Measuring Student Learning Outcomes in Higher Education: Current State, Research Considerations, and An Example of Next Generation Assessment

Ou Lydia Liu, Ph.D.
Managing Senior Research Scientist

Katrina Crotts Roohr, Ed.D.
Associate Research Scientist

Educational Testing Service
Overview

• Introduction to student learning outcomes (SLO) assessment
• Current state of SLO research
• Challenges in implementation and use
• ETS’s approach to next generation assessment
  – Quantitative Literacy
The Context

• Rapid development of higher education
  – 15.9 million to 21.0 million students from 2001 to 2011 (Snyder & Dillow, 2013)

• National goal of higher education
  – By 2020, America should have the highest proportion of college graduates (Obama, 2009)

• Call for quality assurance in higher education
  – “remarkable absence of accountability mechanisms to ensure that colleges succeed in educating students” (U.S. Department of Education, 2006).
Driving Forces

• Accreditation
  – Pressure on institutions to become accountable for student learning

• Accountability calls
  – Voluntary System of Accountability (VSA)
  – Transparency by Design
  – Voluntary Framework of Accountability

• Institutional internal improvement
Current Use of SLO Assessment

• Most institutions had adopted learning outcomes (84%; Kuh et al., 2014).

• Significant more assessment activity now than a few years ago

• Use a variety of tools
Tools to Assess SLO

## Comparison of Assessment Tools

<table>
<thead>
<tr>
<th>Tool</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey</td>
<td>Cost efficient; easy administration; comparison</td>
<td>No direct evidence of student learning</td>
</tr>
<tr>
<td>Locally developed survey</td>
<td>Aligned with instruction; meet institution’s specific needs</td>
<td>No benchmark with other institutions; sometimes lack psychometric quality</td>
</tr>
<tr>
<td>Standardized measures</td>
<td>Comparable across institutions; sufficient validity and reliability evidence</td>
<td>Insufficient alignment with instruction</td>
</tr>
<tr>
<td>Rubrics</td>
<td>Flexibility for adaptation</td>
<td>Poor consistency among users</td>
</tr>
<tr>
<td>Performance assessment</td>
<td>Authentic</td>
<td>Expensive; difficult to implement; poor reliability</td>
</tr>
<tr>
<td>e-portfolio</td>
<td>Offer a range of data</td>
<td>Comparability is an issue</td>
</tr>
</tbody>
</table>
Current Challenges in Learning Outcomes Assessment (Liu, 2011a)

• Insufficient evidence of what learning outcomes assessment predicts

• Design/Methodological issues with value-added research

• The effect of student motivation on test performance
What Does SLO Assessment Predict?

• Traditional success indicators
  – GPA, retention, course completion, graduation (Hendal, 1991; Lakin, Elliott, & Liu, 2012; Marr, 1995)

• Indictors more difficult to obtain
  – Graduate school application, employment, job performance, and life events (Arum, Cho, Kim, & Roksa; 2012; Butler, 2012; Ejiogu, Yang, Trent, & Rose, 2006)

• Choice of criterion depends on the specific learning outcome
Design/Methodological Issues with Value-added Research

• Longitudinal vs. cross-sectional design

• Methodological considerations
  – Choice of statistical models (Liu, 2011b)
  – Unit of analysis
  – Institutional characteristics

• Factor in attrition
Student Motivation in Taking Low-stakes Tests

• Learning outcomes assessment does not have a direct impact on students
  – Low motivation could threaten the validity of the test results

• Ways to monitor student motivation
  – Student self-report
  – Motivation survey: Student Opinion Survey (Sundre & Wise, 2003)
  – Response time effort (Wise & Kong, 2005)
Prior Research on Motivation

• Motivation has an impact on test scores (Liu, Bridgeman, & Adler, 2012; Barry, Horst, Finney, Brown, & Kopp, 2010; Sundre & Wise, 2003; Wise & DeMars, 2005, 2010; Wise & Kong, 2005)

• Students with higher motivation tend to perform better (Braun, Kirsch, & Yamamoto, 2011; Duckworth, Quinn, Lynam, Loeber, & Stouthamer-Loeber, 2011; Kim & McLean, 1995; Liu et al., 2012)

Objectives of an Experimental Motivation Study (Liu et al., 2012)

• Investigate the impact of motivation on low-stakes learning outcomes assessment

• Identify practical motivational strategies that institutions can use
Participants (N=757)

• One four-year research institution
  – n=340, SAT/ACT

• One four-year master’s institution
  – n=299, SAT/ACT

• One community college
  – n=118, placement test scores
Instruments

• ETS Proficiency Profile
  – Multiple-choice test
  – Measures critical thinking, reading, writing, and mathematics
  – Abbreviated version (36 items)

• Essay

• Motivation survey
  – Student Opinion Survey (10 items; Sundre, 1999)
Motivational Conditions

• Created three motivational conditions
• Embedded in regular consent forms
• Random assignment within a testing session
Control Condition

Your answers on the tests and the survey will be used only for research purposes and will not be disclosed to anyone except the research team.

Dear Student:

Educational Testing Service (ETS) is conducting a study of the ETS Proficiency Profile, a test with 36 multiple-choice items and an essay part that take about two and a half hours to complete. Then, you will be asked to fill out the Student Opinion Survey, which consists of ten items. We additionally request your permission to obtain your college admission scores (SAT and/or ACT scores) and high school GPA from the registrar’s office at your institution.

Your answers on the opinion survey will be used only for research purposes and will not be disclosed to anyone except the ETS research team. Your test scores will be averaged with all other students taking the test at your college. Only this average will be reported to your college. This average may be used by employers and educational officials to improve instruction at your college. This information on your institution is viewed and therefore affect the value of your diploma.

You will be given a $50 major credit card gift card in consideration for your time and effort. We thank you for your voluntary participation and ask you to please sign the form below.

Thank you,

Lydia Liu, Ph.D.
Foundational and Validity Research
Educational Testing Service

I hereby accept the terms described above. I understand that group scores from this test may be used by employers and others to evaluate the quality of instruction at my college and that this could affect the value of my degree. I agree to voluntarily participate in the study.

Truman State University
Institution Name

Student ID

Print Name

Signature

Date

Group 3
Institutional Condition

Your test scores will be averaged with all other students taking the test at your institution. Only this average will be reported to your institution. This average may be used by employers and others to evaluate the quality of instruction at your institution. This may affect how your institution is viewed and therefore affect the value of your diploma.
Dear Student:

Educational Testing Service (ETS) is conducting a study of the ETS Proficiency Profile, a test with 36 multiple-choice items and an essay part that take about two and a half hours to complete. Then, you will be asked to fill out the Student Opinion Survey, which consists of ten items. We additionally request your permission to obtain your college admission scores (SAT and/or ACT scores) and high school GPA from the registrar’s office at your institution.

Your answers on the opinion survey will be used only for research purposes and will not be disclosed to anyone except the ETS research team. Your test scores will be averaged with all other students taking the test at your college. Only this average will be reported to your college. This average may be used by employers and others to make judgments about your instruction at your college. This average will reflect how your institution is viewed and therefore affect the value of your diploma.

You will be given a $50 major credit card gift card in recognition for your time and effort. We thank you for your voluntary participation and ask you to please sign the form below.

Thank you,

Lydia Liu, Ph.D.
Foundational and Validity Research
Educational Testing Service

I hereby accept the terms described above. I understand that group scores from this test may be used by employers and others to evaluate the quality of instruction at my college and that this could affect the value of my degree. I agree to voluntarily participate in the study.

Truman State University
Institution Name
Student ID

Print Name

Signature
Date

Group 3

Your test scores may be released to faculty in your college or to potential employers to evaluate your academic ability.
Results

• Motivational instruction has a significant impact on both EPP scores and students’ self-reported motivation
How Motivational Instructions Affected the ETS Proficiency Profile

Control (n=250)  Institutional (n=257)  Personal (n=247)

Scale Score

- Control: 453 (0.26 SD)
- Institutional: 458 (0.41 SD)
- Personal: 462

Copyright © 2014 by Educational Testing Service.
How Motivational Instructions Affected the Essay

- Control (n=250): 4.07 (0.23 SD)
- Institutional (n=257): 4.29 (0.41 SD)
- Personal (n=247): 4.46

Copyright © 2014 by Educational Testing Service.
How Motivational Instructions Affected Self-report Motivation

Control (n=250)  Institutional (n=257)  Personal (n=247)

3.61  3.81  3.89

0.31 SD  0.43 SD
Further Replications

- Examine the effect of a similar motivational instruction (Rios, Liu, & Bridgeman, 2014; Liu, Rios, & Borden, in press).
- How students differ in testing taking behavior
- Effect of motivational filtering
Participants

- College seniors (n=136)
  - From five campuses of a state university system
  - 75% females, 79% Whites, and 76% reporting English as their best language
Experiment/Control Condition

Experiment

“You are about to take the ETS Proficiency Profile. The test takes about 2 hours. Your score on this test may be used in aggregate to evaluate the quality of instruction at xxx (the name of the institution). It may also affect how xxx (the name of the institution) compares to other institutions nationally. The ranking of xxx (the name of the institution) in the comparison may affect the value of your diploma. We strongly encourage you to try your best on this test, regardless of how well you think you can perform, for the sake of xxx’s (the name of the institution) national standing.”

Control

“You are about to take the ETS Proficiency Profile. The test takes about 2 hours. Your score on this test will have no effect on your grades or academic standing, but we do encourage you to try your best.”
# Difference in EPP Performance

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Experimental</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EPP Score</strong></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Total</td>
<td>67</td>
<td>436.25</td>
<td>24.50</td>
<td>63</td>
</tr>
<tr>
<td>Reading</td>
<td>67</td>
<td>114.27</td>
<td>8.74</td>
<td>63</td>
</tr>
<tr>
<td>Writing</td>
<td>67</td>
<td>112.58</td>
<td>5.83</td>
<td>63</td>
</tr>
<tr>
<td>Math</td>
<td>67</td>
<td>111.31</td>
<td>7.05</td>
<td>63</td>
</tr>
<tr>
<td>Critical Thinking</td>
<td>67</td>
<td>110.28</td>
<td>7.61</td>
<td>63</td>
</tr>
</tbody>
</table>

*p < .05
Average Time Spent on Each Item

M_{Control} = 34.36 in seconds, SD_{Control} = 17.00; M_{Experimental} = 49.44, SD_{Experimental} = 22.46

t = 4.36, p < .001, d = .76
Percentage of Not Reached Item

![Graph showing the percentage of not reached item for Control Group and Experimental Group over item numbers. The graph indicates a higher percentage of not reached items in the Control Group compared to the Experimental Group.](image-url)
Unmotivated Students Identified through Item Response Time

Control Group

- Motivated: 42%
- Unmotivated: 58%

Experimental Group

- Motivated: 95%
- Unmotivated: 5%
Performance Difference with/without Filtering

<table>
<thead>
<tr>
<th></th>
<th>No Filtering</th>
<th>Filtering with Response Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0.63</td>
<td>0.23</td>
</tr>
<tr>
<td>Reading</td>
<td>0.72</td>
<td>0.20</td>
</tr>
<tr>
<td>Writing</td>
<td>0.60</td>
<td>0.27</td>
</tr>
<tr>
<td>Math</td>
<td>0.58</td>
<td>0.37</td>
</tr>
<tr>
<td>Critical Thinking</td>
<td>0.53</td>
<td>0.13</td>
</tr>
</tbody>
</table>
ETS’S APPROACH TO NEXT GENERATION SLO ASSESSMENTS
Identifying Core Competencies

- Critical thinking
- Written communication
- Information literacy
- Quantitative literacy
- Civic competency and engagement
- Intercultural competency and diversity
- Oral communication

Qualitative and quantitative market research

Leverage existing R&D capabilities

Input from HEIs and organizations

Research Synthesis

Core Competencies
Considerations of Next Generation Assessment

• Balance between authenticity and psychometric quality
  – Multiple assessment formats

• Consider diversity of higher education population
  – Accessibility
  – Language learner

• Align with instruction
  – Faculty involvement
  – Customization
Current Research on Next Generation Assessment

Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment

Ou Lydia Liu
Lois Frankel
Katrina Crots Roohr

Assessing Quantitative Literacy in Higher Education: An Overview of Existing Research and Assessments With Recommendations for Next-Generation Assessment

Katrina Crots Roohr
Edith Aurora Graf
Ou Lydia Liu

Assessing Written Communication in Higher Education: Review and Recommendations for Next-Generation Assessment

Jesse R. Sparks
Yi Song
Wyman Brantley
Ou Lydia Liu

More research to come soon!

April 2014

September 2014

December 2014
An Example: Quantitative Literacy Framework Development

• Reviewed existing frameworks from
  – National and international organizations
  – Workforce initiatives
  – Higher education institutions and researchers
  – K-12 theorists and practitioners

• Reviewed existing assessments
  – E.g., CAAP mathematics, CLA+ scientific and quantitative reasoning, EPP mathematics
Broad Issues in Assessing Quantitative Literacy

- Mathematics versus quantitative literacy
- General versus domain specific
- Total scores versus subscores
- Student motivation
Theoretical Framework Guiding Assessment Development

• 5 Mathematical Problem-Solving Skills
  – Interpretation, strategic knowledge and reasoning, modeling, computation, and communication

• 4 Mathematical Content Areas
  – Number and operations, algebra, geometry and measurement, probability and statistics

• 3 Real-World Contexts
  – Personal/everyday life, workplace, society
Assessment Structure

- Computer-based assessment
- 45 minute assessment
- 25 test items
- Items cover primary problem-solving skills and content in a variety of real-world contexts
- An on-screen four-function calculator will be provided for the test taker
The chart shows the results of a survey of people at a mall who were asked, “What is your favorite flavor of ice cream, vanilla, chocolate, or strawberry?” Each person selected only one flavor, and every person surveyed had a favorite.

Based on the data shown, indicate which of the following statements are true or false.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/16 of the people surveyed preferred chocolate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50% more women than men preferred strawberry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than half of those surveyed preferred vanilla</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Potential Sources of Construct-Irrelevant Variance

- Accessibility to all students (e.g., students with disabilities and ELs)
  - Need to consider multiple delivery modes, and methods for accessing questions and entering responses
- Technology-enhanced item types
  - Need to have clear directions
  - Should not be over-used
- Computer-based test
  - Possible barrier of completing quantitative items on a computer
- Cognitive reading load
  - Test should measure quantitative skills, not reading ability
References (1)


References (3)


