;” White et al. Even so, as Rickert notes, “rhetoric has so emphasized cognitive content in intention and reception that even in the more robust theories of context, salient variables always take priority, and ambience is relegated to the margins, if dealt with at all” (9). Thus, the carrying out of assessments is often less governed by scholarly nuance and local pedagogical inquisitiveness than by empirical imperatives pragmatically grounded in efficient, “comparable” versions of the holistic method, resulting in quantitative salience that determines or defends a writing program’s value and effectiveness on a campus. In my previous work and here, I suggest that in the many harpings about/for/against assessment we continue to question the very foundations of the holistic method, not because it has been unquestioned or unchallenged or unadjusted by scholars previously, but because this method resists queering, which might include nuanced, marginal, non-salient, or experimental assessments that produce results other than the “institutionally acceptable” kind. As Edward White, Norbert Elliott, and Irvin Peckham posit, “Understanding writing program assessment as an ecology reminds us that we are involved in complexities we both do and do not understand” (32). Holistic rubrics can deliberately suppress the unknown, so that assessments are clearly and readily directed and conducted, often using inexperienced graduate assistants, with a heavy dose of supervisory pressure on (especially untenured) Writing Program Administrators. In our numerous conferences and journals we have the resources to empower interesting and localized teaching/learning relationships generated by provocative debates and disagreements about appropriate assessment. Constantly moving towards the center, towards explicit harmony and sameness via expected standards of social-scientific statistical measurement to determine the “success” of assessment, reinforces for those outside our discipline the primacy of a *correct* methodology over complexly and ecologically hermeneutic meaning and validity, thus maintaining enough legitimacy for administrators to continue to coopt a reductive and possibly irresponsible holistic methodology.

I have encountered—in informal conversations, conference presentations, and manuscript reviews—sincere concerns about the “damaging effects” of these questions I raise, concerns I recognize as genuine forms of “disciplinary piety” as Raul Sanchez calls it. These concerns may also echo strains of conservatism called out by Alexander and Rhodes—emerging from an established center of the assessment subfield, insisting that we establish ourselves in that center before trudging back toward the margins. Even more perniciously, this conservatism entrenches the entire field in a supposed status quo *praxis*; as one reviewer wrote in recommending rejecting this piece: “rubrics are part of the system we agreed to be a part of.” The paradox of disciplinary scholarship invites yet rejects disagreement—the community seeks commonality, which can, as Davis suggests, “demand that the Unthinkable remain unthinkable” (13). Or, as Erin J. Rand states, “the rhetorical agency to resist is paradoxical; even when one seeks to defy the hierarchies of dominant social institutions, one’s agency to speak or act at all is facilitated by the same institutions that are experienced as limits to one’s freedom” (14). I recognize that in affixing my work and perspective to the disciplinary commons, my marginal queer positionality is both facilitated and at risk. In this attempt to adopt and maintain a marginal position, to acknowledge queer possibilities without overtly contradicting, replicating, or painstakingly reviewing previous or concurrent scholarship, I understand the difficulty of the task and recognize its potential in being viewed askance by my colleagues. Yet in order to address the

(im)possibility of queer theory for writing assessment, unexpected (di)stances must address the straightness of our rhetorical practices, which, no matter how socially responsive we attempt to be, remain metonymic to the constraints and necessities of a university discipline: product-based, hyper-meaningful, and legitimacy-seeking methods that disenfranchise difference even as they seek institutional enfranchisement.

In that sense, I believe the rhetoric of agreement in writing assessment practice can be seen as a contributor in altering the fundamental contextual and situated teaching-learning process by producing subjects enframed—though not ensconced—in an explicitly formalized and codified social and intellectual culture, a culture which disciplines and protects itself. Maintaining its power, it acts rhetorically, as Rand describes, exercising agency by deferring “temporarily the possibility of acting or speaking otherwise,” which, she says, “inaugurates the illusion of the intending subject” (23). Thus, as Audre Lorde notes, it becomes less possible to change using “the master’s tools”—the tool, in this case, being reliability as a widely accepted validation of fair, accurate, replicable, and usable data in holistic writing assessments. Rand, following Butler, reminds us that we cannot disassociate ourselves from the reiterations of power in our subjectivity—meaning that we cannot “resist” power outside of that power that defines the forms of resistance. So while the idea of independent resistance, as Lorde advocates, appeals to us, what Rand explains is that “queerness animates resistance within and through the conventions of rhetorical form” (22) by remaining “undecided,” or rather, by acknowledging that “rhetorical agency persists only insofar as the meaning and effects of one’s rhetorical acts are not settled in advance” (23). The challenge for us, then, is to queerly value and to maintain an undecidedness that seems to lack rhetorical assertion, clarity, and governing intention.

But underlying assessment mandates is the very desire for clear persuasion, for predictability, for an establishment of patterns that anticipate what will occur and so determine steps to follow for better “success,” as Charles Harvey notes, paraphrasing Pierre Bourdieu and John Dewey. He states that assessments

are attempts at the measurement and objectification of successful habitus as witnessed in successful practitioners in the fields. Once the successful habitus is objectified, reified, codified, and so on, there is an attempt to make it function as an antecedent to behavior in the hopes that it will produce the consequent behavior that it was originally based upon. (199)

Many writing assessment practices, encouraged by social scientific legitimacy, adopt this “scholastic fallacy” by using consequents of past experience—rubrics—now made antecedents to make the task more efficient by avoiding the rumination required to create the rubric in the first place.<sup>[2]</sup> A rhetoric of agreement guides the process of creating or modeling a rubric, and continues as raters are trained, calibrated, and then expected to commit to a process that ensures agreement by at least two raters on the quantified score of a student paper. The creation and use of a rubric is a rhetorical act, identifying categories of definition—frozen in time and place—for the deliberation of assessing student work. The rubric itself is a medium, not neutral, but also neither requisite nor detrimental to learning. Plenty of us use rubrics or scoring guides that arise out of our own context and practice, and plenty of us reject their use on the basis that each reading of a student paper is its own contextual and situated experience (see Wilson). But the rubric, when generalized beyond its distinct rhetorical moment— and with the assumed necessity of calibration— affects us as experts or developing experts of student writing. Much scholarship on the complexity and contextuality of writing affirms that a calibrated method of writing assessment struggles to match the authentic validity of an individual teacher’s assessment of a student paper during a semester because that teacher alone can take account for the rich complexity of the processes that led to a student’s paper (see Elbow; Gallagher “Being There;” Lynne; O’Neill, Moore, and Huot; Moss; Neal; Purves). Contextual knowledge represents a form of situated expertise, or habitus, both in terms of what the teacher is teaching and how the teacher understands whether students are learning. Such expertise is manifested intuitively, an idea shown by philosophy scholars Hubert and Stuart Dreyfus, the expertise scholar K. Anders Ericsson, and in our field, William Smith. Intuition, of course, is scientifically queer, for it resists the requirement of outside or empirical verification; indeed, it resists replicated verity as validation, proposing instead that extensive experience affords individually nuanced interpretations by multiple individuals that is more valuable in their ecological complexity than multiple individuals arriving at one clear determinate interpretation.

## Dwelling in Disagreement

Importantly, however, intuitive expertise does not mean wholly mastered, nor is it a fixed position. Further, experts are not unerring, and collaborative verification should not be dismissed categorically. In my emphasis here, I queerly “circumscribe,” in Rand’s terms, via excess and indeterminacy the assumption that collaborative calibration among experts leads to a “more correct” answer, especially in assessing writing that is *ambient* by nature: already contextual, hermeneutic, and subjectively unanswerable. And that ambience, according to Rickert, “is given a more vital quality; it is not an impartial medium but an ensemble of variables, forces, and elements that shape things in ways difficult to quantify or specify. These elements are simultaneously present and withdrawn, active and reactive,

and complexly interactive among themselves as much as with human beings” (7). Such an awareness differs starkly from some program-assessment practices; for whether evaluating individual papers or portfolios, the agreement paradigm resonates, disregarding the ambience and foregrounding inter-rater reliability, elevating its representation of impartial validity as the most effective and acceptable way to argue legitimately with our institutional and accreditation administrators. If we engage frequently in calibrated practices, we, as functions of other functions (Davis 23), internalize and are internalized by the context, which often projects calibration as accurate and valid, leading to a state of uncertainty as to our own judgment, possibly diminishing the actual complexity of our work and our dynamic identities. As Harvey said, the codification of complexity through imposed assessment leads to professionals who are “existentially cramped, crippled, and stunted ... ; they are increasingly made incompetent, increasingly bereft of personal judgment sensitive to situational context. They are made, instead, utterly dependent on rules, regulations, and past authorities for the performance of their field activities” (199).

A disturbing consequence of the K-12 common core state standards and its streamlining of testing models is the erosion of individual teachers’ situational judgments, drifting farther away from acknowledging the fluidity of learning, imagination, or the careful observation of “the child in motion” as she “goes about learning or making something” (Himley and Carini 9). The field of rhetoric and composition has unfortunately been an unwitting leader in this development, when by presumed necessity it legitimized assessment practices via social science-inspired methods (Walker). As Geoffrey Sirc notes of the field’s disciplinary transition:

We took out a long term lease on a classier, more institutional setting in which to hold our gatherings, a space much more befitting of our newly disciplined resolve to achieve professional parity with our colleagues, becoming part of the traditional academic enterprise; a “new social scientism” seemed just the thing to de-kookify writing and make our work *just like theirs*. (211)

Perhaps it is time that we reconsider our position by queering the rhetoricity of our unifying practices. Part of this might involve viewing intuitive expertise as a manifestation of *dwelling in disagreement*, a marginal and imaginative alternative to rhetoric’s subjective assertiveness, and slantwise to the normative calibration practices that subsume situational judgment. Disagreement is familiar, for our disciplinary knowledge is produced through generative opposition, dissent, and disputation. And yet those spaces of disagreement and difference are often limited and avoided because they can be uncomfortable and disorderly, meaning that while we may not see our colleagues as “radical alterities,” we nevertheless are less likely to approach differences with a desire to potentially remain in discomfort, as Matthew Heard suggests in connecting rhetoric to attending to the *tone* of interactions. The concept of attunement, Heard writes, “describes less an act of interpretation than a recurring, prolonged dwelling within the complexities of tone” (46). Like queer theory, the examination of attunement and rhetoric together raises questions about fixed identities and emphasizes our situational *actions* in interactions with difference; for “tone is, by nature of its physical properties, uncertain” (Heard 48), and rhetorical attunement embraces “materiality, contingency, emergence, resistance” (Leonard 230). The uncertain, unsettled aspects of difference queerly affirm the value of approaching, attuning to, and dwelling in those fluid spaces that constitute, as Rickert suggests, the ambient and non-linear disagreement and difference of multiple, scaled, agencies—especially those flattened, ignored, or pushed aside by rationalist, masculine, and capitalist imperatives.

## A Failed Study Fails Again

I mentioned near the beginning of this essay a failed assessment study I conducted. The details of the study, which proposed to show that individual instructors could “intuit” valid ratings of student writing as effectively as normed raters, are in the Appendix, but the relevance of that study is that when my hypothesis failed to be validated by statistical reliability, I chose to neither accept nor reject the null hypothesis. In other words, I remained undecided despite the empirical results. While such a position is arguably indefensible, I attempt in the remainder of this essay to further explain how the uncertainty that resulted from my failure to confirm my hypothesis is not a dogmatic stubbornness, but instead an element of a larger failure that I recognize as queer marginality—an ambivalence for “success” and a willingness to dwell in the disagreeable spaces of academic scholarship.

Once I realized that my study had failed, I faced a choice to adjust the study and conduct it again by attentively increasing the reliability or to leave it as a failure. The traditional view of failure, as Halberstam notes, “goes hand in hand with capitalism” (88), employing the cliché that to fail means to try until one succeeds – and the persistent always will. Journal articles describing calibrated-scoring methods for programmatic assessment fit with the capitalistic sense of success—winners can be identified through successful studies employing narrow ranges and definitions of measurement and the losers, well, the losers aren’t published because their studies cannot be validated. In the sense that assessment has become a subject of empirical research within English studies, even adopting APA citation style and requiring the reporting of statistical significance, writing about assessment outside

this empirical frame can be quickly dismissed as being illegitimate for inclusion in the conversation. This gatekeeping function is (necessarily) part of our (academic) culture, making a clearly identifiable distinction in what passes as appropriate (scholarship). Yet, as Ratcliffe suggests in her encouragement for “rhetorical listening,” we have “an ethical responsibility to argue for what we deem fair and just while questioning that which we deem fair and just” (25). In terms of difference in writing assessment, I agree with her that altering our perspective from empirical-based “may help people invent, interpret, and ultimately judge differently in that perhaps we can *hear* things we cannot see” (25). However, such a view persists only from the margins; Halberstam, paraphrasing Scott Sandage’s *History of Failure in America*, reminds us that *seeing* is the ruler of legibility, for “losers leave no records, while winners cannot stop talking about it,” meaning that numerous stories of failure lie “quietly behind every story of success” (88). But that does not mean success is built on top of failure, as is often assumed. Queering failure, as Halberstam does, shifts failure from the capitalistic zero-sum game to a “way of refusing to acquiesce to dominant logics of power and discipline and as a form of critique.” Failure “quietly loses,” says Halberstam, “and in losing it imagines other goals for life, for love, for art, and for being” (88), adding nuance to Samuel Beckett’s well-known aphorism to “fail better.”

My failed study unintentionally queered my view of writing assessment – it was a surprise, but different than if I had sought to *make* it queer by including queer voices or something similarly social-justice oriented. The surprise came in the realization that perhaps we always fail. That realization implies that if I had tinkered with my method to increase the reliability of the rating group, or looked for other ways to validate the study, I would be assimilating into the dominant success narrative, trying to “win” and succeed through a clear path of baseline-to-improvement progress, measured by decontextualized reliable-validity and a rationalist rhetorical lens. Instead, I found through failure not a “lesson learned” for producing a better, successful study, but rather an alternate path of resistance that generates ideas *out of alignment* with my previous understandings of myself as a colleague, scholar, and teacher. I recognize a sensibility that responds to situations more readily than knowledge, and I value that sensibility despite its marginality. For example, I sense that in both the normed raters and the intuitive raters reading my program’s student work, the average score of 3 seems clearly estimable by any attentive writing program administrator or statistician, for that matter; yet this *not-knowing* is administratively unacceptable because it lacks documented empirical evidence. Our field insists on the contextuality and situatedness of writing and writing evaluation, which should alleviate concerns that exercising our expertise-based sensibility will transform into some sort of anti-empirical, free-for-all guessing game about all fields of knowledge. Yet quantified results still hold a superior position, indicating to decision-makers a legitimate, but reductive simplicity: “yes” or “no” on questions of placement, “poor” “fair” or “excellent” in exit portfolios, program effectiveness, and learning outcomes. And the contextual contingencies resulting from that reduction remain ignored by most decision-makers. For example: Will placement decisions using directed self-placement, for example, overwhelm existing and available courses and sections? Will exit portfolio readings stop students from graduating without causing an administrative and parental uproar? What if, as such questions produce a chicken-or-egg-first ambivalence, we decided to accept this unknowing, this “undecidability,” rather than try to overcome it?

As readers might expect, I don’t have answers. But in asking these questions and others on the heels of my failed assessment study, I remain in the marginal space that usually has been quickly abandoned in assessment scholarship. As Bourdieu anticipated, in conducting my study I had been so assimilated into the propriety of calibrated scoring that it did not occur to me that using it as the control for my study would contradict the basis for the study. I am not alone; the resurgent claims of essay-grading software draw on studies comparing computers favorably with calibrated human raters, causing statistical problems (Perelman). Human raters remain complex humans, in various stages of proficiency and expertise, and elevating human readers on the basis of their human-ness may hold back computer scoring for a time. However, calibrated human rating still suppresses the attunement of human-ness, and the degree to which we accept this machine-ing of ourselves affects the professionalism and degree of public trust in teachers as experts (Walker). The proposition to calibrate, to norm, and to suppress the complex differences among us is an accepted problem within the rhetoric of agreement. But “the ideology of consensus,” in Charles Willard’s terms, leads to groups “uniformly priz[ing] interpersonal harmony and ... dependent on a rhetoric of solutions” (145). While proponents of calibration sessions claim to endorse debate and controversy, Willard explains the problematic reality:

Controversy is a way station to somewhere, a temporary setback. We don’t value dissensus so much as we begrudge it a therapeutic effect—like surgery, a painful *rite de passage* through which ideas must pass. The final cause of the passage is harmony, success, and progress. (146)

Indeed, it is queer *not* to embrace harmony, success, and progress through deliberate empirical process. But embracing trust in our expertise requires a circumscription of our reiterative selves toward non-calibrated, fluid ecological beings who are defiantly not “trainable-by-code” machines. This can happen only if we dodge the capitalistic upward trajectory and calibrated agreement—from graduate assistant training to blind peer review—as philosophically beneficial and methodologically pure. Accuracy as a value is not constant—it is a measure within a

construct that has little or no meaning outside that construct, a masculinized myth of order and solution. Our aim as teachers is to facilitate learning, which stubbornly resists accuracy, consistency, generalizability, fairness, efficiency, or any other term that is usually applied to calibrated assessment. And our disciplinary responsibility includes teaching and practicing rhetoric as a “mode of reasoning and decision-making which allows humans to act in the absence of certain, a priori truth” (Jarratt 8).

The connection, or rather, the disconnection among expertise, disagreement, and “unified” writing assessment turns out to be the most interesting aspect of my failed study. If we are expert teachers, or on the way to becoming expert teachers, our pedagogy shifts or leaps constantly because one is responding to the ambience, to the numerous small or large interactions with individuals and texts and offices and classrooms and technologies that continuously alter the way we think and act. We do not need a study to validate this, just as we don’t need a study to validate that professional conversations, workshops, and shared assessment sessions make us more *reflective*, and thus possibly more *effective*, teachers. But the improvement that occurs by such experiences should not mislead us into thinking that agreement and conformity are solely responsible, and thus deserving of becoming political or rhetorical priorities. Expertise is not a culmination of this type of work but rather a close cousin to the idea of attunement: an ongoing process of approaching situations to seek and gain and recognize knowledge, then seeking and gaining and recognizing more, including the excess, within varying contexts and situations. Expertise is not fixed; the intuition assists but does not govern decisions in the same way each time, just as attunement involves an awareness of mood and conscious “rhetorical listening” that are highly dependent on the often unfamiliar cues of the situation. Further, Rickert, contrasting Burke’s and Heidegger’s views on intuition, says:

For Burke, the notion of ‘acting-with’ explains this process: intuitions are caught up in a wider orbit of meanings that make them resound for us as the symbolic animals we are. But as Heidegger intimates, this leaves us with the problem of having to ‘springboard’ back into the world from our experience of it. Heidegger, we might say, simply closes this gap. There is no bare intuition of something; there is only the experience itself already in the perception. (172)

Likewise, because writing’s complexity is not “a Thing” in Latour’s sense (see Lynch), we cannot treat any assessment of it as a solution already found, or dismiss the “experience already in the perception” manifested in writing and evaluating; we must always retain the acknowledgement of writing’s uncertainty, which is understood in intuitive expertise. Full agreement is unlikely among writing experts—again, we fail—so we should resist demanding pseudo-agreement by insisting that experts voluntarily constrain their expertise—or overcome failure—within an imposed frame. I believe that beneficial frames or occasions for expert agreement exist, but at present the importance of agreement is directly related to the accountability, efficient, and ethical value placed upon the assessment, values that have moved beyond the initial development of calibrated-rater models as a defense against models threatening our discipline (Haswell; Herrington and Moran; White; Williamson and Huot; Yancey). More threatening, however, is how the ostensible purpose of assessments—a measure of learning—has been subverted by orderly, tangible, and hyper-meaningful results; results that reduce, quantify, and highlight overly specific learning outcomes to politically compensate for the slippery, unaccountable, messiness of actual student learning.

## Standing On the Table

My aim here is not to undermine writing assessment methods or practices; rather, I hope to highlight an alternate perspective that maintains a healthy uncertainty in our rhetoric. Unfortunately, results-oriented program and accreditation assessments seem to be politically necessary, and they rely on the rhetoric of agreement to assuage the disconnect between the results and student learning. A queer sensibility highlights the danger that political necessity will morph into disciplinary fundamentalism, helping us be actively cognizant of how assessments that suppress or reduce complexity and difference are ontologically suspect. Agreement within any norming group is situational, limited to a temporary construction that will inevitably change when the group convenes another year. Using the same construct again later does not align the results (or “close the loop”) of repeated assessments: artifacts are written by a whole new set of students, and the raters may be different or have another year’s experience that alters their internal negotiation of the construct. Calibration in assessment is credited in obviating those differences. But looking at this from a queer perspective questions the value of that obviation; a sidelong, queer glance sees the divergent space between the end results and the initial calibration sessions as most interesting, because it constitutes the ways those differences are discussed, negotiated, *and* accepted. Yet these spaces remain invisible to the ultimate stakeholders of the information. In other words, outside the calibrated group we do not know how much each individual compromised his or her own experience to norm with the rest of the group. Obviating their differences as a confirmation of the validity (in the institutionally regarded sense) of their decisions ignores the complex processes that fused their varying levels of expertise into assent. Their conversations and disagreements during calibration and rating likely served as valuable professional development, increasing their

experience and expertise, but such growth is hidden from view, obscured completely by “success”—or the reported results.

Thinking along these lines, for me, has spurred ideas for my program to utilize disagreement and difference in a productive and seemingly meaningful way—by involving and trusting all of our program faculty in determining “what we value” and how well their students attain those values—without requiring consensus (Osborne and Walker). That effort relates to “writing program assessment as the process of documenting and reflecting on the impact of the program’s coordinated efforts” (White et al 3) without pressure for any of those efforts to conform. Yet, for many, the frightening result of these efforts to rest unassured keeps writing instructors and writing programs illegible and thus illegitimate (Butler). Cue Alexander and Rhodes calling “queer” composition’s “impossible subject.” As a discipline, we often think ourselves too new and on apparently too shaky of ground to risk the perception that we lack assurance of our value, a way of thinking that seemingly justifies “informed, programmatic practices” to defend against the “Age of Accountability” that appears to threaten writing instruction (White et al. 17). Despite our scholarly insistence that writing is too complex and too situated to fit either quantification or the frame of agreement, the institutional and disciplinary realities compel us to do what is necessary to flee the margins, margins where writing assessment’s appropriateness and validity—and I invoke here Gallagher’s validation heuristic model that is locally determined but guided by disciplinary values (“Assessing”)—could well be gauged by how much *disagreement* and *undecidedness* it produces. As Dreyfus and Dreyfus suggest, disagreement is a hallmark of expertise. If experts do not disagree with each other on some points within a complex field, they are probably not experts. Or, more likely, imposed reliability standards force them to withhold their proficiency for the sake of efficiency. Broad’s dynamic criteria mapping identifies the numerous ways teachers value writing, but even his method attempts to corral those differences by categorization in order to make order out of the subjective chaos of open inquiry. My critique of Broad is soft here—his method has done much to reform some effects of strict, general rubrics on writing assessment. And yet, drawing from the margins of queer theory, I think we should do more than reform; we can instead step aside, which requires us to revel in, not corral, the invigorating differences and ambience within our community of writing instructors and scholars. As Lorde said: “Without community there is no liberation ... But community must not mean a shedding of our differences, nor the pathetic pretense that these differences do not exist” (113).

Paradoxically, in many writing assessment reports, difference and disagreement within a group are deemed fatal to an assessment’s success. Galen Leonhardy and Bill Condon note in their study of “Tier 2” portfolio assessment that raters evaluating student papers from disciplines other than their own disagreed over half the time with raters from the same disciplines as the students (76). Although the assessment’s intent was to “liberate” writing across disciplines by bringing disciplinary communities together to evaluate, the low reliability spurred Leonhardy and Condon to suggest raters come together for more calibration sessions. I bristle at this solution, for it implies that disagreement must be conquered, “acced[ing] to the masculinist myth of Herculean capitalist heroes who mastered the feminine hydra of unruly anarchy” (Halberstam 18). As Heard writes, “Attending to tonality—attunement – describes a complex process of moving, flexing, reading, and responding that *still fails* to capture the ever-modulating resonance of tone generated in contact with others” (49, my emphasis). The university thrives on different disciplinary discourse communities that *fail to agree*. In fact, the more we agree, the more likely we are “seeing like the state” (Scott, qtd in Halberstam 9), which

means to accept the order of things and to internalize them; it means that we begin to deploy and think with the logic of orderliness and that we erase and indeed sacrifice other, more local practices of knowledge, practices moreover that may be less efficient, may yield less marketable results, but may also, in the long term, be more sustaining. (Halberstam 9)

Yet noisy, dominant forms of agreement continue to drown out the ambience of writing assessment.

Consider Peggy O’Neill’s encouragement to collaborate with those with expertise in statistical measures, because “validity and reliability connect to values such as accuracy, consistency, fairness, responsibility, and meaningfulness that we share with others, including psychometricians and measurement specialists” (“Reframing Reliability”). On the surface, these seem to be values we can stand behind, yet underneath those values is a rejection of their excess, as Ratcliffe describes:

Simultaneous recognitions [of commonalities and differences] are important because they afford a place for productively engaging differences, especially those differences that might otherwise be relegated to the status of ‘excess.’ Excess refers to that which is discarded in a culture’s dialogue-as-Hegelian-dialectic; that is, when the thesis and antithesis are put into play, the excess is what is left out of the resulting synthesis. An engagement with differences-as-excesses is important, for as Lorde asserts: ‘It is not those differences between us that are separating us. It is rather our refusal to recognize those differences, and to examine the distortions which result from our misnaming them and their effects upon human behavior and expectation’ (“Age” 115). (95)

The values mentioned by O'Neill represent a collaborative understanding of order that resists questioning (who wants to be labeled as unfair or irresponsible?) as well as troublesome excess, such as whether a consistent and fair assessment that satisfies administrators will be meaningful to teachers or whether a robust, meaningful assessment for teachers is too inconsistent for legislators. And in the larger sense, even if the aforementioned values do connect, our field's strained collaborations with organizations such as ETS, Pearson, or the College Board have only continued to undermine and overwhelm our rich and excessive theories of writing, given that writing assessment in K-12 is arguably fulfilling our worst fears of automated scoring and removing teachers from curriculum and exam development. Yet we are still encouraged to play nicely, and judging from the breadth of assessment work in our discipline, perhaps our prevalent unity *is* our kindness. O'Neill and Linda Adler-Kassner adopt an optimistic non-radical stance, saying that we just need to get involved and engage in conversations regarding assessment (*Reframing Writing*). That is a polite, probably ineffective solution. As Gallagher says in his review of Adler-Kassner's and O'Neill's book, "Reasonable, moderate, cooperative participation—a seat at the 'stakeholders' table—may not be enough" ("Book"). What would be enough? He does not say, but collaboration is not resistance, and the "seat at the table" metaphor does not reimagine anything, only insists on perfunctory access to an already masculine, capitalistic, and entrenched institutional space. Thus, Lorde's admonition that "the master's tools will never dismantle the master's house" reminds us that rhetorical agency's existence requires ideas not settled in advance (Rand), and an invitation to sit at the table is too often merely a gesture.

The queer perspective invoked here reminds us of generative alternate positions—standing on the table? hiding underneath? turning it upside down?—from which we might view differently writing assessment always already providing "an available rhetorical moment" (Yancey). We can be less cooperative and acquiescent and be more disruptive; we can resist troubling trends in K-12(20) assessment by reminding ourselves of Bourdieu's habitus concept neatly summarized by Harvey: "without thinking, without intending to, we reproduce the world that produced us" (197), and which Asao Inoue explains as an underlying factor of inherently racist assessments everywhere (58). Acknowledging this structural and cultural habitus can help us sample or adopt the queer positionality of embracing the excess—trying to see and work from the margins without pulling closer to the always already problematic center. We can continue to deconstruct, to question the very foundations of writing assessment and explore our own presumptions of validity and reliability, perhaps from the perspective of rhetorical attunement, which, according to Rebecca Lorimer Leonard, recognizes rhetoric's influence as valuing and highlighting "instability and contingency, ... political weight and contextual embeddedness" (230). Altering the lens of writing assessment does not undermine previous scholarly efforts in writing assessment, but instead prevents our work from being appropriated and misused by political and corporate opportunists. Basically, I encourage stronger and impolite resistance, a "shattering laughter" (18), as Davis suggests, so that potential and possibility can be *maintained* (Haswell and Haswell 41). I want to embrace the search for "*queer* rhetorical practices—practices that recognize the necessity sometimes of saying 'No,' of saying 'Fuck, no,' offering an impassioned, embodied, and visceral reaction to the practices of normalization that limit not just freedom, but the *imagination* of possibility, of potential" (Alexander and Rhodes 193). Such convulsion likely causes some consternation. Yet not only should we be wary of our own assessment-induced lack of *phronesis*, we should also actively fight it in our students by insisting—to paraphrase Paul Lynch, who draws on Latour, and Anthony Petrucci, who draws on Heidegger and Gadamer—that there is no hidden object that assessment can find, no universal "order" to learning and teaching, and no "ultimate revelation" waiting at the end of a lecture, assignment, or most importantly, after assessment of the predetermined outcomes of pedagogy.

Expertise theory and queer theory together suggest that while our masculine and capitalist society urges us to maintain a heteronormative temporality, striving for tangible antecedents—evidence, explanations, outcomes, deliberation, rules, guidelines—we thrive when we circumvent these through experience and creative difference, leaving more things "undecided" by recognizing that quantitative, calibrated methods promising accuracy and answers are essentially positivism dressed up as optimism. Halberstam's queering of failure offers what I seek in assessment—an alternate outlook that exists *because* of our failure to arrive at those answers:

Not an optimism that relies on positive thinking as an explanatory engine for social order, nor one that insists on the bright side at all costs; rather this is a little ray of sunshine that produces shade and light in equal measure and knows that the meaning of one always depends upon the meaning of the other.  
(5)

Like the non-definitive sex that serves as the loci of queer theory's marginal resistance, teaching and learning actively resist neat explanation and standardization: they are rhetorically ambient rather than conventionally straightforward, messy and fluid, embedded with invigorating complements and disruptive dissonance, with more surprises than answers. According to Rand, because resistance can never be separated from institutional power structures to which it is directed, active resistance always displaces queerness. But queerness cannot fully be excluded, she says, and "it is in this *imperfect* displacement of queerness, the dangerous pleasure of risk" (168) that

failure and negativity and marginality set us in motion. Queerness, like teaching and learning and writing and assessment, holds no elixir qualities. But like anything meaningful, it and other theories and practices consist of “shade and light in equal measure,” always preventing our arrival by keeping us wandering into discomfitive places where we may attend, pause, move around, and perhaps stay for awhile and *dwell* before returning, if we must, if ever, to the well-marked, mainstream paths of perceived certainty and success.

## Appendix: “Failed” Study

In 2010, my university’s administration mandated a large-scale writing assessment in response to pending accreditation, and chartered a holistic scoring team of full-time and part-time English faculty to develop and calibrate to a six-point holistic scoring guide that would assess writing across the entire university. I was a member of the committee (and at the time an untenured faculty member and coordinator of first-year composition) charged with developing an assessment plan and holistic rubric. In our meetings, I repeatedly raised concerns with the process and rubric until I was asked by an administrator to step down in order for the process to move forward quickly. The slight was minor and temporary, but these circumstances led me to approach the local issue with a scholarly exploration of writing assessment with help and support of colleagues.

The mandated assessment team scored 223 first-year composition papers (8-10 pages each) using the established scoring guide. Inter-rater reliability was over 90%, and the average score of the 223 papers hovered right around a 3. With those scores and papers available, I attempted to measure intuitive scoring of the same papers—spurred by curiosity from reading Malcolm Gladwell’s *Blink*. I assembled a group of eight colleagues who were not on the holistic scoring team and had various areas of disciplinary training – literature, creative writing, TESOL—and a range of teaching and writing experience. Most were experienced professors—experts – but one was a graduate assistant who had taught two semesters of first-year composition and another was an undergraduate student. The aim was to measure their judgment of student work when given a short time to do so. Each reader received a folder with 28-30 student papers divided into three sets. They began reading the first paper of the first set, then 45 seconds later, I asked them to stop reading and immediately write down a score between 1 and 6, with 6 representing the best possible score. The group read and scored the first two sets of 10 papers as they had the first paper. For the third set, I instructed them to find the conclusion of each paper and read it first, then any other part of the paper within the 45-second timeframe. Papers were read by only one reader and we completed evaluating the 223 papers in less than 30 minutes.

In my comparison of the two sets of results, as shown in [Table 1](#), the mean score (3.29) of student papers by the “intuitive” raters is slightly higher than the combined score (3.13) of the holistic scorers, though not statistically significant for the sample size. The median score of both groups stands at 3. The reliability between the two groups was 68%, calculated by 72 of 223 scores that differed more than one point between the rating groups. Of those 72 scores, 37 were 1.5 points apart, meaning that they were just beyond the acceptable range of difference between raters but within acceptable range of *one* of the two holistic raters. More significant differences, which in a traditional scoring situation would require a third reader, numbered 35, with only 12 of that group differing 3 or more points from the combined holistic scorers’ score.

**Table 1.** Comparison of average scores from the two assessments

	Holistic Rater 1	Holistic Rater 2	Holistic Raters Combined	“Intuitive” Raters
MEAN	3.09	3.14	3.13	3.29
MEDIAN	3	3	3	3
MODE	3	4	3.5	3

Although the overall reliability reflects fairly positively on the “intuitive” raters, a Pearson correlation analysis showed that the individual scorers did not agree very often. Overall, the “intuitive” raters correlated only 27% of the time to the combined score for the holistic raters. But when the correlation analysis is narrowed to individual fast raters, the relative percentages show a difference for those who had taught 30 sections or more (see [Table 2](#)).

**Table 2.** Correlation Analysis showing relationship to number of courses taught over career

--	--	--	--

FYC Courses Taught in Career	Correlation (Pearson)	Departures from Holistic Team	
0	.08	11	Less-Experienced Raters: 82 of 223 36 Departures (56% agreement)
2	.30	15	
3	.22	10	
16	.37	9	More-Experienced Raters: 141 of 223 36 Departures (75% agreement)
25	-.03	7	
30	.55	7	
38	.46	8	
55	.48	5	

These data illustrate course-taught expertise (Smith). The “intuitive” raters drew upon what they knew—Bob Broad called this “teachers’ special knowledge” (“Reciprocal”)—to make their decisions independent of a common rubric. Because the ostensible accreditation purpose of our holistic program assessment was to identify a baseline average score of a representative sample of student papers, it should be noted that the quick reading of these papers proved just as effective in arriving at the same quantified average as the traditional method, but it used less time and fewer resources (and with no machines or automatons involved). In other words, some sort of validity held without a high reliability figure. However, the correlative reliability as measured by the Pearson test was too low for statistical significance.

## Notes

1. The model is outlined and discussed at length by Osborne and Walker in *Assessing Writing*, but in brief, rather than calibrated raters scoring student writing samples, the writing program is assessed by collecting individual surveys completed by program instructors (trusted as experts), who assess their students collectively—rating their collective performance in meeting program writing objectives, outcomes, and expectations on a 1-5 scale. The survey results afford a snapshot, twice a semester, regarding how students in the program are performing in relation to the program objectives, and while it is possible for teachers to inflate scores, the anonymity of the process, we have found, is more likely to result in teachers frankly assessing their students and evaluating their own activities and role in contributing to the students’ performance. Beyond this, however, the program coordinator is able to identify particular objectives that students across the program are struggling with, which provides immediate opportunities for professional development workshops. ([Return to text.](#))
2. The use of rubrics is often rationalized as student-friendly—to help students know what to expect and how to succeed with an assignment. Aside from the problems with the limited and narrow definitions of success such practices work within, the reality is that rubrics are less student-friendly than belabored-teacher-friendly; they are implemented purposefully to provide an illusion of predictability, accuracy, and fairness to student success with acceptable minimal effort from both teacher and student. Many factors contribute to this, including large class sizes, contingent labor exploitation, and heightened expectations for documenting learning growth, but we should be careful to recognize what we and our students lose, if, by using any rubric, we skip over the difficult continuous “rumination” as a reader when we evaluate individual papers. ([Return to text.](#))

## Works Cited

- Adler-Kassner, Linda and Peggy O'Neill. *Reframing Writing Assessment to Improve Teaching and Learning*. Utah State UP, 2010.
- Alexander, Jonathan, and Jacqueline Rhodes. "Queer: An Impossible Subject for Composition." *JAC: A Journal of Composition Theory*, vol. 31, no. 1, 2011, pp. 177-206.
- Booth, Wayne. *Modern Dogma and the Rhetoric of Assent*. University of Chicago Press, 1974.
- Bourdieu, Pierre. *Outline of a Theory of Practice*. Translated by Richard Nice, Cambridge UP, 1990.
- Broad, Bob. "Reciprocal Authorities in Communal Writing Assessment: Constructing Textual Value within a 'New Politics of Inquiry.'" *Assessing Writing*, vol. 4, no. 2, pp. 133-167.
- . *What We Really Value: Beyond Rubrics in Teaching and Assessing Writing*. Utah State UP, 2003.
- Carr, Nicholas. *The Shallows: What the Internet is Doing to Our Brains*. W.W. Norton, 2011.
- Condon, William. "Large-Scale Assessment, Locally Developed Measures, and Automated Scoring of Essays: Fishing for Red Herrings?" *Assessing Writing*, vol. 18, no.1, 2013, pp. 100-108.
- Davis, D. Diane. *Breaking Up (At) Totality: A Rhetoric of Laughter*. Southern Illinois UP, 2000.
- Dreyfus, Hubert and Stuart Dreyfus. *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. The Free Press, 1986.
- Elbow, Peter. "Ranking, Evaluating, and Liking: Sorting Out Three Forms of Judgment." *College English*, vol. 55, no. 2, 1993, pp. 187-206.
- Ericsson, K. Anders, et al. *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge University Press, 2006.
- Gallagher, Chris W. "All Writing Assessment is Local." *CCC*, vol. 65, no. 3, 2014, pp. 486-505.
- . "Assessing Locally, Validate Globally: Heuristics for Validating Local Writing Assessments." *Writing Program Administration*, vol. 34, no. 1, 2010, pp. 10-32.
- . "Being There: (Re)Making the Assessment Scene." *College Composition and Communication*, vol. 62, no. 3, 2011, pp. 450-476.
- . "Book Review: Adler-Kassner's and O'Neill's Reframing Writing Assessment." *Present Tense*, vol. 2, no. 1, 2011.
- Gladwell, Malcolm. *Blink: The Power of Thinking Without Thinking*. Little, Brown, 2005.
- Halberstam, Judith. *The Queer Art of Failure*. Duke UP, 2011.
- Harvey, Charles. "Making Hollow Men." *Educational Theory*, vol. 60, no. 2, 2010, pp. 189-201.
- Haswell, Janis and Richard Haswell. *Authoring*. Utah State UP, 2010. Print.
- Haswell, Richard H. "Automatons and Automated Scoring: Drudges, Black Boxes, and Dei Ex Machina." *Machine Scoring of Student Essays: Truth and Consequences*, edited by Ericsson, Patricia Freitag and Richard H. Haswell, Utah State University Press, 2006. 57-78.
- Heard, Matthew. "Tonality and Ethos." *Philosophy and Rhetoric*, vol. 46, no. 1, 2013, pp. 44-64.
- Herrington, Anne and Charles Moran. "What Happens When Machines Read Our Students' Writing?" *College English*, vol. 63, no. 4, 2001, pp. 480-499.
- Himley, Margaret and Carini, Patricia F. *From Another Angle: Children's Strengths and School Standards : the Prospect Center's Descriptive Review of the Child*. Teachers College Press, 2000.
- Huot, Brian A. "The Influence of Holistic Scoring Procedures on Reading and Rating Student Essays." *Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations*, edited by Williamson, Michael

- M. and Brian A. Huot, Hampton Press, 1993, pp. 206-236.
- . *(Re)Articulating Writing Assessment for Teaching and Learning*. Utah State UP, 2002.
- Inoue, Asao B. *Antiracist Writing Assessment Ecologies: Teaching and Assessing Writing for a Socially Just Future*. WAC Clearinghouse/Parlor, 2015.
- Jarratt, Susan. *Rereading the Sophists: Classical Rhetoric Refigured*. Southern Illinois UP, 1991.
- Leonard, Rebecca Lorimer. "Multilingual Writing as Rhetorical Attunement." *College English*, vol. 76, no. 3, 2014, pp. 227-247.
- Leonhardy, Galen and William Condon. "Exploring the Difficult Cases: In the Cracks of Writing Assessments." *Beyond Outcomes: Assessment and Instruction Within a University Writing Program*, edited by Richard H. Haswell, Ablex, 2001.
- Lorde, Audre. "The Master's Tools Will Never Dismantle the Master's House." 1984. *Sister Outsider: Essays and Speeches*. Ed. Berkeley, CA: Crossing Press. 110-114. 2007.
- Lynch, Paul. *After Pedagogy: The Experience of Teaching*. NCTE/CCCC, 2013.
- . "Composition's New Thing: Bruno Latour and the Apocalyptic Turn." *College English*, vol. 74, no. 5, 2012, pp. 458-476.
- Lynne, Patricia. *Coming to Terms: A Theory of Writing Assessment*. Utah State UP, 2004.
- McLuhan, Marshall. *Understanding Media*. 1966. Gingko Press, 2003.
- Moss, Pamela. "Can There be Validity Without Reliability?" *Educational Researcher*, vol. 23, no. 4, 1994, pp. 5-12.
- Neal, Michael. *Writing Assessment and the Revolution in Digital Texts and Technologies*. Teachers College Press, 2011.
- O'Neill, Peggy. "Reframing Reliability for Writing Assessment." *Journal of Writing Assessment*, vol. 4, no. 1, 2011.
- O'Neill, Peggy, Cindy Moore, and Brian Huot. *A Guide to College Writing Assessment*. Utah State UP, 2009.
- Osborne, Jeff and Paul Walker. "Just Ask Teachers: Building Expertise, Trusting Subjectivity, and Valuing Difference in Writing Assessment." *Assessing Writing*, vol. 22, 2014, pp. 33-47.
- Perelman, Les. "Critique of Mark D. Shermis & Ben Hammer, 'Contrasting State-of-the-Art Automated Scoring of Essays: Analysis.'" *Journal of Writing Assessment*, vol. 6, 2013.
- Petruzzi, Anthony. "Articulating a Hermeneutic Theory of Writing Assessment." *Assessing Writing*, vol. 13, 2008, pp. 219-242.
- Purves, Alan. "Reflections on Research and Assessment in Written Composition." *Research in the Teaching of English*, vol. 26, no. 1, 1992, pp. 108-122.
- Rand, Erin J. *Reclaiming Queer: Activist & Academic Rhetorics of Resistance*. University of Alabama Press, 2014.
- Ratcliffe, Krista. *Rhetorical Listening: Identification, Gender, Whiteness*. Southern Illinois UP, 2006.
- Rickert, Thomas. *Ambient Rhetoric*. University of Pittsburgh Press, 2013.
- Sanchez, Raul. "First, A Word." *Beyond PostProcess*, edited by Dobrin, Sidney I., J.A. Rice, and Michael Vastola, Utah State UP, 2011, pp. 183-194.
- Santayana, George. *The Sense of Beauty: Being the Outline of Aesthetic Theory*. Dover, 1955.
- Scott, Tony and Lil Brannon. "Democracy, Struggle, and the Praxis of Assessment." *College Composition and Communication*, vol. 65, no. 2, 2013, pp. 273-298.
- Sirc, Geoffrey. "The Salon of 2010." *Beyond PostProcess*, edited by Dobrin, Sidney I., J.A. Rice, and Michael

Vastola, Utah State UP, 2011, pp. 195-218.

Smith, William L. "Assessing the Reliability and Adequacy of Using Holistic Scoring of Essays as a College Composition Placement Technique." *Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations*, edited by Williamson, Michael M. and Brian A. Huot, Hampton Press, 1993, pp. 142-205.

---. "The Importance of Teacher Knowledge in College Composition Placement Testing." *Reading Empirical Research Studies: The Rhetoric of Research*, edited by John R. Hayes et al., Lawrence Erlbaum Associates, 1992, pp. 289-316.

Walker, Paul. "Composition's Akrasia: The Devaluing of Intuitive Expertise in Writing Assessment." *enculturation*, vol.15, 2013. <http://enculturation.net/compositions-akrasia>.

White, Edward et al. *Very Like a Whale: The Assessment of Writing Programs*. Utah State UP, 2015.

Willard, Charles Arthur. "Valuing Dissensus." *Argumentation: Across the Lines of Disciplines*, edited by De Gruyter, Mouton, Dordrecht, Foris Publications, 1986, pp. 145-158.

Williamson, Michael and Brian A. Huot, eds. *Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations*. Hampton Press, 1993.

Wilson, Maja. *Rethinking Rubrics in Writing Assessment*. Heinemann. 2006.

Yancey, Kathleen Blake. "Looking Back As We Look Forward: Historicizing Writing Assessment." *The Norton Book of Composition Studies*, edited by Susan Miller, Norton, 2009, pp. 1186-1204.

"Let's Disagree (to Agree)" from *Composition Forum* 35 (Spring 2017)

© Copyright 2017 Paul Walker.

Licensed under a [Creative Commons Attribution-Share Alike License](#).



Return to [Composition Forum 35 table of contents](#).