

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Journal of the National Collegiate Honors Council -
-Online Archive

National Collegiate Honors Council

2010

Learning Outcomes Assessment in Honors: An Appropriate Practice?

Scott Carnicom

Middle Tennessee State University, carnicom@mtsu.edu

Christopher A. Snyder

Marymount University

Follow this and additional works at: <http://digitalcommons.unl.edu/nhcjournal>



Part of the [Higher Education Administration Commons](#)

Carnicom, Scott and Snyder, Christopher A., "Learning Outcomes Assessment in Honors: An Appropriate Practice?" (2010). *Journal of the National Collegiate Honors Council --Online Archive*. 279.

<http://digitalcommons.unl.edu/nhcjournal/279>

This Article is brought to you for free and open access by the National Collegiate Honors Council at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Journal of the National Collegiate Honors Council --Online Archive by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Learning Outcomes Assessment in Honors: An Appropriate Practice?

SCOTT CARNICOM

MIDDLE TENNESSEE STATE UNIVERSITY

CHRISTOPHER A. SNYDER

MARYMOUNT UNIVERSITY

In its ideal form, systematic assessment is a legitimate way for honors programs and colleges to gauge strengths and weaknesses, measure the effect of various learning environments, and evoke positive institutional change based on objective, empirical data. Such assessment can take two main forms. *Programmatic assessment* (also known as program evaluation) is an extremely useful tool for gathering evidence and evaluating whether an honors program embodies the NCHC's basic characteristics (Sederberg 159) and/or meets its own institutional goals, e.g., higher rates of retention, graduation, graduate/professional school acceptance, and successful competition for national fellowships. Furthermore, as Otero and Spurrier argue (5), this often-required process can offer honors programs a way to improve, tailor the assessment mechanism, demonstrate program strengths, and garner financial support. Like it or not, in these competitive times programmatic assessment has become a part of American higher education, and honors programs or colleges that do not engage in it, or at least shape it to their own purpose and design, risk alienating accountability-driven entities on and off campus.

While honors programs are certainly not immune to such self-interested concerns, our true bottom line is providing students with an enriched education that cultivates learning at the very highest scholarly levels. To this end, the second main type of assessment, *learning outcomes assessment*, attempts to measure what college students learn as a result of participation in honors and also to distinguish the unique characteristics of an honors education. This essay will focus on the second type, highlighting some limitations to the assessment of learning in honors. First, we will examine limitations in the methodology and logic of learning assessment from a behavioral science perspective, raising concerns about what we are truly measuring and how we are evaluating, interpreting, and applying this information. Second, we will raise

important professional concerns about the necessity of learning assessment and the impact, if any, it has on the basic tenet of academic freedom.

GENERAL LIMITATIONS TO THE MEASUREMENT OF LEARNING

Human learning is a complex set of intertwined neural processes; it is a vast, adaptive, higher-order cognitive mechanism that consists of numerous levels of serial and parallel information encoding, processing, consolidation, retention, manipulation, and recall. For over a hundred years, behavioral scientists from Ebbinghaus to Skinner to Kandel have dealt with the immensity and ambiguity of learning by dissecting it and measuring discrete and often simple behaviors potentially dependent on learning, e.g., syllable recall, key pecks, and gill withdrawal. Scientists are reductionists. Breaking extremely complicated phenomena such as learning into smaller parts allows for easier, systematic study. It is understood, though, that such limited measurements only hint at the tip of a very big iceberg. This focus on small details in order to approach large, complicated topics partially explains why learning has been of such great scientific interest for well over a hundred years to countless researchers across a wide variety of fields and disciplines. Learning possesses numerous, intricate layers and can't be limited to any one single measure.

Furthermore, by focusing on publicly observable behaviors, behavioral scientists can empirically measure concrete data. Indeed, the classic behaviorist definition of learning is a relatively permanent change in behavior. Learning is an invisible psychological construct that can only be measured through related behavior. Unfortunately, we can't pop the hood and directly observe the mental process of learning. Even when using advanced visualization technology such as magnetic resonance imaging, what we are really measuring is regional blood flow, another related behavior. A behavioral measurement is a stand-in for some underlying cognitive process and may only encapsulate a small portion of this process.

We can only indirectly witness learning via behavior. The trick is using our professional judgment and logic to pick the best set of behaviors to measure, which raises the issue of measurement validity. Validity, in its psychometric sense, describes how accurately a tool measures what it purports to measure (Moss et al. 112). For example, if you attempt to measure a person's height by taking his temperature with a thermometer, you will have a measurement very low in validity. Alternatively, if you use a bathroom scale to measure height, your validity will go up slightly because taller people *tend* to weigh more than shorter people. Validity is a matter of degree ranging from 0 to 100% accuracy. Using a bathroom scale to measure height isn't perfect, but it will get you closer to the ballpark than a thermometer.

Invoking another example, an IQ test may tap into some aspects of intelligence and therefore be partially valid, but it is not a 100% accurate index of intellectual prowess and potential, unaffected by say, socioeconomic status. In other words, validity is the extent to which the behavior we have chosen correlates with the underlying construct, but correlation does not imply causation. Thus, any simple measurement of learning is unlikely to be 100% complete or valid because all tests of learning have a built-in level of error resulting in less-than-perfect accuracy.

Thus, the reductionist/behaviorist approach is a tradeoff, providing greater internal validity but less generalizability. Behavioral scientists gain insight and objectivity by dissecting learning into smaller parts and studying related behaviors, but they knowingly lose focus on the bigger picture. In other words, while single scientific measures of learning are useful and interesting, by definition they provide a limited and potentially inaccurate view of overall learning. This is not a criticism of behavioral science but rather a logical premise and limitation of experimental methodology that is often overlooked when standardized tests or rubrics are developed as expedient measures of collegiate learning. According to Gardner, the developmental psychologist who has advocated for multiple intelligences, there is

. . . a bias towards focusing on those human abilities or approaches that are readily testable. If it can't be tested, it sometimes seems, it is not worth paying attention to. My feeling is that assessment can be much broader, much more humane than it is now and that psychologists should spend less time ranking people and more time trying to help them. (23)

Accurately summarizing learning—especially the breadth and depth of learning that occurs across four years in a collegiate honors program—in a few simple quantitative measures is a difficult task to say the least and one that we can and probably should avoid.

LIMITATIONS TO STANDARDIZED ASSESSMENT

Given the basic limitations of all learning measurements noted above, we should cautiously and skeptically view any single or standardized assessment of “collegiate learning” that doesn’t include, at bare minimum, a wide variety of observations across several years. However, such standardized tests not only exist but have been promoted by officials within both the Bush and Obama administrations (Spellings Commission Report; “Assessment disconnect”), and are voluntarily used by some institutions to reveal “learning,” e.g., the Measure of Academic Proficiency and Progress or MAPP (renamed the ETS Proficiency Profile in December 2009). The basic rules of behavioral

science tend to be forgotten as the job of assessment is foisted on beleaguered institutions and unprepared faculty (most of whom are not trained social scientists) trying to satisfy the demands of external agencies questionably seeking transparency in the collegiate learning process.

While a standardized measurement may satisfy these pressures, it oversimplifies learning in a one-size-fits-all package. Another way to look at this problem, if we may borrow from Long, is that if you condense or distill learning down to something very simple, all you are going to measure is simple learning, and this is what such standardized forms of learning assessment tend to do by definition and design. Tests of this nature tend to suffer from construct under-representation; they are too narrow, failing to fully encapsulate all facets of learning, especially the higher forms of learning aspired to by honors faculty and their students. Nonetheless, some universities require pending graduates to complete such tests assessing students' knowledge within their subject area and also within the general education core. Honors students are expected to score higher on both tests than their non-honors colleagues, but such low, quantitative standards seem especially antithetical to honors education.

In addition to representing merely basic levels of learning, such tests can become a political football. Which knowledge and skills are important? What is important to memorize and what isn't? Who sets the agenda? To what extent are faculty members involved in the decision? Again, any simple test is extremely limited in its ability to assess overall learning at a wide variety of institutions with a high degree of validity. Interestingly, though, the degree of validity of such tests (which is often not known by those using them) is usually estimated via comparison to a real world measurement of learning, namely students' grades, a measure that external agencies don't accept (Educational Testing Service, "Validity . . .").

Not only do tests of this type tend to measure merely factual knowledge (as compared to understanding, reasoning, or creative ability), but they do so in a manner that lacks meaning to the student. Because these tests are usually given at the end of a collegiate career and usually have no impact on grade point average, students have little extrinsic incentive to perform well on the test and little or no opportunity to learn from it. Low scores on such standardized forms of assessment might indicate poor learning or the frustration of seniors who have to pay their parking tickets, not to mention library fines, and take two more tests before they can get their diploma in the mail. Thus, in addition to not fully gauging learning, such measures are subject to confounding variables that are unrelated to learning but can greatly affect so-called "learning" scores.

Additionally, many of the standardized assessments that seemingly possess high face validity are given at the end of a collegiate career without the benefit of a comparison pre-test. Such post-tests might help institutions or external agencies categorize or sort students based on preexisting abilities but provide little or no glimpse of what was acquired during college. In this scenario, one can assess whether students are meeting some limited, static, absolute standard but learn nothing about the growth or change in the students across their academic career or beyond. Furthermore, this less than ideal practice raises concerns regarding temporal validity; even if a single test could demonstrate the complicated array of student learning from start to finish with 100% accuracy, it could make no claims of validity five or ten or fifty years from now. In other words, such a measure cannot gauge continued reflection, intellectual growth, and curiosity across the lifespan of students, nor can an assessment snapshot taken at the end of a collegiate career accurately predict the effect of a college education at the age of twenty-two, much less seventy-two.

The goal of a good education is a lifetime of continuous learning and critical thinking. Education isn't just about the memorization of narrow, immediate, transitory knowledge and vocational skills; it is a habit of the mind that keeps us intellectually nimble, offers a way to adapt to an ever-changing environment, and prepares us for a vocational landscape that doesn't currently exist. Education, especially the form to which we aspire within the honors community, should obviously go far beyond basic competency and literacy. While standardized tests might appease external agencies and ease the pain of their assessment mandates, this streamlining dilutes and belittles the significance of learning.

LIMITATIONS TO OTHER FORMS OF ASSESSMENT

Our objections to simplified or standardized forms of learning outcomes assessment are not new and should come as no surprise to those of us in the honors community. Indeed, some in the honors community have opted for the use of more holistic, qualitative, customized forms of assessment such as learning portfolios and theses (Zubizarreta). Such autonomously designed forms of evaluation may provide a viable alternative to standardized assessment, yielding more meaningful data (to professors and students) and allowing faculty members to rethink and reformulate their teaching approach. By integrating and examining multiple forms of evidence, as in a learning portfolio, we provide a more robust picture of higher forms of learning. From a behavioral science standpoint, multiple measures of learning (especially across time, disciplines, and observers) greatly increases reliability and

allows for a more complete and potentially more accurate prediction of learning.

But, with or without portfolios, assessment is still often objectionable to some professors outside (and within) the social sciences. Many humanists have concerns about such nascent assessment techniques, arguing that Homer's epic verse does not lend itself to such measurement, that no one should subject Shakespeare to a rubric, that the purpose of studying philosophy and theology at the highest levels is not to generate "learning outcomes," and that a student's interpretation and understanding of such work cannot be condensed into a simple measurement. Assessment ideas are alien to many professionals in the humanities, who may view the human condition as Hamlet did (or parodied):

What [a] piece of work is a man, how noble in reason, how infinite in faculties, in form and moving, how express and admirable in action, how like an angel in apprehension, how like a god! the beauty of the world; the paragon of animals . . .
(II.2.303–307)

Wrong we may be (and Hamlet certainly had his doubts), but to force the creative arts, the humanities, and perhaps even the natural sciences into a social science paradigm is to privilege one view in the university and do disservice to the others. Even customized assessment techniques, which admirably attempt to collect more direct evidence of learning, fail to capture the essence of learning in some fields, rendering them no better if not worse than traditional grading practices. Objective assessment rubrics that define student deliverables may work very well in some situations or disciplines (and faculty members should be free to try them in an effort to improve instruction and optimize learning), but whether they are better than traditional techniques is highly questionable.

Critics might argue that no one is forcing a unitary form of assessment upon faculty. Rather, accrediting agencies are encouraging faculty to tailor or create more objective assessment techniques that yield more visible evidence of student learning. Such an approach, even if not mandated, has pitfalls, though, because higher forms of learning yield less overtly and immediately to demonstration and are thus assessed less frequently or accurately. Therefore, as Gardner suggests (23), such accountability practices potentially could decrease standards over time, insidiously lowering the bar of expectations we have for our graduates to basic skills and knowledge. In other words, by gradually simplifying our rubrics or measurements to better capture "learning," we ignore higher and more elusive forms of learning vital to scholarship. It's a case of the tail (or test) wagging the dog.

Despite these methodological concerns, large portions of the academic world seem to have accepted the inevitability of assessment. Institutions are scrambling to deal with these new assessment demands, forming committees, hiring consultants and gurus, creating new administrative posts, and spending countless dollars on this thorny issue. An obvious question is: why reinvent the wheel? Across eight semesters, an honors student will take approximately forty different courses with forty different grades comprising their grade point average, thus providing forty opportunities for the collective wisdom of a variety of different professors to reliably measure the many faces of the behemoth that we call learning. When viewed in this light, assessment data are redundant at best and limited in their usefulness.

However, instead of using grade point averages, accrediting agencies seeking increased transparency in the evaluation of student learning demand “objective” student learning outcome data. They want to see tangible evidence of learning demonstrated directly by the student. Why is this necessary? For many, many years, teachers at all levels have engaged in assessment of student work, codified with grades. Yet, accrediting agencies do not accept course grades as evidence of learning even though they are used as a yardstick to assess the validity of standardized tests like the MAPP. Instead, they want institutions to demonstrate that the course grades given by professors are based on objective tests. Do the accrediting agencies go too far? Is the demand for objective tests an infringement on academic freedom or a reasonable request given grade inflation and the proliferation of dubious for-profit universities?

The use of grades as an assessment tool is certainly not without controversy. Some in the pro-assessment camp imply that professors are unable to grade objectively because of a built-in conflict of interest since student learning is linked to evaluation of teaching effectiveness. This is an interesting argument because it too touches on a basic tenet of behavioral methodology: experimenter bias. The behavior and evaluation of subjects (or, in this case, students) can be greatly affected by the bias or expectations of the experimenter (in this case, professor). Thus, one must strive to remove this potentially confounding variable (and the temptation to inflate grades) when measuring learning. However, linking teacher effectiveness to assessment scores of student learning firmly places the professor in this quandary. Traditionally, professors used to be third-party guides, providing students with learning opportunities and feedback along the way. By directly linking assessment and evaluation of teaching, you make the instructor partial and, at worst, create a system where it is advantageous to teach to some limited test or rubric customized by the teacher to “objectively” demonstrate learning (or a simplified version thereof). In this case, the problem of teacher bias is created by the

requirement of assessment. Grades may not be the optimal way to assess learning, but neither are assessment techniques.

Furthermore, the task of assessment distracts faculty from their primary scholarly duties: teaching and research. Ironically, this kind of distraction runs counter to the business-like demands for increased faculty accountability and productivity. If faculty are constantly engaged in assessment exercises (or even, as here, pointing out flaws with assessment), they are by definition spending less time preparing for class and doing research within their specific field.

Finally, even if assessment data of student learning outcomes were completely valid, comprehensive, and unique, it still would not provide specific data about teaching. Even if we can accurately measure learning from multiple, diverse sources via grades or formal learning outcomes assessments, we still do not have a direct measure of teaching. While professors certainly do have some control over student learning, it is only indirect; what we have much greater control over is our approach in the classroom, how we teach. We would be remiss if we didn't admit that there is a potentially strong relationship between student learning and teaching, but our point is that assessment in its current mandated forms doesn't directly measure teaching effectiveness.

Identifying a deficit in student learning does not necessarily indicate a problem with instruction. Furthermore, if we do see a valid problem with student learning related to poor instruction, our assessment tool tells us nothing about how to change teaching to improve the situation; it advocates trial and error based on questionable data. We often hear from external agencies that we need to focus on student learning, but, in doing so, we mostly focus on the assessment mechanism instead of teaching. Good assessment is not the same as good teaching. We shouldn't confuse the two and we shouldn't neglect the one behavior we have direct control over: how we teach. Instead of customizing assessment rubrics to better reveal "learning," we should focus on optimizing teaching and maintaining the highest scholarly standards in the classroom.

The goal is, or should be, improving teaching to optimize learning. No single approach to teaching works equally well in all situations, at all times, with all students, at every institution, so professional judgment and teacher autonomy are crucial. Statistical analyses can take a dizzying amount of data and reveal patterns and trends; descriptive statistics of learning outcomes data can reveal that students are not meeting certain basic goals or standards in a class; but statistics compress data, providing information about some imaginary average person well after semester's end. Many would argue that the best teachers vary and tailor their approach to different students during the

term, respecting the diversity of perspectives and learning styles. Probably most do, and probably most succeed, so we need to question the so-called crisis in higher education and seek out any direct evidence that American higher education is broken as a result of poor teaching.

ASSESSMENT AND ACADEMIC FREEDOM

As scholars, the *idea* of assessment, evaluation, and improvement should not sound surprising or threatening because this is what we all normally do every day. No matter what our discipline or academic field, we share our ideas and approaches with others, offering them up for public scrutiny and challenge via conference presentations, exhibitions, performances, peer-reviewed journals, and books. As professors, we engage in a similar, albeit sometimes less formal, practice in the classroom; we tailor and strive to improve our teaching based on interaction with and feedback from students, peers, and the literature.

For hundreds of years, academics in search of truth and wisdom have engaged in this traditional process of dissemination, critique, reflection, and revision; as a result, the academy and many of our fields have progressed without the benefit of external oversight. Even the critical Spellings Commission Report admits in its preamble that “most Americans don’t see colleges and universities as a trouble spot in our educational system. After all, American higher education has been the envy of the world for years” (vi). The report cites, as evidence of success, the number and variety of U.S. institutions of higher learning, the increasingly open access to their campuses, their role in advancing the frontiers of knowledge through research discoveries, the new forms of teaching and learning which emerge from them, and the number of Nobel Prizes and Rhodes Scholarships earned by Americans.

Nonetheless, the *practice* of assessment emerged in the mid-1980s and has transitioned away from an individual academic exercise in self-reflection and improvement to an institutional requirement, mandated by external accrediting agencies (Miller). In programs such as nursing or accounting, perhaps assessment is a vital aspect of quality control overseen by the discipline. However, in the liberal arts, in which most honors programs are based, the need for assessment is not necessarily a matter of health or wealth. Even so, honors programs are increasingly being required to engage in outcomes-based learning assessment.

Universities used to generate new ideas and create models that were adopted by those outside the ivory tower, from art and entertainment to industry and politics. However, the modern university, perhaps lacking its old confidence, turns again and again to the corporate world for many of its practices, including so-called accountability. Politicians, claiming to speak for the

“consumers” of higher education who spend ever-increasing sums for college tuition (as state contributions dwindle), have in many cases required colleges and universities that receive state and federal funding, which means just about every institution of higher learning, to show transparency and accountability, and the schools, urged by accreditation agencies, have decided that assessment of student learning is the best response to critics and consumers alike. Through reaccreditation, budgeting decisions, curriculum approval, and other means, pressure has been exerted on academy members to embrace the culture and practice of assessment.

At the root of this accountability and assessment movement is a lack of trust in faculty and an erosion of academic freedom. For most faculty members, academic freedom is seen as the right, earned through the long and rigorous process of tenure review, of a professor to present potentially unpopular or controversial material and arguments in our classes and research without censure from university authorities. When academic freedom was first defined for American institutions of higher learning in the early twentieth century, academic leaders attempted to break away from the master-servant model that had previously characterized the relationship between administrators and faculty. In the United States, academic freedom was first formally defined in 1915 by the American Association of University Professors’ (AAUP) *Declaration of Principles on Academic Freedom and Academic Tenure*. According to this document, university faculties are “appointees” of the legal governing authority “but not in any proper sense” its “employees” (295). “[O]nce appointed, the scholar has professional functions to perform in which the appointing authorities have neither competency nor moral right to intervene” (295). The definition was revised and reissued in 1940 by the AAUP and the Association of American Colleges as the *Statement of Principles on Academic Freedom and Tenure*. It states in part:

Institutions of higher education are conducted for the common good and not to further the interest of either the individual teacher or the institution as a whole. The common good depends upon the free search for truth and its free exposition.

Academic freedom is essential to these purposes and applies to both teaching and research. Freedom in research is fundamental to the advancement of truth. Academic freedom in its teaching aspect is fundamental for the protection of the rights of the teacher in teaching and of the student to freedom in learning. It carries with it duties correlative with rights. (3)

These definitions mainly concern First Amendment free speech protection, itself vulnerable after recent court decisions (AAUP, “Protecting”). However,

many of us assume that these protections extend to content, method, and evaluation within our courses. Instructors define assignments, evaluate student work in accord with fairness and the practices of our disciplines, and assign a final grade according to a scale established by our institutions. Under course-based assessment, however, instructors are advised by assessment officers or committees to adopt certain types of assignments, to devise rubrics for evaluating these assignments, and then to use the data to measure student learning. We have pointed out some of the concerns that many non-social scientists have with this approach. But our point is that, even if you believe in the value of such an approach for your course, the individual instructor should be the one to make the determination, not an administrator, committee, or outside agency. Imposition of a specific educational philosophy or practice from outside a discipline—be it from a politician, an administrator, or a faculty colleague—is an infringement of academic freedom.

CONCLUSION

Our position is that program metrics provide tangible evidence on which to base institutional change, but similarly acquired evidence about *learning* is not nearly as useful given limitations of validity. Objective assessments of learning, borrowed from the social sciences, are interesting theoretical tools that shouldn't necessarily cross over to the practice of teaching or policy-making in all disciplines. Learning outcomes assessment data can suffer from low validity; couple this problem with misinterpretation, and you can end up with a skewed view of learning at your institution over time. Given the limited nature of some forms of assessment data, the academy should resist the outside pressure to assimilate and adopt easy measures of learning that fail to capture the complete essence of our complex fields of study.

Methodologically questionable assessment techniques ranging from standardized tests to course-based rubrics have been hastily and redundantly adopted to reveal learning already demonstrated through traditional academic work. As a profession, we should use the best evidence available to improve teaching and learning, which are our passion, our calling. While many see assessment as a panacea, a perfectly accurate index of learning, we have argued that all tests vary in their degree of validity; institutions and accrediting bodies must avoid making rash decisions based on them. If we base changes to teaching on oversimplified measurements of learning, then we are by definition teaching to a sub-standard test. If the measurements are flawed or merely tap into simple types of learning, we may think learning has increased when, in reality, we've abandoned common sense, logic, and autonomous professional judgment in favor of an externally applied practice we don't know how to use responsibly.

LEARNING OUTCOMES ASSESSMENT IN HONORS

Furthermore, simple, untested, and questionably valid assessment techniques are not a good measure of, or substitute for, good teaching. Professors should continue to strive for excellent teaching and optimal learning for each individual student, but they should be free to do this in an appropriate professional manner consistent with best practices in their discipline. We should continue fostering intellectual diversity and creativity instead of worrying about lowest-common-denominator accountability scores. We should retain the academic freedom that has contributed to the great success of the American higher education system. Innovation is the product of freedom and diversity. Requiring faculty to assess learning in a specific way is the same as telling faculty how to teach within their respective disciplines; it is a slippery slope and an erosion of academic freedom undermining the best traditions and continued success of the academy.

While the practice of learning outcomes assessment is fraught with limitations and controversy, we don't want to leave the reader with the impression that we are against progressive ideas leading to academic innovation. Indeed, we hope that this paper will stimulate further inquiry on this topic within the honors community. Assessment is an interesting idea worth additional discussion and exploration, but, as often practiced, it is subject to many flaws, restricts our understanding of learning, has not been empirically demonstrated to lead to optimal instruction, and unnecessarily imposes a particular pedagogical approach on professionals in a wide variety of disciplines.

ACKNOWLEDGEMENTS

This paper is partially based on a roundtable discussion held at the 2009 NCHC conference in Washington, DC (Carnicom, S., Snyder, C., Bruce, K., Engel, S., Lanier, G., & Salas, A. 2009. Honors assessment: A valid exercise? Annual meeting of the National Collegiate Honors Council, Washington, DC). Many thanks are due to our co-authors and participants for a thoughtful and stimulating exchange. Many thanks are due as well to Phil Mathis, professor emeritus and former dean of the honors college at MTSU, for his helpful comments.

REFERENCES

American Association of University Professors. 1940 Statement of principles on academic freedom and tenure. Accessed from: <<http://www.aaup.org/NR/rdonlyres/EBB1B330-33D3-4A51-B534-CEE0C7A90DAB/0/1940StatementofPrinciplesonAcademicFreedomandTenure.pdf>> on February 2, 2010.

- . 1915 Declaration on academic freedom and academic tenure. Accessed from: <<http://www.aaup.org/NR/rdonlyres/A6520A9D-0A9A-47B3-B550-C006B5B224E7/0/1915Declaration.pdf>> on February 24, 2010.
- . Protecting an independent faculty voice: Academic freedom after *Garcetti v. Ceballos*. Nov/Dec 2009. Accessed from: <<http://www.aaup.org/NR/rdonlyres/B3991F98-98D5-4CC0-9102-ED26A7AA2892/0/Garcetti.pdf>> on February 2, 2010.
- Assessment disconnect. *Inside Higher Ed*. January 27, 2010. Accessed from: <<http://www.insidehighered.com/news/2010/01/27/aacu>> on February 24, 2010.
- Ebbinghaus, H., Ruger, H., & Bussenius, C. (1913). Rapidity of learning series of syllables as a function of their length. *Memory: A Contribution to Experimental Psychology*. New York, NY: Teachers College Press.
- Educational Testing Service. Measurement of Academic Proficiency and Progress. Accessed from: <<http://www.ets.org/portal/site/ets/menu.item.1488512ecfd5b8849a77b13bc3921509/?vgnextoid=ff3aaf5e44df4010VgnVCM10000022f95190RCRD&vgnnextchannel=f98546f1674f4010VgnVCM10000022f95190RCRD>> on February 22, 2010.
- Educational Testing Service. Validity of the measure of academic proficiency and progress (MAPP) test. A report commissioned by the Educational Testing Service. Accessed from: <http://www.ets.org/Media/Tests/MAPP/pdf/5018_mapp.pdf> on February 19, 2010.
- Gardner, H. (2006). *Multiple Intelligences*. New York, NY: Basic Books.
- Kandel, E. (2006). *In Search of Memory: The Emergence of a New Science of Mind*. New York, NY: W W Norton & Co.
- Long, A. (2009). How would Athena score on the ACT? Wisdom and knowledge in honors education. Presentation in Middle Tennessee State University Honors College Lecture Series. March, 2009. Murfreesboro, TN.
- Miller, M.A. (2006). The legitimacy of assessment. *The Chronicle of Higher Education*. September 22, 2006.
- Moss, P.A., Girard, B.J., & Haniford, L.C. (2006). Validity in educational assessment. *Review of Research in Education*, 30, 109–162.
- Otero, R. & Spurrier, R. (2005). *Assessing and Evaluating Honors Programs and Honors Colleges: A Practical Handbook*. Lincoln, NE: National Collegiate Honors Council Monograph Series.
- Sederberg, P.C. (2008). *The Honors College Phenomenon*. Lincoln, NE: National Collegiate Honors Council Monograph Series.
- Shakespeare, W. (1997) *Hamlet, Prince of Denmark*. From *The Riverside Shakespeare*. Ed. G. Blakemore Evans et al. 2nd ed. Boston: Houghton Mifflin.

LEARNING OUTCOMES ASSESSMENT IN HONORS

- Skinner, B. (1938). *The Behavior of Organisms: An Experimental Analysis*. Oxford, England: Appleton-Century.
- Spellings Commission Report. A test of leadership: Charting the future of US higher education. Commissioned by Secretary of Education Margaret Spellings. Pre-publication copy, September, 2006. Accessed from: <http://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/pre-pub-report.pdf> on February 2, 2010.
- Zubizarreta, J. (2009). *The Learning Portfolio: Reflective Practice for Improving Student Learning*. 2nd ed. San Francisco, CA: Jossey-Bass.

The authors may be contacted at
carnicom@mtsu.edu.