



## An Annotated Bibliography of Writing Assessment

**PEGGY O'NEILL**

*Loyola College*

**MICHAEL NEAL**

*Clemson University*

**ELLEN SCHENDEL**

*Grand Valley State College*

**BRIAN HUOT**

*University of Louisville*

Welcome to the first installment of an annotated bibliography on writing assessment. Over the next several issues, we will be publishing different sections of the bibliography such as Theory, Classroom Response, and Portfolios. Our intent is to provide a resource for all those who work in writing assessment (K-college) or are looking for writing assessment scholarship. In categorizing and sectioning the scholarship, we have used theories and practices of writing assessment instead of educational level because we see a need for more crossover, especially between K-12 and college, in research, theory, and practice.

Although our goal is to make the bibliography as comprehensive as possible, we are omitting unpublished sources, such as dissertations and ERIC documents because our purpose is to give a sense of the published scholarship of the field (and because those who might want to access other materials can use the appropriate search engines for those references). However, even with this more narrowed scope, we realize that we are bound to have omissions. To this end, we invite readers to send us entries and annotations that we miss. In the future, we hope to publish the entire bibliography in one place, and your additions will help produce the most comprehensive bibliography possible.

### The History of Writing Assessment

The first installment of the bibliography focuses on writing assessment history. We have chosen the various entries for the section on history based on two different criteria. First, we have included work that is historical in nature and talks about early notions of writing assessment and their development. Second, we have

included early work on writing assessment that has historical value even though its focus may not have been on history per se (such as the early work on holistic scoring). For edited collections, we list particularly important chapters separately even if we include the entire text as a separate entry.

Applebee, A.N. English language arts assessment. Lessons from the past. *English Journal*, 84(4), 40-46.

Offers six lessons “based primarily” on Applebee’s involvement in the National Assessment of Educational Progress (NAEP): (a) student involvement has risen over the long term, (b) curricular issues are the most important issues in shaping assessment, (c) traditional assessments ignore process-related skills, (d) all tasks are not equal, (e) standards are not absolute, and (f) portfolios offer false hope and salvation. Concludes with five principles for literacy assessments.

Camp, R. (1993). Changing the model for the direct assessment of writing. In Williamson and Huot (Eds.), 45-78.

Reviews the traditional approach to writing assessment that emphasizes narrow definitions of validity and reliability and that supports the use of multiple-choice tests and impromptu writing samples. Argues that this approach conflicts with the current view of writing as a complex, multifaceted activity that is contextually dependent. Proposes that writing assessment be informed by newer measurement theories, in which validity is seen as a unitary concept that includes the consequences of the test results, to develop contextualized assessments that are theoretically compatible with current conceptions of writing.

Cody, S. Scientific principles in the teaching of composition. *English Journal*, 4, 161-172.

Provides a rationale for an approach to teaching composition based on the “modern scientific method” of formulating an hypothesis and testing it for results. Evaluation and response play a significant role in this course of study because the author “knows of no other way in which English composition can be taught successfully than by this careful, sympathetic, strictly personal criticism of the work prepared under favorable conditions.” Advocates students learn self-criticism and revision and with feedback to individuals given in a whole-class setting (similar to workshops).

Cooper, C.R. & Odell, L. (Eds.). (1977). *Evaluating writing: Describing, measuring and judging*. Urbana, IL: National Council of Teachers of English.

Gives an overview of assessment concerns and approaches important to writing studies in the mid-1970s. The text is practical, describing and illustrating a number of assessment techniques, and far-reaching in scope, including essays such as Charles R. Cooper’s “Holistic Evaluation of Writing”; Richard Lloyd Jones’ “Primary Trait Scoring”; Lee Odell’s “Measuring Changes in Intellectual Processes as One Dimension of Growth in Writing”; and Patrick J. Finn’s “Computer-Aided Description of Mature Word Choices in Writing”.

Cooper, P. (1984). *The assessment of writing ability: A review of research*. Princeton, NJ: Educational Testing Service. GREB No. 82-15R.

Reviews the literature on direct and indirect writing assessment up to the early 1980s. Provides an important, comprehensive source of information on scholarship that established the efficacy of direct writing assessment. An invaluable resource for understanding the historical and theoretical issues that helped to shape the still dominant forms of direct writing assessment.

Davis, B. G., Scriven, M. & Thomas, S. (1987). *The evaluation of composition instruction* (2nd ed.). New York: Teachers College Press.

The second edition of a handbook for writing assessment developed through the evalua-

tion of The Bay Area Writing Project. Furnishes a historical view of early writing assessment, while at the same time it provides a resource of procedures, terms, and principles for measurement specialists and writing teachers and administrators. Writing specialists should find the glossary of measurement terms especially helpful.

Diederich, P.B., French, J.W & Carlton, S.T. (1961). *Factors in judgments of writing quality*. Princeton: Educational Testing Service. RB No. 61-15. ERIC ED 002 172.

Germinal study involving the scoring of 300 essays on a 9-point scale by 53 readers who represented six professional fields. Ninety-four percent of the papers received at least seven different scores. Over 11,000 comments on 3,500 papers were used in a factor analysis that isolated five main types of responses. These responses were used to develop the initial analytic scoring guideline.

Diederich, P. B. (1974). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.

This manual for the direct assessment of writing begins, "The principle task of this booklet will be to suggest ways of improving the reliability of grades on essays." Documents how to insure that raters agree on their judgments of student writing. Using an analytical scoring procedure developed from the factor analysis of 53 raters' scores on 300 student papers in the Diederich, French, and Carleton study, this book describes how to read student writing reliably.

Fuess, C. (1967). *The College Board: Its first fifty years*. New York: College Entrance Examination Board.

This book chronicles the first 50 years of the College Entrance Examination Board. Although much of the text does not relate to writing assessment per se, there is important information about the ways in which writing was assessed. It is particularly interesting for the chronicle of the changes, challenges, and rebuttals about writing assessment brought about through the adoption of multiple-choice-type testing in the late 1930s and early 1940s.

Godshalk, F. I., Swineford, F. & Coffman, W. E. (1966). *The measurement of writing ability*. Princeton, NJ: Educational Testing Service. CEEB RM No. 6.

Watershed study that established the viability of direct writing assessment. The procedures used to attain acceptable rates of interrater reliability are still practiced today. Although it would be hard to accept the study's claim that "this problem [measuring student ability to write] has at long last been solved," this is an important study historically and practically because it helped to establish holistic scoring as an acceptable method for assessing student writing.

Greenberg, K.L., Wiener, H.S & Donovan, R.A. (Eds.). (1986). *Writing assessment: Issues and strategies*. New York: Longman.

A collection of essays that comes from the National Testing Network in Writing, the purpose of which was "to enable teachers, administrators, and researchers to pool resources, exchange ideas, and review data" (p. xii). Essays include A. Lunsford's "The Past—and Future—of Writing Assessment"; S.P. Witte, M. Trachsel, and K. Walters' "Literacy and the Direct Assessment of Writing: A Diachronic Perspective"; and E.M. White's "Pitfalls in the Testing of Writing."

Hanson, F. A. (1993). *Testing testing: Social consequences of the examined life*. Berkeley: University of California Press.

An anthropological study of how testing has come to play such an influential role in American life, with half the book focusing on qualifying tests that measure aptitude and competency. Chapter 7, "The Forest of Pencils," includes a discussion on the history of

written exams as well as the development of the science of testing. Although not focused specifically on writing assessments, Hansen's analysis, which relies on Foucault's theories about discipline, knowledge, and power, provides readers with a larger sociocultural framework for understanding literacy and standardized educational testing.

Huot, B. A. (1990.) The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60, 237-263.

A comprehensive review of the scholarship of direct writing assessment through the 1980s. Huot identifies three focal points in the literature: (a) topic development and task selection, (b) text and writing quality, and (c) influences on rater judgment on writing quality. From the literature review, Huot concludes that research in writing assessment has been neglected, especially by composition scholars with assessment practices outstripping theory.

Huot, B. (2002) *(Re)Articulating writing assessment for teaching and learning*. Logan: Utah State University Press.

Draws on social, historical, and disciplinary factors that have influenced writing assessment. Makes the case for a new configuration of the field that includes integrating work from K-college and emphasizes the role of assessment in teaching and learning. Historical information is woven throughout the book as Huot reviews past trends in classroom-based and large-scale assessment methods and argues for reconceptualizing writing assessment theories and practices.

Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. New York: Farrar, Straus and Giroux.

A fascinating social history about the development of the SAT, the people and personalities who conceived it, and its original purpose. Although the narrative tends to wander at times, it tells the story of how the SAT has come to be so influential in American society. Written for a general audience and not specific to writing assessment but does provide valuable insight into the SAT and the development of assessment as a field.

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), 76-94.

Explains that all types of evaluation are founded in the same technology—eliciting sample behavior from a larger domain of interests; making inferences about a person's probable performance relative to the domain; and classifying, describing, and making decisions by the individual or institution. Argues that in western society, technologies have been blindly accepted as representations of progress, but because technologies are the 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.

Palmer, O. (1960). Sixty years of English testing. *College Board*, 42, 8-14.

Provides an overview of English testing by the College Board through the first 50 years of its history. The emphasis is on the progressive nature of CEEB English testing and its move toward more technically sophisticated measures that helped to establish writing assessment as a scientifically defensible practice. Interesting historically and culturally, as it positions English teachers and direct writing assessment as resistant to the scientific progress achieved in English testing.

Starch, D. & Elliott, E.C. (1912). Reliability of the grading of high-school work in English. *School Review*, 20, 442-457.

Describes an early study that depicts the problems with rater consistency in the marking of high school themes. Not much available here about interrater reliability that hasn't been covered elsewhere, but this is an interesting historical document that describes early efforts at understanding the problems with consistent scoring of student writing.

Thorndike, E.L. (1913). Notes on the significance and use of the Hillegas scale for measuring the quality of English composition. *English Journal*, 12, 551-561.

An early piece that uses principles of scientific measurement to evaluate student work. Examines the difference between high school students' writing performance and those of "recognized masters of English prose" and "measures the amount of error to be expected in grading specimens of English writing" when using a scale. Concludes that "errors will be large," but that "they will diminish with practice" and that with practice the errors will "at least . . . be smaller than the errors now made by teachers in grading paragraph-writing."

Trachsel, M.(1992). *Institutionalizing literacy: The historical role of college entrance examinations*. Carbondale: Southern Illinois University Press.

Traces the development of standardized college entrance exams with a focus on the verbal portion of the SAT. Analyzes the role that entrance exams played in the bifurcation of academic literacy into reading (literary studies) and writing (composition studies) as well as the influence of corporate and bureaucratic influences on the formation of academic disciplines. Although not exclusively focused on the assessment of writing, Chapters 2 and 3 address the role of testing in the appropriation and standardization of literacy.

Valentine, J.A. *The College Board and the school curriculum: A history of the College Board's influence on the substance and standards of American education 1900-1980*. New York: College Entrance Examination Board.

Details from a College Board perspective about developments and changes in testing over its first 80 years. Provides an interesting account of the move from essay testing to multiple-choice tests, including the backlash from English teachers and the response to this backlash by College Board administrators.

White, E.M. (1993). Holistic scoring: Past triumphs and future challenges. In Williamson and Huot (Eds.), 79-108.

White provides a history of battles in which he and others engaged to change the entrenched practice of assessing writing via the identification of correctness within multiple-choice tests. Those propagating the use of holistic scoring in the 1970s were confident in their answers to the many difficulties of writing tests, including problems with flexibility, consistency, and accuracy. In retrospect, holistic scoring has inevitably produced new problems, especially in relationship to portfolio assessment. The greatest triumph of holistic scoring is the control over problems associated with reliability and validity; however, portfolios have yet to successfully address problems associated with fairness and responsibility.

White, E. M. (1994). *Teaching and assessing writing* (2nd ed.). San Francisco: Jossey Bass. Provides an overview of classroom-based and large-scale direct assessment methods, starting with the premise that writing tests are best when they support teaching and learning by involving teachers in test design and producing results that can be used by teachers in the writing classroom. Although only one chapter directly assesses history ("The Politics of Assessment: Past and Future), historical information is included throughout. The text has been considered an important text in writing assessment, especially at the college level.

Williamson, M. M. (1993.) An introduction to holistic scoring: The social, historical, and theoretical context for writing assessment. In Williamson and Huot (Eds.), 1-43.

Provides an historical overview of holistic scoring in order to “raise some basic problems with the validity of holistic scoring that have yet to be addressed in any validation research” (p. 2). Discusses the tensions between assessment practices developed through psychometric assumptions and those that are designed with theories of writing: Williamson shows how the theories and values undergirding particular assessment methods influence the data and interpretation of those data that are collected, explaining how scientific and psychological thought have influenced the way we think about assessment technologies and their validity. Williamson concludes that future writing assessments must necessarily be constructed with theories of writing in mind, for he imagines that such assessments will be more democratic in nature and more representative of and sensitive to the many uses of writing.

Williamson, M. M. & Huot, B.A. (Eds.). (1993). *Validating holistic scoring for writing assessment: Theoretical and empirical foundations*. Cresskill, NJ: Hampton Press.

The focus is not on history, but this collection includes several chapters that provide historical reviews of the theoretical and practical development of holistic scoring. Chapters that are particularly helpful in providing historical perspectives include Michael Williamson’s “An Introduction to Holistic Scoring: The Social, Historical and Theoretical Context for Writing Assessment”; Roberta Camp’s “Changing the Model for Direct Writing Assessment”; and Edward M. White’s “Holistic Scoring: Past Triumphs, Future Challenges” (all separate entries). Taken as a whole, the book provides an overview of the development, uses, and issues of holistic scoring.

Witte, S.P. & Faigley, L. (1983). *Evaluating college writing programs*. Carbondale: Southern Illinois University Press.

Reviews four different writing program evaluations, asking questions and critiquing approaches to build a conceptual framework for understanding the complexities and importance of writing program evaluation. Provides an interesting, historical perspective on particular writing programs and the development of writing program assessment, while at the same time making important contributions to our understanding of theory and practice.

Witte, S. P., Trachsel, M. & Walters, K. (1986). Literacy and the direct assessment of writing: A diachronic approach. In Greenberg, Wiener, and Donovan (Eds.), 13-34.

Shows how literacy and assessment have come to be defined throughout history in order to “help foster an awareness of current assumptions” (p. 14). What constitutes literacy, they show, has been influenced throughout history by both religious and political ideologies that intertwine literacy with moral purpose. They demonstrate that the assessment of literacy has been tied to assumptions about writing as a transparent view of students’ intellectual capabilities. Witte, Trachsel, and Walters conclude that “we need to begin examining the assumptions being made in writing tasks about what it means to be literate” (p. 31).

Yancey, K. B.(1999.) Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication*, 50, 483-503.

Identifies three “overlapping waves” in the history of college large-scale writing assessment—represented by objective tests, holistic scoring of impromptu essays, and portfolios—with each wave feeding into the other but not displacing the earlier one. In discussing the development of these trends, Yancey also locates three significant issues that point to the future: the role of the self in assessment, the role of program assessment, and the knowledge-making potential of assessment.