



ETS Research Spotlight

Measuring Learning Outcomes in Higher Education: The Role of Motivation

ISSUE 9, AUGUST 2013

Foreword

ETS Research Spotlight

ISSUE 9

AUGUST 2013

Senior Editor: Jeff Johnson

Contributing Editor: Brenda Whiteman

Copy Editor: Eileen Kerrigan

Layout: Sally Acquaviva

Cover Design: George Barrett

Views expressed in this report are those of the authors and do not necessarily reflect the views of the officers and trustees of Educational Testing Service.

Digital copies of this report can be downloaded free of charge from:
www.ets.org/research/spotlight9

To comment on any article in this issue, write to: RDWeb@ets.org



Copyright © 2013 by Educational Testing Service. All rights reserved. ETS, the ETS logo, LISTENING. LEARNING. LEADING., GRE, TOEFL, TOEFL IBT and TOEIC are registered trademarks of Educational Testing Service (ETS). THE PRAXIS SERIES is a trademark of ETS. All other trademarks are the property of their respective owners. 23297

Preparing students to succeed at all stages of education is central to ETS's mission of advancing quality and equity in education. Assessment can serve as one among many tools in pursuing that mission — but only if assessment results support valid interpretations of what students know and can do. Recently, ETS researchers conducted an important study of the effect that test-taker motivation can have on the validity of assessment results. This study, by ETS researchers Ou Lydia Liu, Brent Bridgeman, and Rachel Adler, is the focus of the featured research synopsis in this issue of *ETS Research Spotlight*. The synopsis summarizes an article that appeared earlier in 2013 in the journal *Educational Researcher*.

In particular, Liu and her colleagues were interested in the role that motivation plays in low-stakes higher education learning outcomes assessments. Here, student motivation is of concern because results from such assessments — while they may affect the curriculum planning of higher education institutions or the public's perceptions of these institutions — generally have no direct consequences for individual students.

If students who take these assessments aren't motivated to try their best, the results could underestimate how much students have learned, and thus lead to invalid conclusions about the effectiveness of higher education institutions.

The study raises major questions of concern to educators and policymakers who, in their quest to improve the quality of higher education and demonstrate accountability, are looking at multiple ways — including standardized assessments — to evaluate student learning outcomes. The findings suggest that motivation does, in fact, play a significant role in performance on higher education outcomes assessments.

Findings such as these are an important contribution to the constant, ongoing process of assessment validation. The work of Liu and her colleagues is a part of a larger research agenda. At ETS, we recognize that the field of assessment is evolving on many levels, from the purposes of assessments, to the technologies used to deliver them, to the populations who take them. Even as this evolution takes place, being able to draw valid conclusions about what students know and can do remains critical.

We hope you find this issue of *ETS Research Spotlight* to be an informative look at our commitment to the validity of assessment results.

A handwritten signature in black ink, appearing to read "Ida Lawrence".

Ida Lawrence,
Senior Vice President,
Research & Development

FEATURED RESEARCH SYNOPSIS

Measuring Learning Outcomes in Higher Education: The Role of Motivation

This issue of *ETS Research Spotlight* features an article by ETS researchers Ou Lydia Liu, Brent Bridgeman, and Rachel Adler, who examined the role that motivation plays in university students' performance on low-stakes, standardized outcome assessments such as the ETS® Proficiency Profile (EPP).¹ The EPP measures university-level skills in such areas as critical thinking, reading, writing, and problem solving. According to the authors, the outcomes of these assessments are intended to yield evidence of what students are or are not learning. Institutions, in turn, use these instruments to meet their accreditation requirements and fulfill accountability demands.

The article presents findings from the authors' study, which investigated the effect of students' motivation on low-stakes assessments, because, by nature, such tests have no direct consequences for students. In essence, the authors claim that these assessments, unlike university admission examinations, do not usually require one to pass or achieve a certain score to graduate or advance, personally or professionally.

The authors cited existing research findings, most notably those presented in the book *Academically Adrift* by Richard Arum and Josipa Roksa,² who claimed that university students are not learning enough academically. Liu et al. believed that Arum and Roksa's findings failed to consider the role of motivation when students took the Collegiate Learning Assessment (CLA), which measures the value an institution contributes to student learning. Very few studies, Liu et al. noted, have manipulated motivation in an attempt to better understand its possible effect on test scores.

According to the authors, higher education institutions have been considering assessments such as the ETS Proficiency Profile because they are under growing political pressure to show accountability for educating their students. As evidence of this pressure, the authors cite the Spellings Commission,³ formed in 2005, which was charged with developing a comprehensive national strategy for postsecondary education in order to improve the quality of higher education for U.S. students, with the goal of making them more competitive with their peers on a global scale.

Editor's note: The full reference list and technical details regarding this research appeared in the original work, which was:

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Educational Researcher*, 41(9), 352-362.

¹ <http://www.ets.org/proficiencyprofile/about/>

² Arum, R., & Roksa, J. (2011). *Academically adrift*. Chicago, IL: University of Chicago Press.

³ <http://www2.ed.gov/about/bdscomm/list/hiedfuture/index.html>

In the study, the authors examined the use of outcomes assessment scores as one of many tools to measure *learning gain* or *value-added* — what students may have learned as a result of instruction. They wanted to determine which evidence of student learning is scalable or can be replicated and compared at other institutions. The authors contend that the data they collected are of particular importance to institutions for such purposes as improving curriculum and overall institutional effectiveness. Their aim in conducting this study was to improve the validity of interpretations possible from the results of outcomes assessments.

Approach and Intervention Methods

The authors addressed three research questions:

- 1) What is the relationship between students' self-reported motivation and test scores?
- 2) Do motivational instructions affect student motivation and performance?
- 3) Do conclusions drawn about college learning gains change with test format (i.e., multiple choice vs. essay) and motivational instruction?

The authors created three motivational conditions or settings to manipulate the level of motivation that students had to perform as well as possible on the EPP. In doing so, they divided the students into three groups:

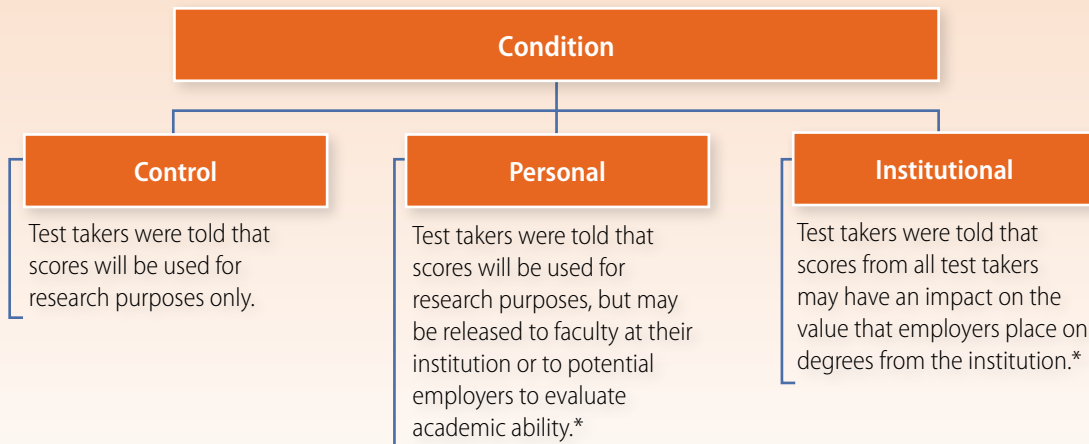
- 1) Control Condition (the authors told the students that their test scores will be used for research purposes only)
- 2) Personal Condition (the authors told the students that their scores will be used for research purposes but also may be released to faculty at their institution or to potential employers to evaluate their academic ability)
- 3) Institutional Condition (the authors told the students that their collective test scores may have an impact on the value that employers place on degrees from their institution)

According to the authors, higher education institutions have been considering assessments such as the ETS Proficiency Profile because they are under growing political pressure to show accountability for educating their students.

The authors recruited 757 students from three higher education institutions in three states to take the tests. These institutions included: 1) one four-year university with very high research activity; 2) one institution where the highest degree offered is a master's degree; and 3) one community college. The authors recruited students with at least one year of college experience at the four-year institution and community college students who had taken at least three courses. The researchers did not include first-year students at four-year institutions because they presumed them to be naturally more motivated due to their eagerness to fit into college life.

Motivational Conditions

In order to examine the possible effect of motivation, the researchers divided the test takers into three groups, or *motivational conditions*. Each group received different information about how their scores would be used.



* After students in these groups were finished with the test, the institutions assured them that, despite what they read in the instructions, their scores would be used only for research.

The authors administered the abbreviated online version of the EPP, which includes an optional 30-minute essay that measures college-level writing ability. The authors also asked the students to fill out the Student Opinion Survey (SOS), a widely used survey that researchers have developed to measure levels of motivation while taking low-stakes tests. ETS paid the students \$50 each to complete the tests and survey. The institutions debriefed the students in the Personal Condition and Institutional Condition groups after they took the test to assure them that, despite what they read in the instructions, their scores would not be shared with anyone other than their own research teams after the completion of the data collection.

In the study, the authors examined the use of outcomes assessment scores as one of many tools to measure learning gain or value-added — what students may have learned as a result of instruction.

Analysis and Results

The authors used statistical analyses known as multiple linear regression (MLR), analysis of variance (ANOVA), and analysis of covariance (ANCOVA) in order to investigate the relationship between the students' interpretation of their own level of motivation and their test scores. The authors conducted separate analyses for students at each institution. In addition to the scores from the ETS Proficiency Profile with its optional essay and the SOS survey, the authors also included admissions scores (i.e., SAT®, ACT®) for students at four-year institutions and placement test scores for students from the community college.

The authors found that the motivational condition or setting had a significant impact on the ETS Proficiency Profile and essay scores. They saw that students in the Personal and Institutional Conditions

The authors contend that motivation — or the lack of it — has a significant impact on test scores.

experimental groups reported higher levels of motivation than those in the Control Condition group. Likewise, the authors found that those students who were in the experimental groups had test scores that were significantly higher than those who were in the Control group. Simply put, the authors concluded that students — when motivated — can significantly improve their test performance.

The specific test format (i.e., multiple-choice or constructed response), the authors maintained, also played a key role toward building evidence needed to support their claim that motivated students perform better on low-stakes tests. The researchers observed more evidence of learning growth on multiple-choice tests compared with essay-format tests. Although they could not confirm the possible reasons, the authors believed that it may take more effort for a student to create an essay, surmising it may be a less-motivating task.

Conclusions/Recommendations

The authors contend that motivation — or lack of it — has a significant impact on test scores. The results of their study confirmed researchers' concerns that students do not exert their best effort in taking low-stakes assessments when they know that it has no direct consequence for the test takers. Students in the Personal and Institutional Condition groups had test scores that were significantly higher than the scores of students in the Control Condition group. Although there was no significant difference between the Personal and Institutional Condition groups, students in the Personal Condition group had slightly higher test scores.

The authors contended in their study that students with higher levels of motivation are more likely to score higher on low-stakes outcome assessments. Equally important, they stated that the format of the test (i.e., multiple-choice or constructed response) also has an impact on motivation.

The authors caution that it is dangerous to draw conclusions about the quality or effectiveness of a higher education institution based on learning outcomes assessment data without considering the role of motivation. They state that, without effective motivation strategies, students' abilities can be underestimated in low-stakes assessments.

From the authors' perspective, nonmonetary incentives, such as stressing the importance of test results to institutions and to students, are all cost-efficient and feasible ways to help boost students' motivation. The study demonstrated that students do not have to be monetarily enticed to boost their motivation to perform well on these assessments. Instead, the authors created settings that would help motivate the students to do their best.

The authors point out that the implications of low-stakes outcomes assessments extend globally. The Assessment of Higher Education Learning Outcomes (AHELO) project, sponsored by the Organisation for Economic Co-operation and Development (OECD), aims to test what graduates of higher education institutions know and can do in the areas of critical thinking, writing, and problem solving. With such insight, the authors maintain that researchers will be able to broaden their findings by comparing results across countries once the data are available. Taking a finding from *Academically Adrift* that called attention to the limited learning of university students, the authors note that varying levels of motivation among international students will more than likely play a key role in determining test score performance as the U.S. higher education system is compared to its international counterparts.

Further Discussion

With many factors to consider, the authors write that they need to further analyze the function of the Institutional Condition. To recap, in the case of the Institutional Condition, the authors told the students that their test scores would be used for research purposes and that their scores will be averaged with the scores of other students taking the test at their college. They assert that gaining more insight in this area would be particularly helpful to institutions to help improve their strategies to better engage students in taking low-stakes tests. In other examples, they recommend that institutions provide students with certificates to give to potential employers to validate their performance and overall academic ability. The certificates, they maintain, are likely to improve students' motivation in taking the test as the test results bear personal consequences for students.

Although the authors gathered consistent findings from diverse student populations, they caution that the analysis of students from the three institutions is limited in scope because the data pool is small. The authors conclude that acquiring further knowledge and practical strategies to motivate students in taking low-stakes outcomes assessment in higher education will help improve the quality of data gathered for accountability and institutional improvement purposes.



The aim in conducting this study was to improve the validity of interpretations of the results of higher education learning outcomes assessments. The researchers were concerned about the possible effect of motivation on the results of such assessments.

You May Also Be Interested In ...

Listed below is a selection of other recent ETS research publications related to the valid use of test results. Learn more about these and other publications — or download an electronic copy, where available — from the ETS ReSEARCHER Database at <http://search.ets.org/researcher>.

PUBLICATIONS

An Argument Approach to Observation Protocol Validity

Authors: C. A. Bell, D. H. Gitomer, D. F. McCaffrey, B. K. Hamre, R. C. Pianta, & Y. Qi
Educational Assessment: Measurement Issues in the Study of Teaching,
 Vol. 17, Nos. 2–3, pp. 62–87
 2012

Why Bother with Research When We Have Common Sense?

Author: B. Bridgeman
R&D Connections, No. 20
 September 2012

Is There Any Interaction Between Background Knowledge and Language Proficiency That Affects TOEFL iBT® Reading Performance?

Authors: Y. Z. Hill & O. L. Liu
ETS Research Report No. RR-12-22
 October 2012

Validating the Interpretations and Uses of Test Scores

Author: M. T. Kane
Journal of Educational Measurement, Vol. 50, No. 1, pp. 1–73
 Spring 2013

Validation as a Pragmatic, Scientific Activity

Author: M. T. Kane
Journal of Educational Measurement, Vol. 50, No. 1, pp. 115–122
 Spring 2013

Writing Essays on a Laptop or a Desktop Computer: Does It Matter?

Authors: G. Ling & B. Bridgeman

International Journal of Testing, Vol. 13, No. 2, pp. 105–122

April-June 2013

Standard Setting Lessons Learned in the South African Context: Implications for International Implementation

Authors: M. Pitoniak & N. Yeld

International Journal of Testing, Vol. 13, No. 1, pp. 19–31

January-March 2013

Validating Automated Essay Scoring for Online Writing Placement

Author: C. Ramineni

Assessing Writing, Online First

November 2012

Automated Essay Scoring: Psychometric Guidelines and Practices

Authors: C. Ramineni & D. M. Williamson

Assessing Writing, Online First

November 2012

The Effects of Glossary and Read-Aloud Accommodations on English Language Learners' Performance on a Mathematics Assessment

Authors: M. K. Wolf, J. Kim, & J. Kao

Applied Measurement in Education, Vol. 25, No. 4, pp. 347–374

October 2012

Contrasting Automated and Human Scoring of Essays

Author: M. Zhang

R&D Connections, No. 21

March 2013

About ETS

At ETS, we advance quality and equity in education for people worldwide by creating assessments based on rigorous research. ETS serves individuals, educational institutions and government agencies by providing customized solutions for teacher certification, English language learning, and elementary, secondary and post-secondary education, as well as conducting education research, analysis and policy studies. Founded as a nonprofit in 1947, ETS develops, administers and scores more than 50 million tests annually — including the *TOEFL*® and *TOEIC*® tests, the *GRE*® tests and *The Praxis Series*™ assessments — in more than 180 countries, at over 9,000 locations worldwide.



Scan this code to read the publication
on your mobile device.



Listening. Learning. Leading.®

www.ets.org