

INTRODUCTION TO VOLUME 1, TECHNICAL AND POLITICAL CONTEXTS

Diane Kelly-Riley

University of Idaho

Ti Macklin

Boise State University

Carl Whithaus

University of California, Davis

Considering Students, Teachers, and Writing Assessment, Volumes 1 and 2 focus on the increasing importance of students' and teachers' lived experiences within the development and use of writing assessments. Together, the pieces in these volumes reflect upon how writing assessment research has contributed to five major themes: (1) technical psychometric issues, particularly reliability and validity; (2) politics and public policies around large scale writing assessments; (3) the evolution of—and debates around—automated scoring of writing; (4) the major theoretical changes elevating fairness within educational measurement and writing assessment; and (5) the importance of considering the lived experiences of the humans involved in the assessment ecology.

The Journal of Writing Assessment (JWA) has been a primary scholarly forum that has chronicled this evolution. These two volumes examine key themes from scholarship published in *JWA* in the past twenty years. Each section is introduced by current scholars in writing assessment who provide a retrospective for the issues of the past and these authors comment on the ways in which these issues continue to unfold. As such, they also represent generations of scholars in conversation with each other providing a model necessary as we continue to navigate the unfolding complexities of writing assessment situated in society. That is this field, in particular, benefits from revisiting issues and controversies of the past to see how our responses informed the practices of the present.

Volume 1 explores the dynamic issues connected to reliability and validity and how writing assessment contributed to the evolution of these concepts, the

shifting political context of writing assessment and the rise of automated scoring of writing. Volume 2 explores the evolutions in theory and practice related to fairness and writing assessment and then the ways in which the people who teach and learn in these spaces shape writing assessment practice.

TECHNICAL EVOLUTIONS

The first volume focuses on technical and political issues. The rise of local considerations in writing assessment emerges most fully in the articles published in *JWA* during the first two decades of the twenty-first century. Rather than excluding the lived experiences of students and teachers, *JWA* has taken the lead in documenting how contextual, situated, and localized forms of writing assessment may provide fuller—more valid, reliable, and fairer—pictures of students' writing. The history of this move valuing localized forms of writing assessment has not been fully told. This movement reaches back to Edward White's (1978) early advocacy for direct assessment of students' writing rather than a reliance on indirect forms of writing assessment. It also echoes—perhaps even amplifies—Kathleen Yancey's (1999) and others' work (Calfee & Perfumo, 1996; Elbow & Belanoff, 1997; Hawisher & Selfe, 1997; Herman et al., 1993; Herter, 1991) on writing portfolios in the 1990s reflecting on students' emerging knowledge about writing, their writing processes, and their development as writers. Since its inception, *JWA* has published scholarship from the unique angle of how local contexts inform writing assessments.

Revisions to the *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2014) shifted discussions around the core educational measurement constructs of validity and reliability and drove changes in these scholarly areas. The *Standards* govern much of the thinking about standardized assessments, particularly true within the psychometric and educational measurement sides of writing assessment. The *Standards* is a living document open to revision, and the changes between the 4th and 5th editions in 1999 shifted discussions within writing assessment away from a singular focus on the importance of reliability to an understanding that validity is the most important consideration in writing assessment systems and is situated in particular contexts. The published discussions between Richard Haswell (1998) and Pamela Moss (1998) foreground how debates around the concept of validity assumed an increasingly important role in writing assessment. Once the focus of validity changed, teachers had a clearer role in determining and contributing to meaningful assessment. The increased emphasis on validity enabled teachers to push back against the limitations of standardized tests, opening up a new area of research that involved local contexts and faculty expertise. *JWA*'s establishment in 2003 provided a

venue for writing teachers and educational researchers to explore the implications of considering local contexts on writing assessments.

PROGRAMMATIC IMPLICATIONS

A frequently told origin story of writing assessment in North American postsecondary education points toward 1874 and the addition of an extemporaneous writing sample in the Harvard entrance examination. Norbert Elliot details (2005) how the Harvard exam was used to place students into its curricula. More than half of the students required remedial coursework and additional support setting up the tension between assessment and instruction. Elizabeth A. Wright, Suzanne Bordenlon, and S. Michael Halloran (2020), however, offer a corrective to this historicizing of writing assessment. In “‘Available Means’ of Rhetorical Instruction,” they take up Royster and Kirsch’s call to explore “the lessons taught to those students unable to attend those schools for elite white men” (p. 245). Wright, Bordenlon, and Halloran point out how late 19th-century rhetorical education and writing instruction took place in a wide variety of secondary and postsecondary educational contexts including Catholic institutions, women’s colleges, historically Black universities and colleges, as well as within the often repressive contexts of boarding schools for indigenous children (pp. 254-257). Thus, there is a broader history of the structures and lasting impacts of writing assessment yet to be explored.

Across all of these contexts, writing placement mechanisms grew more profoundly as standardized tests became more widely available. Many of these placement exams attempted to capture students’ readiness to enter postsecondary study, but the means of the exams often did not correspond to the curricular realities in the classrooms. Haswell (2004) notes that the 1900’s “saw testing firms grow ever more influential and departments of English grow ever more divided between using ready made goods, running their own placement examinations, or foregoing placement altogether.” (para. 3) Faculty in English departments devised their own assessment systems. The English Equivalency Exam (EEE) was used by the California State University and Colleges between 1973 and 1981; later, it was replaced by the English Placement Test in 1977 developed by Edward White and his colleagues. Haswell and Elliot (2017) observe “the few scholars and test administrators who were using holistic scoring were using all their energies to confront the problems of cost and scoring reliability, as practical aspects of the large testing programs they were supervising.” (White, 1993, p. 82) The EEE went beyond that, as White (1984) himself declared in his essay, “Holisticism.” The method of holistic scoring may have achieved some pragmatic ends making “the direct testing of writing practical and relatively

reliable” (White, 1984, p. 408), and it may have achieved some indirect social ends, bringing “together English teachers to talk about the goals of writing instruction” (p. 408), but beyond that “it embodies a concept of writing that is responsible in the widest sense . . .” (p. 408). It was responsible for its product, which was responsible for its advertised use.

This move toward localization continued in the late 1980s when an area of research emerged from the lived experiences of teachers and students in composition courses in response to accreditation and accountability mandates. Moore, O’Neill, and Crow (2016) detail this extensive history of compositionists “using assessment to improve student learning before it was emphasized so much by accreditors . . . [because they] understood the link between learning assessment and teaching improvement before accreditors made the connection explicit” (p. 20). Many of these teacher-researchers struggled with the day-to-day implications of the theoretical constructs of validity and reliability. As they grappled with these constructs in their contexts, new practices and research paths emerged. Early examples are detailed by Moore et al. (2016) demonstrating the field of composition’s historical response to external assessment mandates. The first was Elbow and Belanoff’s (1997) portfolio system which replaced a mandated university proficiency exam. Another system, developed and implemented at Washington State University, included an entry-level Writing Placement Exam and junior Writing Portfolio developed by Richard Haswell and his colleagues (2001). This program entwined formative writing assessment with disciplinarily situated writing instruction across the entire undergraduate curriculum. At all levels, writing teachers were involved in the assessments, and a comprehensive writing center provided support for students, including required small group sessions concurrently supporting students in upper-division disciplinary writing courses for those who did not pass the mid-career assessment (Haswell, 2001).

The core educational measurement constructs of validity and reliability continued to undergo major reconceptualization. In 1999, major revisions to the *Standards for Educational and Psychological Testing* were jointly authored by the American Psychological Association (APA), American Education Research Association (AERA), and the National Council of Measurement in Education (NCME)—professional organizations which guide and govern best practices in assessment and measurement. In this revision, validity was cast as the most important consideration above all. Now, tests or assessments were no longer considered stand-alone entities that needed to adhere to standards of technical qualities of reliability or validity. Instead, a major philosophical understanding of assessment shifted to see these measurements in social contexts in which the uses and interpretations of scores must be considered in each and every setting. This was a revolutionary shift.

The late 1990's also saw significant educational reform in the US with assessment playing a key role in these public and political arenas. During this time, writing studies teachers pushed back against the standardized test movement which attempted to represent and measure writing ability through knowledge of grammar and other writing rules (Bloom et al., 1996). In standardized testing, multiple choice test items were used as a way to measure the quality of students' writing. Writing teachers and researchers resisted these indirect, decontextualized forms of evaluating students' writing abilities. From their positions in the classroom, compositionists knew this evaluation did not serve the instructional needs of either students or faculty, and they advocated for locally-developed assessment measures attentive to classroom contexts and actual student learning outcomes. Thus, portfolio assessment developed out of the work of postsecondary writing teachers. This process is described in White et al.'s 1996 collection, *Assessment of Writing: Politics, Policies, Practices*. By the mid 90s, ways of measuring the construct of writing became more nuanced. Compositionists realized that writing is socially situated and began to publish research findings supporting this position. Understanding the people who designed and participated in the assessments and the multiple ways in which they were enacted across different institutional sites became a key component of writing assessment.

The increased accountability context within educational settings in North America resulted in innovative programmatic responses. The Council of Writing Program Administrators (CWPA) started to discuss and collaborate on whether "a pithy and effective list of objectives for writing [and] programs existed" (as cited in Harrington et al., 2003, p. xv). These conversations among members at all levels of expertise were enabled by many compositionists joining the then newly created WPA-L email listserv in the late 1990s. The members of this group recognized the multiple stakeholders who were invested in the outcomes of first-year composition. This exigence resulted in the development of the *WPA Outcomes Statement for First-Year Composition (WPA OS)* "a statement . . . plain enough to speak to those outside the discipline, yet rooted in disciplinary language enough to have status in the field" (Harrington et al., 2003, p. xvi). The WPA OS is a consensus document detailing the expectations for first-year writing common to most postsecondary institutions in North America (Harrington et al., 2001). Kathleen Yancey (2003) says that the *WPA OS* was intentionally written as outcomes and not standards for performance that needed to be achieved.

By framing and modeling curricular and assessment work as driven by faculty and local contexts, the collaborators of the *WPA OS* also began to formalize a new area of research. This new area of local programmatic response to assessment had several offshoots as contextually situated responses to assessment

and accountability mandates. In *Reclaiming Assessment: A Better Alternative to the Accountability Agenda*, Chris Gallagher (2007) describes his locally focused efforts with colleagues across K-12 programs in Nebraska. Gallagher and his collaborators argue that accreditation programs developed by teachers with students and learning in mind result in the best programs. Others like Christine Farris (2014) from Indiana University led the Writing and Reading Alignment Project which intended “to help teachers examine their current instructional practices and goals for student learning and develop new strategies to promote skills in critical reading, evidence-based writing and discussion as expected in college-level coursework.” (Indiana University, 2014)

Wendy Sharer and her colleagues (2016) describe their efforts at Eastern Carolina University to reclaim accountability and assessment for postsecondary settings. In their edited collection, *Reclaiming Accountability: Improving Writing Programs through Accreditation and Large-Scale Assessments*, they provide models responding to the call of a 2007 WPA Executive Board letter that proclaimed “those who teach writing and those who administer writing programs need to be involved in defining the terms and setting the parameters of large-scale writing assessment so that any changes implemented in response to assessment are in keeping with what research and practice have demonstrated to be truly effective in helping student writers” (p. 3). In Behm’s edited collection (2013), *The WPA Outcomes Statement: A Decade Later*, the effect of this situated research agenda is apparent. Topics in the book cover personal identity, its application to writing across the curriculum and disciplines, extensions into global settings, use with second language approaches, and impacts on technology. Much of the research on writing assessment explored its connections to instruction.

THE PROMISE AND PERILS OF TECHNOLOGY

Beginning with No Child Left Behind (NCLB) in the early 2000s, large scale assessment and accountability efforts moved testing and accountability to the center of educational policy and practice. As a result, the challenge of testing hundreds and thousands of students across institutions became a reality. Initially, technology was seen as the remedy to manage such a large-scale endeavor. But this effort was hampered by limitations within the technologies to reliably and validly evaluate a significant amount of writing as a socially situated, complex construct. Later these educational reform efforts morphed into the Common Core State Standards Initiative (CCSSI) further elevating the role for writing throughout the K-12 curriculum.

CCSSI marked a new period in the assessment landscape in which educational reformers, largely nonprofit and philanthropic organizations like the

Gates and Lumina foundations, had strong legislative support to reshape American education into one focused on the preparation of workers to advance the American economy. To measure their progress, these efforts partnered with testing companies like Pearson and ETS. With complex assessments being implemented on such a large scale, the possibilities of machine or computer scoring of writing ascended to the forefront. The challenges for assessing learning through writing remained and were amplified in these large-scale assessments.

These emerging curricular efforts recognized the importance of teaching writing as situated within disciplinary genres from the beginning of school. As accountability efforts moved from NCLB to the Smarter Balanced Assessment Consortium (SBAC) and Partnership for Assessment of Readiness for College and Careers (PARCC) national assessments, the initial response was to use technology, particularly the potentials for automated essay scoring, to support the integration of writing across the K-12 curriculum. These large-scale assessments have meant that writing assessment has taken a much more central role in accountability efforts. The challenge remains to develop computer-based scoring that represents the complexity of writing taught and assessed in the classroom.

PART ONE. TECHNICAL ISSUES IN THE ASSESSMENT OF WRITING: RELIABILITY AND VALIDITY

Part One of this collection focuses on technical issues in the assessment of writing, particularly reliability and validity as published in the *Journal of Writing Assessment*. Raters' approaches to texts are one of the most vexing issues in writing assessment. Controlling for individual raters' idiosyncrasies is one of the longest running issues in writing assessment reaching back to Paul Diederich's work at ETS in the 1960s. Writing assessment researchers' work on reliability and validity has taken many forms. In his piece, David H. Slomp, editor of *Assessing Writing*, contextualizes and responds to these changes in the technical constructs in writing assessment. Slomp's response focuses on the ways in which *JWA*'s legacy bridges the gap between educational measurement and writing studies. He explores the implications for research and practice that emerge from dialogues between these two fields. Slomp frames and responds to the following key articles from *JWA*.

Peggy O'Neill's "Reframing Reliability for Writing Assessment" (2011) shifts away from traditional discussions about inter-rater reliability as the ultimate goal—the single most important form of reliability within writing assessment at the time. She argues that both writing studies and psychometrics offer multiple forms of reliability that need to be attended to in the building of writing assessment systems and in research about writing assessment. Drawing on Lakoff's

(2002, 2004, 2006) work, O’Neill suggests that by moving discussions of reliability in writing assessment beyond inter-rater reliability, more nuanced, and more accurate, forms of writing assessment can be developed. These emerging forms of writing assessment might not only acknowledge but also account for—in a psychometrically rigorous way—variations across readers and variations across tests in the ways that Pamela Moss (1998) and Richard Haswell (1998) recognize as hermeneutic or rhetorical practices.

Diane Kelly-Riley’s “Validity Inquiry of Race and Shared Evaluation Practices in a Large-Scale, University-wide Writing Portfolio Assessment” (2011) advances the field’s understanding not only of the balance between reliability and validity but also brings into the conversation vital contextual elements involving race and racism. Her article takes on the question of race—and in more subtle ways racism—by looking at the implementation of a locally-developed, context-rich writing portfolio assessment system. Kelly-Riley’s article is a precursor to the consideration of fairness and antiracist practices in writing assessment by providing an empirical study that looked at how raters understand and apply race in an assessment context. Writing assessment has struggled to develop an operational definition of race. Race is often defined by government agencies that collect data on race, but the experience in the writing classroom calls for more nuanced representations of race.

O’Neill and Kelly-Riley’s work lead toward approaches outlined in Elliot et al.’s “Three Interpretative Frameworks: Assessment of English Language Arts-Writing in the Common Core State Standards Initiative” (2015). Elliot, Rupp, and Williamson examine how standards-based definitions of validity, reliability/precision, and fairness were integrated into the Smarter Balanced Assessment Consortium (SBAC) and Partnership for Assessment of Readiness for College and Career (PARCC) English Language Arts – writing assessments. They encourage stakeholders to be informed consumers when interpreting and using SBAC or PARCC scores about students’ writing. Their work foreshadows a move within writing assessment research and practice encouraging stakeholders (WPAs, students, teachers, and parents) to not just accept the scores from large-scale state or national-level writing assessments at face value but to integrate how they will be used, to examine their meaning and their use value.

PART TWO. POLITICS AND PUBLIC POLICY OF LARGE-SCALE WRITING ASSESSMENT

Part Two explores the political dimensions of writing assessment. In her contextualization of this section, Carolyn Calhoon-Dillahunt, Yakima Valley College, Past Chair of the Conference on College Composition and Communication

and Past President of the Two-Year College Association, synthesizes these major educational reform movements and how they impact writing assessment scholarship and practices. This section highlights work by Edward M. White, Arthur N. Applebee, Hammond and Garcia, and Toth et al. All of these authors anticipate and wrestle with large scale writing assessment in terms of political and policy issues. Political changes across educational reform movements have both shaped and responded to assessment issues. The critiques of both placement in two-year colleges and of AES have centered around how students' writing must be considered and evaluated as contextual, rather than stripped of context for a placement decision afforded by the cost savings of having software evaluate a piece of writing.

Edward M. White's "The Misuse of Writing Assessment for Political Purposes" (2005) and Arthur N. Applebee's "Issues in Large-Scale Writing Assessment: Perspectives from the National Assessment of Educational Progress" (2007) set the stage for early political discussions around writing assessment. White argues that many large-scale writing assessments are motivated "by political rather than educational, administrative, [or] professional concerns." For White, *No Child Left Behind* and its reliance on testing "without the resources and leadership for students to achieve the skills they will be tested on" is a crucially flawed educational policy and a misuse of writing assessments based on politicians' misunderstanding of what educational testing can tell us. He considers a wide range of mandated, large-scale writing assessments ranging from required state-level testing of secondary students through placement exams for incoming college students to graduation requirements for college students. He suggests that the misuses of writing assessments "[are derived] from an exaggerated, even a credulous misunderstanding, of what particular kinds of assessments can accomplish." Such observations continue to underscore the misuse of assessments in educational settings.

In contrast to White's critique of assessment as gatekeeping, Applebee's "Issues in Large-Scale Writing Assessment: Perspectives from the National Assessment of Educational Progress" focuses on the contributions of the large-scale, national-level programmatic assessment conducted through the National Assessment of Educational Progress (NAEP). Applebee suggests the NAEP writing assessment is valuable because it is not tied to the assessment of individual students, but rather a way of looking at how students' writing is developing and comparing achievements in writing across states. White's and Applebee's works are both polemic, but research like J. W. Hammond and Merideth Garcia's show the legacy of informed and principled approaches documenting the effects of large-scale assessment on teachers.

Writing assessment, politics, and public policies in the first two decades of the twenty-first century requires that we address the effects of *No Child Left*

Behind (NCLB) and the Common Core State Standards (CCSS). Hammond and Garcia's "The Micropolitics of Pathways: Teacher Education, Writing Assessment, and the Common Core" (2017) and Toth et al.'s "Writing Assessment, Placement, and the Two-Year College" (2019) describe the impacts of these initiatives on teachers and students. Hammond and Garcia take the Common Core State Standards (CCSS) as their point of departure. They examine how teacher education programs frame the CCSS for their teachers-in-training. Their works suggest postsecondary faculty, teachers, and teachers-in-training "micropolitically interpret" the Common Core. In fact, Hammond and Garcia suggest that writing teachers and writing teachers-in-training foreground their own local writing assessments since teachers seem most focused on curriculum and instruction issues and secondarily on CCSS and pathway-related reforms to education. One of the key findings from Hammond and Garcia's work is the value of adopting a micropolitical perspective when considering writing curricula, instruction, and assessment.

The emphasis on learning pathways was championed in the educational reforms promoted through CCSS. As such, pathway-based reforms had a dramatic effect on community colleges, Toth et al.'s introduction to the *Journal of Writing Assessment's* Special Issue on Placement and Two-year Colleges takes up the overlapping issues of educational reform and how writing assessments have been used in placement decisions. In their ambitious and wide-ranging article, Toth, Nastal, Hassel, and Giordano review the history of two-year colleges within American higher education. They attend to the ways in which this history and pathway-based educational reform movement intersects with models for assessing and placing students within ESL, basic writing, or first-year composition courses. They then extend their discussion by turning to questions around the validity and the uses for writing assessment and placement systems. Toth et al.'s attention to sociocultural factors highlights the ways in which questions of writing assessment are being looked at at a systems level rather than only at the level of individual students.

PART THREE. IMPLICATIONS OF AUTOMATED SCORING OF WRITING

In Part Three, key pieces published in *JWA* explore possibilities and pitfalls with technology and writing assessment. Large-scale assessments became more commonplace as the accountability movement gained traction in public educational settings. During the late 1990s, No Child Left Behind was implemented across K-12 public school systems and the challenge of assessing each and every student became a reality. As accountability systems evolved, partnerships between

testing companies, educational reform nonprofit organizations with strong legislative support, and textbook publishing companies evolved into new initiatives connected to career and college readiness and capitalized upon the economic investment in public education. A focus emerged on secondary and post-secondary education to prepare students in economic terms. This resulted in more complex curricula—such as the Common Core State Standards Initiative—which emphasized students’ readiness for workplace or college challenges. The backbone of this curriculum was writing—where it became embedded across multiple disciplines across grade levels. Writing was a primary means to demonstrate and assess student proficiency across disciplinary areas. To meet the challenge of assessing student performance across the country, test developers and researchers turned to automated scoring of student writing. Assessing the construct of writing when it is socially situated presents new challenges difficult for technology to address alone. This more robust, and socially situated construct of writing better represents what occurs in classroom settings, but it also requires the development of writing assessment systems that connect human readers and writing technologies.

In her introduction to Part Three, Laura Aull, Associate Professor and Writing Program Director at the University of Michigan, responds to the major developments in Automated Essay Scoring (AES) and contextualizes the issues in relation to key publications from *JWA* from the past twenty years. She discusses the major possibilities and limitations of these technologies, particularly as they relate to the ongoing implications of socially situating writing assessment. Her response highlights the intersections of Artificial Intelligence and AES in education and measurement, learning analytics, and user-centered design. Aull considers the impacts of these emerging writing assessment technologies on educational equity. As this book was going to press, ChatGPT had recently emerged and reignited the importance of this scholarship to the writing assessment community.

In “Validity of Automated Scoring: Prologue for a Continuing Discussion of Machine Scoring Student Writing”, Michael Williamson (2003) lays out the tension between the field of writing studies and educational measurement as automated scoring of writing took hold. Williamson encourages writing studies scholars and practitioners to learn the language of educational measurement in order to weigh in on these evolving conversations which would inevitably bend toward a socially situated context because of the consideration of the use of test results and their impacts on test takers. Williamson notes that automated scoring of writing would likely become prevalent given the millions of pieces of writing that required evaluation in the large-scale assessment systems.

Next, Les Perelman’s critique of AES software and its uses in assessing writing challenge the widespread adoption of this technology. His work points to

the important distinction that writing is a complex, socially situated activity and the necessary reductions that must occur to the construct of writing when it is assessed by computer software and algorithms. In “Critique of Mark D. Shermis & Ben Hamner, ‘Contrasting State-of-the-Art Automated Scoring of Essays: Analysis,’” Perelman (2011) notes that Shermis and Hamner reported high reliability between human and machine readings of student work, but Perelman argues that the samples are very short and were written in response to literary analysis or reading comprehension passages. Such writing samples are hardly representative of complex writing situated in context, genre, and social circumstance.

Finally, Jordan Canzonetta and Vani Kannan explore these writing assessment technologies in “Globalizing Plagiarism & Writing Assessment: A Case Study of Turnitin” (2016). This piece highlights the additional uses of automated essay scoring and its integration within learning management systems for plagiarism detection or online writing support. In order to be reliable, automated essay reading and scoring systems must operate with narrowly defined constructs of writing. Canzonetta and Kannan explore the implications of importing the US construct of writing into other cultures. In their view, the global reach of Turnitin privileges western academic writing and stigmatizes nonwestern writing. As a result, the plagiarism software reinforces western values about authorship not necessarily representative in other places. As Artificial Intelligence and systems like ChatGPT become more prevalent, it’s important for writing assessment researchers and scholars to document and understand the possibilities and limitations within them.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Behm, N. N., Glau, G. R., Holdstein, D. H., Roen, D., & White, E. M. (Eds.). (2013). *The WPA Outcomes Statement: A decade later*. Parlor Press.
- Calfee, R. C., & Perfumo, P. (Eds.). (1997). *Writing portfolios in the classroom: Policy and practice, promise and peril*. Routledge.
- Elbow, P. & Belanoff, P. (1997). Reflections on an explosion: Portfolios in the ‘90s and beyond. In K. B. Yancey & I. Weiser (Eds.), *Situating portfolios: Four perspectives* (pp. 21-33). Utah State University Press.
- Elliot, N. (2005). *On a Scale: A social history of writing assessment in America*. Peter Lang.
- Gallagher, C. (2007). *Reclaiming assessment: A better alternative to the accountability agenda*. Heinemann.

- Harrington, S., Rhodes, K., Fischer, R. O., & Malenczyk, R. (2003). Introduction: Celebrating and complicating the Outcomes Statement. In S. Harrington, K. Rhodes, R. O. Fischer, & R. Malenczyk (Eds.), *The outcomes book: Debate and consensus after the WPA Outcomes Statement* (pp. xv-xix). Utah State University Press. <https://doi.org/10.2307/j.ctt46nwgw>
- Harrington, S., Malenczyk, R., Peckham, I., Rhodes, K., & Yancey, K. B. (2001). WPA outcomes statement for first-year composition. *College English*, 63(3), 321-325.
- Haswell, R. H. (Ed.). (2001). *Beyond outcomes: Assessment and instruction within a university writing program*. Ablex.
- Haswell, R. H. (1998). Multiple inquiry in the validation of writing tests. *Assessing Writing*, 5(1), 89-110.
- Haswell, R. H. (2004). Post-secondary entry writing placement: A brief synopsis, CompPile.org. <https://wac.colostate.edu/docs/compfile/pd/writingplacementresearch.htm>
- Haswell, R. H. & Elliot, N. (2017). Innovation and the California State University and Colleges English Equivalency Examination, 1973-1981: An Organizational Perspective, *Journal of Writing Assessment*, 10(1), <https://escholarship.org/uc/item/7rt5v9p2>
- Hawisher, G. E., & Selfe, C. L. (1997). Wedding the technologies of writing portfolios and computers. In K. B. Yancey & I. Weiser (Eds.), *Situating portfolios: Four perspectives* (pp. 305-321). Utah State University Press.
- Herman, J. L., Gearhart, M., & Baker, E. L. (1993). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment*, 1(3), 201-224.
- Herter, R. J. (1991). Writing portfolios: Alternatives to testing. *English Journal*, 80(1), 90.
- Indiana University. (2014). IU to collaborate with high school teachers on Writing and Reading Alignment Project. <https://education.indiana.edu/news/2014/2014-06-11-01.html>
- Lakoff, G. (2002). *Moral politics: How liberals and conservatives think* (2nd ed.). University of Chicago Press.
- Lakoff, G. (2004). *Don't think of an elephant! Know your values and frame the debate*. Chelsea Publishing.
- Lakoff, G. (2006). Simple framing. *Rockridge Institute*. http://www.rockridgeinstitute.org/projects/strategic/simple_framing
- Moore, C., O'Neill, P., & Crow, A. (2016). Assessing for learning in an age of comparability, In W. Sharer, T. A. Morse, M. F. Eble, & W. B. Banks (Eds.), *Reclaiming accountability: Improving writing programs through accreditation and large-scale assessments* (pp. 17-35). Utah State University Press.
- Moss, P. (1998). Testing the test of the test: A response to "Multiple Inquiry in the Validation of Writing Tests" [by Richard H. Haswell]. *Assessing Writing* 5(1), 111-122.
- Sharer, W., Morse, T. A., Eble, M. F., & Banks, W. P. (2016). Introduction: Accreditation and assessment as opportunity. In W. Sharer, T. A., Morse, M. F. Eble, & W. B. Banks (Eds.), *Reclaiming accountability: Improving writing programs through accreditation and large-scale assessments* (pp. 3-13). Utah State University Press.

- White, E. M., Lutz, W. D., & Kamusikiri, S. (1996). *Assessment of writing: Politics, policies, practices*. Modern Language Association.
- White, E. M. (1978). Mass testing of individual writing: The California model. *Journal of Basic Writing*, 1(4), 18-38. <https://doi.org/10.37514/JBW-J.1978.1.4.03>
- Wright, E. A., Bordelon, S., & Halloran, M. (2020). "Available means of rhetorical instruction": "Broadening perspectives" on rhetorical education prior to 1900. In J. J. Murphy & C. Thaiss, (Eds.), *A short history of writing instruction: From ancient Greece to the modern United States* (pp. 244-271). Routledge.
- Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication*, 50(3), 483-503.
- Yancey, K. B. (2003). Standards, outcomes, and all that jazz. In S. Harrington, K. Rhodes, R. O. Fischer, & R. Malenczyk (Eds.), *Debate and consensus after the WPA Outcomes Statement* (pp. 18-23). Utah State University Press.