

Q&A with Norbert Elliot

by David Eubanks



Norbert Elliot, Ph.D., is Professor Emeritus of English at New Jersey Institute of Technology. A specialist in writing assessment, his most recent research focuses on student performance in digital environments as it is understood in cognitive, intrapersonal, intrapersonal, and neurological domains. With Edward M. White and Irvin Peckham, Norbert has recently won the 2015 CPTSC Award for Excellence in Program Assessment for Very Like a Whale: The Assessment of Writing Programs (Logan: Utah State University Press, 2015) for its significant contribution to the field of educational assessment of writing programs, in the area of research. Dr. Elliot can be reached at elliott@njit.edu.

Q: What are some of the most important things we've learned from researching writing assessment?

A: More than anything else, we have learned that writing is among the most complex of human activities and, as such, we must be cautious about drawing conclusions of student ability from a single evaluative episode, no matter how robust. As assessment specialists, we can make only limited inferences about ability based on a narrow slice of performance, and we must be humble about the inferences we make and the score use we advocate.

In this environment of uncertainty, I continue to be astonished at the sheer number of assessments we conduct. A 2014 study from the Center for American Progress found urban high school students spend 266 percent more time taking district-level exams than their suburban counterparts. While we often think that testing initiatives are largely federal, the fact is that state and regional tests bombard our schools—often, with duplicative or unclear aims. On the college level, tests that clearly under-represent the writing construct are used to remediate students at the very beginning of their journeys. The Complete College America project found that more than 50 percent of students entering two-year colleges and nearly 20 percent of those entering four-year universities are placed in basic skills classes that do not carry credit. In this environment, the simple truth that an admitted student should be a qualified student has been lost.

The need for high-quality assessment practices, including empirical analysis that is fair, valid, and reliable, is therefore enormous. Folks know that tests are not telling the whole story—or even a very good one. Talented, hardworking teachers daily encounter intelligent, earnest students. The disjuncture between the two worlds—those of the test and those of the classroom—is expanding, and there is a need for information that makes sense to the many stakeholders of American education.

Q: Can you talk about validity of assessment and what challenges we face there?

A: Historically, validity evidence has been seen as the effort to gather information from a family of analytic methods tied to construct representation and reliability/precision. More recently, validity has been interpreted as the degree to which proposed interpretations and uses of test scores are justified. The work of Samuel J. Messick and Michael T. Kane is of paramount importance here. In terms of integrated evidential systems, Robert J. Mislevy has brought us the system of evidence-centered design. His work inspired the concept of Design for Assessment discussed below. In this landscape, we face two challenges.

The first is absence of what the authors of the 2014 *Standards for Educational and Psychological Testing* term assessment literacy. Networks of assessment stakeholders—students and guardians, teachers and administrators, legislators, and workforce leaders—need good counsel to understand the complex world of assessment. For many, validity remains an up-or-down vote—and nothing remains further from the truth. If I had one wish, it would be that the assessment of the Common Core Standards Initiative for our schools would have been better explained to everyone in a national campaign. So much good would have come from such an educative effort, and so much misunderstanding has resulted from its absence.

(Continued on page 7)

Q&A with Norbert Elliot

Continued from page 6

The second challenge is the need for validity and reliability to be seen in terms of fairness. When the new Trinitarian model of these foundational concepts is framed, it is too often narrowly constrained. Needed is a comprehensive view of fairness that incorporates validity and reliability/precision evidence within an integrative framework. We see the need for new paradigms when we reflect on the seismic demographic shifts occurring across the United States. While enrollment for White students will decrease 6 percent between 2011 and 2022, for example, there will be an increase of 33 percent between 2011 and 2022 for Hispanic students. As the *Projections for Education Statistics to 2020* documents, this increase will also be accompanied by a 20 percent increase for Asian/Pacific Islander students and by a 44 percent increase for students who are two or more races. That which has served a primarily White population of the past will not necessarily serve a heterogeneous population of the future, and new ways of conceptualizing validity in terms of fairness are needed.

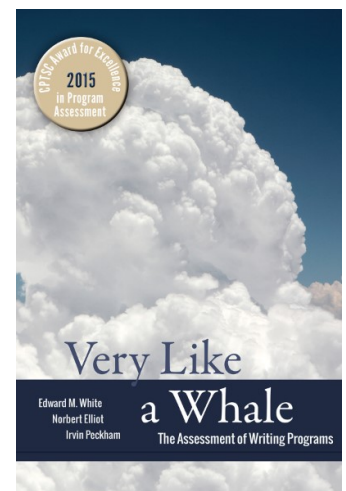
Q: What do you think the role of machine scoring will be, looking into the future?

Machine scoring will play a very important role in writing assessment in the future. While present systems are often concerned with knowledge of conventions and organizational principles associated with the academic essay, future systems will focus on a wide-range of writing genres and the ability to provide real-time feedback to students, teachers, and administrators. Additionally, while present systems have focused on the cognitive dimension, future systems will address sentiment analysis and other intrapersonal domains. Research by Jill Burstein and Beata Beigman Klebanov is already pointing the way to a new era of machine scoring.

As well, we will learn more about how to integrate these automated writing evaluation (AWE) systems into digital learning environments. Web-based course platforms can tell us an enormous amount about student performance and can therefore allow us to help students through rapid reporting. Each click a student makes, each search conducted, reveals a domain of student ability. Using analytic methods associated with Big Data analysis, real-time information can be gathered in order to increase the opportunity to learn. In this area, Joe Moxley at University of South Florida is leading the way with the digital platform My Reviewers. In the world I envision, protest against the first-generation of AWE will diminish as folks recognize that these systems, used for feedback in digital learning environments, will help students obtain the literacies so needed in the global, increasingly diverse world to that is rapidly emerging before us.

Q: You've just come out with a new book with Edward M. White and Irvin Peckham. Can you tell us about what you hoped to accomplish in the book?

A: The book was written as an update to *Evaluating College Writing Programs* by Stephen P. Witte, and Lester Faigley (Carbondale: Southern Illinois University Press, 1983). A classic, that volume addressed the importance of empirical evidence in the evaluation of post-secondary writing programs. In 2015, regional accreditation has become increasingly powerful, and writing program administrators must be able to provide information that documents opportunities for student learning and achievement. As senior researchers (i.e., old men) who have spent their careers in assessment, we wrote a book that continued the rich contributions of writing studies to program evaluation, updating foundational concepts (such as current views of validity) and advancing new pedagogies (such as writing in digital environments).



(Continued on page 8)

Q&A with Norbert Elliot

Continued from page 7

To address this current context, we created a new evaluative model for writing programs: Design for Assessment (DFA). As an accountability framework and form of relational modeling, DFA helps writing program administrators identify the variables that impact the writing program—and to ecologically model the variables to increase student success.

Q: Can you tell us more about the concept of Design for Assessment?

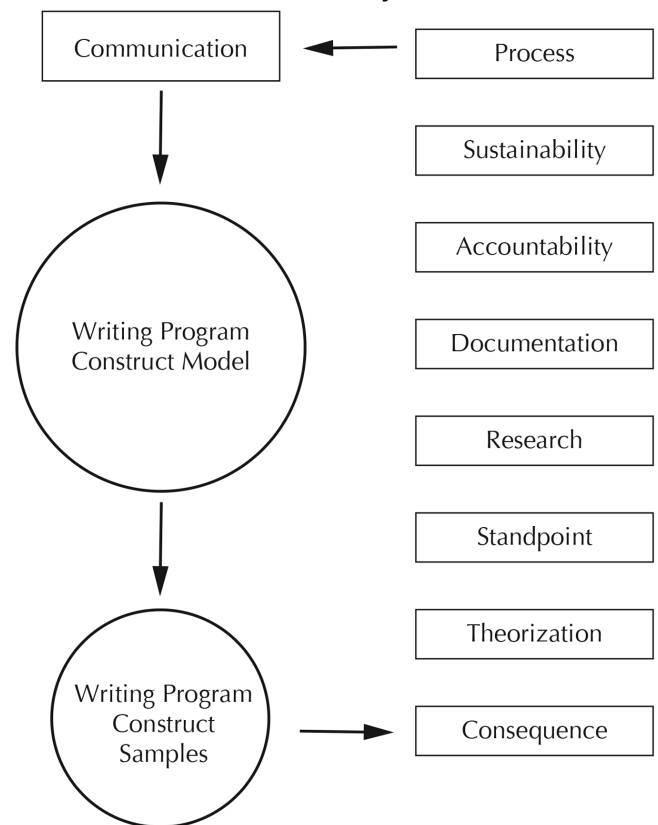
A: The DFA framework advances a component design that, in turn, allows the identification of a variety of sources of evidence. These sources range from anticipation of consequences associated with the assessment to communication with stakeholders about assessment processes, findings, and future directions.

The difference between our model and more traditional ones is that we propose to flip the idea of assessment so that, in advance of the assessment, the entire spectrum of the assessment is planned in exacting detail. As is the case with evidence-centered design, this technique allows a practical approach to program assessment that focuses on planning and accountability from the earliest stages.

As well, the component design allows a principled framework that focuses first on consequences. Why, we ask, are we undertaking this assessment, and what are the consequences associated with it? This concentration on impact brings forward the concept of fairness. Often associated with as justifiable score use for population subgroups and individuals within them, DFA advances fairness as associated with the opportunity to learn. As such, DFA allows us to anticipate which student groups may be disadvantaged by the assessment and to identify the reasons for that possible disadvantage. DFA thus allows administrators and instructors to plan—again, in advance—the kinds of resource allocation necessary to leverage success of all students.

Because the model advocates programmatic approaches, its focus is on the design, development, and assessment of writing programs—not solely on distinct courses and isolated assessment episodes. Students learn in all kinds of ways in all kinds of places, and the DFA approach is designed to encourage writing across the curriculum and writing in the disciplines.

If DFA is the main idea of the book, its central ethos is to ensure that stakeholder networks—from students to workforce leaders, from parents to classroom instructors—have an important voice in instructional and assessment efforts. From advisory boards to students themselves, each group is a rich resource and should be included in considerations ranging from the comprehensiveness of the curriculum to the reporting of assessment outcomes. In this emphasis on multiple voices, our emphasis remains squarely on fairness as the integrative foundational principle of assessment.



Design for Assessment framework for the assessment of writing programs. (From Very Like a Whale, © 2015 University Press of Colorado. Reprinted by permission.)

Further Reading

Assessing Empathy through Historical Role-Playing Games *from page 2*

Carnes, Mark C. *Minds on Fire: How Role-Immersion Games Transform College*. Cambridge, Mass.: Harvard University Press, 2014.

Konrath, Sarah H, Edward H. O'Brien, and Courtney Hsing. "Changes in Dispositional Empathy in American College Students over Time: A Meta-Analysis." *Personality and Social Psychology Review* 15, no. 2 (2011): 180-98.

Lang, James M. "Stop Blaming Students for Your Listless Classroom." *Chronicle of Higher Education*. September 29, 2014.

Lidinsky, April. "'Reacting to the Past' to be Proactive in the Present: Feminist Roots of High-Impact Practices." *Feminist Teacher* 24, no. 3 (2014): 208-12.

Lightcap, Tracy. "Creating Political Order: Maintaining Student Engagement through *Reacting to the Past*." *PS: Political Science and Politics* 42, no. 1 (2009): 175-79.

Olwell, Russell and Azibo Stevens. "'I had to double check my thoughts: How the Reacting to the Past Methodology Impacts First-Year College Student Engagement, Retention, and Historical Thinking.'" *History Teacher* 48, no. 3 (2015): 561-72.

Zembylas, Michalinos. "Pedagogies of Strategic Empathy: Navigating through the Emotional Complexities of Anti-racism in Higher Education." *Teaching in Higher Education* 17, no. 2 (2012): 113-125.

Shaping the Campus Conversation, Part One *from pages 3-4*

References about implicit and explicit categorizations of student work, from page 3:

P. Maki (2010). *Assessing for Learning: Building a sustainable commitment across the institution*. 2nd edition. Stylus Publishing, Sterling VA. Pp 219-253.

L. Suskie (2009). *Assessing Student Learning: A common sense guide*. 2nd edition. Jossey-Bass Publishers, San Francisco. Chapter 9.

Using a rubric to produce both grades and assessment data, from the University of Virginia's Office of Institutional Assessment and Studies (<http://avillage.web.virginia.edu/iaas/assess/tools/assessment-graphic.pdf>).

M. J. Goggins Selke (2013). *Rubric Assessment Goes to College: Objective comprehensive evaluation of student work*. Rowan & Littlefield, Lanham MD.

B. E. Walvoord & V. J. Anderson (2009). *Effective Grading: A Tool for Learning and Assessment*. 2nd edition. Jossey-Bass, San Francisco.

B. Lovitts (2007). *Making the Implicit Explicit: Creating performance expectations for the dissertation*. Stylus Publishing, Sterling VA.

References about academic freedom, from page 4:

AAUP 1940 statement & 1970 interpretive comments (<http://www.aup.org/report/1940-statement-principles-academic-freedom-and-tenure>).

C. Nelson (Dec 2010), *Defining academic freedom* (http://www.insidehighered.com/views/2010/12/21/nelson_on_academic_freedom).

M.W. Finkin & R.C. Post (2009). *For the common good: Principles of American academic freedom*. Yale University Press.