

# Assessment UPdate

January–February 2007  
Volume 19, Number 1



Progress, Trends, and Practices in Higher Education

## Revisiting the Blind Alley of Value Added

Trudy W. Banta, Gary R. Pike

THE COMMISSION ON THE FUTURE OF HIGHER EDUCATION APPOINTED by Secretary of Education Margaret Spellings issued a report in September 2006 entitled *A Test of Leadership: Charting the Future of U.S. Higher Education* (U.S. Department of Education, 2006). Two key commission recommendations are stated: “Higher education institutions should measure student learning using quality assessment data” and “The results of student learning assessments, including value-added measurements that indicate how much students’ skills have improved over time, should be made available to students and reported in the aggregate publicly” (p. 23). Anticipating such recommendations, in its Spring 2006 issue of *Perspectives*, the American Association of State Colleges and Universities issued the following statement in an article entitled “Value-Added Assessment: Accountability’s New Frontier”: “Value-added assessment allows true comparisons of the difference college makes to students across institutions and institutional types, instead of simply reflecting institutional resources and/or reputation” (p. 3). Interest in measuring what students learn in college and using one or a limited number of standardized instruments (thus permitting institutional comparisons) to do that has probably never been more intense. Our purpose in this article is to raise some questions about the capacity of standardized instruments to measure gain, or value added. We describe the ways in which change or growth and development can be measured and summarize what experts over the past four decades have concluded about the reliability and usability of these approaches.

### Research-Based Conclusions About Value Added

In a brief, accessible article in a 1984 issue of the *AAHE Bulletin* entitled “The Blind Alley of Value Added,” Jonathan Warren, who had spent a portion of his career at the Educational Testing Service (ETS), summed up the thinking of measurement specialists at that time: “In the abstract, the logic of value added gives it great appeal. In practice, I’ll argue, it seldom leads anywhere. Its results are too often trivial, difficult to make sense of, and peripheral to most instructional purposes. An undue focus on value added

## CONTENTS

### ARTICLES

- Revisiting the Blind Alley of Value Added  
*Trudy W. Banta, Gary R. Pike* 1
- Using Questionnaires in a College of Pharmacy and Health Sciences  
*Richard L. Slaughter, Susan Ann Talley* 3
- Unite and Conquer: A Collaborative Approach to Faculty Development  
*Thomas W. Paradis, Kathleen L. Smalldon* 6
- The College President’s Critical Role in Modeling Assessment: A Lesson in Leadership by Example  
*R. Ray D. Somera* 9

### COLUMN

- From the States  
*Peter T. Ewell* 11

### FEATURES

- Resources  
*Susan Kahn* 5
- Resources  
*John A. Muffo* 13

# Assessment Update

## Progress, Trends, and Practices in Higher Education

January–February 2007  
Volume 19, Number 1

### Editor

Trudy W. Banta, *professor of higher education and senior advisor to the chancellor, Indiana University–Purdue University Indianapolis (IUPUI)*

### Managing Editor

Karen Elaine Black, *director of program review, IUPUI*

### Assistant Editor

Frances W. Oblander, *educational consultant*

### Book Review Editor

Susan Kahn, *director of institutional effectiveness, IUPUI*

### Consulting Editors

Peter T. Ewell, *National Center for Higher Education Management Systems*

Thomas Anthony Angelo, *Victoria University of Wellington, Wellington, New Zealand*

T. Dary Erwin, *James Madison University*

Cecilia L. Lopez, *Harold Washington College*

Marcia Mentkowski, *Alverno College*

Jeffrey A. Seybert, *Johnson County Community College*

Peter J. Gray, *United States Naval Academy*

Gary R. Pike, *IUPUI*

### Designer

Kerry Golemon Schwartz, *Insight Out!*

**Assessment Update: Progress, Trends, and Practices in Higher Education** (Print ISSN 1041-6099; online ISSN 1536-0725 at Wiley Interscience, [www.interscience.wiley.com](http://www.interscience.wiley.com)) is published bimonthly by Wiley Subscription Services, Inc., A Wiley Company, at Jossey-Bass, 989 Market St., San Francisco, CA 94103-1741. Periodicals Postage Paid at San Francisco, CA, and additional mailing offices. Individual subscriptions are \$127 per year (institutional \$152). Back issues are available in limited supply at \$29 per issue. To order, phone toll-free (888) 378-2537 or fax (888) 481-2665. Visit our Web site at [www.josseybass.com](http://www.josseybass.com). *Postmaster:* Send address changes to Assessment Update, Jossey-Bass, 989 Market St., San Francisco, CA 94103-1741.

Copyright © 2007 Wiley Periodicals, Inc., A Wiley Company. All rights reserved. Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, c/o John Wiley & Sons, Inc., 111 River St., Hoboken, NJ 07030; (201) 748-8789, fax (201) 748-6326, [www.wiley.com/go/permissions](http://www.wiley.com/go/permissions).

may mean that today's opportunity to improve quality will slip away from us" (p. 10). "For practical purposes, value added isn't a workable concept and we have to get on with alternative ways of looking at the effect of what we do" (p. 12).

In a careful review of the literature on value-added assessment published in 1988, Leonard Baird, also a measurement expert employed for a time at ETS, concluded, "If executed thoughtfully, value-added assessment has some potential for the improvement of instruction at the program level. It is much less appropriate or useful at the institutional level of analysis. It is, above all, not a panacea, or even a solution to be recommended widely" (p. 215).

**Pike investigated the reliability of gain (value added), using three methods: gain scores, residual scores, and repeated measures.**

**He found that all three methods of representing change had serious shortcomings.**



For a decade that began in the mid-1980s, we were responsible for testing several thousand freshmen and seniors at the University of Tennessee, Knoxville (UTK) in response to the state's performance funding mandate. In 1992, Gary Pike published the results of an empirical study involving 722 students who had taken the ACT College Outcome Measures Project (COMP) exam, first as freshmen, then again as seniors at UTK. Pike investigated the reliability of gain (value added), using three methods: gain scores, residual

*(continued on page 14)*

## Call for Contributions

The editor welcomes short articles and news items for *Assessment Update*. Guidelines follow for those who would like to contribute articles on outcomes assessment in higher education.

- **Content:** Please send an account of your experience with assessment in higher education. Include concrete examples of practice and results.
- **Audience:** *Assessment Update* readers are academic administrators, campus assessment practitioners, institutional researchers, and faculty from a variety of fields. All types of institutions are represented in the readership.
- **Style:** A report, essay, news story, or letter to the editor would be welcome. Limited references can be printed; however, extensive tables cannot be included.
- **Format:** In addition to standard manuscripts, news may be contributed via letter, telephone, or fax (317) 274-4651. The standard manuscript format is a 60-space line with 25 lines per page. If word processing is used, please submit a 3½" diskette and three paper copies of your article. Word is preferred. Articles may also be sent to <kblack@iupui.edu> as a Microsoft Word attachment.
- **Length:** Articles should be four to eight typed, double-spaced pages (1,000–2,000 words). Short news items and content for the Calendar and Book Review sections may be 100–500 words in length. Annotations of recent publications for the Resources feature should be about 50–100 words long.
- **Copyright:** Articles shall not have been registered for copyright or published elsewhere prior to publication in *Assessment Update*.
- **Deadlines:** Each issue is typically planned four months before its publication. Future deadlines for submitting articles are April 1 (July–August 2007 issue), June 1 (September–October 2007 issue, and August 1 (November–December 2007 issue).

Please address mailed contributions and comments to Trudy W. Banta, Editor, *Assessment Update*, Rm. 140 Administration Bldg., 355 N. Lansing St., Indianapolis, IN 46202-2896. ■

## Revisiting the Blind Alley of Value Added

(continued from page 2)

scores, and repeated measures. He found that all three methods of representing change had serious shortcomings.

In the early 1990s, measurement specialists apparently believed they had said all that was needed to put the concept of value added in higher education to rest and turned their attention to other topics. In 2005, however, attempts to use student score gain to measure effectiveness of K–12 teachers prompted officials at ETS to commission Henry Braun to write a primer on value-added models. Braun (2005) observed that “the statistical models underlying [value-added measures] were originally developed for use in settings, such as agriculture, in which randomized experiments and sufficient data are the norm. [In these settings,] endowing statistical estimates with causal interpretations is relatively straightforward” (p. 16). But school systems do not operate by assigning students, then teachers, randomly to classes. For this and other reasons, Braun concluded, “reliance on a single statistical measure cannot be recommended” (p. 16). And in his October 2005 letter to individuals receiving a copy of Braun’s primer, ETS president and CEO Kurt Landgraf stated, “Results from value-added models should not serve as the primary basis for making consequential decisions. Other measures must be included in any fair and valid teacher evaluation system.”

As compelling as the concept of measuring student growth and development in college, or value added, may be, research does not support the use of standardized tests for this purpose.

### A Brief Explanation of Value-Added Concerns

There are two basic methods of representing change during college. The first method involves calculating difference scores. If a test is administered at entry

and again at graduation, the difference between scores can be used to represent value added. This approach is used in Tennessee as part of the state’s performance funding program. The second method of representing value added is to regress exiting scores on entering scores. The difference between the expected score produced by the regression model and the actual score at exit (that is, the residual) can also be used to represent value added. This is a form of analysis of covariance and is the approach employed by the Collegiate Learning Assessment .

Three key questions must be answered about the utility of value-added measures and the confidence we can place in them. Are the measures reliable? Do the measures accurately represent what students have learned? What measure should be used?

Questions about the reliability of difference scores were a focus of early criticisms of value added. Carl Bereiter (1963) and Lee Cronbach and Lita Furby (1970) observed that the reliability of difference scores is extremely low. Moreover, as the correlation between measures at entry and exit increases, the reliability of the difference score decreases. Ross Traub (1967) observed that the reliability of residual scores is also quite low and tends to decrease as the correlation between entry and exit scores increases. Not surprisingly, Pike’s study (1992) of ACT comp scores found that the reliability for a difference score was 0.14 and the reliability for a residual score was 0.17—a very poor showing, indeed.

The second question that arises in discussions of value added is whether the measures accurately represent student learning. Scholars such as Robert L. Thorndike (1966) have criticized difference scores because pretest-posttest differences are almost always negatively correlated with entering scores. That is, students with low scores at entry gain

more than students with high entering scores. This feature of difference scores is an artifact of floor and ceiling effects for the tests and the fact that the measurement error in entering test scores is included in the difference score, but with the sign reversed. Residual scores also have been criticized for not representing student learning. In fact, Baird (1988) argued that residual scores were not a measure of change. Because regression analysis removes all of the variance in exiting scores that is linearly related to entering scores, residual scores will underestimate student learning if there is any linear relationship between learning and entering ability. Concerns related to the first two questions suggest that value-added analysis will not yield accurate and appropriate information about student learning within institutions.

Last but certainly not least, the decision about which approach to use is not trivial. Frederic Lord (1967, 1969) examined weight gain in males and females during college, using difference scores and analysis of covariance (that is, residual scores). When the analyses used difference scores, Lord found no differences in weight gain for males and females. However, when the same data were analyzed using residual scores, there was a significant difference in weight gain for males and females. Known as “Lord’s Paradox,” this phenomenon raises serious concerns about the ability to compare institutions or groups within institutions using a value-added standard. In fact, Lord (1967) concluded, “there simply is no logical or statistical procedure that can be counted on to make proper allowances for uncontrolled pre-existing differences between groups” (p. 305).

### Alternative Measures

In his 1988 review article, Baird observed, “The more tests assess general characteristics, the less sensitive they are to change due to educational programs. That is, the tests become so general as to assess relatively stable characteristics of students. In the cognitive area, the more general tests

border on measures of general intelligence” (pp. 206–207). This statement strongly suggests that testing students’ discipline-specific knowledge and skills as well as their generic skills (writing, speaking, critical thinking) as applied in their major fields of study provides a much more promising avenue for assessing growth and development in college than do today’s most commonly discussed tests of generic skills alone. We support a focus on major field assessment, with an emphasis on using student electronic portfolios as the most authentic instrument for demonstrating growth over time.

In yet another ETS publication, *A Culture of Evidence: Postsecondary Assessment and Learning Outcomes*, Carol Dwyer, Catherine Millett, and David Payne (2006) also support a discipline-based approach, recommending that “the United States create a system that will assess workforce readiness” (p. 19). They argue that “from an economic perspective [workforce readiness] has direct links to the economic competitiveness of the United States in a global economy” (p. 19). Dwyer, Millett, and Payne observe that professional fields such as education, social work, nursing, law, and medicine already have measures of workforce readiness or domain-specific knowledge sufficient for workforce entry. They suggest that “appropriate academic organizations” (p. 20) in fields currently without such measures be charged “with the task of articulating a set of expected learning outcomes” (p. 20) and a set of pre- and post-measures for assessing growth in those fields.

## Conclusion

Given the weight of evidence offered by prominent measurement experts over the past forty years, we must conclude that pursuing the concept of value added condemns us to revisit a blind alley. We offer this very brief review of literature and the lists of references and additional resources that follow in the hope that they will provide colleagues in higher education with

some reasoned arguments to use in addressing the current press to assess with a standardized test and to compare institutions by using value-added measures. ■

## References

- American Association of State Colleges and Universities. “Value-Added Assessment: Accountability’s New Frontier.” *Perspectives*, Spring 2006.
- Baird, L. L. “Value-Added: Using Student Gains as Yardsticks of Learning.” In C. Adelman (ed.), *Performance and Judgment: Essays on Principles and Practice in the Assessment of College Student Learning*. Washington, D.C.: U.S. Government Printing Office, 1988.
- Bereiter, C. “Some Persisting Dilemmas in the Measurement of Change.” In C. Harris (ed.), *Problems with Measuring Change*. Madison: University of Wisconsin Press, 1963.
- Braun, H. I. *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models*. Princeton, N.J.: Policy Evaluation and Research Center, Educational Testing Service, 2005.
- Cronbach, L. J., and Furby, L. “How Should We Measure Change—or Should We?” *Psychological Bulletin*, 1970, 74, 68–80.
- Dwyer, C. A., Millett, C. M., and Payne, D. G. *A Culture of Evidence: Postsecondary Assessment and Learning Outcomes*. Princeton, N.J.: Educational Testing Service, 2006.
- Lord, F. M. “A Paradox in the Interpretation of Group Comparisons.” *Psychological Bulletin*, 1967, 68, 304–305.
- Lord, F. M. “Statistical Adjustments When Comparing Preexisting Groups.” *Psychological Bulletin*, 1969, 72, 336–337.
- Pike, G. R. “Lies, Damn Lies, and Statistics Revisited: A Comparison of Three Methods of Representing Change.” *Research in Higher Education*, 1992, 33, 71–84.
- Thorndike, R. L. “Intellectual Status and Intellectual Growth.” *Journal of Educational Psychology*, 1966, 57, 121–127.
- Traub, R. E. “A Note on the Reliability of Residual Change Scores.” *Journal of Educational Measurement*, 1967, 4, 253–256.
- U.S. Department of Education. *A Test of Leadership: Charting the Future of U.S.*

*Higher Education*. Washington, D.C.: U.S. Department of Education, 2006. Also available at [<http://www.ed.gov/about/bdscomm/list/hiedfuture/reports/pre-pub-report.pdf>].

Warren, J. “The Blind Alley of Value Added.” *AAHE Bulletin*, 1984, 37(1), 10–13.

## Additional Resources

- Hanson, G. R. “Critical Issues in the Assessment of Value Added in Education.” In T. W. Banta (ed.), *Implementing Outcomes Assessment: Promise and Perils*. New Directions for Institutional Research, no. 59. San Francisco: Jossey-Bass, 1988.
- Harris, C. W. *Problems in Measuring Change*. Madison: University of Wisconsin Press, 1963.
- Holland, P. W., and Rubin, D. B. “On Lord’s Paradox.” In H. Wainer and S. Messick (eds.), *Principals of Modern Psychological Measurement: A Festschrift for Frederic M. Lord*. Hillsdale, N.J.: Erlbaum, 1983.
- Linn, R. L., and Slinde, J. A. “The Determination of the Significance of Change Between Pre- and Post-Testing Periods.” *Review of Educational Research*, 1977, 47, 121–150.
- Terenzini, P. T. “Measuring the Value of College: Prospects and Problems.” In C. Fincher (ed.), *Assessing Institutional Effectiveness: Issues, Methods, and Management*. Athens: University of Georgia Press, 1989.
- Willett, J. B. “Questions and Answers in the Measurement of Change.” In E. Z. Rothkopf (ed.), *Review of Research in Education*, 1988, 15, 345–422. Washington, D.C.: American Educational Research Association.

*Trudy W. Banta is professor of higher education and senior advisor to the chancellor for academic planning and evaluation and Gary R. Pike is associate professor of higher education and executive director of information management and institutional research at Indiana University—Purdue University Indianapolis.*

Copyright of *Assessment Update* is the property of John Wiley & Sons, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.