

UC Davis

Journal of Writing Assessment

Title

Introduction to Volume 4

Permalink

<https://escholarship.org/uc/item/0546n464>

Journal

Journal of Writing Assessment, 4(1)

Author

Huot, Brian

Publication Date

2011

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Introduction to Volume 4

by Brian Huot

Introduction

The first time I wrote an introduction for a writing assessment journal was for the inaugural issue of *Assessing Writing* in 1994. Since 1994 I have written or co-authored twenty or so introductions, with this being (by my count) the twenty-first. This introduction is different in many ways. I write not to introduce a single issue but a whole volume, my last volume as editor of the *Journal of Writing Assessment (JWA)*. Such an occasion provides an opportunity to address the changes in editorship and delivery and to address the overall position of *JWA* within the current context of writing assessment scholarship. In other words, how does the current volume contribute to ongoing conversations? Does it begin new conversations or resurrect moribund but important issues? Is there anything distinctive, disruptive or contrary to currently published work on writing assessment? I see this introduction as a way to introduce the four articles of the volume and to situate them and *JWA* as it moves to digital delivery under new editorial control. In addition to taking the field's temperature with this volume's thermometer, I will, of course, summarize the main features of the articles, their connection to each other and their possible contributions overall. I also want to consider the two reviews in the volume written by two graduate students studying writing assessment. Last, but not least, I want to introduce the new editors and the changes in format and delivery for *JWA*. I am thrilled to be passing *JWA* on to Diane Kelly-Riley and Peggy O'Neill. Their tenure as editors ensures that *JWA* will continue to publish a wide range of articles employing various methods, considering a range of important issues in various fields and professions interested in writing assessment for a multiple purposes.

Volume Four

My last sentence about the range of scholarship in writing assessment seems like a good segue into looking at the articles and reviews that comprise this volume of the journal. Three of the articles report on empirical studies, and the fourth is an essay on reliability. The reviews address two different books, one published fairly recently and one from earlier in the decade. The authors of two of the articles are affiliated with the National Writing Project (NWP), whose role in writing assessment goes back to some of the first holistic scoring sessions outside of the Educational Testing Service (ETS) or the College Entrance Examination Board (CEEB), who were instrumental in developing holistic and other means of what was then called "direct writing assessment." An evaluation of the Bay Area NWP resulted in one of the few full-length books on writing assessment during that time, *The Evaluation of Composition Instruction in Writing* (Davis, Scriven, & Thomas, 1981). The other two articles are authored by the incoming editors, giving the volume a forward-looking aspect, since Diane and Peggy will influence the shape, substance and future of *JWA* in the same ways their research has already had a shaping influence on writing assessment. I wish I could say I had planned it this way, that I had invited the incoming editors to submit work that would frame their perspectives as editors and scholars. In reality, Peggy's essay was part of a panel I heard and invited as a possible special issue of *JWA* on reliability in writing assessment. I had also heard Diane speak about her work and invited her to submit. My luck, or is it ongoing persistence in finding the most interesting work, then, becomes the journal's gain. The two reviews represent the work of a new generation of writing assessment scholars. Kristen Getchell recently defended her dissertation on students' perceptions of their writing placement and Elliot Knowles is in the process of collecting data about the role of assessment in the college writing classroom. It's instructive that both of these new generation scholars are focusing on the student, a necessary and largely unexplored territory in much writing assessment scholarship.

Arranging the order of the articles for this full volume proved to be instructive of the ways in which the articles connect to and bounce off of each other, creating a statement or at the very least a position from which to think about writing assessment. I was tempted just to put the two NWP articles together, though in which order I wasn't sure. However, if you look beyond the authorship of these two articles and beyond the fact that both used writing samples from middle school writers, the studies not only address different issues--one focuses on the influences on interrater reliability and the other on identifying the prominent features of student writing--they also stem from a different research agenda. Nancy Robb Singer and Paul LeMahieu's study of reliability answers an ongoing need for basic research into the procedures we use for assessing writing. Gere's (1980) and Faigley, Cherry Jolliffe and Skinner's (1985) contention that the need for the implementation of writing assessment practices outruns its theory and research is as true today as it was twenty-odd years ago. On the other hand, Sherry, Richard, and David report on a study in which teachers read not to assign a numerical score but to derive "numerical values from specific rhetorical features," pushing the envelope of traditional writing assessment by looking not at how to make teachers agree but rather to focus teacher/readers on the rhetorical features of student writing that at the same time have a statistical relationship with state-mandated writing tests. In a sense the two NWP studies push against each other, one providing necessary information for those who would use holistic and analytic scoring in, primarily, large-scale assessment, and the other article highlighting the textual features salient to a group of teachers. The NWP should be applauded for adopting such a wide scope of inquiry. Both of these studies contribute important, necessary, and very different insights for writing assessment.

Thinking past the NWP connection I saw no strong reason to order the two NWP pieces together and decided to begin the volume with Nancy and Paul's study, "The Effect of Scoring Order on the Independence of Holistic and Analytic Scores." This study examines four possible conditions for teachers' reading as holistic or analytic scorers, examining the common-sense idea that

holistic raters would be affected by reading analytically first. Their study addresses an important basic concern for any writing assessment program involving both analytic and holistic scoring. The difference in scores depending upon the condition not only indicates a level of statistical significance, but these differences also establish scoring order and the use of one or more scoring schemes as an important theoretical and methodological decision that can affect the raters' scores.

Nancy and Paul establish that the use of holistic and analytic scoring can be the source of a recognizable influence (or error). A systematic rather than random source of error potentially possible in the use of multiple scoring procedures should be an ongoing concern for all who use multiple scoring systems in the same assessment. In addition to contributing to best practices in writing assessment, this study also challenges a common conception of reliability in writing assessment as just an interrater reliability coefficient. How can scores from the three different conditions be used to make an argument for reliability across scoring systems when we know that the order of scoring can significantly affect interrater reliability? In other words, all scores are not equally meaningful and, as Peggy O'Neill details in her essay, "Reframing Reliability for Writing Assessment," writing assessment reliability is an integral, complex, and largely ignored and misrepresented concept. The insights generated by Nancy and Paul's work provide an interesting segue into Peggy's article.

Peggy begins her essay by noting that "Writing an essay about reliability and writing assessment presents several challenges." The initial challenge is the diversity of people who work in writing assessment. She notes that K-12 educators, college educators, educational measurement professionals, some employed by colleges and universities and some who work for various testing companies, and those who work at the state or federal level can all potentially have input into the assessment overall. This diverse input can "frame" reliability in many profound ways. My use of quotation marks around frame are meant to acknowledge Peggy's use of the concept of framing as defined by George Lakoff, in which framing is "a conceptual structure used in thinking," and where frames can be invoked through specific language use. Using the concept of framing, Peggy examines many of the major conceptions and misconceptions of reliability in writing assessment, and illustrates how specific frames limit an understanding and use of reliability to design and conduct more meaningful writing assessments. Reframing reliability as a way to ensure the dependability, consistency, accuracy, and precision of the interpretations and decisions made from evidence gathered from assessing student writing connects it to the validation arguments important to create and maintain any writing assessment. Ultimately, Peggy's essay establishes the importance of reliability for writing assessment beyond current conceptions or frames of reliability as merely a statistical artifact. These two articles cover much important ground on reliability in writing assessment. While "The Effect of Scoring Order on the Independence of Holistic and Analytic Scores" provides basic information about the use of holistic and analytic scoring in the same assessment, it also alerts us to the possibility of systematic differences in scores, opening up the possibility for understanding reliability beyond just in a statistical framework. "Reframing Reliability for College Writing Assessment" picks up the thread from the previous article, using the "frame" as a theoretical construct to explore traditional and other possible ways we can ask questions about, understand, and use reliability to generate assessments that promote teaching and learning.

The next two articles share some connection to the first two in that they address salient textual and racial issues that influence scores of student writing. Both of these studies recognize that scores given in writing assessments are a part of the fluent reading process in which experience, expectation, and a host of other factors can affect a reader's interpretation of the text, not to mention its evaluation. Like the first two, these studies provide a larger, more complete picture when understood together as a part of an overall agenda to further understand what can influence rater scores. While Nancy and Paul alert us to the possible influences in the use of holistic and analytic scoring, the two remaining pieces examine various textual and important social influences on rater judgment and student scores.

Diane Kelly-Riley's study, "Validity Inquiry of Race and Shared Evaluation Practices in a Large-Scale, University-Wide Writing Portfolio Assessment," the third piece of the volume, focuses on a specific kind of rater influence. Diane's work also illustrates the ongoing nature of validation study for any writing assessment (Moss, 1997). Just as important as this study's data design, collection, and analysis is the recollection of the "story" of the African American woman's query about possible racial bias in the university's writing assessment. Instead of explaining away this woman's concern, Diane uses it to further investigate possible systematic influences or error within a writing assessment that had already undergone extensive validation research. Diane's decision to study the impact of race not only provides us with important information and new questions, it also models the best practice of anyone who is conducting writing assessments on a regular basis.

Ultimately, "Validity Inquiry of Race and Shared Evaluation Practices in a Large-Scale, University-Wide Writing Portfolio Assessment" ends up asking more questions about the influence of race on scores in Washington State's portfolio writing assessment than it answers. While the study concludes that race is not an influence on readers, it also notes that students of color routinely perform at a statistically significant lower rate. So, while Diane is able to confirm the young African American woman's contention about race and writing assessment, the cause of African American students' lower scores is unclear. Diane's article concludes with a range of possible inquiries into these findings, including examining the support structures for student writers and the very structure of the assessment itself. While these ongoing investigations are outside the scope of the study, it is *this* study that models the recursive nature of validation research that has helped Diane and her colleagues identify new questions in their ongoing validation/reflective practice, typifying a model for the kind of validation processes necessary for anyone using writing assessment to

make important decisions about students.

Sherry Swain, Richard L. Graves, and David T. Morse's "A Prominent Feature Analysis of Seventh Grade Writing" is the final article in the volume and provides important information about textual features and writing assessment scores. Sherry, Richard, and David address the overall question of what textual features are important to teachers and what relationship these features might have to scores generated in state-mandated writing assessments. In addition, this study creates a method for exploring textual features in student writing for a variety of purposes, pulling together an overall theme of the volume that addresses questions about where writing assessment scores come from. Where Nancy and Paul look at word order, Peggy examines the role of current frames about reliability, and Diane explores the influence of race, Sherry, Richard, and David look at the texts students write and which features of these texts trigger certain kinds of evaluation and scores.

While Diederich French and Carlton (1961), Broad (2003), and others (Broad et al., 2009; see Huot, 1990 for a summary of other, early studies) have studied what characteristics of student writing teachers value, Sherry, Richard, and David connect teacher-values to student writing, unearthing which textual features positively or negatively affect students' scores on state-mandated assessments. Prominent feature analysis provides a clearer picture of what state-mandated assessments value in terms specific to the texts students write. The relationship between textual features identified by teachers and scores from state-mandated writing assessments can be established and used as a way to rhetorically and pedagogically translate the values of standardized assessment for the middle school writing classroom. Like the other three articles in the volume Sherry, Richard, and David's research provide a new method for looking at student writing and writing assessment overall. This kind of information helps teachers understand what will be valued in their students' writing on the ubiquitous state-mandated writing assessments, helping to situate classroom practices aimed at improving test scores on recognizable, valuable textual practices. In addition, this study pushes test developers to include prominent feature analysis as important evidence for any argument made about the content validity evidence of a writing assessment. It should also encourage college writing programs interested in exploring teacher values to anchor these values within the texts students write. Perhaps most importantly, this study affirms the volume's focus on student work, creating a new, student text-generated criteria for evaluating writing.

Elliot Knowles' review of *Machine Scoring of Student Writing: Truth and Consequences* provides the usual summary of articles and foci within a volume devoted to the use of computer software to evaluate student writing, pinpointing several of the reoccurring themes throughout the volume. In addition, Elliot grounds his review within technology theory that holds the construction of technology is never innocent or neutral. People create technologies out of a sense of purpose to satisfy a particular need of a specific group of people. Using this theoretical lens, Elliot focuses on how individual chapters and ultimately the entire volume examine "how such technologies are put into use." Elliot concludes his review with a section titled "Moving Forward" in which he situates automated scoring and this volume as a specific type of response to a particular technology, reminding us that the lessons learned from this volume and a study of technology overall will be useful as future technologies are integrated into the field of writing assessment.

I invited Kristen Getchell to review *The Guide to College Writing Assessment* after having a conversation with her about the book, even though I am one of its authors, because our conversation provided me with new insight into a text I had helped create. Acknowledging that "This book is unique in its design," Kristen structures her review in interesting and instructive ways. She distinguishes between the first three chapters as theoretically and historically foundational and the remaining chapters as more practical in focus, devoting an entire section to appendices. What's instructive about this approach is that ultimately Kristen identifies a holistic perspective with which to understand and evaluate the book: "I realized that the book, the appendix section included, is a strong demonstration of the intersection between theory and practice." Through this lens of "the intersection of theory and practice," the various elements Kristen identifies become part of a unified approach. In addition to Gerry McNenny's contention that *The Guide to College Writing Assessment* is light on theory and heavy on practice, Kristen provides another way to view the volume's navigation through theory and practice in which the two interact and are enacted.

Taken as a whole, the articles and reviews in this volume challenge many current notions and practices about assessment. Although the past ten years has seen the publication of nearly twenty books on writing assessment, none of these books focus on the theoretical and empirical evidence examined throughout the volume. While much of the field's scholarship examines the promises and pitfalls of technology or new teacher-led procedures for assessment like Dynamic Criteria Marking or Directed Self-Placement, this volume helps provide a sort of infrastructure that asks and answers basic questions relevant to popularly used assessment procedures. For example, no matter what assessment procedures we might favor, validation requires that we make an argument for the dependability of the information being used to make educational decisions. In addition, teacher-led assessments still require ongoing examination and validation that answer stakeholder questions about possible influences on the assessment itself. This volume, then, can be seen as a valuable launching point for a host of new assessments available now and in the future scholarship of writing assessment.

In Conclusion

Beginning with the fifth volume, Peggy O'Neill of Loyola University of Maryland and Diane Kelly-Riley of Washington State University become *JWA* editors and continue their first issue online and free in January 2012. I am pleased to be handing over *JWA* to

assessment scholars who have not only made important contributions to the field but who continue to think as inclusively and expansively as possible about crucial, rudimentary principles like reliability (see O'Neill this volume) and about important unintended writing assessment consequences for specific, stigmatized minority groups like African Americans (see Kelly-Riley this volume).

JWA was founded after El Sevier, who came to own *Assessing Writing* (see the introduction of the first volume for the rest of the story), dismissed Kathi Yancey and me as editors. Barbara Bernstein, who owns and manages Hampton Press, came to our assistance and agreed to publish the journal. After the first volume or so, I became the sole editor, and over a now eight-year period, *JWA* published four volumes, including this, my last and the journal's first online volume. Clearly, I have failed as a production manager and editor, since a journal needs to appear in a regular, timely manner. While I note my failures as an editor, I must also recognize the successes and contributions of the authors of the scholarship that has appeared in *JWA*. The twenty or so articles that comprise the first four volumes have featured some of the most prominent scholars in writing assessment as well as some of the field's most promising young voices. In addition, *JWA* has published the work of teachers who were dissatisfied about the assessment procedures used at their institutions and did something about it. I have no doubt that the first four volumes of *JWA* will continue to make a sound contribution to many important current and future conversations about writing assessment. I am pleased to acknowledge that the first three volumes will join the fourth, online and free.

I have to admit that as I get the last volume ready to leave my hard drive, I feel huge relief. I am relieved that for the first time in almost two decades I will no longer be responsible for producing a journal on a regular basis. I am also relieved that *JWA* will survive my own incompetence. I will miss working with authors and getting to read the hundreds of manuscripts that have come my way over the years. I will miss the unique perspective on a growing field that being an editor for two writing assessment journals has given me. I look forward to being a member of *JWA*'s Editorial Board and a part of its new online presence under Diane and Peggy's direction.

References

- Broad, B. (2003). *What We Really Value: Beyond Rubrics in Teaching and Assessing Writing*. Logan, UT: Utah State University Press.
- Broad, B., Adler-Kassner, L., Alford, B., Detweiler, J., Estrem, H., Harrington, S., McBride, M., Stalions, E., & Weede, S. (2009). *Organic Writing Assessment: Dynamic Criteria Mapping in Action*. Logan, UT: Utah State University Press.
- Condon, W. (2011). Reinventing writing assessment: How the conversation is shifting. *WPA: Writing Program Administration*, 34(2).
- Davis, B. G., Scriven, M., & Thomas, S. (1981). *The Evaluation of Composition Instruction*. Inverness, CA: Edgepress.
- Diederich, P., French, J., & Carlton, S. (1961). *Factors in judgment of writing quality* (Report No. 61-15). Princeton, NJ: Educational Testing Service.
- Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60, 237-263.
- McNenny, G. (2011). In P. O'Neill, C. Moore, and B. Huot (Eds.). *A Guide to College Writing Assessment*. Logan, UT: Utah State University Press. Retrieved from

Copyright © 2021 - *The Journal of Writing Assessment* - All Rights Reserved.