

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 Crofts 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

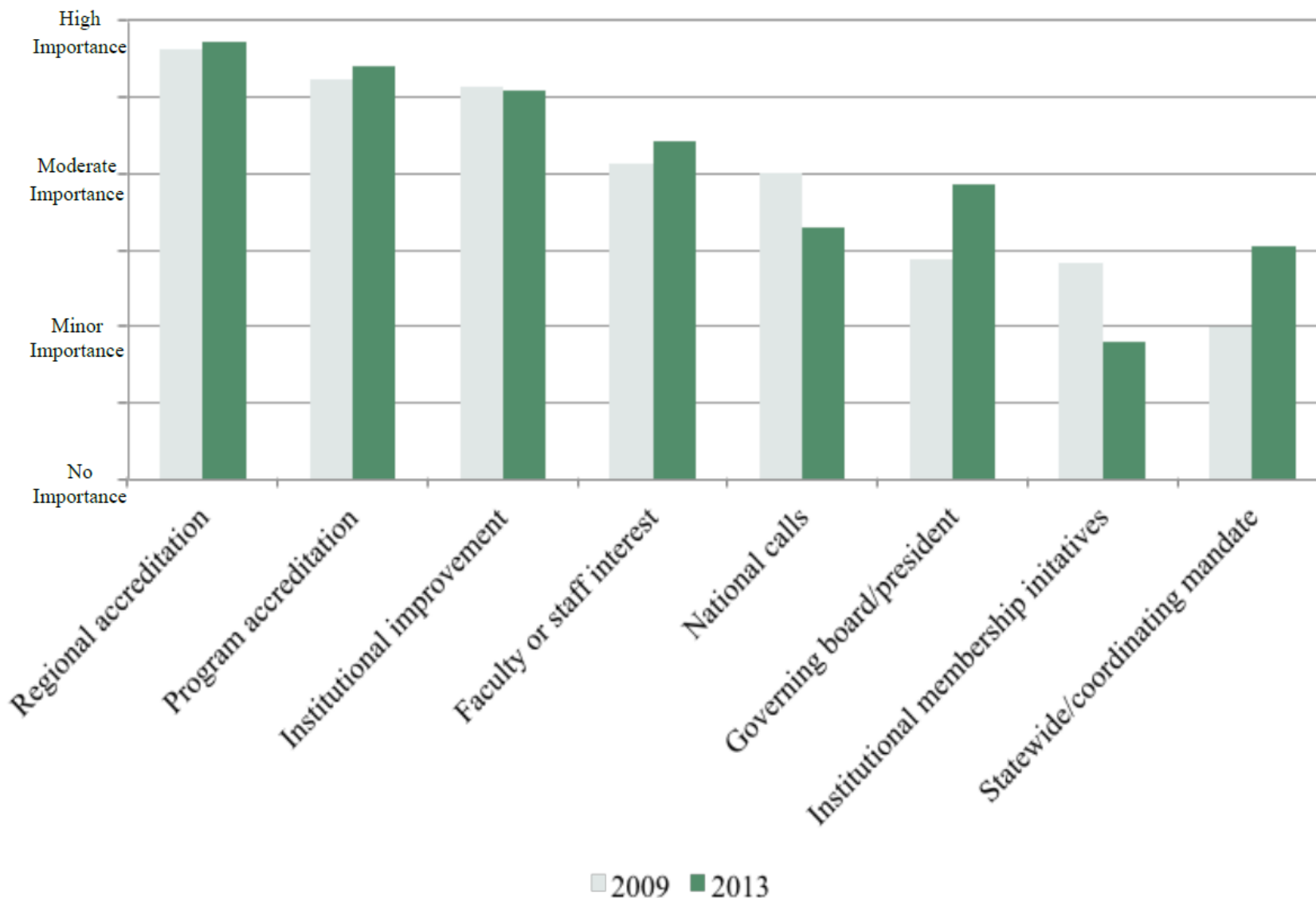
- Rapid development of higher education
 - 15.9 million to 21.0 million students from 2001 to 2011 (Snyder & Dillow, 2013)
- National goal of higher education
 - By 2020, America should have the highest proportion of college graduates (Obama, 2009)
- Call for quality assurance in higher education
 - “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





Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014, p.11). *Knowing what students know and can do: The current state of student learning outcomes assessment in U.S. colleges and universities*. Champaign, IL: National Institute for Learning Outcomes Assessment.

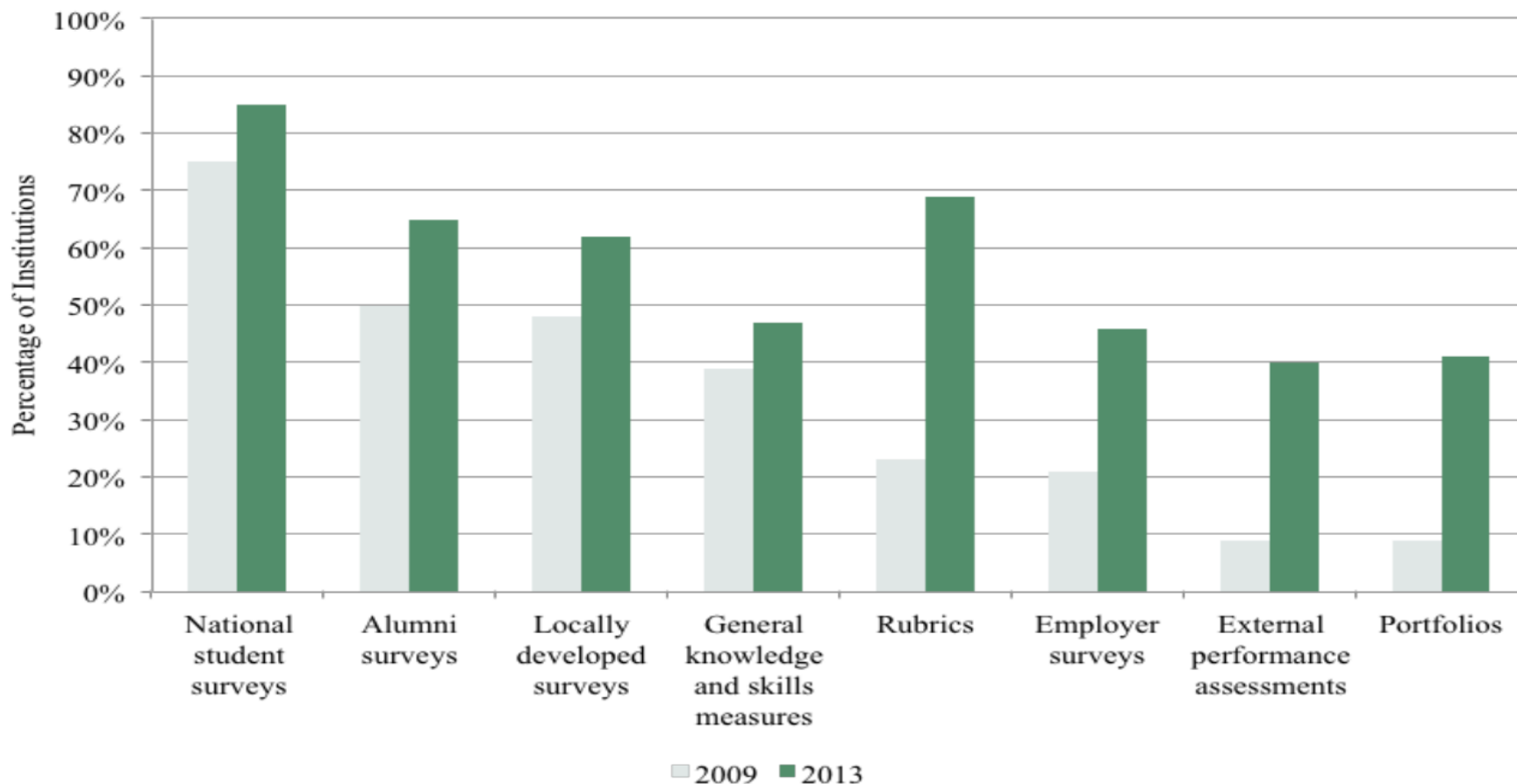


Current Use of SLO Assessment

- 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



Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014, p.14). *Knowing what students know and can do: The current state of student learning outcomes assessment in U.S. colleges and universities*. Champaign, IL: National Institute for Learning Outcomes Assessment.



Comparison of Assessment Tools

Tool	Advantages	Disadvantages
Survey	Cost efficient; easy administration; comparison	No direct evidence of student learning
Locally developed survey	Aligned with instruction; meet institution's specific needs	No benchmark with other institutions; sometimes lack psychometric quality
Standardized measures	Comparable across institutions; sufficient validity and reliability evidence	Insufficient alignment with instruction
Rubrics	Flexibility for adaptation	Poor consistency among users
Performance assessment	Authentic	Expensive; difficult to implement; poor reliability
e-portfolio	Offer a range of data	Comparability is an issue



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)
- Indicators 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)
- Strategies of varying effectiveness (Braun et al., 2011; Kim & McLean, 1995; Liu et al., 2012; O'Neil, Sugrue, & Baker, 1995/1996)



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

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 evaluate the quality of instruction at your college. This may affect 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 compensation for your time and effort. We thank you for your voluntary participation and ask you to please fill out 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 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.



Institutional Condition

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 evaluate the quality of instruction at your college. This may affect 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 compensation 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 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.



Personal Condition

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 evaluate the quality of instruction at your college. This may affect 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 compensation for your time and effort. We thank you for your voluntary participation and ask you to please fill out 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.

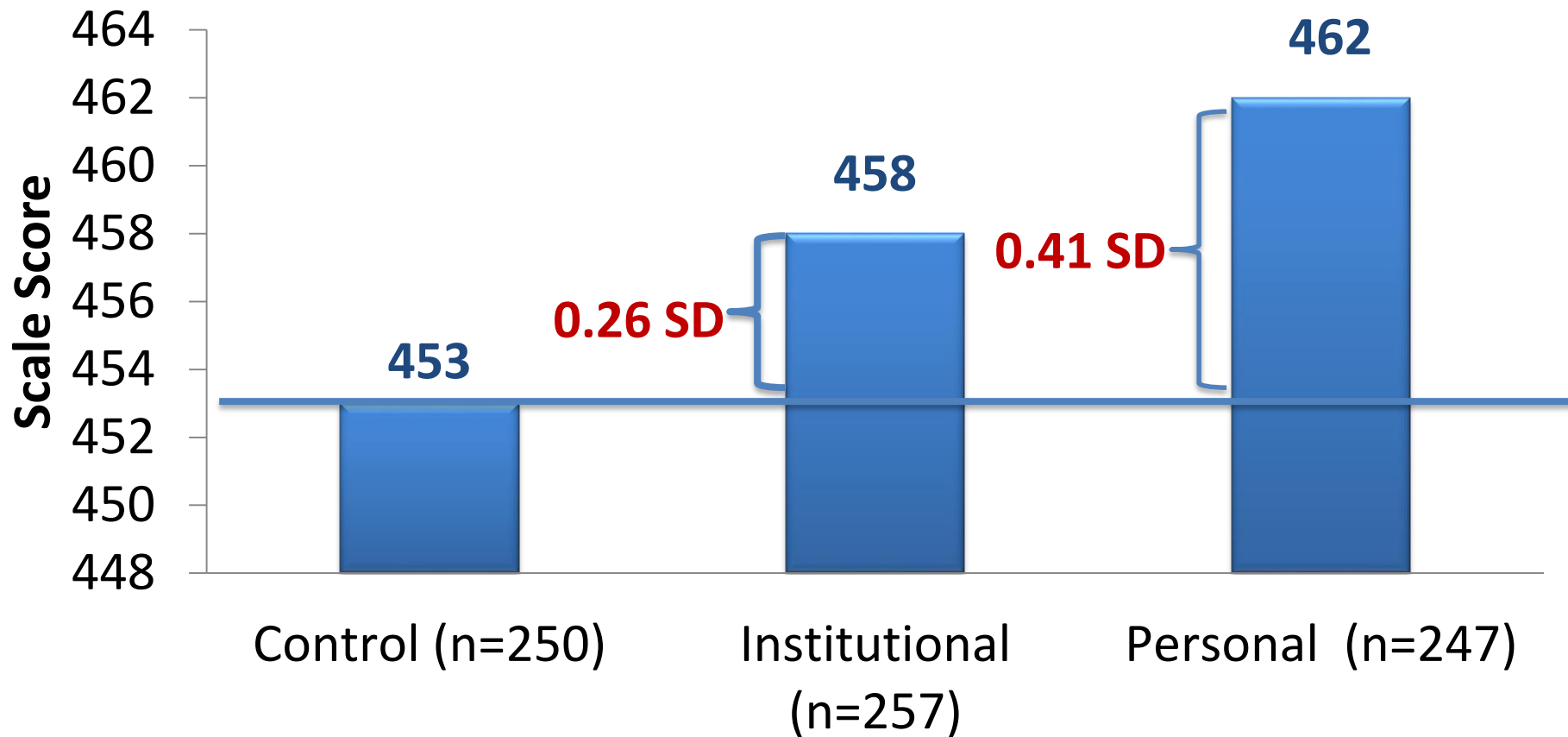


Results

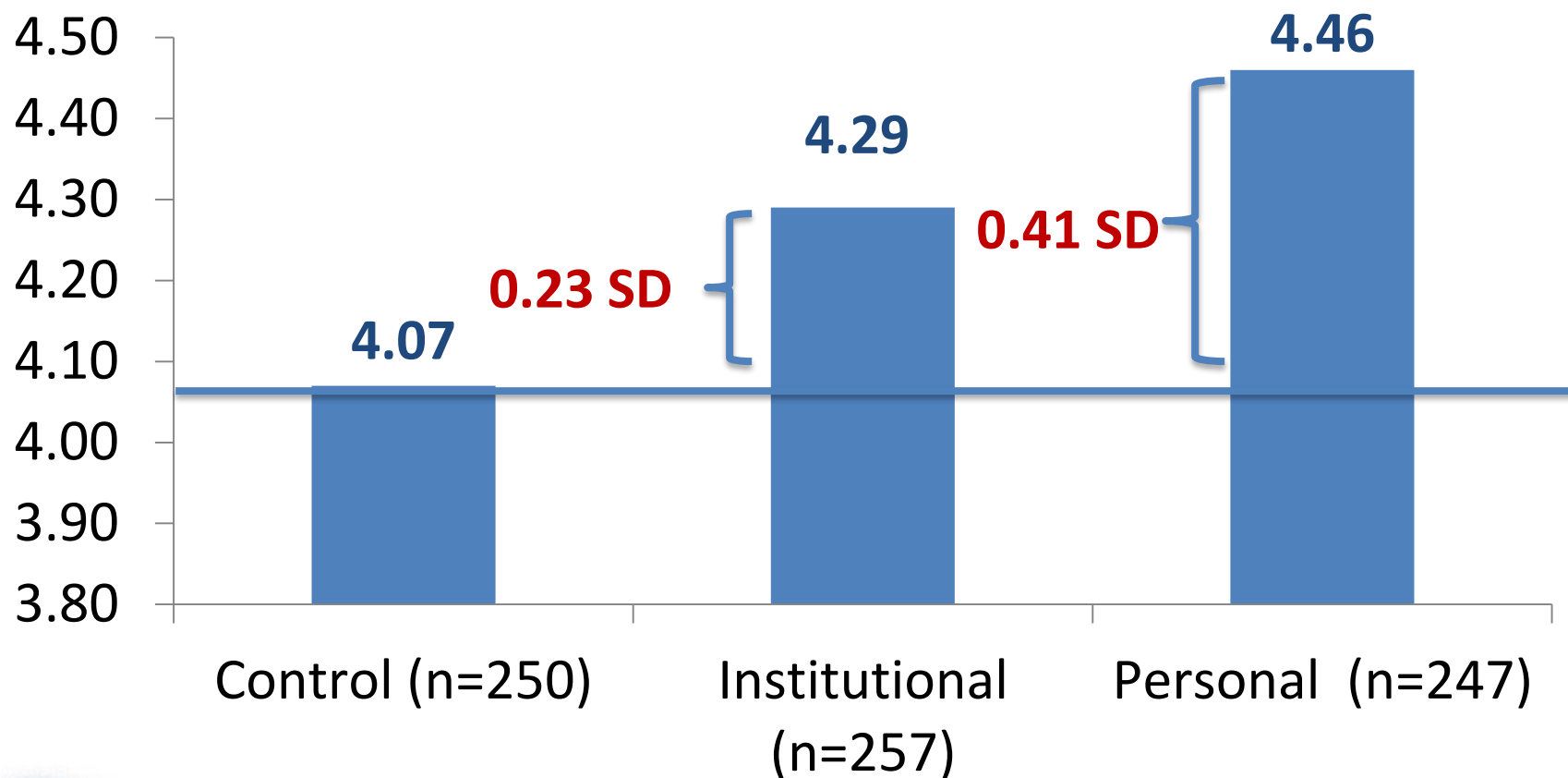
- Motivational instruction has a significant impact on both EPP scores and students' self-reported motivation



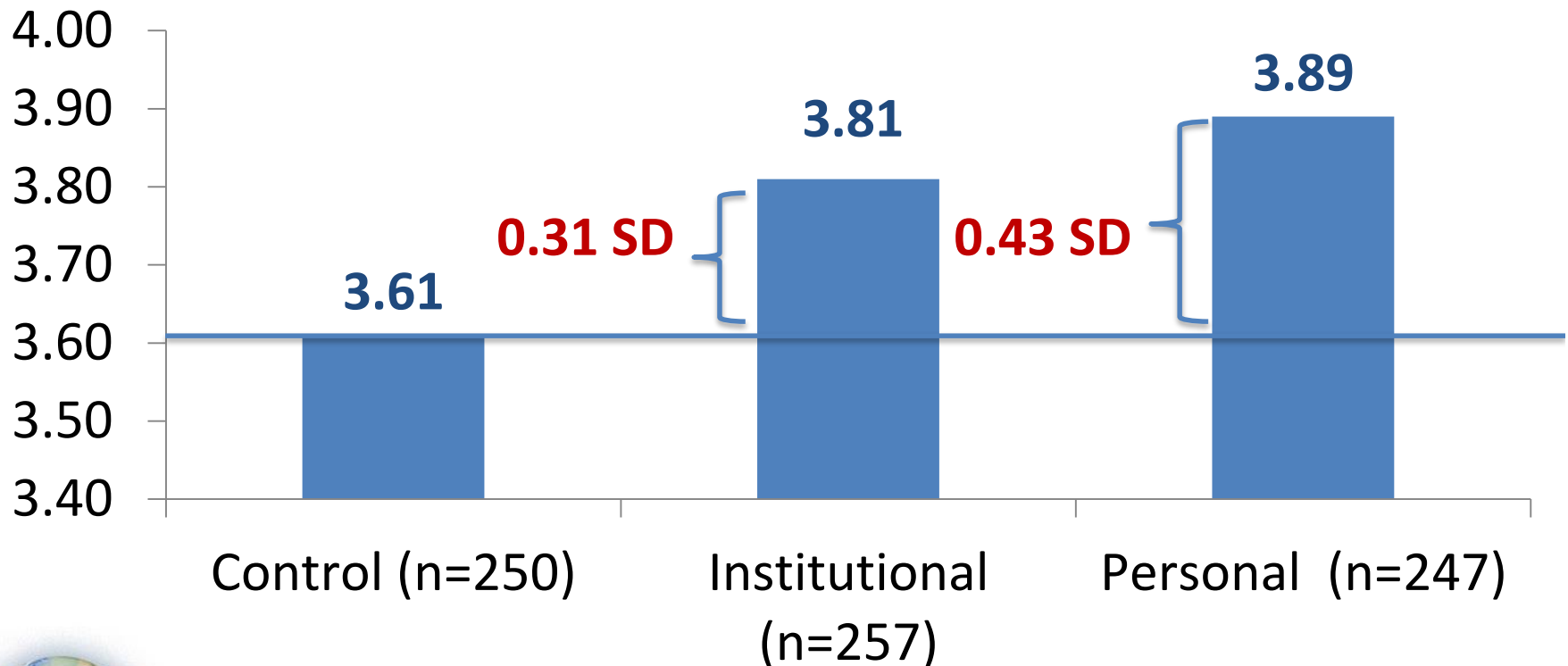
How Motivational Instructions Affected the ETS Proficiency Profile



How Motivational Instructions Affected the Essay



How Motivational Instructions Affected Self-report Motivation



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.”



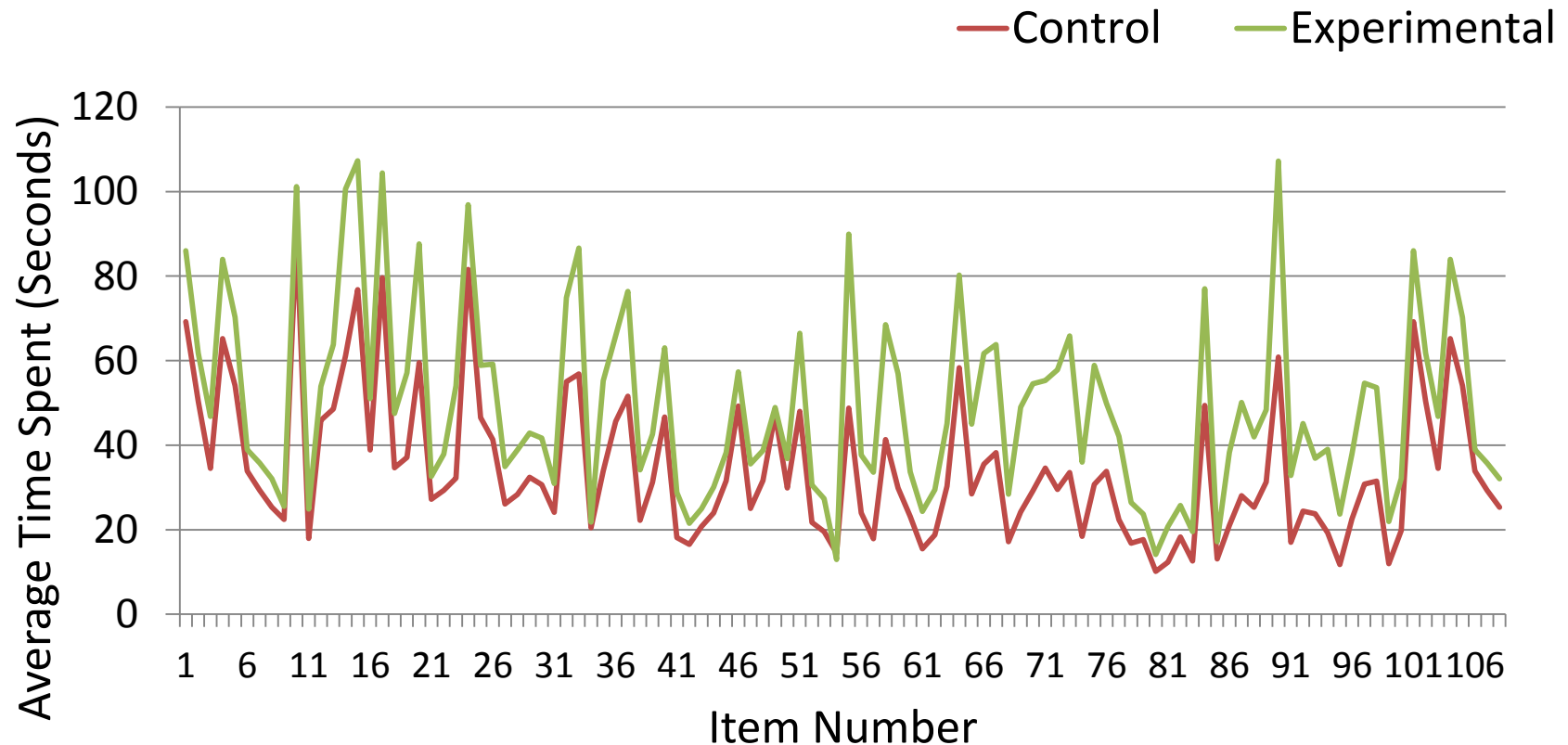
Difference in EPP Performance

	Control				Experimental				
EPP Score	N	Mean	SD		N	Mean	SD	t	d
Total	67	436.25	24.50		63	450.40	20.32	3.59*	0.63
Reading	67	114.27	8.74		63	119.83	6.50	4.13*	0.73
Writing	67	112.58	5.83		63	115.71	4.57	3.42*	0.60
Math	67	111.31	7.05		63	115.37	6.94	3.30*	0.58
Critical Thinking	67	110.28	7.61		63	114.11	6.68	3.05*	0.54

** $p < .05$*



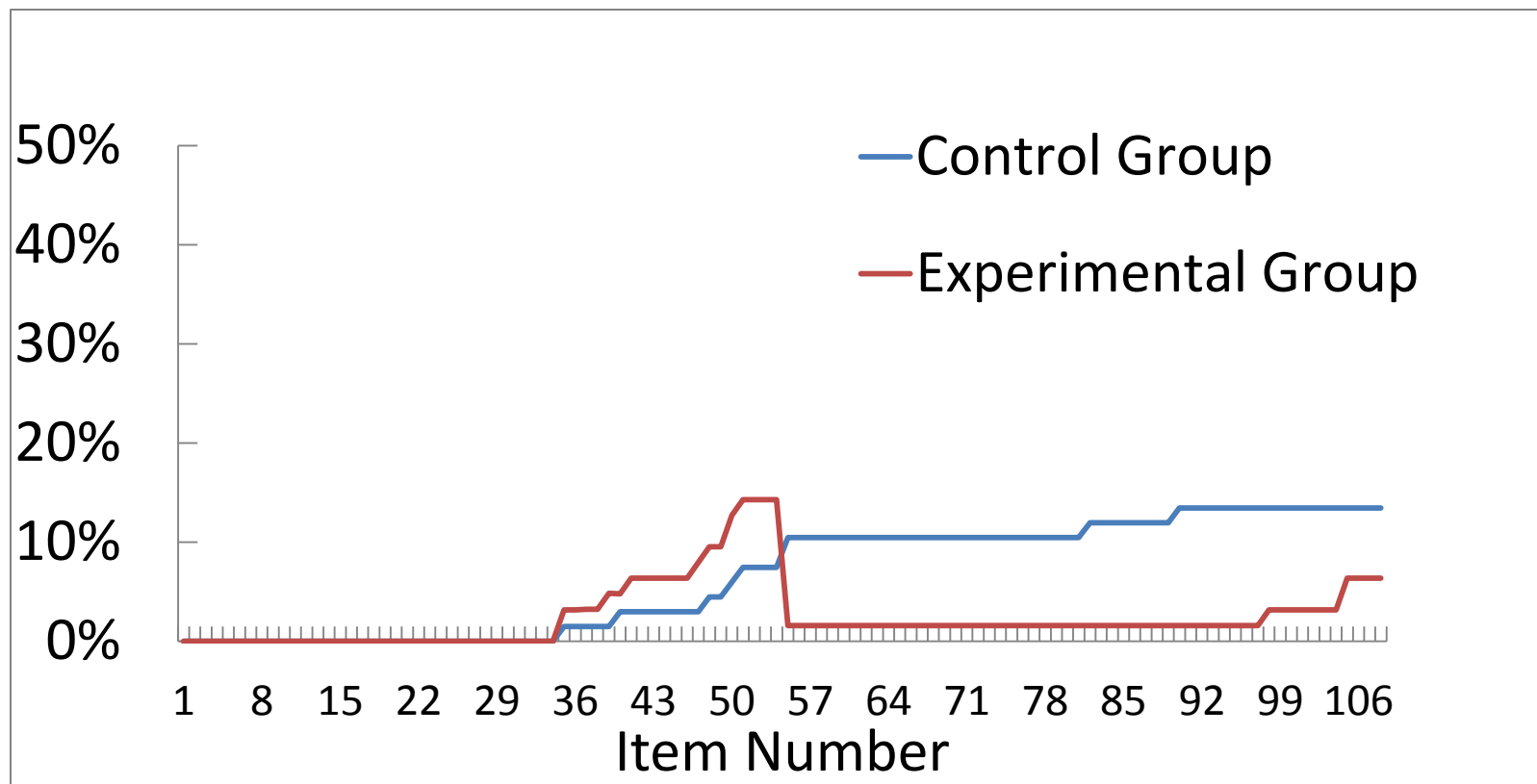
Average Time Spent on Each Item



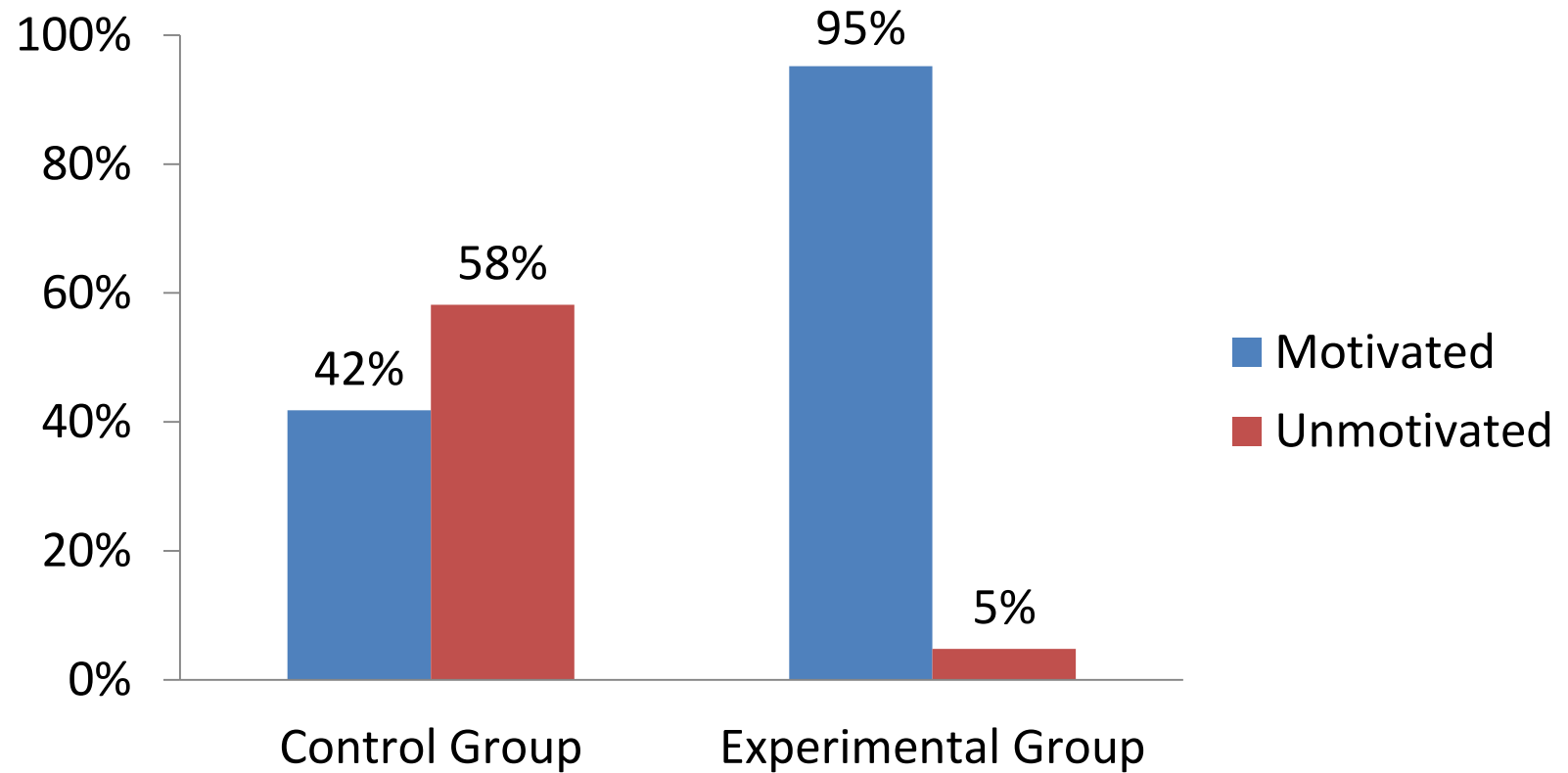
$M_{Control} = 34.36$ in seconds, $SD_{Control} = 17.00$; $M_{Experimental} = 49.44$, $SD_{Experimental} = 22.46$
 $t = 4.36$, $p < .001$, $d = .76$



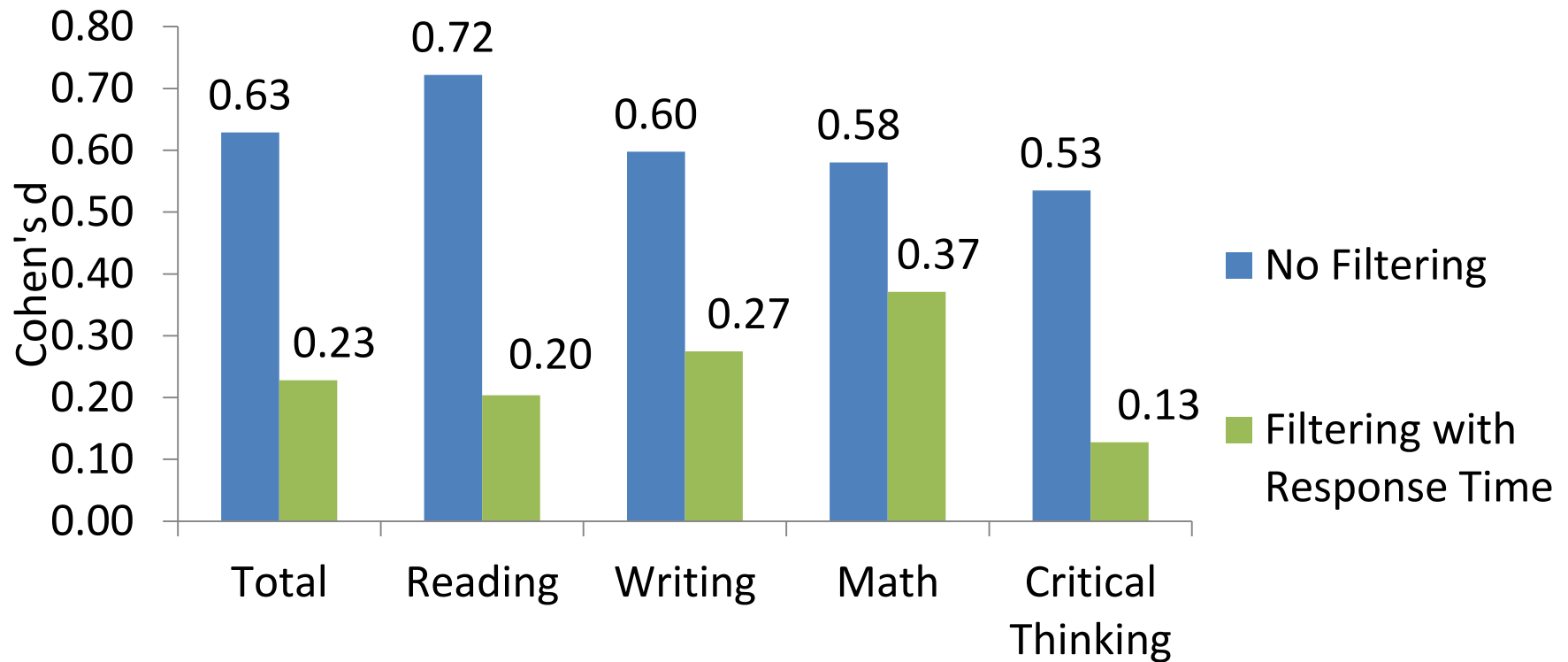
Percentage of Not Reached Item



Unmotivated Students Identified through Item Response Time



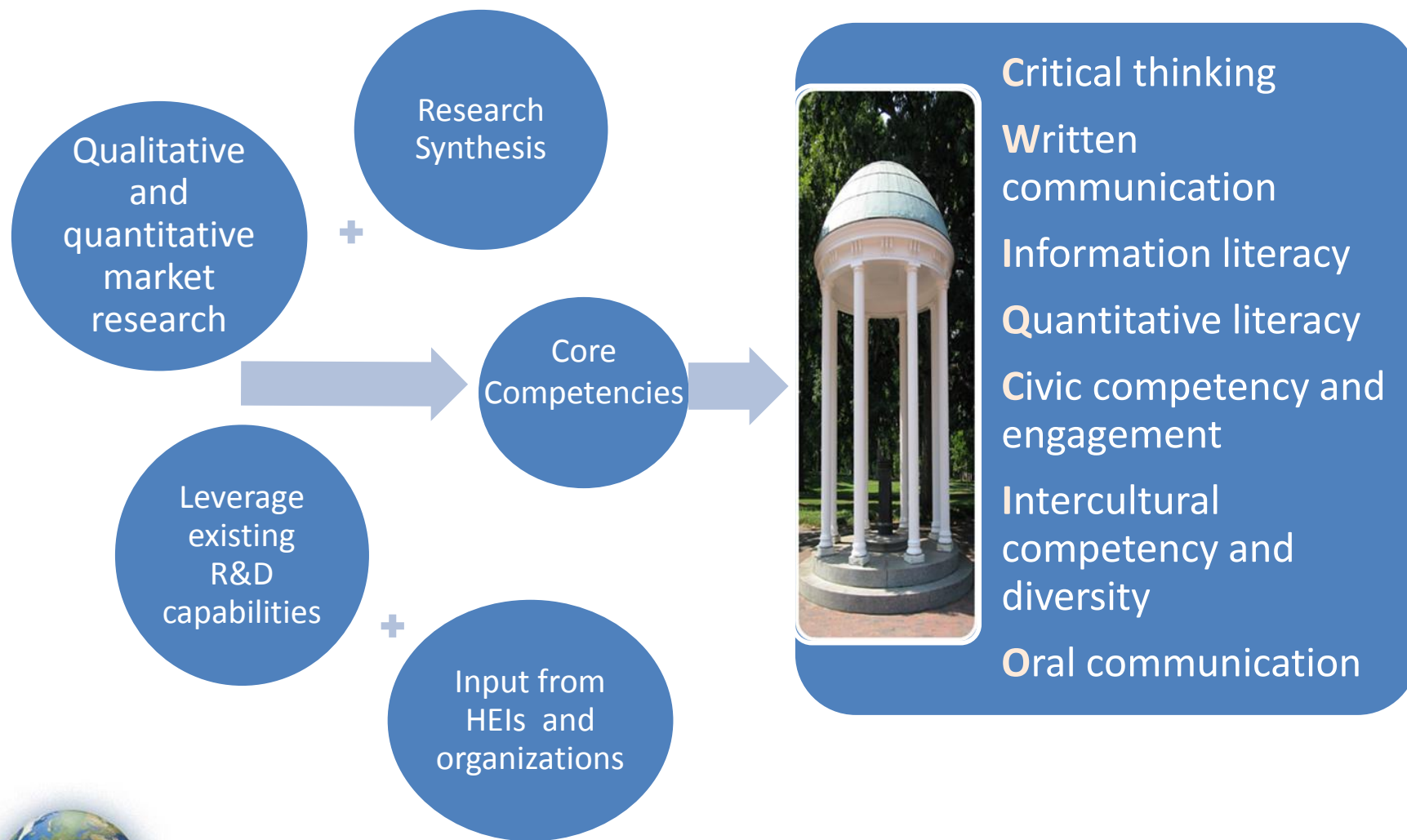
Performance Difference with/without Filtering



ETS'S APPROACH TO NEXT GENERATION SLO ASSESSMENTS



Identifying 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



Listening. Learning. L



Listening. Learning. Lead



Listening. Learning. Leading.®

Research Report ETS RR-14-10

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

Ou Lydia Liu
Lois Frankel
Katrina Crotts Roohr

Research Report ETS RR-14-22

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

Katrina Crotts Roohr
Edith Aurora Graf
Ou Lydia Liu

Research Report ETS RR-14-37

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

Jesse R. Sparks
Yi Song
Wyman Brantley
Ou Lydia Liu

April 2014

Discover this journal at
Wiley Online Library
wileyonlinelibrary.com

September 2014

Discover this journal at
Wiley Online Library
wileyonlinelibrary.com

December 2014

Discover this journal at
Wiley Online Library
wileyonlinelibrary.com

More research to come soon!



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

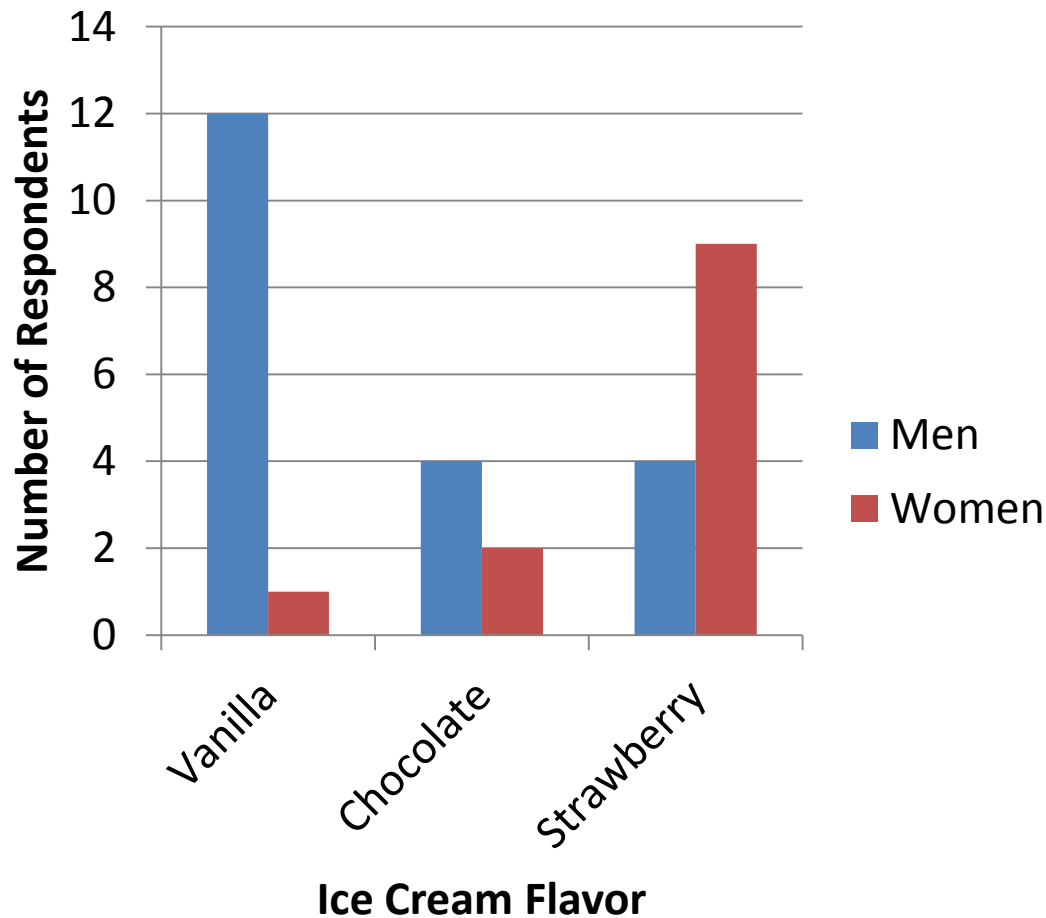


Assessment Structure

- 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



Sample Item



Statement	True	False
3/16 of the people surveyed preferred chocolate		
50% more women than men preferred strawberry		
Less than half of those surveyed preferred vanilla		

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)

- Arum, R., Cho, E., Kim, J., & Roksa, J. (2012). *Documenting uncertain times: Post-graduate transitions of the Academically Adrift cohort*. New York, NY: Social Science Research Council.
- Arum, R., & Roksa, J. (2011). *Academically Adrift*. Chicago: University of Chicago Press Books.
- Barry, C. L., Horst, S. J., Finney, S. J., Brown, A. R., & Kopp, J. P. (2010). Do examinees have similar test-taking effort? A high-stakes question for low-stakes testing. *International Journal of Testing*, 10, 342–363.
- Braun, H., Kirsch, I., & Yamamoto, K. (2010). *An experimental study of the effects of monetary incentives on performance on the 12th grade NAEP reading assessment*. Princeton, NJ: Educational Testing Service.
- Butler, H. A. (2012). Halpern Critical Thinking Assessment predicts real-world outcomes of critical thinking. *Applied Cognitive Psychology*, 25(5), 721–729.
- Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences*, 108, 7716–7720.
- Ejiogu, K. C., Yang, Z., Trent, J., & Rose, M. (2006, May). *Understanding the relationship between critical thinking and job performance*. Poster presented at the 21st annual conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Hendel, D. D. (1991). Evidence of convergent and discriminant validity in three measures of college outcomes. *Educational and Psychological Measurement*, 51, 351–358.
- Kim, J. G., & McLean, J. E. (1995, April). *The influence of examinee test-taking motivation in computerized adaptive testing*. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, CA.
- Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014). *Knowing what students know and can do: The current state of student learning outcomes assessment in U.S. colleges and universities*. Champaign, IL: National Institute for Learning Outcomes Assessment.



References (2)

- Lakin, J., Elliott, D., & Liu, O. L. (2012). Investigating the impact of ELL status on higher education outcomes assessment. *Educational and Psychological Measurement*, 72, 734-753.
- Liu, O. L., Rios, J. A., & Borden, V. (accepted). The effects of motivational instruction on college students' performance on low-stakes assessment. *Educational Assessment*.
- Liu, O. L. (2011a). Outcomes assessment in higher education: Challenges and future research in the context of Voluntary System of Accountability. *Educational Measurement: Issues and Practice*, 30(3), 2-9.
- Liu, O. L. (2011b). Value-added assessment in higher education: A comparison of two methods. *Higher Education*, 61(4), 445-461.
- Liu, O. L., Bridgeman, B., & Adler, R. (2012). Measuring learning outcomes assessment in higher education: Motivation matters. *Educational Researcher*, 41, 352 – 362.
- Liu, O. L., Frankel, L., & Roohr, K. C. (2014). *Assessing critical thinking in higher education: Current state and directions for next-generation assessment* (ETS RR-14-10). Princeton, NJ: Educational Testing Service.
- Markle, R., Brenneman, M., Jackson, T., Burrus, J., & Robbins, S. (2013). *Synthesizing frameworks of higher education student learning outcomes* (ETS RR-13-22). Princeton, NJ: Educational Testing Service.
- Marr, D. (1995). *Validity of the academic profile*. Princeton, NJ: Educational Testing Service.
- O'Neil, Jr. H. F., Sugrue, B., & Baker, E. L. (1995/1996). Effects of motivational interventions on the National Assessment of Educational Progress mathematics performance. *Educational Assessment*, 3, 135-157.
- Obama, B. (2009). *President Obama's Address to Congress*. Retrieved from http://www.nytimes.com/2009/02/24/us/politics/24obama-text.html?_r=2 on Feb 20 2010.



References (3)

- Rios, J., Liu, O.L., & Bridgeman, B. (2014). Identifying unmotivated examinees on student learning outcomes assessment: A comparison of two approaches. *New Directions for Institutional Research*, 161, 69-82.
- Roohr, K. C., Graf, E. A., & Liu, O. L. (2014). *Assessing quantitative literacy in higher education: An overview of existing research and assessments with recommendations for next-generation assessment* (ETS Research Report No. RR-14-22). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12024.
- Sparks, J. R., Song, Y., Brantley, W., & Liu, O. L. (2014). *Assessing written communication in higher education: Review and recommendations for next-generation assessment* (ETS RR-14-37). Princeton, NJ: Educational Testing Service.
- Snyder, T. D., & Dillow, S. A. (2013). *Digest of Education Statistics 2012* (NCES 2014-015). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington D.C.
- Sundre, D.L. (1999). *Does examinee motivation moderate the relationship between test consequences and test performance?* (Report No. TM029964). Harrisonburg, Virginia: James Madison University. (ERIC Document Reproduction Service No.ED432588).
- Sundre, D. L., & Wise, S. L. (2003, April). *Motivation filtering: An exploration of the impact of low examinee motivation on the psychometric quality of tests*. Paper presented at the meeting of the National Council on Measurement in Education, Chicago, IL.
- U.S. Department of Education (2006). *A test of leadership: Charting the future of American higher education*. (Report of the commission appointed by Secretary of Education Margaret Spellings). Washington, DC: Author.
- Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10, 1-17.
- Wise, S. L., & DeMars, C. E. (2010). Examinee noneffort and the validity of program assessment results. *Educational Assessment*, 15, 27-41.
- Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18, 163-183.
- Wise, S. L., & Ma, L. (2012, April). *Setting response time thresholds for a CAT item pool: The normative threshold method*. Paper presented at the meeting of the National Council on Measurement in Education, Vancouver, Canada.

