



Comments on the Commentaries about 'Seven Red Herrings'

October 2012

Roger Benjamin

I am pleased to accept the invitation to briefly respond to some of the points made by those who commented on my "Seven Red Herrings" paper which appeared in the September 2012 issue of the NILOA monthly newsletter. In his Foreword, Peter Ewell predicted that the merits and role of standardized testing will almost certainly continue to be debated. With this in mind, I also offer a few thoughts about what to expect in the future.

Trudy Banta, Gary Pike, and Terrel Rhodes view the promise and potential of standardized testing differently than Margaret Miller and Gordon Davies. Miller sees standardized measures as essential, because the field demands highly reliable and valid assessment tools. At the same time, she believes formative assessment is important as well, albeit for different purposes. Davies goes a step further by saying that colleges and universities must use standardized student learning outcomes measures to assure the public of that these institutions continue to make meaningful, valued contributions both to individuals and the larger society.

Banta and Pike represent the formative end of the assessment continuum. Most of the arguments they presented in their commentary about standardized assessment measures, particularly the Collegiate Learning Assessment (CLA), have appeared previously. Many of their points have been addressed by CLA staff, the Educational Testing Service (ETS), and other researchers, including a summary of approximately 90 studies (Benjamin, et al. 2012). Although my paper was not about the CLA per se, it is worth summarizing several cogent responses available elsewhere to the Banta and Pike's main arguments.

For example, average CLA value-added scores are highly reliable especially at the institution level (freshmen=.94; seniors=.86). Aggregate student motivation is not a significant predictor of aggregate CLA performance, and does not invalidate the comparison of colleges based upon CLA scores. Moreover, the types of incentives that students seem to prefer are not related to motivation and performance.

Although we continue to believe that a no-stakes approach is appropriate for the value-added model in higher education, motivation is a problem for individual student results. CAE (Council for Aid to Education) now offers a version of the CLA protocol, CLA+, which is reliable and valid for individual student performance, as does the Education Testing Service with its Proficiency Profile, and the American College Testing Program with its Collegiate Assessment of Academic Progress. It may well be appropriate in the future to attach stakes to the CLA, which, in turn, likely will increase student motivation to do well.

There is no interaction between CLA task content and field of study. Our researchers find that the CLA protocol

Viewpoint

www.learningoutcomesassessment.org

measures 30% of the knowledge and skills faculty desire. Results are improved significantly if a representative sample is drawn. Finally, that the CLA is highly correlated with the SAT does not mean the two tests measure the same thing. High school grades combined with the CLA predict freshmen and senior GPA at about the same level as the SAT alone. High school grades plus the SAT and CLA generate a higher prediction than either test alone. This would not be true if the SAT and CLA measured the same thing.

Banta and Pike are correct in advocating a focus on disciplines, but stray off track by rejecting that standardized test can accurately measure generic cognitive skills (Benjamin et al. 2012). The mean size effect of the growth in student learning outcomes for all colleges testing annually for the past eight years is approximately .73 standard deviations, demonstrating that college attendance is associated with improving these skills.

Banta and Pike suggest there is qualitative evidence to buttress their claims. It would be helpful to know the evidence to which they refer. Measurement scientists privilege statistical-based evidence. This makes conversation between the two groups difficult. Elsewhere I (Benjamin, 2012) explained what I call the assumption of the equality of fields of inquiry. Faculty members are reluctant to question the legitimacy of fields of inquiry that they may not be familiar with. There are solid reasons for this assumption. For example, an obscure field of molecular biology in veterinary medicine focusing on retroviruses in monkeys was critical in helping researchers develop treatments for AIDS. Breakthroughs in one scientific field may lead to startling breakthroughs in others. Measurement science is a field of inquiry that is too well established to be dismissed by colleagues arguing for formative assessments only. For example, Banta and Pike and Rhodes make good arguments for using e-portfolios to assess student learning. However, e-portfolios do not yet pass muster as tools that are sufficiently reliable and valid to obviate the need for appropriate standardized tests for decisions with stakes attached.

Both Davies and Miller want testing organizations to make public student outcome test results. What I should have said was that external demands *will require* institutions to make their student learning outcomes transparent and that peer review principles aligned with core values of the academy will provide foundational support for higher education leaders creating assessment reporting systems

Peter Ewell noted that faculty prefer to keep assessment results confidential, for internal use only. It is worth noting that testing organizations can achieve greater economies of scale in test development which lowers the price of individual assessments. Aided by recent developments in education technology, there appears to be a burst of innovation in creating assessments for direct use by faculty as instructional tools. Finally, samples of students tested at individual institutions are seldom large enough for the results to be considered sufficiently reliable. More widely used standardized assessments can boost confidence in the results found at individual institutions.

What We Can Expect

The competency-based model now gaining considerable traction will require assessments that corroborate the efficacy of the student learning claimed. Many of those assessments will be standardized tests. There is and will continue to be ample room for formative and standardized tests in postsecondary education. The issue is how to better leverage the virtues of both, for the benefit of improved teaching and learning for the larger societal goals

Faculty members are reluctant to question the legitimacy of fields of inquiry that they may not be familiar with.

Davies posited.

This, then, is not the time to defend the status quo. Many colleagues may be comfortable in defending positions that marginalize assessment in postsecondary education. Because increasing numbers of private and public leaders believe human capital is the nation's principal resource, debates about how to improve education will continue to grow. The rise of Internet-based education and concerns for the quality of higher education provided by more traditional means are fueling external demands for increased transparency, restructuring, and accountability.

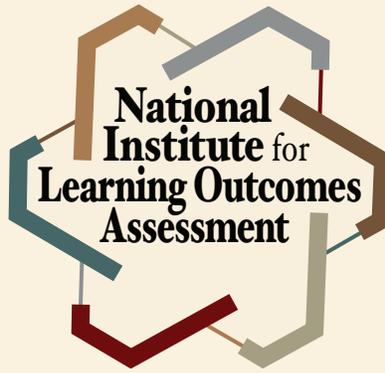
External demands for benchmarking student learning outcomes are destined to increase. However, higher education institutions possess a high level of legitimacy and relative autonomy anchored by department-based governance. The initial challenges for increased transparency of student learning outcomes will come from external forces. Responses to these demands will be developed by innovators within the higher education community. We need all hands on deck to experiment with ways to improve teaching and learning.

Finally, higher education institutions must respond to persistent external demands for more systematic evidence about student learning outcomes. In doing so, the enterprise must also maintain faculty autonomy in determining appropriate assessment approaches; reject college and university ranking systems; privilege efforts to improve student learning; develop assessment protocols that combine standardized and formative assessments; and adhere to peer review principles when constructing accountability systems. About this last observation, there seems little to debate.

References

- Benjamin, R. (2012). *The new limits of education policy: Avoiding a tragedy of the commons*. London: Edward Elgar.
- Benjamin, R. Elliot S., Klein S., Steedle, J., Zahner, D., & Patterson, J. (2012). *The case for generic skills and performance assessment in the United States and international settings*. New York: Council for Advancement of Education.

We need all hands on deck to experiment with ways to improve teaching and learning.



Please Cite As:

Benjamin, R. (2012, October). *Comments on the commentaries about 'seven red herrings'*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Follow us on social media:



@NILOA_web



@LearningOutcomesAssessment

Sign up to receive our monthly **NILOA Newsletter** and stay up to date with our research and publications.

