

FOREWORD

Kathleen Blake Yancey

Florida State University

Brian Huot

Kent State University

It's a truism to note that writing assessment has come into its own during the last several decades, and one of the factors propelling that growth is the *Journal of Writing Assessment (JWA)*. As the selected articles-now-chapters presented here suggest, writing assessment is both more and different than what it seems to be. While it can simply appear as a rudimentary exercise in evaluating writing, writing assessment, as the authors here have documented, researched, and theorized, is at least twofold: (1) an exercise of considerable sophistication and complexity operating within a context (2) that can overwhelm, and sometimes sabotage, the exercise itself. These twin observations informed our goal when we created *JWA*, a new journal focused on writing assessment that would circulate scholarship taking up questions about how to best assess writing as well as about the contextual factors, often invisible, that shape and, too often, mis-shape writing assessment. Put in the current vernacular, with *JWA* we hoped to make writing assessment—and its many dimensions—transparent.

The articles in our first issue of *JWA* made this goal visible. In “Moving Beyond Holistic Scoring through Validity Inquiry,” for instance, Peggy O’Neill (2003) focused on validity, a key issue in writing assessment; her article is included here. Turning to context, George Hillocks (2003) addressed the impact of state assessments in his “How State Assessments Lead to Vacuous Thinking and Writing.” Sandra Murphy did likewise, in her case looking not at the impact of writing assessment on students, but rather at its impact on teachers in one state; such teachers support students’ writing development as they practice assessment within their classrooms.

That first issue of *JWA* concluded with an annotated bibliography; compiled by Peggy O’Neill, Michael Neal, Ellen Schendel, and Brian Huot (2003), it too spoke to *JWA*’s vision. Three bibliographic entries in particular articulate *JWA*’s goal and its importance while forecasting the kinds of research, theory, and practice published in *JWA* during the last 14 years, as sampled in this edited collection. The first bibliographic entry, Nicholas Lemann’s 1999 book *The Big Test: The Secret History of the American Meritocracy*, details a social and cultural history of the SAT.

Although the stated purpose of the SAT was to change the college admissions process by eliciting relevant information from college applicants so as to predict their success in college, it also clearly intended to shift college admissions from one based in legacies to one based in merit. The Lemann account also clarifies how the SAT both succeeded and failed in that intention, demonstrating that assessment, even when informed by the science of tests and measurements, is always contextualized, always enacting a policy, whether visible or not.

A second item in the bibliography, O. Palmer's College Board Report 42, "Sixty Years of English Testing," (1960) argues that the science informing the College Entrance Examination Board (CEEB) English testing contributes to such testing "as a scientifically defensible practice" (O'Neill et al., 2003). Again, here too science plays a role, not so much to forward a kind of democracy, however, but rather to defend the practices of a growing assessment industrial complex. In the CEEB model Palmer defends, both teachers and direct writing assessment are positioned as opponents of CEEB, as "resistant to the scientific progress achieved in English testing" (O'Neill et al., 2003). What teachers, rooted in the everyday of the human classroom, may have understood better than measurement experts is how writing assessment, regardless of the science, cannot be cleaved from the contexts and complications accompanying it. As important, seeing students day in and day out, teachers also understood how very contingent any decision based on assessment is.

In his 1994 "A Technological and Historical Consideration of Equity Issues Associated with Proposals to Change the Nation's Testing Policy," George Madaus seems to agree with teachers. Approaching what we might call the assessment problem philosophically in this third bibliographic entry, with a view informed by both phenomenology and practicality, Madaus observes that certain principles define assessment. All evaluations, he notes, rely on samples of behavior; all evaluations make inferences "about a person's probable performance relative to the domain" (Madaus, 1994); and all assessments render decisions by individual or institution. Moreover, the technologies don't operate apart from the culture of their origin. Instead, as

products of a culture, they often extend, shape, and reproduce the same culture. The values that underlie testing are utilitarianism, economic competition, technological optimism, objectivity, bureaucratic control and accountability, numerical precision, efficiency, standardization, and conformity. (O'Neill et al., 2003)

It's worth noting that while such values, including standardization, conformity, and economic competition, may locate the US, its testing industry, and its schools, they are much less likely to be the values motivating teachers.

The articles in the two volumes of this edited collection carry these issues of assessment and context forward, especially as they have been raised and considered over time. In Volume 1, the collection's first section, *Technical Issues in the Assessment of Writing—Reliability and Validity*, speaks to issues articulated by both Peggy O'Neill and George Madaus, issues inherent in assessment that, as both O'Neill and Madaus demonstrate, are *not apart from* larger human issues, but are rather *a part of* them. The second section, *Politics and Public Policy of Large-Scale Writing Assessment*, calls to mind the article by George Hillocks and the history of college admissions provided by *The Big Test*. The third section, *Implications of Automated Scoring of Writing*, questions how the evolution of automated essay scoring extends the dangerous logic of a "true" score as valid and reliable across contexts. In Volume 2, the fourth section, *Theoretical Evolutions—Towards Fairness and Aspiring to Justice*, again calls to mind the equity issues and analysis developed by Madaus. And the fifth section, *Students' and Teachers' Lived Experiences*, evokes the line of inquiry pursued by Sandra Murphy. As astute readers have already noted, it's also fair to observe that in this set of correspondences between the introductory issue of *JWA* and the current collection's chapters, one section in the collection, *Implications of Automated Scoring of Writing*, is left out: our first issue of *JWA* did not provide for the important questions about writing assessment raised by digital technologies. Still, apprehending that they were on the horizon, we made a start in the very next issue, courtesy of Michael Williamson's (2003) "Validity of Automated Scoring: Prologue for a Continuing Discussion of Machine Scoring Student Writing."

All of which is *not* to say that we anticipated all of the rich writing assessment scholarship of the next decade and a half: our correspondences, of course, are not predictive. But it is to say that the chapters here extend and elaborate what we had hoped for in creating *JWA*, in the process refiguring continuing issues, sounding new notes, and pointing us to new futures. For example, one chapter argues that the divide between the educational measurement and the writing assessment communities might be bridged with a "unified field of writing assessment." The construct of writing, another chapter explains, can no longer ignore "the role of commonly available tools such as word processing software." And yet another chapter brings together science and the law in a shared inquiry into the results and subsequent effects of writing assessment, employing a disparate impact analysis framework contributing to a better, more human, more humane, and more equitable assessment. Threaded throughout are the technical issues and principles of writing assessment, the writing assessments themselves, and the contexts in which they are embedded.

Remembering the recent history of writing assessment, focused on assessments and their contexts as we prepare for a better writing assessment future, we

are very pleased to be learning from and with the authors included here. We feel confident that you will be as well.

REFERENCES

- Hillocks, G. (2003). How state assessments lead to vacuous thinking and writing. *Journal of Writing Assessment*, 1(1), 5-21. <https://escholarship.org/uc/item/33k2v0w5>
- Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. Farrar, Straus and Giroux.
- Madaus, G. F. (1994). A technological and historical consideration of equity issues associated with proposals to change the nation's testing policy. *Harvard Educational Review*, 64(1), 76.
- Murphy, S. (2003). That was then, this is now: The impact of changing assessment policies on teachers and the teaching of writing in California. *Journal of Writing Assessment*, 1(1), 23-45. <https://escholarship.org/uc/item/1fg1531r>
- O'Neill, P. (2003). Moving beyond holistic scoring through validity inquiry. *Journal of Writing Assessment*, 1(1), 47-65. <https://escholarship.org/uc/item/4qp611b4>
- O'Neill, P., Neal, M., Schendel, E., & Huot, B. (2003). An annotated bibliography of writing assessment. *Journal of Writing Assessment*, 1(2), 73-78. <https://escholarship.org/uc/item/78q9v569>
- Palmer, O. (1960). Sixty years of English testing. *College Board*, 42, 8-14.
- Williamson, M. M. (2003). Validity of automated scoring: prologue for a continuing discussion of machine scoring student writing. *Journal of Writing Assessment*, 1(2), 85-104. <https://escholarship.org/uc/item/8nv3w3w8>