

RETROSPECTIVE.

TOWARD FAIRNESS IN WRITING ASSESSMENT

Diane Kelly-Riley

University of Idaho

Ti Macklin

Boise State University

Carl Whithaus

University of California, Davis

SHIFTING AWAY FROM RELIABILITY AND VALIDITY TOWARDS FAIRNESS AND EQUITY

Looking back at the development of writing assessment over the last twenty years, we see the field increasingly attending to the importance of students' and teachers' lived experiences. The development of writing assessment instruments is still a prominent focus of the field, but the objective of accurately measuring the "true score" for a student's overall, generalized writing ability that would hold across contexts has diminished. It has been replaced by questions about contexts and the nuances around writers' backgrounds and the writing tasks they are being asked to engage in. Fairness has become a vital third consideration on par with validity and reliability. In fact, if we trace a forty-year historical arc from 1960 through 2000, writing assessments moved from indirect writing assessments to direct, timed writing assessments to portfolio-based writing assessments. Beyond 2000, they have continued to evolve. The interest in writing contexts that the use of portfolios promoted within the field has led researchers to ask more and more pointed questions about how situational elements may be included rather than excluded in writing assessment activities. This move has been expressed powerfully in Asao Inoue's (2015) emphasis on the importance of considering the entire ecology around a writing assessment and in Anne Ruggles Gere et al.'s emphasis on the importance of "communal justicing" (2021, p. 384). Within the pages of the *Journal of Writing Assessment*, we have seen these moves toward developing and studying situated forms of writing assessment such as directed self-placement or labor-based, contract grading. Overall,

the field has shifted away from focusing on methods, such as inter-rater reliability and construct validity in large-scale writing assessments, and embraced questions about learning differences, working to create more just educational systems, and mitigating the impacts that present obstacles to equity, such as racism, ableism, and poverty. As a result, the role of fairness has increased and has become a major consideration informing the field's work.

Many of the writing assessment studies that have looked closely at the lived experiences of students and teachers have examined effects of racism. These studies have helped open the door to later work that includes a wider and more complex representation of fairness. Mya Poe and John Aloysius Cogan, Jr.'s "Civil Rights and Writing Assessment" as well as Wood's "Engaging in Resistant Genres as Antiracist Teacher Response" have been vital articles in the field's critique of racist assessment practices and the development of antiracist methods of writing assessment. Their work brought fairness into mainstream conversations about how writing program directors at community colleges, state colleges, and research universities should develop assessment practices to create the conditions for more equitable educational outcomes. At the same time that Poe, Cogan, and Wood have pushed forward the conversation about combating systemic racism within writing assessment systems, Leslie Henson and Katie Hern's "Let Them In: Increasing Access, Completion, and Equity in English Placement Policies at a Two-Year College in California" has utilized a disparate impact analysis to document how refinements to writing placement systems can be a powerful lever for reducing racial and ethnic gaps in terms of course completion outcomes. Their work takes a serious look at writing within the community college context and does so in a way that emphasizes how writing assessment may be reformed to increase equitable outcomes. Considerations of fairness also need to include the institutional context in which students and teachers work.

These local considerations should also include conversations about individuals' learning needs. In "Neurodivergence and Intersectionality in Labor-Based Grading Contracts," Kathleen Kryger and Griffin X. Zimmerman zero in on these questions around accessibility. They challenge racist and classist linguistic ideologies and ask how labor-based grading contracts may be used to honor neurodivergence and intersectional student identities. Their work digs into how student experiences and identities cannot be separated from a writing assessment and the way an assessment constructs and defines value (i.e., what is good writing). Kryger and Zimmerman's article embraces the possibilities for situated writing assessments, particularly labor-based grading contracts, to enhance fairness and make room for more nuanced readings and valuing of student writing. Shane Wood's "Engaging in Resistant Genres as Antiracist Teacher Response" provides a unique teacher perspective by focusing on the genre of

teacher response to students' writing within a contract grading assessment ecology, something not frequently discussed in scholarship on the increasingly popular practice of contract grading. Wood challenges teachers to carefully consider how their response practices can—and do—reinforce White language supremacy, despite their best intentions, thus causing harm to students. By challenging scholars and practitioners to reconsider one of the most important and frequent sites of student-teacher interaction, Wood reframes and reconceptualizes the practice of teacher response to student writing. Like Kryger and Zimmerman's work, Wood's essay considers the intersections of antiracist praxis and teachers' assessment of student writing. These two works underscore the increased importance of fairness in relationship to writing assessment; they also ground writing assessment practices in students' and teachers' lived experiences rather than privileging the contexts of large-scale writing assessments.

Taken together Poe and Cogan's, Henson and Hern's, Kryger and Zimmerman's, and Wood's essays embody the field's shift away from studies that privilege reliability and validity without addressing questions of fairness and equity. Students' and teachers' contexts matter for these researchers as they embrace questions about learning differences, develop techniques for fairer writing assessment, and work to create more equitable educational outcomes for diverse student populations. These questions around fairness and equity are leading into more detailed discussions about how contract grading functions. For instance, Ellen Carillo's *The Hidden Inequities in Labor-Based Contract Grading* (2021) has taken up questions around labor-based grading contracts and how they make assumptions about normative achievements being tied to time spent working on a task. Her development of engagement-based grading contracts suggests ways in which situated assessment practices are being challenged and refined. That is, the development of questions about fairness and equity in writing assessment has not achieved a determined final form (i.e., the best practices are writing portfolios, or the best practices are labor-based grading contracts, or the best practices are engagement-based grading contracts). Rather the turn in writing assessment work to questions about fairness and equity is just beginning. Exploring the debates that run through Poe and Cogan's, Henson and Hern's, Kryger and Zimmerman's, and Wood's works help sketch out the contours on which further inquiries can be built. Studying these debates can also highlight the ways in which students' and teachers' lived experiences may become more central to research into writing assessment practices.

Considering questions about learning differences speaks not only to issues in contract grading but also to issues in writing assessment more broadly. Kryger and Zimmerman's work draws on a wealth of sources about neurodiversity, and engagement with these sources suggests the ways in which writing assessment

scholarship may develop more nuanced and contextualized ways of considering the value of students' writing and the ways in which learning is represented in writing samples. These types of moves towards more situated understandings of how learning and knowledge are embedded within writing samples, reflective texts about writing processes and goals, and logs about labor or engagement may also mitigate the impacts of racism and discrimination as obstacles to student success. That is, the context-sensitive, situated forms of writing assessment championed by researchers considering neurodiversity may also prove beneficial when researchers, writing program administrators, and educational policymakers work to create more equitable educational systems. Poe and Cogan's, Wood's, and Henson and Hern's articles reflect how the field of writing assessment has confronted—and is working to address—inequitable learning outcomes driven by seemingly facially neutral, institutionalized forms of discrimination. Their works suggest that students' and instructors' lived experiences are valuable when designing writing assessment systems that range from the classroom-level to institutional-level and even to the state-level. Evaluating how writing assessments promote, or limit, access for diverse students is part of the work that writing assessment researchers need to engage in. The work becomes particularly meaningful when questions about students' and teachers' lived experiences are considered in detail and inform how writing assessments are designed or modified.

FAIRNESS: CONSIDERING LIVED EXPERIENCES AS WAYS TO MITIGATE DISPARATE IMPACTS

In the *Journal of Writing Assessment's* Special Issue on a Theory of Ethics for Writing Assessment, Mya Poe and John Aloysius Cogan, Jr. detail the importance of a flexible, integrative framework to consider unintended consequences on demographic groups through writing assessment practices. This Special Issue was dedicated to the exploration of fairness more broadly and the fact that fairness had been underplayed in the research literature, in particular, on writing assessment. The group of authors aimed to extend the significant evolutions of educational measurement theory articulated in the 2014 revision of the *Standards for Educational and Psychological Testing* in which fairness was added as a foundational consideration. Poe and Cogan Jr. utilize and adapt the legal framework articulated in the Civil Rights Act of 1964—aimed to address intentional and unintentional discrimination—to writing assessment theory and practice.

In “Civil Rights and Writing Assessment: Using the Disparate Impact Approach as a Fairness Methodology to Evaluate Social Impact,” Poe and Cogan highlight a model for considering the effects of assessment practices on discrete groups of students using concepts from the Civil Rights legislation. Writing

assessment practices used by postsecondary programs, they argue, utilize seemingly facially neutral testing practices, but their inquiry demonstrated that there is no such thing as a neutral testing practice. Use of tests and the interpretation of their scores must be thoughtfully considered and if unintentional bias occurs, the program using the test must have a way to mitigate the disparate outcomes. By systematically reviewing student performance by disaggregated data, they were able to determine that their particular site indeed had an unintended, but still negative effect on a particular demographic group of students. That is, their testing practice—while on the surface appeared methodologically sound—actually disadvantaged the educational outcomes for a particular group of students. Their study provides a model grounded in empirical data to review the impacts of students and writing assessment tests within particular settings. As they note:

In the end, if equitability is to be valued, it must be seen. Fairness in theory cannot be an afterthought to validity or reliability. Fairness in action demands local attention in which we repeatedly question how we can achieve equitable results with less adverse impact. . . . Test scores may reflect social inequality, but the use of test scores works to create that social inequality. Racial isolation and structural inequality are not merely reflective of such social mechanisms; social mechanisms work to sustain invisibility, racialized isolation, and structural inequality. The creation of opportunity structures through approaches such as disparate impact analysis holds the potential to provide visibility, community, and equity. (p. 151)

Poe and Cogan's work provides us with a concrete and practical way to situate the consideration of fairness. They acknowledge that tests and scores may result in disparate impacts on different demographic groups. That is not a reason to discard the test; rather, they advocate for a thoughtful way to mitigate the impact of the bias through other programmatic means. That is, no test will ever be perfect. We need to have programmatic ways to account for their limitations and to do so we must first know how the tests are operating.

In another *Journal of Writing Assessment* Special Issue on Two-Year College Placement, Leslie Henson and Katie Hern explore the ways in which disparate impact studies can be used to evaluate how legacy writing assessment systems have inequitable impacts on students' lives. Their project at Butte College in Northern California highlights how established writing assessment systems may have persistent, unintended consequences on particular demographic groups. Using a disparate impact analysis, Henson and Hern document how achievement gaps along racial and ethnic lines may be reinforced by the structure of

a writing assessment and placement system. Henson and Hern contextualize their consideration of how writing assessment and placement systems work at Butte College within the larger data set of California's Community College System. They note that statewide most California community college students are considered "unprepared" with "more than 80% of incoming students [being required to] enroll in one or more developmental courses." Butte College's writing placement practices exist within this statewide system of placement and Henson and Hern show how the legacy of standardized tests has negatively affected practices at Butte. Their article critiques how "the standardized tests community colleges rely on to assess college readiness are a large contributor to the problem." But they also move beyond only a critique of current inequities based on the continued reliance on standardized, legacy forms of assessment and discuss how Butte College's new model of placements is leading to more equitable outcomes.

Based on multiple years of work, Henson and Hern trace the changes at Butte College through four different phases that include examining not only success in basic or first-year writing courses but also student success in later courses. Their work was part of a larger conversation within California about remediation at community colleges, and in particular, about concerns of the impacts of extensive levels of remediation being required for students of color. In 2018, the California legislature passed AB 705, a law that aligned with the writing assessment and placement practices Henson and Hern discuss. AB 705 requires community colleges to allow students to place into college-level (i.e., first-year composition rather than remedial English) as long as their writing assessments do not indicate that they are "highly unlikely to succeed." This state-wide policy shift addresses issues of fairness and highlights the ways in which debates around writing assessment systems can impact large numbers of students. It is indeed these relationships between writing assessment practices at particular colleges (e.g., Butte in this case) and larger assessment systems that provide a key area for considering the impact of fairness as an emerging concern for writing assessment scholars. These concerns are not only at play between the level of a single institution and state-wide policies. They may also be areas of investigation that connect individual classrooms and instructors' writing assessment practices with larger conversations in the field, such as neurodiversity.

Another *Journal of Writing Assessment* Special Issue, this one on contract grading, yielded two articles that focused specifically on teachers' lived experiences and issues of fairness that arise at the classroom level. Kathleen Kryger and Griffin X. Zimmerman's "Neurodivergence and Intersectionality in Labor-Based Grading Contracts" confront issues with the practice of contract grading while offering suggestions for more deliberately using grading contracts as a means of combating ableism experienced by students. Shane Wood's "Engaging in

Resistant Genres as Antiracist Teacher Response” draws on teachers’ experiences to examine how response patterns may replicate White language supremacy. Moving beyond this observation, he identifies teacher response as a dynamic genre that can help build anti-racist forms of response. These studies remind us that it is important to interrogate contract grading as an assessment instrument. While contract grading shifts many pedagogical practices at the root, it makes those changes based around a new model of writing assessment. It is a conversation about what we value that connects assessment with pedagogical practices in ways that impact teachers’ and students’ lives.

Kryger and Zimmerman’s chapter focuses on student experience. They confront issues of learning differences by challenging the notion that labor-based grading contracts are good for all, or even most, students by viewing this practice through the lens of neurodivergence. The authors remind us that both students and teachers represent a wide variety of learning experiences, styles, and preferences in writing classes and, although well-meaning, labor-based grading systems can and do result in the same unintentional discrimination that Poe and Cogan and Henson and Hern illustrate in their articles. The authors specifically address issues of fairness and equity by suggesting that the requirement of time logging in many labor-based grading systems is ableist and that this practice requires a more intersectional approach to classroom assessment. By complicating this increasingly popular grading system, like Wood, Kryger, and Zimmerman force readers to reconsider their understanding of a widely-accepted practice, focusing squarely on fairness as a priority in writing assessment practices. In this way, Kryger and Zimmerman set a foundation for continued work on labor-based grading to be a more inclusive and equitable approach to assessment while offering unique insight into the assessment experience for both neurodiverse students and teachers.

In his article, Shane Wood examines the ways in which both teacher and student response to student writing perpetuates White language supremacy. Wood calls for teachers and students to interrogate response to writing in order to disrupt the invisible reinforcement of linguistic racism. Specifically, the practice of response, one of the most common points of student-teacher interaction in writing classes, is taken to task for creating an inequitable learning environment, even in classes that practice seemingly antiracist writing assessment ecologies such as grading contracts. Like Poe and Cogan, Henson and Hern, and Kryger and Zimmerman, Wood identifies response to student writing as a site of (often unintentional) racist teaching practices. Wood’s framework for this interrogation is situated on teacher and student lived experiences as a deeply reflective exercise, requiring students and teachers to identify the genre of response, consider the purpose and nature of response, analyze and identify how White language practices inform the response, and finally reflect on how response can resist the

circulation of White language supremacy. This collaborative framework facilitates productive conversations surrounding language and power using the familiar genre of response as the site of study. This article illustrates how antiracist writing assessment work can and should be done at the class level as a partnership between teachers and students. All told, these four articles representing the lived experiences of teachers and students illustrate that, as Wood points out, “[g]ood intentions can still have violent consequences” (p. 233). Having a diversity of students and teaching practices at place in colleges across the United States requires us to have—and to interrogate—these new approaches. Examining them through the lens of teachers’ and students’ experiences may lead towards more fair and equitable learning outcomes.

SHAPING THE FUTURE OF CONSIDERATIONS OF FAIRNESS

The future of writing assessment lies in evolving ways that we may consider and accommodate the complex identities of students, faculty, and the institutions in which work is assessed. These articles provide an important blueprint for the way forward. Gere et al.’s lens of “communal justicing” (2021, p. 384). demonstrates the importance of a thorough examination of disciplinary infrastructure. Gere et al. argue that “to change the disciplinary infrastructure that shapes assessment, justicing must be *communal*: we all need to participate in the revision of the pasts, policies, and publications on which writing assessment depends” (p. 385). This means that we need to consider the entire ecology, to use Asao Inoue’s terminology, that surrounds the assessment of writing.

Such an effort has been underway for decades in educational measurement practices through the major revisions resulting in first the substantive philosophical reconceptualization of validity in 1998 and subsequently resulted in the 2014 revision to include consideration of fairness. These writing assessment practices have been evolving to be more expansive and inclusive in considering student performance and how we measure it. At this juncture—during a time of racial and cultural reckoning in the early 2020s—we argue that it is important to maintain the expansiveness in the consideration of fairness to protect this evolution in our practices. The approaches highlighted in this section point to an important path forward: one deeply committed to considering the lived experiences of students and faculty who inhabit the multiple institutional sites in which we teach, learn, and assess.

As these articles detail, we must consider the multiple and discrete ways in which students come to our institutions and demonstrate their writing abilities. Disciplinary identity is enacted through written communication, and that

language-informed identity makes us rethink traditional views of instruction and assessment. The lens of fairness is the means through which writing assessment practices may continue to necessarily evolve. Fairness while it is variously defined can be unified under principles of equity and opportunity to learn. Such aims necessarily need to be contextualized within specific institutional sites, where attention is paid to the ability of admitted students to access knowledge, skills, and attitudes of their particular fields of study with special attention paid to the affordances and barriers that accompany intersectionality (of socioeconomic status, ability, gender, race, and other individual differences). Students move in and out of identities that may advantage or disadvantage them in particular contexts. Our writing assessment practices need to accommodate the complexities with which our students present themselves. No longer can we assume a monolithic identity that represents a “college student.”

Likewise, we cannot and should not assume that a college student has the same experience at different institutional sites. Our practices need to reflect the particular missions of the postsecondary institutions and the faculty who teach at them should also be supported to assess students’ writing in ways that are valuable and meaningful to the people in their courses. Writing assessment practices have been evolving parallel to educational measurement practices. We’ve moved from the emphasis on method (holistic scoring and an emphasis on reliability) to a more situationally-based writing assessment practice. Directed self-placement and contract grading underscore the adaptability and flexibility of writing assessment practices to be attentive to the diverse needs of students in postsecondary courses and to adjust to the situational needs of an institution.

The shift in writing assessment has been a move from a high focus on methodology questions towards a more expansive conversation about how assessment practices can benefit students. *The Journal of Writing Assessment* has helped the field advance that change. The field of writing assessment has pushed for moves away from indirect writing assessments to direct writing assessments, from direct writing assessments to portfolio-based writing assessment as a way of capturing how writers develop over time, and now towards more situated forms of writing assessment that consider social contexts, their complexities, and ultimately the impacts on students. Articles published in *JWA* have pushed for more complexity in how colleges placed students, for portfolio-based assessment, for student involvement in directed self-placement, and now for better representation in how students come to our classes.

As the final section of *Considering Students, Teachers and Writing Assessment*, these chapters look to the future and provide us with a path forward. For a moment, it’s worth attending to, even meditating on, on what we want to see as writing assessment practices continue to develop. Even though the field has

shifted the locus of where writing assessment happens, particularly in the development of contract grading practices—we're still obligated to interrogate them. Writing assessment practices are not intrinsically good because they are new. Changes to writing assessment systems address deficiencies in current practice. These changes to practice have been rooted in categories that are visible—race and gender, for example – but we need to continue moving towards ways in which to account for things ‘unseen’ – learning differences, economic background, sexual orientation, and other considerations and how they might play out in our assessment of students’ writing.

What might we expect to see in terms of fairness and emerging research? Researchers might take a more community-based approach to their data collection and studies. These approaches could lead not only to more diverse student and teacher voices being included within writing assessment studies but could also increase the diversity among researchers. These shifts would require changes in methodologies and the guidelines for these types of studies. How, for instance, will studies of contract grading evolve so that they speak across institutions? Will researchers continue to work on alignment between shared empirical practices and the complexity of local contexts? Will researchers be able to develop studies that are replicated across contexts? What will be the dynamics among the categories of reliability, validity, and fairness? If, as we have argued in this collection, there has been a shift towards including fairness and looking at equitable outcomes, then what shifts will occur within large-scale writing assessment practices as well as local writing assessment practices? Will studies consider different scales and different scopes of writing assessment systems? That is, will questions about how writing assessment policies work at local, institutional, state, and national level develop in ways that continue to balance reliability, validity, and fairness?

REFERENCES

- American Educational Research Association, American Psychological Association, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Carillo, E. (2021). *The hidden inequities in labor-based contract grading*. Utah State University Press.
- Gere, A. R., Curzan, A., Hammond, J. W., Hughes, S., Li, R. Moos, A., Smith, K., VanZanen, K., Wheeler, K. L., & Zanders, C. (2021). Communal justicing: Writing assessment, disciplinary infrastructure, and the case for critical language awareness. *College Composition and Communication*, 72(3), 384-411.
- Inoue, A. B. (2015). *Antiracist writing assessment ecologies: Teaching and assessing writing for a socially just future*. The WAC Clearinghouse; Parlor Press. <https://doi.org/10.37514/PER-B.2015.0698>