AALHE
Association for the Assessment of Learning in Higher Education

Intersection

October 2015

CONTENTS

Page 1
From the Editor

Page 2
Assessing Empathy through Historical Role-Playing Games

Pages 3-4
Shaping the Campus Conversation, Part One

Page 5
Teaching to the Test: Is it Really so Bad?

Pages 6-8
Q&A with Norbert Elliot

Page 9
Further Reading

From the Editor

Welcome to the fall edition of the Intersection. This issue marks something of a turning point for the publication in that the editorial board has a more distributed work flow for finding new authors and content, vetting, editing, and preparing these pages for publication. The idea has been to turn the AALHE newsletter—where we began—into a professional but informal place for the exchange of ideas between assessment practitioners. The variety of activities and scope of intellectual endeavor that assessment encompasses continues to amaze me. In this issue we touch on the assessment of empathy, the importance of matching message and audience, the authenticity of testing and its implications, and an interview with a scholar of writing assessment that touches on many topics in three short pages. And more, of course. If advertisers have taught us anything, it is that there is always more.

As a theme to weave these various topics together I offer an insight from working on inter-rater agreement calculations. There are, of course, many of those statistics, with arcane names like "Krippendorff's Alpha," which could be straight out of *Harry Potter,* but they all calculate agreement between raters in one way or another. When we have agreement about the meaning of evidence, it creates what we might call a provisional reality. We can read in *Nature* this month that the definition of a kilogram finally has agreement, due to the convergence of three different physical approaches. In assessment we do not have that level of precision, but the idea is the same.

Socially constructed reality is alive and well in our everyday work with assessment. If faculty members don't agree on what assessment results mean (or worse: agree that the results are meaningless) our efforts may be wasted. We can hardly go wrong then, if we begin with that idea in mind and work backwards. This attitude induces a certain humility on the part of assessment leaders. We don't have all the answers because the answers—and the questions—have to be socially constructed. You will see variations of this idea in the articles that follow.

Between now and the January edition, there are other ways to engage with AALHE. Mark you calendars for these free webinars; more information will arrive in your inbox.

Friday, October 16, 2015, 2:00-3:30pm Central Time
Faculty Perspectives on Cultures of Assessment
Matthew Fuller, Sam Houston State University
Register: https://attendee.gotowebinar.com/register/3180659909157253377

Friday, December 11, 2015 11:00am Central Time
Easy Predictors: Quickly Explore Your Data to Find Trends, Relationships, and Predictors
David Eubanks, Furman University
Register: https://attendee.gotowebinar.com/register/669880027553370369

Intersection is the quarterly publication of the Association for Assessment of Learning in Higher Education
Editorial board: David Eubanks (editor), Fashaad Crawford, David Gliem, Lisa Mehlig, David Onder
1080 Export St, Suite 180, Lexington, KY, 40504 | 859-388-0855 | info@aalhe.org
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Assessing Empathy through Historical Role-Playing Games

Tamara Venit Shelton, Jonathan Hanna, and Daniela Blei

None of us imagined that historical role-playing games would transform our approach to teaching. We were not players of Dungeons and Dragons or World of Warcraft. Yet, as historians, we have joined hundreds of faculty in using role-playing games to revitalize the classroom. The most organized and widely known historical role-playing curriculum is Reacting to the Past (RTTP), developed by Barnard College professor Mark Carnes. With ten published games and more in the pipeline, RTTP has been adopted at some 400 colleges and universities in the U.S. and internationally. In RTTP, students immerse themselves in major moments from the past, as Kentucky state legislators contemplating secession in the spring of 1861, or as Athenians rebuilding a democratic society in 403 B.C. Primary sources inform their actions. Much of the analysis of RTTP has focused on its potential to enhance student engagement. See for example, Lightcap, 2009; Olwell and Stevens, 2015; and James Lang's 2014 series in *The Chronicle of Higher Education* (see list of references on page 9). In our own classrooms, where we have taught U.S. and European history, role-playing games have done much more. Our experience is that historical role-playing games teach empathy.

Once considered a personality trait, we now know that empathy is an important social and cognitive skill that can be taught and learned. Social scientists have suggested that empathetic students make better communicators, citizens, and leaders, but researchers at both the University of Michigan and the University of Rochester have found millennials to be less empathetic than previous generations, pointing to a "crisis of empathy" across college campuses. With its emphasis on interpreting past lives, history is well positioned to teach empathy. Students practice the flexibility, imagination, and compassion that historical thinking requires. When RTTP commissioned a study of student learning, the data showed a strong positive correlation between its role-immersion game curriculum and empathy. On the Balanced Emotional Empathy Scale (BEES), first-year seminar students who completed a RTTP class scored higher than a control group. April Lidinsky, a professor of Women and Gender Studies, observed students in her RTTP class engaging in "playful intellectual risk-taking" that led to their involvement with campus feminist groups and other forms of political activism.

We see in our classes how imagining the past through the eyes of others allows students to be daring; to express opinions and new ideas and, most critically, to adopt and act on a worldview entirely not their own. In an environmental history class, one student felt that a game on global climate change compelled her to "listen to [my opponents] and try to find a middle ground… You had to understand how they thought, not just what they thought." Cultivating empathy in the classroom is not always comfortable or easy, for students and instructors. We may naturally empathize with individuals we admire, but what about others we find reprehensible or repugnant? In a class on race and ethnicity in U.S. history, one of our students balked at deploying racist logic to create a character who was, well, racist. His personal ethics were guided by tolerance and egalitarianism. "I don't want to empathize with a racist," he declared. Yet coming to a deep understanding of racism as an ideology with all its uncomfortable and tragic implications was the key to unlocking the contradictions of American political and social history. Michalinos Zembylas, an education professor at the Open University of Cyprus, recommends that instructors invite emotional response and coach students through such difficult moments with the goal of achieving what Zembylas calls "the reconciliatory perspective of empathy," or "shared reflective engagement with the other's emotional life."

RTTP is an innovative approach to student assessment that cultivates both intellectual and emotional engagement. Empathy matters, science has shown, and teaching empathy requires expanding our imaginations and our arsenal of assessment tools. For further reading, see the references on page 9.

Tamara Venit-Shelton is Associate Professor of History at Claremont McKenna College and can be reached at tvenit@cmc.edu; Daniela Blei is an editor and independent Scholar in San Francisco and can be reached at daniela.h.blei@gmail.com; Jonathan Hanna, C. Phil., History, Claremont Graduate University can be reached at jonathan.hanna@cgu.edu .

Shaping the Campus Conversation, Part One

Ephraim Schechter

At a campus consult a few years ago, the institution's president introduced my opening presentation to the faculty by saying "We all know why we're here – the accreditors are coming!" Think about the message that sent to the faculty and staff – "Assessment isn't for our own purposes, it's for somebody else."

The president wanted to generate faculty buy-in. But what he told them was "this is a bureaucratic task that you have to do whether you want to or not." That's not what he wanted to accomplish, especially since many of his faculty already assumed that assessment was something imposed from outside that didn't have any real implications for their day-to-day work as faculty members.

There's a lesson here about clear communication. Shaping your campus conversation involves paying attention to who you're talking with, what you're trying to say to them, and how what they already think will affect how they interpret what you're saying.

This two-part article reviews some assumptions about assessment and the (mis)communications they can lead to, and suggests some things to think about for talking with various audiences – faculty, administrators, boards, accreditors, students, and the public.

Assessment serves two purposes: (a) gathering information for planning and (b) accountability reporting. It's common for people to emphasize one purpose and de-emphasize the other when they think about assessment. Or even to think that you can't have both, that you have to treat assessment as either for planning or for accountability. (I'll come back to this point in Part 2. First, let's take the two purposes separately and look at some common assumptions.)

Start with the idea that assessment is primarily for gathering information to help us plan programs, or courses, or even individual assignments. For planning, we tend to want "rich, raw details" – lots of information about what's going on and why. And by "we" here, I mean us – the people doing the planning and implementing the plans. In other words, assessment is for our own purposes.

You may hear some people say that what we're really trying to accomplish can't be measured. Assessment guru Peter Ewell calls this the "ineffability trap" – "I know good work when I see it but I can't tell you exactly what it is." But if I know it when I see it, I can start digging into how I know it, what I'm looking for.

Think about sorting students' work into piles, for grading. "This paper's pretty good, it goes in the B pile; this one's really good, it goes with the As" and so on. To do this we must have made implicit decisions about (a) what we want the students to know and be able to do, and (b) what defines the levels for each of the things we want them to be able to know/do. In other words, there's an implicit rubric, a categorization scheme. Notice that explicit "outcomes-based grading" isn't something new; it's more a matter of paying closer attention to the details of what we already do. It can give us more precise information for planning by (a) focusing on one outcome at a time and (b) clarifying the evaluation process. Benefits include (a) consistency, and (b) making it easier to explain what we're doing, to students, other faculty, or any other audience (which helps the students understand better what we're looking for and what their grades mean).

This approach can seem simple and rough. What about reliability and validity and statistical significance and double-blind experimental designs with control groups? Some argue that "only full-fledged random-assignment tightly-controlled experimental designs can produce results I'd believe, and then only if they're statistically significant. We don't have time/money/opportunity to do that kind of research, so let's not do anything." But the real choice isn't "full experimental design or nothing," it's "full experimental design or the kind of casual ad-hoc process we use now." So tightening the process up even a bit really is a useful

(Continued on page 4)

Shaping the Campus Conversation, Part One

Continued from page 3

improvement. Outcomes assessment in higher education is usually pretty pragmatic, and not designed for publication in your discipline's primary journals. That's ok. Decisions are going to be made one way or another. Assessment is a way to feed information into the process.

Particularly when all this is different from how your faculty see what they already do, you may run into the assumption that assessment violates academic freedom. Not according to AAUP's definition of academic freedom. Finkin & Post (2009), examining AAUP policy statements and the 100 years of associated "case law," conclude that "Academic freedom is not the freedom to speak or to teach just as one wishes. It is the freedom to pursue the scholarly profession, inside and outside the classroom, according to the norms and standards of that profession" (p 149). The norms and standards include local and discipline views of curriculum and pedagogy. Defining learning outcomes, assessing student performance in the outcome areas, and using the results to inform planning seems well within the definitions of "professional norms & standards."

Even given all that, remember that I started with an example of how easy it is to communicate that assessment really isn't about gathering information for planning. We've already looked at "we do it for the accreditors" (or for somebody else who isn't us). Here are some more examples that communicate, maybe unintentionally, that assessment isn't really part of our own planning process.

When a program's assessment report is sent up the line, to an Assessment Committee, for example, it too often disappears into the bureaucratic "black hole." There's no sense that it served any real purpose except reporting *per se*. This doesn't encourage the departments to invest much in the process. The message is that it's just a report, not part of our real planning and budgeting process.

To send the message that assessment really is important and "part of the way we do things at our institution," administrators need to show that they want and use the results, by continually asking "How do you know? What's the evidence behind that decision? Behind that request?" – and by telling the departments how they used that information in their decisions. Without that feedback, the message is "this isn't really for us (the department, the college, the university). We do it because somebody else wants it."

Another way to confuse the connection between assessment and program planning is to say that assessment is all about change, that the important thing is the change the assessment results led to. This one is subtle. Yes, planning can involve making changes. So can "improvement" (as in "planning & improvement"). Asking only about changes is too easily misinterpreted as indicating that the purpose of assessment is to "fix what's wrong." But the real question is "what did you do as a result of what you found out?", and "we didn't change that because the data say we shouldn't" is a legitimate response. It's important to look at both the things that need attention and the things that are ok.

This part of the article dealt with assumptions you may encounter about "assessment for planning." Part Two, in the January edition, will look at some assumptions about "assessment for accountability," and why assessment for planning and assessment for accountability aren't necessarily incompatible. For further reading see the references on page 9.

Ephraim Schechter is a founding board member of AALHE and well-known consultant in assessment. He can be reached at eischech@higheredassessment.com.
Photo credit: Jacquelyn Schechter, http://www.mindfulphotography.com/home.

Want to Contribute?

We invite your ideas and expertise for the *Intersection* as well as *Emerging Dialogues*, a new AALHE web site for short articles and comments. Email ideas or short pieces to intersection@aahle.org or for Emerging Dialogues to jdowns2@delmar.edu or ddirlam@vwc.edu.

Teaching to the Test: Is it Really so Bad?

Catherine Wehlburg

In discussions with many people (both inside and outside academia), I have found that the phrase "teaching to the test" is seen by most as a negative statement. I've heard those who scoff at the concept and who see this as a cop-out in teaching. "She isn't really teaching them what they need to know, she's just teaching to the test." But let's think about this phrase for a moment. Is it really such a bad thing to strive to do in education?

Of course, the answer to that question depends on the quality of the test. If the test is a high stakes test that doesn't really focus on things that many think are important in learning, then teaching to the test is, indeed, a bad thing – perhaps even something that is unethical to do in education. But, if the test is an accurate representation of what students need to learn—like, say, learning to land an airplane—then teaching to the test is exactly what we should try to do.

The list of the purposes of testing is quite long – some are for accountability, some focused on measuring individual student learning, and some with an eye towards standardization and comparisons. These include providing feedback about student learning, providing feedback to students about their own learning, providing feedback to a faculty member about her/his teaching, providing a mechanism for assigning grades, and can be used to demonstrate the attainment of competencies

If the test is well-written and reasonably valid and the scoring is meaningful with little measurement error, it can provide a great deal of information to a department (or individual faculty member) about what students actually know and what they may not yet understand. The key to making this work is to ensure that the "test" or the measure of student learning is actually reflective of what students should be learning.

James Popham (2001) stated "if a teacher directs instruction toward the body of knowledge or skills that a test represents, we applaud that teacher's efforts. This kind of instruction teaches to the knowledge or skills represented by a test. But if a teacher uses the actual test items in classroom activities or uses items similar to the test items, the teacher is engaging in a very different kind of teaching" (see http://goo.gl/ZOj9wN). Having a teacher (or an overall educational program) focus on what Popham calls "curriculum teaching" means that faculty focus their teaching on a "specific body of content knowledge or a specific set of cognitive skills." Thus, the test can represent the overall body of knowledge and the faculty member teaches from that overall content area.

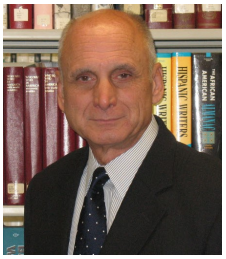
In order to do this well, faculty must be able to construct and to use clearly stated (and measurable) student learning outcomes. Assessment offices across higher education have been working toward this for decades. Some faculty and programs are understanding this and view the creation of learning outcomes as a path to improve learning, teaching, and educational attainment. Some, however, are still seeing this process as limiting and constraining. These faculty haven't yet seen that we can measure learning – even messy, difficult to see learning. This is our continued task in our assessment offices – engaging faculty with the processes of good assessment practices so that faculty have the information and the evidence about real student learning. (And, of course, we can then use that information for accountability – but this should be the secondary reason for assessment).

The purpose of education is not to prepare students to take tests, as we all know. But methods for measuring student learning (including testing) are part of discovering what students have learned and is crucial to improving and enhancing higher education. We have to carefully determine what "the test" is and how to teach toward it.

Catherine Wehlburg is Associate Provost for Institutional Effectiveness at Texas Christian University, and President-Elect of AALHE. She can be reached at c.wehlburg@tcu.edu.

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* elliot@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

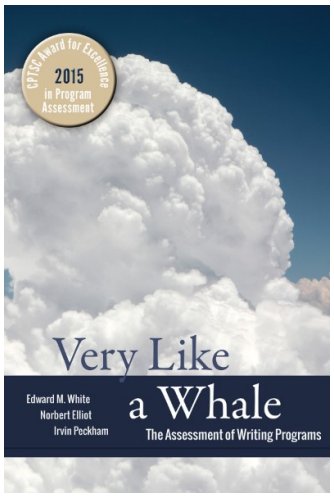
The second challenge is the need for validity and reliability to be seen in terms of fairness. When the new Trinitarian model of these foundational concepts is framed, it is too often narrowly constrained. Needed is a comprehensive view of fairness that incorporates validity and reliability/precision 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?

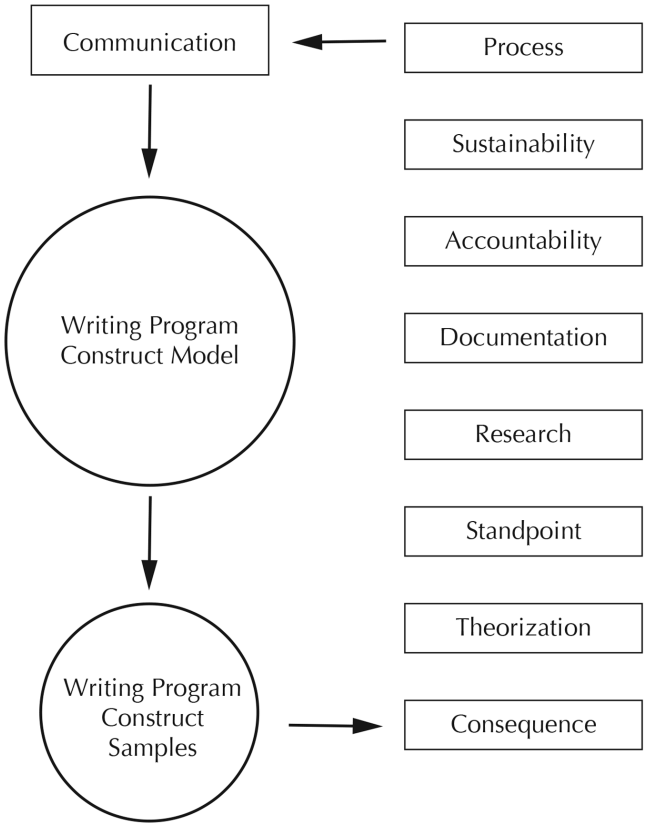
A: The DFA framework advances a component design that, in turn, allows the identification of a variety of sources of evidence. These sources range from anticipation of consequences associated with the assessment to communication with stakeholders about assessment processes, findings, and future directions.

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.aaup.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.