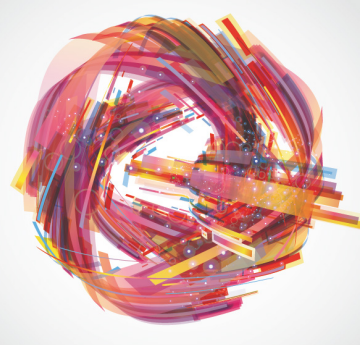

Abstract

Increased demands for accountability have placed an emphasis on assessment of student learning outcomes. At the post-secondary level, many of the assessments are considered low-stakes, as student performance is linked to few, if any, individual consequences. Given the prevalence of low-stakes assessment of student learning, research that investigates the relationship between student motivation, effort, and performance on low-stakes tests is warranted as these tests are increasingly being used to make judgments about the quality of student learning. This quasi-experimental study was conducted at a public mid-sized university with 87 undergraduate students enrolled in four 100-level general education courses. The researchers examined the effects of motivational prompts on student motivation, effort, and performance on a low-stakes test. Results indicated that motivational condition had a significant effect on students' performance as measured by total mean scores on a low-stakes standardized test. Students in the personal motivational condition outperformed students in the other conditions. However, motivational prompts were not found to affect students' critical thinking subscores or self-reported effort and importance scores.

AUTHORS

Katrice A. Hawthorne, Ph.D.
Norfolk State University

Linda Bol, Ph.D.
Old Dominion University

Shana Pribesh, Ph.D.
Old Dominion University

Yonghee Suh, Ph.D.
Old Dominion University

Effects of Motivational Prompts on Motivation, Effort, and Performance on a Low-Stakes Standardized Test

Increased demands for accountability affect education at elementary, secondary, and post-secondary levels and have placed an emphasis on assessment of student learning outcomes (Wise & DeMars, 2005), often via standardized tests. Notable shifts from changing student demographics to new delivery formats (e.g., distance learning and massive open online courses) are also occurring throughout higher education in the United States. Accountability in higher education has "received unprecedented attention" as a result of these and other shifts, which have called into question the ambiguous accountability and assessment methods of colleges and universities (Liu, 2011, p. 21). In addition, a number of recent reports (Arum & Roksa, 2011; Baer, Cook, & Baldi, 2006) have led policymakers and stakeholders to question student learning in higher education. Institutions generally respond to questions about quality and accountability by providing evidence of graduation rates, licensure pass rates, and graduate and professional school admissions rates; however, these data fail to provide even an overview of what students are actually learning (Millett, Payne, Dwyer, Stickler, & Alexiou, 2008).

CORRESPONDENCE

Email
khawt002@odu.edu

In accordance with K-12 accountability efforts, conclusions about the quality of higher education are increasingly being based on learning outcomes assessment data. At the post-secondary level, many of the assessments used are considered low-stakes. Tests that have minimal or no consequences for the individual test taker are generally considered non-consequential or low-stakes, while tests that affect grades, admissions, or graduation are often referred to as consequential or high-stakes (Waskiewicz, 2011).

Previous research has examined K-12 student performance on national and international standardized assessments (O'Neil, Sugrue, & Baker, 1996), but much of the research in higher education relies on graded versus ungraded instructor developed pre- and post-tests (Boyas, Bryan, & Lee, 2012; Sundre & Kitsantas, 2004). In addition, much of the

research on performance differences between motivated and unmotivated test takers examines motivation through the use of incentives, including monetary compensation and extra credit points (O'Neil et al., 1996; Wise & DeMars, 2005). Without consequences or incentives, it is assumed that students will not perform to the best of their ability on low-stakes tests; thus, the results are not valid indicators of their knowledge and abilities (Wise & DeMars, 2005). The research on university students' motivation and performance on low-stakes tests suggests that students who are motivated and invest effort score higher than those who do not (Cole, Bergin, & Whitaker, 2008). The use of locally developed instruments raises questions about whether the findings linking motivation and student performance can be extended to include the use of standardized tests in low-stakes contexts in the college classroom (Liu, Bridgeman, & Adler, 2012). Few studies (e.g., Liu et al., 2012; Waskiewicz, 2011) have examined university students' motivation using standardized outcomes assessment instruments.

Standardized tests of general academic competencies (i.e., writing and critical thinking) are increasingly being used in higher education as evidence of student learning (Hoyt, 2001; Liu, 2011). According to a report by ETS®, nearly 1,400 institutions of higher education have used at least one standardized outcomes assessment test (Liu, 2011). The results from these tests are reported with the "implicit assumption that the scores represent the best effort the student[s] could put forth" (Wolf & Smith, 1995, p. 227). Yet despite widespread use of standardized outcomes assessment tests and the low stakes connected to student performance, there is little empirical evidence on the interpretation of these assessment results (Liu, 2011). A primary concern regarding the implementation and interpretation of standardized outcomes measures is the low-stakes nature of the task and the resultant lack of motivation and effort on the part of students to perform to the best of their ability (Hoyt, 2001; Liu, 2011; Wise & DeMars, 2005).

These assessments may have significant implications for institutions, yet many students may fail to see the individual consequences as the tests do not directly affect course grades or their standing at the university.

Accountability and Low-Stakes Assessment

Outcomes assessment is now required by all regional higher education accrediting associations (Hoyt, 2001) and by many discipline-specific associations (Boyas et al., 2012; Waskiewicz, 2011). Publicly funded institutions of higher education, which have traditionally relied on enrollment-driven funding (Hoyt, 2001), are increasingly being asked to demonstrate student learning and to justify expenditures of taxpayer dollars based upon the results from low-stakes assessments (Wise & DeMars, 2005; Wise & Kong, 2005). The use of standardized low-stakes assessments is growing despite widespread concern that low-stakes assessments may underestimate student learning (Baumert & Demmrich, 2001). These assessments may have significant implications for institutions, yet many students may fail to see the individual consequences as the tests do not directly affect course grades or their standing at the university. Thus, research that examines the conditions that affect motivation and effort on student performance on low-stakes tests is warranted as these tests are increasingly being used to make judgments about the quality of student learning.

Expectancy-Value Theory

Motivation is a dynamic, multifaceted phenomenon that is situated, contextual, and domain-specific (Linnenbrink & Pintrich, 2002; Pintrich, 1989). Expectancy-value theory offers an important view of the nature of achievement motivation (Wigfield, 1994). The expectancy-value theory of achievement developed initially by Eccles in 1983 and later refined by Wigfield and Eccles (2000) serves as the framework for this study and much of the research on student motivation and performance on low-stakes tests (e.g., Swerdzewski, Harmes, & Finney, 2009; Waskiewicz, 2011). In expectancy-value models of achievement motivation, *expectancy* is defined as a student's belief that he or she can complete a task successfully, and *value* is defined as a student's perceptions about why he or she should complete a task (Wigfield & Eccles, 2000). Task value beliefs are defined in terms of intrinsic value (i.e., interest), utility, importance, and cost (Wigfield, 1994).

Expectancy-value theorists argue that "student choice, persistence, and performance can be explained by their expectations about how well they will do on the activity and the extent to which they value the activity" (Wigfield & Eccles, 2000, p. 68). Expectancies and values are assumed to influence performance, effort, persistence, and achievement choices (Wigfield & Eccles, 2000). In the context of low-stakes assessment, expectancy-value models

are expanded to include not only a student's perception of success and value (or importance), but also the perceived level of effort he or she must expend and the intrinsic value or enjoyment gained from completing the task (Wise & DeMars, 2005).

It was hypothesized that personalized motivational prompts would elicit different levels of motivation than generic motivational prompts. Students in both of the personalized conditions (personal and combined) reported higher importance and effort scores than students in the other conditions.

According to Wolf, Smith, and Birnbaum (1995), the expectancy component of the expectancy value model can be extended in testing situations to include students' perceptions of the mental effort necessary to complete the task. Thus, test-taking motivation, which is linked to a specific task (i.e., motivation to perform well on a given test), can be considered a form of achievement motivation (Eklöf, 2010). Studies on test-taking motivation have consistently found motivation to be correlated with test performance and test consequence (Cole et al., 2008).

Motivational Interventions

Nevo (1995) contended that the manipulation of variables related to non-psychometric properties of the test, such as the testing conditions, the face validity of the test, the clarity of test instructions, and the behavior of proctors, can result in improvement of scores among examinees. Much of the research on performance differences between motivated and unmotivated examinees attempts to alter motivation through the use of incentives, specifically monetary compensation and graded versus ungraded assignments (Boyas et al., 2012; O'Neil et al., 1996; Wise & DeMars, 2005). Waskiewicz (2011) examined pharmacy students' test taking motivation on a low-stakes standardized test by randomly assigning students to two groups and providing them with letters from the dean of the school of pharmacy. The letters of the students in the experimental group were personalized and highlighted the need for students to put forth their best effort as the results would help improve curriculum. In contrast, the letters of the students in the control group were not personalized and briefly described how the test would identify limitations in students' knowledge. The experimental group reported putting forth more effort than the control group (Waskiewicz, 2011). Without consequences or incentives, it is assumed that students will not perform to the best of their ability on low-stakes tests and therefore, the results are not valid indicators of their knowledge and abilities (Wise & DeMars, 2005).

In the present study, we used a quasi-experimental design to investigate whether motivational prompts affected student motivation and effort on a standardized low-stakes test. One proctor administered the test ETS® Proficiency Profile (ETS® PP) and the Student Opinion Scale (SOS) in four 100-level general education courses. Additionally, this research addressed whether students' performance was affected by receiving motivational prompts.

Method

Participants

The participants were 87 undergraduate students enrolled in four 100-level general education courses. An email, detailing the study's purpose, was sent to faculty teaching 100- and 200-level general education courses. Four faculty members agreed to have the test and survey administered in their 100-level general education courses. The courses sampled were BIO 100: Introduction to Biological Science, IUL 100: Introduction to University Life, PED 100: Fundamentals of Fitness for Life, and SCI 101: Introduction to Physical Science. The four courses sampled are all 100-level courses included in tiers one and two of the university's three-tiered general education curriculum. The sample consisted of 34 male students (39.1%) and 53 female students (60.9%). Nearly 84% of students were lower-division students (freshmen and sophomores). There was no significant difference in students' ability as measured by SAT critical reading and mathematics scores.

Instruments

The participants completed both the abbreviated ETS® Proficiency Profile (ETS® PP) and the Student Opinion Scale (SOS) between May 2013 and August 2013. The four courses were assigned to one of four conditions: (a) a control condition, (b) a university condition, (c) a personal condition, and (d) a combined university/personal condition. Students were administered the abbreviated version of the ETS® PP and completed the SOS immediately after.

The abbreviated paper-pencil form of the ETS® PP assesses four core area skills (critical thinking, reading, writing, and mathematics) in the context of the humanities, the natural sciences, and the social sciences (Young, 2007). The ETS® PP is a 36-item, 40-minute timed multiple-choice test. The critical thinking subscore was used as critical thinking questions generally require more cognitive effort than other items. The internal consistency reliability for the ETS® PP ranges from .80 to .89 (Liu, 2008). ETS® PP total scores range from 400 to 500, while subscores range from 100 to 130.

The SOS is a 10-item, Likert-type instrument that measures examinee motivation (Sundre, 2007). The SOS consists of two subscales, importance and effort, and the items are measured on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Internal reliability for use in general education programs was evaluated using Cronbach's alpha and consistent scores were obtained for the importance subscale, .82, and the effort subscale, .86. Possible scores for both the importance and effort subscales range from 5 to 25 (Sundre, 2007).

Procedures

This study was modeled after a study conducted by Liu et al. (2012), which used the online abbreviated version of the ETS® PP and the SOS. However, unlike Liu et al., the test and survey were administered in intact classrooms and included an additional condition, referred to as the combined university/personal condition. Students were told that their test performance would not be linked to their course grade or affect their standing at the university, but they were asked to include their university student identification number on the ETS® PP and the SOS. The four classrooms were assigned to one of four conditions, and students received motivational prompts verbally from the proctor and in writing. An analysis of variance (ANOVA) was conducted to determine whether students' reported effort and importance on the SOS as well as students' performance on the ETS® PP differed based on the receipt of motivational prompts. An analysis of covariance (ANCOVA) was also conducted to help control for the effect of prior student ability on test performance. Since this study randomly assigned intact groups to one of the four conditions, the use of ANCOVA is appropriate as it reduces bias associated with initial chance differences between the groups (Huitema, 2007).

This finding suggests that altering instructions to include personalized motivational prompts may positively impact students' performance on standardized tests.

Results

The first research question addressed was, "Is there a difference in performance for students who received test instructions with motivational prompts compared to students who did not receive test instructions with motivational prompts?" As indicated in Table 1, students in the personal condition received higher total mean scores and higher critical thinking subscores on the ETS® PP than students in the other conditions. The total mean score for all students tested nationally is 441.6, and the critical thinking subscore is 111.2. Therefore, while the mean score is higher for students in the personal condition, nationally, these scores place students in the 44th percentile and the 41st percentile, respectively.

Table 1
Total Mean Score and Mean Critical Thinking Subscore by Condition

Condition	*n*	Total *M (SD)*	Critical Thinking *M (SD)*
Control	20	429.30 *(14.254)*	107.50 *(3.763)*
University	23	426.04 *(13.907)*	106.70 *(5.040)*
Personal	20	437.40 *(14.412)*	109.40 *(5.529)*
Combined	24	425.88 *(13.829)*	107.25 *(5.542)*

An ANOVA was conducted to investigate the effect of the motivational conditions on test performance. It was hypothesized that the motivation of students who received personalized motivational prompts would be different from the motivation of students who did not receive personalized motivational prompts. Results from the one-way ANOVA indicated that the motivational prompts, and as a consequence, condition had a significant effect on the total mean ETS® PP score, $F(3, 83) = 3.035$, $p < .05$, $\eta^2 = .099$. The mean difference between the personal condition and the combined condition was 11.42.

Table 2
ANOVA Table for Total Mean Score

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	1805.352	3	601.784	3.035	.034
Within Groups	16459.982	83	198.313		
Total	18265.333	86			

To determine if there were differences in student ability between the four groups, an ANCOVA was conducted. In the absence of a pre-test, SAT critical reading and math scores were used to determine if students' performance was due to ability. SAT scores were not available for the entire sample; however, the results of 69 students with SAT scores, ETS® PP scores, and SOS scores suggest that students' prior ability was unrelated to student performance on the ETS® PP, $F(3,61) = .364$, $p > .05$, $R^2 = .335$.

For institutions and assessment professionals, this study provides evidence that motivational prompts may impact student performance on low-stakes tests, as students in the personal condition received significantly higher mean ETS© PP scores than students in the other conditions.

The results failed to support the main effect of condition on students' critical thinking subscores, $F(3, 83) = 1.131$, $p > .05$, $\eta^2 = .039$. While students in the personal condition outperformed students in the other three conditions, students in the combined condition received the lowest total mean score and the second lowest critical thinking subscore.

To determine if there were differences in student ability between the four groups, an ANCOVA was conducted. SAT critical reading and math scores were used to determine if students' critical thinking subscores were due to ability. SAT scores were not available for the entire sample; however, the results of 69 students with SAT scores and ETS® PP scores suggest that students' prior ability was unrelated to students' critical thinking performance on the ETS® PP, $F(3,61) = .323$, $p > .05$, $R^2 = .223$.

The second research question addressed was "Is there a difference in motivation for students who received test instructions with motivational prompts compared to students who did not receive test instructions with motivational prompts?" Descriptive statistics of students' motivation by condition as measured by the importance and effort scales of the SOS are presented in Table 3.

Table 3
SOS Importance and Effort Mean Scores by Condition

Condition	*n*	Importance *M (SD)*	Effort *M (SD)*
Control	20	15.90 *(2.972)*	15.80 *(4.112)*
University	23	15.96 *(3.561)*	15.00 *(2.876)*
Personal	20	17.25 *(4.141)*	17.30 *(2.638)*
Combined	24	16.08 *(3.309)*	15.88 *(3.069)*

Average raw SOS importance and effort scores for first-year students in a low-stakes general education assessment context were 14.94 and 17.62, respectively (Sundre, 2007). In this study students in the personal condition reported higher mean importance and effort scores; however, when compared to the normed scores of freshmen, their scores place them in the 70th and the 42nd percentile.

It was hypothesized that personalized motivational prompts would elicit different levels of motivation than generic motivational prompts. Students in both of the personalized conditions (personal and combined) reported higher importance and effort scores than students in the other conditions. However, while students in the personal condition indicated higher mean importance scores than students in the other conditions, the results from the one-way ANOVA indicated that motivational condition had no significant effect on students' importance scores, $F(3, 83) = .676$, $p > .05$, $\eta^2 = .023$. Students in the personal condition also indicated higher mean effort scores than students in the other conditions; however, the difference did not reach statistical significance, $F(3, 83) = 1.877$, $p > .05$, $\eta^2 = .064$.

To determine if students' motivation was related to differences in student ability, an ANCOVA was conducted. SAT scores were not available for the entire sample; however, the results of 69 students with SAT scores, ETS® PP scores, and SOS scores suggest that students' prior ability was unrelated to effort, $F(3,61) = .810$, $p > .05$, $R^2 = .167$, or importance, $F(3,61) =$

.107, $p > .05$, $R^2 = .045$. Thus, the relationship between motivation and performance was not due to students' prior ability.

Discussion

The university has used the ETS® PP since 2009 to assess its general education learning outcomes. General education course instructors have been encouraged to use the results to identify student strengths and weaknesses and to evaluate and inform teaching and learning. However, the low-stakes nature of the test had raised questions regarding the validity of the test results, and concomitantly the soundness of altering instruction or curriculum based on such results.

The purpose of the current study was to explore the use of motivational prompts to motivate and communicate to students the usefulness of low-stakes assessment. The varying instructions were designed to manipulate student motivation by appeals to their "academic citizenship" (i.e., the values and behaviors expected of university students; Macfarlane, 2007). The expectation was that personalized motivational prompts would impact students' motivation to perform well on the ETS® PP despite the test's low-stakes nature. Previous research (Baumert & Demmrich, 2001; Liu et al., 2012; O'Neil et al., 1995; Waskiewicz, 2011) suggests that altering test instructions in low-stakes testing contexts might appeal to students' varying goal orientations. In addition, studies that examine the use of practical strategies to motivate students are needed as they have the potential to allow researchers to better isolate the variables that affect motivation and to develop testing models that best demonstrate student learning in low-stakes contexts.

This study, while modeled after Liu et al. (2012), included notable differences that may explain the mixed results. Unlike participants in the Liu et al. study, students in this study did not receive a monetary incentive to participate, and the test was embedded in the course. As a result, while the instructors volunteered to have the test embedded in the course, students did not self-select to participate. Although the test was not connected to the course grade, students were obligated to participate.

Monetary incentives, particularly performance contingent financial rewards, are often used in studies on student motivation and low-stakes tests (Baumert & Demmrich, 2001; O'Neil et al., 1995). Liu et al. (2012) administered the test and survey to over 750 students at three institutions, and students received $50 for their participation. However, interventions that include changes to motivating instructions are often considered more desirable as they are easier to implement (Liu et al., 2012; O'Neil et al., 1995). Such interventions also advance notions about learning that are not clouded by monetary incentives.

The extent to which test scores can be trusted to reflect students' actual abilities, the more valid inferences about student learning are and the more useful the evidence derived from these tests becomes.

In addition, a fourth condition, which included a combined personal and institutional prompt, was added to this study in an attempt to parse out any differences among conditions. Significant differences in performance were found by Liu et al. (2012) for students in the two treatment conditions (i.e., institutional and personal) when compared to the control condition; however, there were no statistically significant performance differences between the two treatment conditions. Waskiewicz (2011) found that students who received a personalized incentive in the form of individualized letters reported putting forth more effort on a low-stakes test than students in the control group who received generic letters. In this study, motivational condition had a significant impact on the total mean ETS® PP scores. Similar to Liu et al. and Waskiewicz, students in the personal condition performed significantly better than those in the other groups. This finding suggests that altering instructions to include personalized motivational prompts may positively impact students' performance on standardized tests.

Limitations

One potential limitation of this study is the sample. This study was limited to students enrolled in four 100-level general education courses at one institution. Although the courses were randomly assigned group membership, additional implementation and testing of the treatment in other courses and at other universities is needed before the results can be generalized. Moreover, the small sample size prevents firm conclusions from being drawn. Nevertheless, the study's design may be easily replicated.

The use of personalized motivational prompts provides low-stakes testing programs with a practical, sustainable, and low-cost strategy to enhance student performance.

An additional limitation may have been the homogeneity of the treatment conditions. An attempt to parse out differences in the treatments by adding a combined condition may have led to a lack of distinctiveness in the motivational prompts. Therefore, it is likely that the combined condition was too similar in nature to the other conditions to have a significant effect on student performance or motivation. In addition, it is difficult to determine if students attended to the motivational prompts. The prompt in the combined condition was longer than the other prompts, which may have led to student fatigue. To ascertain if students ingested the prompt, it might be necessary to have students sign the motivational prompt, indicating that they have read it. It might also be necessary to survey students after the administration of the survey to determine if they can identify the instructional prompt they received.

Finally, the SOS is a self-report measure of motivation; thus, its usefulness depends on the sincerity of students' responses. Students may have indicated that they expended high or low effort or that the test was of high or low importance when the opposite was true. Eklöf (2010) maintains that students who lack motivation to perform on an assessment may also lack motivation to accurately answer questions regarding their motivation. Just as multiple measures should be used to measure students' learning outcomes, multiple measures should also be used to measure motivation (Eklöf, 2010; Wise & Kong, 2005).

Implications

While test consequence and various incentives have been used as proxies for student motivation, the most appropriate source of information about a student's motivation is the student, yet minimal research has been conducted on student motivation and their perceptions of low-stakes tests (Nevo, 1995). Research on instruments that examine test-taker motivation on low-stakes tests is growing, but more is needed to fill the existing gap in the literature regarding examinee reactions to tests and the test conditions that affect performance and motivation.

For institutions and assessment professionals, this study provides evidence that motivational prompts may impact student performance on low-stakes tests, as students in the personal condition received significantly higher mean ETS© PP scores than students in the other conditions. This university uses low-stakes tests to measure student learning across the general education program and to make corresponding improvements in curriculum and instruction. Low student motivation prompts questions about whether the data collected are valid measures of student achievement (Abdelfattah, 2010). The extent to which test scores can be trusted to reflect students' actual abilities, the more valid inferences about student learning are and the more useful the evidence derived from these tests becomes.

The students in this study reported above average importance scores, yet their test performance was well below the mean. This suggests a paradox that requires additional investigation as it relates to similar populations of students. This inconsistency is relevant as it relates to expectancy-value theory in that previous research suggests that students' expectancy and efficacy perceptions are influenced by the difficulty level of the task and students' familiarity with the material (Pintrich, 1989). If some students lack clarity about their ability, as Aronson and Inzlicht (2004) suggest, then the cognitive-motivational component of expectancy-value theory should be explored in greater detail to determine the link between cognitive strategies and motivational components.

Furthermore, these results suggest that assessment does not have to be high-stakes to motivate students to perform. The use of personalized motivational prompts provides low-stakes testing programs with a practical, sustainable, and low-cost strategy to enhance student performance. In addition, motivating students to perform to the best of their ability on low-stakes tests may acculturate students to assessment for learning instead of assessment for grades. Future research could extend this line of inquiry by using students' names to enhance motivation as well as accountability.

References

Abdelfattah, F. (2010). The relationship between motivation and achievement in low-stakes examinations. *Social Behavior and Personality, 38*, 159-168. doi: 10.2224/sbp.2010.38.2.159

Aronson, J., & Inzlicht, M. (2004). The ups and downs of attributional ambiguity: Stereotype vulnerability and the academic self-knowledge of African American college students. *Psychological Science, 15*, 829-836. doi: 10.1111/j.0956-7976.2004.00763.x

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Baer, J. D., Cook, A. L., & Baldi, S. (2006). *The literacy of America's college students*. Washington, DC. Retrieved from American Institutes for Research http://www.air.org/filesThe20Literacy20of20/Americas20College 20Students_final20report.pdf

Baumert, J., & Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education, 16*, 441-462.

Boyas, E., Bryan, L. D., & Lee, T. (2012). Conditions affecting the usefulness of pre- and post-tests for assessment purposes. *Assessment & Evaluation in Higher Education, 37*, 427-437. doi: 10.1080/02602938.2010.538665

Cole, J. S., Bergin, D. A., & Whittaker, T. A. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology, 33*, 609-624. doi: 10.1016/j.cedpsych.2007.10.002

Eklöf, H. (2010). Skill and will: Test taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice, 17*, 345-356. doi: 10.1080/0969594X.2010.516569

Hoyt, J. E. (2001). Performance funding in higher education: The effects of student motivation on the use of outcomes test to measure institutional effectiveness. *Research in Higher Education, 41*, 71-85.

Huitema, B. E. (2007). Analysis of covariance (ANCOVA). In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 30-33). Thousand Oaks, CA: Sage Publications. doi: 10.4135/9781412952644.n18

Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review, 31*, 313-327.

Liu, O. L. (2008). *Measuring learning outcomes in higher education using the Measure of Academic Proficiency and Progress*. (ETS RR-08-47). Princeton, NJ: Educational Testing Service.

Liu, O. L. (2011). *Examining American post-secondary education*. (Report No. ETS RR-11-22). Retrieved from http:// www.ets.org/Media/Research/pdf/RR-11-22.pdf

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

Macfarlane, B. (2007). Defining and rewarding academic citizenship: The implications for university promotions policy. *Journal of Higher Education Policy and Management, 29*, 261-273.

Millet, C. M., Payne, D. G., Dwyer, C. A., Stickler, L. M., & Alexiou, J. J. (2008). *A culture of evidence: An evidence-centered approach to accountability for student learning outcomes*. Retrieved from http://www.ets.org/Media/ Education_Topics/pdf/COEIII_report.pdf

Nevo, B. (1995). Examine feedback questionnaire: Reliability and validity measures. *Educational and Psychological Measurement, 55*, 499-504. doi: 10.1177/0013164495055003017

O'Neil, H. F., Sugrue, B., & Baker, E. L. (1996). Effects of motivational interventions on the national assessment of educational progress mathematics performance. *Educational Assessment, 3*, 135-157.

Pintrich, P. R. (1989). The dynamic interplay of student motivation and cognition in the college classroom. In C. Ames & M. Maehr (Eds.), *Advances in motivation and achievement: motivation enhancing environments* (Vol. 6, pp. 117-160). Greenwich, CT: JAI Press.

Sundre, D. L. (2007). *The Student Opinion Scale (SOS): A measure of examinee motivation. Test Manual*. Harrisonburg, VA: The Center for Assessment & Research Studies, James Madison University.

Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology, 29*, 6-26. doi: 10.1016/S0361-476X(02)00063-2

Swerdzewski, P. J., Harmes, C. J., & Finney, S. J. (2009). Skipping the test: Using empirical evidence to inform policy related to students who avoid taking low-stakes assessments in college. *The Journal of General Education, 58*, 167-195. doi: 10.1353/jge.0.0043

Waskiewicz, R. A. (2011). Pharmacy students' test-taking motivation-effort on a low-stakes standardized test. American *Journal of Pharmaceutical Education, 75*(3), 1-8.

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*, 49-78.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*, 68-81. doi: 10.1016/ceps.1999.1015

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., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*, 163-183. doi: 10.1207/s15324818ame1802_2

Wolf, L. F., & Smith, J. K. (1995). The consequences of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*, 227-242.

Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test, motivation, and mentally taxing items. *Applied Measurement in Education, 8*, 341-351.

Young, J. W. (2007). *Validity of the Measure of Academic Proficiency and Progress (MAPP) test*. Princeton, NJ: Educational Testing Service.