Title

An Annotated Bibliography of Writing Assessment: Minority Issues in Writing Assessment

Permalink

https://escholarship.org/uc/item/9hg9v5p4

Journal

Journal of Writing Assessment, 3(2)

Authors

Edgington, Anthony
Hester, Vicki
O'Neill, Peggy
et al.

Publication Date

2007

Copyright Information

Peer reviewed

Minority Issues in Writing Assessment
An Annotated Bibliography

ANTHONY EDGINGTON
University of Toledo

VICKI HESTER
St. Mary's University

PEGGY O'NEILL
Loyola College, Maryland

ELLEN SCHENDEL
Grand Valley State University

This annotated bibliography focuses on issues surrounding minorities and writing assessment, including issues associated with various ethnic groups as well as those issues associated with gender studies and with minorities in special education. In addition to reporting on minorities and testing consequences, several selections make recommendations for revising or constructing testing instruments to address some of the special issues minorities face with classroom and large-scale assessments. Several selections also pose important questions that classroom instructors, school administrators, and writing assessment specialists might ask before conducting assessments that involve minority students. Over the past 30 years, studies of SAT, ACT, and other high-stakes tests report that test scores continue to lag behind for certain minority groups. Several selections that we annotate below offer reasons for these differences and fresh insights and analyses into trends for minority test scores.

Our goal with this abridged collection of sources is to provide multiple perspectives on the issues surrounding minorities and writing assessment and to provide examinations of some of the main theoretical and pedagogical issues concerning minority issues and writing assessment. We have focused our attention on published scholarship while omitting unpublished sources such as ERIC documents and dissertations. During our attempts to locate resources, we believe we also located a need for further research in this area of minority issues and writing assessment, though if we omitted sources we hope our readers will call this to our attention.

Ball, A. F. (1997). Expanding the dialogue on culture as a critical component when assessing writing. *Assessing Writing, 4*(2), 169-202.

Ball reports on two studies that examine the rhetorical and linguistic features of texts and how they contribute to holistic assessment of student writing. She addresses issues of teacher evaluation of writing produced by ethnically diverse students. The first study reports the results of a group of four Euro-American teachers evaluating a set of 23 texts written by students identified as ethnically diverse from working and lower class backgrounds. The second study replicates the first with a group of four African-American teachers. Ball notes that in comparing the scores of the two groups, the texts written in "academically oriented organizational patterns" (p. 178) were rated higher than those written in nonacademic patterns, with Euro-American students receiving the highest scores, African-American students receiving moderate ratings, and Hispanic-American students receiving the lowest ratings. However, Ball notes that the Euro-American teachers consistently rated the writing by Euro-American students higher than the texts by non-Euro-American students, whereas African-American teachers gave more moderate ratings to Euro-American student texts. The Euro-American teachers gave higher holistic ratings to all texts than the African-American teachers. The third section of the article reports on a discussion with the four African-American teachers about issues associated with ethnic and linguistic diversity, pedagogy, assessment, and policymaking. Ball concludes that it is critical to include voices of teachers with diverse backgrounds in discussions of writing assessment.

Ball, A. F. (1999). Evaluating the writing of culturally and linguistically diverse students: The case of the African American vernacular English speaker. In C. R. Cooper & L. Odell (Eds.), *Evaluating writing: The role of teachers' knowledge about text, learning, and culture* (pp. 225-248). Urbana, IL: National Council of Teachers of English.

Ball provides a close textual analysis of a student text written in African American Vernacular English (AAVE) identifying key syntactic, semantic, phonological, and stylistic features of AAVE. She explains that historically AAVE has been viewed from four perspectives: deficit, difference, proficiency, and resource. From her analysis, she offers five key principles for teachers of AAVE speakers aimed at providing effective instruction and evaluation for all students.

Cross, T., & Slater, R. B. (2006). A large black-white scoring gap persists on the SAT. *The Journal of Blacks in Higher Education, 53*. Retrieved March 10, 2007, from http://www.jbhe.com/features/53_SAT.html.

This article shows that since 1976 African-American scores on the SAT test have risen slowly but steadily—except for a short time in the late 1980s—whereas White SAT scores have dropped, reducing the gap between Black and White scores. In 1976, the average Black score was 240 points (20%) below the average White score but is now 200 points below. The mean score on the writing section of the SAT in 2006 was 428 for Blacks and 519 for Whites, suggesting that the new writing component of the SAT will lessen the gap even further. Although Black scores continue to increase, a new study by the College Board estimates that a large racial gap in standardized test scores will still exist 22 years from now although Justice O'Connor expressed a goal for eliminating affirmative action within 25 years. These writers discuss reasons for the scoring gap, one being that Black students attend schools that remain underfunded and poorly staffed. Students are not provided the same quality of education offered most White students. Also, Black students do not follow the same academic track as White students in secondary education; Black students are often taught an Afrocentric curriculum that may increase self-esteem and

Black pride but does little to prepare students for college entrance exams, and students are often tracked into vocational programs. This article could be useful to program administrators when deciding whether or not (or how) to factor in SAT scores for placement into first-year writing courses. This article also provides useful statistics and analyses from the College Board comparing Black and White SAT scores since 1976.

Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom.* New York: The New Press.

Delpit's book focuses on the problems teachers often encounter when teaching and assessing "other people's children," those students who arrive from backgrounds, cultures, and communities that differ drastically with that of the teacher. She suggests a blending of process- and skills-based education, valuing home dialects and languages, understanding the importance of context, and the need to assess language and literacy through eyes other than her own. Most pertinent to assessment researchers is the first article in Part 3: "Cross-Cultural Confusions in Teacher Assessment." Delpit criticizes past and current teacher assessments for becoming obstacles to teachers of color, "contributing to wholesale elimination of people of color from the teaching force" (p. 136) while ignoring the culturally different teaching and communication styles of many instructors. She suggests providing multiple contexts for teachers to discuss their pedagogy (with assessors, peers, students), seeking feedback from peers and supervisors, and allowing teachers the chance to select those who will assess or interview them. This book would be beneficial to those conducting studies on large scale and/or authentic assessments, teacher education, and the influence of race on pedagogy and student-teacher communication.

Garcia, G. E., & Pearson, P. D. (1991). The role of assessment in a diverse society. In E. H. Hiebert (Ed.), *Literacy for a diverse society* (pp. 253-278). New York: Teachers College Press.

Garcia and Pearson argue for more teacher involvement in large-scale assessments and for an improved teacher understanding of the degree to which assessment materials can "distort or reflect the literacy development of students from diverse linguistic, cultural, and/or economic backgrounds" (p. 254). The writers suggest a type of situated programmatic assessment if assessments are to drive the curriculum, suggesting approaches that include classroom site visitations, interviews with teachers, students, parents, administrators, as well as an examination of a wide range of teaching and learning artifacts (p. 269). Several other changes they recommend foreground teacher-student relationships and teacher-parent relationships so that teachers better understand the cultural norms (pp. 262-264). Garcia and Pearson address reliability, validity, objectivity and efficiency in order to consider problems of ethnic bias and test interpretation. This article posits that for assessments to be trustworthy, authentic, and instructional for diverse student populations, multiple indices of student development are necessary. This chapter can be useful for test development and for a historical review of constructivist reading and writing assessment since 1980.

Harry, B., Klinger, J. K., Sturges, K. M., & Moore, R. F. (2002). Of rocks and soft places: Using qualitative methods to investigate disproportionality. In D. J. Losen & G. Orfield (Eds.), *Racial inequity in special education* (pp. 71- 92). Cambridge, MA: Harvard Education Press.

Although this study looks at research practices that contribute to inappropriately placed minorities into special education rather than specifically at writing assessment, it remains

useful because the research questions relate to issues of minorities and assessment. The writers provide research information specific to gender and ethnicity. This work is useful for future teachers and school administrators as well as those who instruct them. The writers include research studies and arguments for triangulating data concerning placement decisions for all minorities, as for all students, because of the complex social construction of all students (pp. 83-87).

Heubert, J. P. (2002). Disability, race, and high-stakes testing of students. In D. J. Losen & G. Orfield (Eds.), *Racial inequity in special education* (pp. 137-165). Cambridge, MA: Harvard Education Press.

According to Heubert, research shows that states with high-stakes testing also have high enrollments in special education for minority students, and many of these states do not provide instruction in the subject matter and skills for which they test these students for placement (p. 137). The consequences for those placed—those who do not graduate high school but who should have been able to graduate—include much lower earnings in the workforce and reduced opportunities for further education (p. 149). Low-achieving minority students with disabilities are often retained in their current grade, and although controversy exists about the causes for dropout rates among these students, many scholars agree that large-scale testing and graduation testing increase the dropout rates among these students (pp. 146-147). This chapter provides some background and summary of the *No Child Left Behind Act*, and gives statistics for graduation tests in many states with discussions of such issues as improper exemptions, exclusions, and absences of minority students with disabilities from test-taking populations (p. 147). This chapter is useful to those researching large-scale testing, providing theory-based arguments concerning testing use, arguing that tests should not be used to penalize individual students but to promote education, and that educators should use multiple measures when making important decisions about placement for individuals (pp. 150-156). The last section makes eight detailed recommendations for constructing high-stakes testing for public schools with a closing discussion about the reasons for appropriate test use, among them an argument against using tests for tracking or retention (pp. 156-159).

Howell, K.W., Bigelow, S.S., Moore, E.L., & Evoy, A.M. (1993). Bias in authentic assessment. *Diagnostique, 19*(1), 387-400.

The researchers studied the creation and implementation of a statewide authentic assessment Arizona. They summarize current views on validity theory and acknowledge that this information guided their research project. The study was comprised of two phases. The first phase involved a questionnaire sent to teachers, administrators, and researchers involved with the state assessment. This phase found that the test was not productive or accommodating for students with disabilities, produced some bias among raters toward the content in essays (at times written by minority students), and displayed an overall negative view of the format and content of the test. The second phase expanded on the discussion of fairness and bias addressed in the first phase, as the researchers created two essays (one with distinct cultural markers) in order to gauge whether racial and/or cultural content in essays influenced readers/raters. The research found that among 147 respondents, many lowered their expectations for the paper containing minority passages, thus raising the average ranking. As the authors argue, "the findings from Phase Two demonstrate that assessment problems do not reside exclusively within tests. . . . Not all bias is found in test items" (p. 398). The researchers conclude that "it may be that large scale exercises in authentic assessment will need to be preceded by efforts to screen those who administer the assessment for instructional skills, and to screen those who score the

samples for bias" (p. 399). This study is useful to those exploring the influence of bias on large-scale assessments and for those who are currently considering implementing investigations into state-wide assessment practices.

Kynard, C. (2006). "Y'all are killin' me up in here": Response theory from a newjack composition instructor/sistahgurl meeting her students on the page. *Teaching English in the Two-Year College, 33*(4), 361-387.
Kynard's article focuses on teacher and student identities and how these identities are played out through written texts and responses. Kynard is especially interested in her response-identity as an African-American woman or "sistahgurl." First, she acknowledges the unequal, labor-intensive position she inhabits as a writing teacher, but insists that if she wants her students to succeed, she needs to "meet them on the page" regardless of her working conditions. Luckily, her educational environment (a community-based, predominantly Black college with mostly working-class students) allowed her and her students' identities to evolve, especially through their writing. In their written exchanges, Kynard and her students come to value language and lexicon while writing about personal topics and concerns central to their racial and local communities. Along the way, Kynard and her students began to understand how their own backgrounds, biases, and experiences influenced how they read and respond to each other (via journals, written letters, and peer review sessions). She compares this to a reading session with fellow teachers, who resisted reading through their own social lenses. As Kynard writes, "the major difference between me and my colleagues…is I am aware of the situatedness of my responses and their locations simply because they are not hegemonic, dominant, and thereby, normalized" (p. 374). This article would work well with research studies on the influence of race on assessment (in conjunction with the work of Arnetha Ball) or in studies of the connection between social factors and assessment.

Lacelle-Peterson, M. W., & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review, 64*(1), 55-75.
This article argues that English language learners (ELLs) will continue to fall behind students who are native speakers of English—despite the promise of educational reform—unless assessment of ELLs is more carefully thought out and implemented. Even though educational reform is geared toward equality and fairness, argue Lacelle-Peterson and Rivera, the "perspectives and needs" of ELLs have been largely unacknowledged and unaddressed (p. 56). The article then carefully details the particular educational and assessment needs of ELLs before describing four principles that should be followed when designing assessments of ELLs: that the assessments should "be comprehensive in the sense that they seek to provide an integrated account of all that ELLs are learning, both in language and in academic content areas" (p 68); use "multiple indicators . . . to assess ELLs progress in languages and academic areas" (p. 68); "focus on students' progress over time toward established goals, rather than on comparisons" (p. 69); and "be sensitive to the particular needs of the groups of students locally" (p. 69). The article ends with guidelines for a process of designing assessments based on these principles. Despite the article being 13 years old, it is still timely, as many of the problems relating to educational reform as described by Lacelle-Peterson and Rivera are still apparent in 2007. It is a particularly useful article in that it applies sound educational measurement principles to assessment and describes how changes to assessment systems can be made at the local level.

McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing, 18*(4), 333-349.

McNamara opens with a discussion of the recent influence of social theory on language studies, in general, and upon language testing, specifically. He includes a detailed overview of Messick's work in the area of validity, including Messick's belief that "all interpretations of test scores involve questions of value, that is, that we have no 'objective', 'scientific', value-free basis for this activity" (p. 335). McNamara connects this social view of language testing to the work of Butler, who argues that social and gender identity is directly tied to social environments and actions. Often, however, these environments and actions tend to be invisible or unconscious, which influences how much we are cognizant of how our identities are "performed." Connecting the two theorists, McNamara wonders "are the constructs of language testing susceptible to a performative unveiling, to be revealed as social constructs serving social ideologies?" (p. 339). He then moves to the classroom environment, arguing that current assessment practices are at odds with the needs of learners and teachers, the "most at risk" groups. He finds that, while large-scale assessment initiatives are touted as a benefit to teachers and learners, these assessments more often become a barrier or extra burden that must be endured. He calls for a greater focus on creating assessment methods that will assist teachers and learners, mainly through incorporating more critical reflection into our assessment practices. He concludes with ideas for how to connect critical reflection more to classroom assessment. This work is useful for those unaware of current validity theory, those who are seeking ways to add a critical reflective component to assessment, and for those interested in the social view of assessment, especially in relation to how various social groups are affected by assessment decisions.

Matzen, R. N., & Hoyt, J. E. (2004). Basic writing placement with holistically scored essays: Research evidence. *Journal of Developmental Education, 28*(1), 2-34.

This article reports on a university-level study in which students placed by COMPASS and DRP (Degrees of Reading Power) tests into two basic writing courses were given the task of writing a timed essay at the beginning of the semester. The essays were scored holistically by English teachers in the program, and students scoring well enough could move from the first basic writing course into the second. Surveys were also administered at the end of the term to find out about students' experiences: whether they visited the writing center, attended class regularly, encountered any nonacademic events that affected their performance in school, and tried their best on the day of the writing test. The findings support those of previous research: direct assessment of writing was more accurate than either of the tests—or the tests in combination—in predicting student success. There was a low correlation among COMPASS, DRP, and holistic essay scores, particularly among minority students, who scored disproportionately low on the standardized tests. More useful than the article's research findings themselves are the examples of how the research process and results were beneficial to the program and provided useful data for faculty to argue for resources to incorporate best practices into the program's design.

Mountford, R. (1999). Let them experiment: Accommodating diverse discourse practices in large-scale writing assessment. In C. R. Cooper & L. Odell (Eds.), *Evaluating writing: The role of teachers' knowledge about text, learning, and culture* (pp. 366-396). Urbana, IL: National Council of Teachers of English.

Mountford challenges the narrow, linear definition of academic writing associated with most composition programs, arguing that we should broaden our vision of academic discourse based on the diverse approaches actually used by academics on behalf of the

diverse people in our classes. She acknowledges that this more inclusive approach requires changes in how we evaluate writing. She concludes with implications for large-scale writing assessments, making four recommendations for procedures to accommodate students from nontraditional backgrounds: (1) use multimodal assessments (2) educate evaluators on features of writing of nontraditional groups, (3) avoid autobiographical essays if only one sample is evaluated, and (4) design prompts to evoke information available to any person regardless of cultural background if only one sample is evaluated.

Peterson, S. (1998). "Evaluation and teachers' perceptions of gender in sixth-grade student writing." *Research in the Teaching of English, 33*(2), 181-208.
This study examines sixth-grade teachers' perceptions of gender and the implications these perceptions have for large-scale writing assessment, not only for elementary school but also secondary and college-level writing assessment. The teachers involved in Peterson's study typically characterized the writing of girls as more sophisticated, articulate, and outward focused with higher levels of awareness of audience (p. 197). Boys' writing, on the other hand, was described as choppy, less concrete, drawing on "superficial super-heroes who interacted using violence" (pp. 197-198). When one student exhibited cross-gender writing characteristics, and teachers knew the writer's gender, teachers consistently gave the student lower scores than did teachers who did not know the student's gender of this piece. Teachers exhibited gender expectations, and when students failed to fulfill those expectations, scores suffered, even in this anonymous grading situation. This study will be useful for those researching or designing large-scale assessments as it demonstrates the sociocultural influences on assessment processes and the difficulties involved in producing unbiased test scores of student writing. This study also provides a strong bibliography for other gender studies of writing assessment.

Roever, C., & McNamara, T. (2006). Language testing: the social dimension. *International Journal of Applied Linguistics, 16*(2), 242-258.
The authors preview their book, *Language Testing: The Social Dimension*, which includes first a study of validity theory, connecting works by such theorists as Messick and Cronbach. The writers ask about the social values and social consequences of assessments and argue for both psychometric and socially analytical approaches to assessment (p. 245). This article reviews the book's coverage of psychometric approaches to assessment and ethics, assessment and social identity, as well as some discussion of the unintended consequences to minorities and English as a Second Language students of the *No Child Left Behind Act* (NCLB). "For example . . . fewer teaching resources are devoted to areas that are not tested in the NCLB assessment . . ., and schools may be tempted to encourage weaker language minority students to drop out in order to raise the average proficiency level" (p. 254). These writers suggest that studies need to focus explicitly on the consequences of language assessments. This article may be useful to those researching large-scale assessment theories as well as writing program administrators involved in test development who want to read more about ethics and fairness regarding minority issues and who want to consider the kinds of questions that should be asked before constructing large-scale assessment instruments.

Steele, C. (2003). Stereotype threat and African American achievement. In T. Perry, C. Steele, & A. Hilliard, III (Eds.), *Young, gifted and black: Promoting high achievement among African-American students* (pp. 109-130). New York: Beacon Press.
Steele reports on several studies he and his colleagues administered that show how "stereotype threat," the fear of being seen from a negative stereotype, caused significant-

ly lowered performance on the verbal and math sections of the Graduate Record Exam. Steele found that the test scores of achievement-oriented students suffered the most from stereotype threat. On the other hand, students with lower academic goals suffered less from stereotype threat when tested. Steele also discusses how test administrators and classroom teachers can reduce stereotype threat. This chapter offers insight into some of the ways that large scale assessments and classroom assessments may fail to accurately measure minority student performance, especially the performance of high-achieving minority students. This research would be useful to teachers as well as assessment specialists as it provides insight into the apprehension and distrust of many achievement-oriented, African-American students and the influence of this distrust on student scores.

White, E., & Thomas, L.L. (1981). Racial minorities and writing skills assessment in the California state university and colleges. *College English, 43*(3), 276-283.
White and Thomas conduct a comparison study to see what effects different types of testing have on racial minorities. The two tests used are the Test of Standard Written English (TSWE)—a multiple-choice test focusing on grammar and usage—and the California State University and Colleges English Placement Test (EPT)—comprised of four sections, including a single-sample essay, multiple-choice reading test, and two multiple-choice tests on "sentence construction" and "logic and organization." At the heart of the study are questions concerning the validity of using multiple-choice exams to measure writing ability. The researchers studied samples from 10,719 students from four different racial categories: White, Black, Mexican-American, and Asian-American. The researchers found that while there was consistency in the scores among White students, there was greater inconsistency among minority student sample scores. As the researchers state, "the TSWE usage test . . . rendered a much more negative judgment of these students' use of English than did the evaluators of their writing" via the EPT essay test (p. 281). White and Thomas warn those who use multiple-choice tests to gauge writing ability to be aware of these possible discrepancies and call for continued research on the topic. The research would be useful for those involved in placement testing (especially where multiple-choice tests are used for placement) and for those investigating the effects of various tests on minority populations and their writing ability.