

CAE

A Strategy for Increasing Student Motivation on Low-Stakes Assessments

Zachary G. C. Kornhauser
Jillian Minahan
Karen L. Siedlecki
Jeffrey T. Steedle



Abstract

Low-stakes assessments are increasingly being used by institutions of higher education as measures of student learning. One issue facing the validity of these assessments is that of student motivation, as research indicates that student motivation may play a large role in influencing test scores. To address this issue, the current study sought to increase motivation and, in turn, performance by appealing to students' sense of academic citizenship. Results of the study reveal that this manipulation does not appear to lead to increases in motivation or improved performance on low-stakes assessments.

Objectives

1. To determine whether student motivation on a low-stakes exam can be increased by activating a sense of academic citizenship amongst students.
2. To determine whether increased student motivation leads to greater performance on a low-stakes exam.

Perspective

Recently, a call for higher education reform has led to increased pressure on institutions of higher education to more precisely document student learning outcomes (Cole & Osterlind, 2008). Consequently, institutions of higher education are increasingly being held accountable for their students' learning as measured by scores on standardized, low-stakes assessments (Cole, 2007). Low-stakes assessments are typically understood as exams that do not affect the individual test-taker, but may have real consequences for the institution administering the exam. For example, student scores may have important implications for the fate of the institution when those scores are used to justify the receipt of taxpayer money for funding and to maintain or be granted accreditation (Wise & DeMars, 2010). Thus, the validity of the test scores may have significant consequences for institutions of higher education.

One element that may influence a student's test scores in general, and on low-stakes assessments in particular, is student motivation. Motivation is defined as "the process whereby goal-directed activity is instigated and sustained," so, within an assessment context, goal-directed activity is putting forth one's best effort on a test with the goal of accurately representing one's knowledge and capabilities in the content area that is being tested (Pintrich & Schunk, 2002, p.5; Wise & DeMars, 2003). Wise and DeMars (2005) define test-taking effort as "a student's engagement and expenditure of energy toward the goal of attaining the highest possible score on the test" (p. 2). If a student sustains this level of effort throughout the test, his or her scores may be interpreted validly as his or her true level of proficiency based upon test performance (Wise & DeMars, 2003).

Student motivation can be understood through expectancy-value models of academic motivation, which view a student's motivation to complete an academic task as a function of (a) expectancy, or the student's beliefs that he or she can complete a task, and (b) value, the beliefs that a student holds regarding why he or she should complete a task (Pintrich & Schunk, 2002). In addition, Wigfield and Eccles (2000) suggest that a student's value beliefs are influenced by (a) how important the student thinks performing well on the task is and (b) what he or she gave up to do the task. Because low-stakes assessments typically have no consequences for individual test-takers, these models would suggest that students would not be highly motivated to exert effort on low-stakes exams.

Wise and Smith (2012) extend the above models by developing the demands-capacity model of examinee test-taking effort. The authors assert that for any examinee-item encounter, the effort given by the examinee to the particular item is a function of two model constructs: 1) *resource demands* (RD), which represents the effort required to answer the item correctly and 2) *effort capacity* (EC), which is an examinee characteristic that represents how much effort he or she is willing to expend to answer the items. Determinants of RD include mental taxation, cognitive processing, and item difficulty. EC is determined by several factors with external, personal consequences being the most influential. If there are no test consequences of a particular exam (e.g., low-stakes exams), then EC is determined by internally motivating factors, such as desire to please teachers, competitiveness, and academic citizenship. The authors elaborate by giving the example that if there are no personal consequences for examinees, but they are aware that the results of a low-stakes assessment have significant implications for a group to which they belong, such as their institution of higher education, then "they may still exhibit strong collective effort due to feelings of citizenship or competitiveness" (Wise & Smith, 2012, p.150). The authors stress, however, that this model is preliminary at this time and has yet to be empirically tested.

As Wise and DeMars (2005) note, low student motivation is one of the most persistent threats to the validity of low-stakes exams. Specifically, students who are not motivated on low-stakes exams perform an average of .58 standard deviations worse on exams than those who are more motivated (Wise & DeMars, 2005). In order to address the issue of low student motivation, researchers have investigated methods of increasing student motivation including offering financial incentives (e.g., Cole, 2007; O'Neil, Sugrue, & Baker, 1996) and raising the stakes of an exam (e.g. Sundre, 1999; Wolf & Smith, 1996). These solutions, although appealing, may not feasibly be applied to the low-stakes assessments commonly administered in institutions of higher education. For instance, offering financial incentives to students is likely too costly to implement sensibly, and raising the stakes of a low-stakes exam will likely lead to student resentment and test-taking anxiety, which may also systematically influence the validity of results (Wise & DeMars, 2005).

Therefore, the necessity for a strategy to increase motivation in institutions of higher education that can be practically applied remains. As Wise and Smith (2012) hypothesize, examinees may be motivated internally by appealing to their sense of 'academic citizenship.' Academic citizenship has been framed in the context of students helping for the overall benefit of their school (e.g., Schmitt et al., 2007), and taking responsibility for a task whether or not it is asked of them (e.g., Gore, Kiefner, & Combs, 2012). Such a solution may involve explaining to students that the welfare of their institution may depend in part on their performance on a particular exam. This explanation may appeal to students' sense of shared responsibility as individual stakeholders who are part of a larger institution.

Although such a solution to increase student motivation has yet to be applied in institutions of higher education in the United States, it has been employed successfully in studies of students in elementary schools. For instance, Brown and Walberg (1993) experimentally manipulated the sense of shared responsibility amongst elementary school students in Chicago by telling them prior to taking a test that their scores would be used to evaluate the quality of their schools and their teachers. The results of the study demonstrated that students who received such instructions performed better than controls. In addition, Baumert and Demrich (2001) conducted a similar study in German elementary schools and concluded that "accentuating the societal utility value of the test, and thus inducing situational interest, is in itself a sufficient condition for the generation of test motivation" (p. 458).

Despite all of the previous research on low-stakes assessments, no study to date has attempted to increase student motivation for a low-stakes exam by activating a sense of shared responsibility or appealing to academic citizenship amongst college students in the United States. The purpose of the current study is to determine (a) whether student motivation on a low-stakes exam can be increased by appealing to students' academic citizenship, and (b) whether increased student motivation leads to higher scores on a low-stakes exam.

Method

Participants

Participants included full-time undergraduates between the ages of 18 and 30 years at a university in the New York City area. 149 undergraduate students participated in this study. Participants were an average of 21 years old and spent an average of 4 semesters in college by the time of their participation in study. 73% of the participants identified as Caucasian, 16% identified as Latino, 12% identified as Asian, and 6% identified with multiple ethnicities. Participants in the study had an average GPA of 3.30 (SD = .53), and an average SAT score of 1877 (SD = 211).

Materials

The low-stakes assessment used in this study, and the outcome measure, is a shortened, 35 question multiple-choice version of the Collegiate Learning Assessment (CLA), which is a test of critical thinking, analytical reasoning, and problem solving. The CLA was created to focus on these broad abilities because they cut across academic majors, and they are commonly mentioned in the mission statements of institutions of higher education (Klein, Benjamin, Shavelson, & Bolus, 2007).

At present, the CLA has been administered to more than 250,000 students in over 500 institutions of higher education. The CLA is an attractive low-stakes exam for many reasons, including that it provides a measure of 'value added' that reflects institutional contributions to student learning (Klein, Benjamin, Shavelson, & Bolus, 2007). In addition, the abilities the CLA assesses are also thought to be valuable for the rapidly changing twenty-first century workplace (Silva, 2008). Furthermore, the CLA has predictive power, as research has found it to be an effective predictor of first-year college GPA (Arum, Roksa, & Velez, 2008; Zahner, Ramsaran, & Steedle, 2012).

Participants in the study also completed three surveys. The first survey, The Student Opinion Survey (SOS; Sundre, 1999), is a ten item Likert-scale that measures participants' self-reported effort and self-reported importance of a task completed. Both of these constructs are considered to be important for test-taking motivation. Examples of an item related to effort is "I engaged in good effort throughout this test," and an example of an item related to importance is "Doing well on this test was important to me." Both subscales can be computed separately when calculating participants' motivation scores.

The second survey completed by the participants was the Sense of Community Index-2 (SCI-2; Chavis, Lee, & Acosta, 2008), which is a 25 item Likert-scale that measures participants' level of feelings and connectedness to their institution. The initial item, referred to as the "The Community Referent," asks participants to indicate how important it is for them to feel a sense of community with other community members. The next 24 items are referred to as the "Total Sense of Community Index." Examples of these items include "I can recognize most members of this community," "Fitting into this community is important to me," and "I feel hopeful about the future of this community." This survey can be scored by adding the 24 items.

The third survey administered is a demographic questionnaire that assesses ethnicity, gender, year in school, and GPA. Participants also completed the Raven's Progressive Matrices (Raven, 1962), which is a test of reasoning that is designed to be independent of knowledge in order to examine reasoning as a potential moderator.

Procedure

Participants who provided their consent to participate in the study were randomized into either a control ($n = 69$) or experimental group ($n = 76$). Following this randomization, participants were given the CLA with an accompanying instruction sheet. The instruction sheet and CLA were the same for both groups, with one important difference: those participants randomized to the experimental group were told that "As a member of the Fordham University student community, it is important that you put forth your best effort on this exam as your results will have real consequences for Fordham University. Remember that you are representing Fordham University by taking this exam, so please try to perform to the best of your ability." In contrast, participants in the control group were instead told, "you will not have any personal consequences for performing well or poorly on this exam, and your confidentiality will be protected." After completion of the CLA, participants were given the SOS, SCI-2, and demographic questionnaire.

Results

Descriptive statistics of the experimental and control groups appear in Table 1 (below)

Table 1: Descriptive statistics comparing experimental and control groups

	N	GPA	Matrix	CLA	Effort	Importance	Connectedness
Experimental	76	3.28 (.42)	8.53 (2.22)	24.7 (3.56)	19.32 (3.47)	17.73 (3.18)	66.05 (12.8)
Control	69	3.32 (.61)	8.27 (2.35)	24.6 (3.21)	18.78 (3.97)	17.2 (3.63)	68.35 (13)
Total	145	3.30 (.53)	8.40 (2.28)	24.6 (3.37)	19.04 (3.22)	17.45 (3.42)	67.26 (12.91)

Note: Experimental and control groups did not differ on any of the above variables.

Pearson correlations were conducted to determine if there was a relationship between CLA score and self-reported effort and importance of the exam, and between self-reported motivation and connectedness to participants' institutions of higher education. When the sample was collapsed across conditions, CLA scores were positively correlated with self-reported effort $r(139) = .271, p < .01$. Connectedness scores were also positively correlated with both effort $r(141) = .210, p < .05$, and self-reported importance of the exam $r(144) = .16, p < .05$.

When the sample was divided across experimental and control conditions, stronger correlations were found amongst the constructs in the experimental group than in the control group across a number of constructs. For instance the correlation between CLA and effort was stronger for the experimental group $r(68) = .28, p < .05$ than it was for the control group $r(71) = .256, p < .05$. Further, the correlation between effort and connectedness was moderate for the experimental group $r(68) = .35, p < .01$, but found to be non-significant for the control group.

Analyses were conducted to compare the performance and levels of motivation amongst participants in the experimental group to participants in the control group. Results of the analyses did not find any significant differences between the experimental and control group in regards to CLA score $t(141) = .20, p = .83$, effort $t(139) = .99, p = .41$, importance of the exam score $t(142) = .94, p = .34$, or connectedness to institution score $t(143) = 1.07, p = .28$.

Analyses were conducted to determine if participants in the experimental group were more likely than those in the control group to report that their scores would be compared to those of students at other institutions (i.e., a manipulation check). Results of a chi-square analysis indicated that participants in the experimental group were more likely to answer "true" when asked whether their scores would be compared to those of students at other institutions, and participants in the control group were more likely to answer "don't know" or "false" to this same question ($\chi^2 = 60.42, p < .01$).

Further analyses were conducted to determine if participants in the experimental group who reported that their scores would be compared to students of other institutions would score higher on the motivation scale and CLA than participants in the control who did not think or were not sure if their scores were being compared to students at other institutions (Table 2). Participants in the experimental group were marginally significantly more likely to report higher levels of motivation than participants in the control $t(112) = 1.859, p = .06$. In addition, a small effect was detected with participants in the experimental group reporting higher levels of motivation than those in the control group ($d = .35$). In contrast, there were no differences between the experimental and control on the effort $t(113) = 1.442, p = .15$ and academic importance subscales $t(115) = .736, p = .46$ of the motivation scale. In addition, participants in the experimental group did not significantly outperform participants in the control group on the CLA $t(114) = .429, p = .66$, and the effect size was small ($d = .08$).

Table 2: Descriptive statistics of participants who passed the manipulation check

	N	GPA	Motivation	CLA	Effort	Importance	Connectedness
Experimental	59	3.28 (.41)	41.27 (5.86)	24.7 (3.56)	19.32 (3.40)	17.71 (3.13)	66.49 (13.11)
Control	49	3.30 (.68)	39.23 (5.81)	24.43 (3.17)	18.77 (3.23)	17.24 (3.81)	67.80 (13.51)
Total	108	3.30 (.55)	40.34 (5.90)	24.57 (3.29)	19.24 (3.24)	17.49 (3.46)	67.10 (13.26)

Additional analyses of covariance were conducted to determine if matrix reasoning and connectedness might serve as covariates in the relationship between experimental group and CLA score. Results revealed that the covariates of matrix reasoning and institutional connectedness were not related to CLA score, $F(1, 113) = 3.244, p = .07$; $F(1, 114) = .811, p = .37$.

Regressions were conducted to determine if self-reported effort predicted participants' test performance, and if connectedness to institution predicted effort. Self-reported motivation significantly predicted test performance $b = .24, t(136) = 9.91, p < .001$ and explained 5% of the variance in test score $R^2 = .05$; $F(1, 136) = 8.08, p < .001$. In addition, connectedness to the institution significantly predicted self-reported effort $b = .29, t(138) = 11.42, p < .001$ and explained 8% of the variance in motivation $R^2 = .08$; $F(1, 138) = 13.5, p < .001$.

Finally, it should be pointed out that based on the standardized regression coefficient between motivation and test performance ($b = .24$) and the difference in effect size between the control and experimental groups in motivation ($d = .35$), the small effect size found between the control and experimental groups in terms of CLA performance ($d = .08$) would be expected based on the change in motivation scores.

Discussion

The results of this study indicate that the manipulation employed was partially successful in accomplishing its intended objectives. Although participants in the experimental group were more likely to report that their scores on the CLA would be compared to students at other institutions, they did not report higher levels of motivation or perform better on the CLA itself.

These findings have both theoretical and practical applications. From a theoretical standpoint, these findings reveal that, in contrast with Wise and Smith's (2012) hypothesis, appealing to students' sense of academic citizenship was not in itself enough to lead students to put more effort on a low-stakes examination. There are, of course, a number of reasons why the findings from the current study were not in line with what Wise & Smith (2012) predicted. In the first place, motivation is a multi-faceted construct that can fluctuate during a given examination (Wise & Smith, 2012). Thus, it is perhaps not surprising that the manipulation utilized in this current study was not effective on its own in raising examinee performance. Further, and perhaps more importantly, the manipulation used in this study was very subtle; as mentioned earlier, the only differences between the control and experimental groups was that participants in the experimental group were told in the instructions that their results would be compared to those of students at other institutions, while the experimental group was not given this prompt. The non-significant increase in motivation might be due in part to the subtlety of the manipulation.

Another factor worth mentioning is the setting in which the experiment was administered. Standardized, low-stakes assessments are typically given to students in large numbers, oftentimes in the presence of college faculty and administrators, and have a distinctly serious 'feel' to them. In the current study, although every effort was made to replicate the setting in which low-stakes assessments are typically given, participants participated in the study in small numbers for logistical reasons, and ecological validity is thus a concern.

Nevertheless, since there was a small effect found for the level of motivation of participants in the experimental group, it is possible that this effect would increase in size if the manipulation was more prominent and ecologically valid. In addition, there was a significant positive correlation between motivation and CLA performance, and results from the regression analysis and effect sizes reveal that the magnitude of increase in CLA scores is consistent with what would be expected based on the increase in participants' level of motivation. If the manipulation was more effective in increasing participants' motivation, it is likely that the scores would increase in magnitude as well. A further finding of note is that participants who feel more connected to their institution reported higher levels of motivation, which was what Wise and Smith (2012) predicted.

Future research should seek to determine whether performance on low-stakes exams may be improved by building on the manipulation utilized in this study by altering the manipulation or coupling it with other types of intrinsic motivators. For instance, perhaps this current

manipulation would have been more successful if students were continuously reminded that their scores would be compared to those of students at other institutions throughout the exam rather than just being told this at the beginning of the exam. Or, maybe if other intrinsic motivators were utilized, such as offering students feedback on their exam performance, the effect of multiple intrinsic motivators might have enough of a combined effect to increase student motivation. As the calls for assessment and accountability in higher education increase in fervor, it appears likely that the use of low-stakes assessments will only grow in prevalence. Therefore, it is apparent that now, more than ever, a workable solution is needed for handling the pressing issue of examinee motivation on low-stakes examinations.

References

- Arum, R., Roksa, J., & Velez, M. (2008). Learning to Reason and Communicate in College: Initial Report of Findings from the CLA Longitudinal Study. New York: Social Science Research Council.
- 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(3), 441-462.
- Brown, S. M. & Walberg, H. J. (1993). Motivational effects on test scores of elementary students. *The Journal of Educational Research*, 86, 233-236.
- Chavis, D. M., Lee, K. S., & Acosta, J. D. (2008). *The Sense of Community (SCI) Revised: The Reliability and Validity of the SCI-2. Paper presented at the 2nd International Community Psychology Conference*, Lisboa, Portugal.
- Cole, J. S. (2007). *Motivation to do well on low-stakes tests*. University of Missouri-Columbia, Columbia, MO.
- Cole, J.S. & Osterlind, S.J. (2008). Investigating Differences between Low- and High-Stakes Test Performance on a General Education Exam. *The Journal of General Education*, 57(2), 119-130.
- Gore, J.S., Kiefner, A.E., & Combs, K.M. (2012). Personality Traits that Predict Academic Citizenship Behavior. *Journal of Applied Social Psychology*, 42(10), 2433-2456. doi: 10.1111/j.1559-1816.2012.00948.x
- Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. *Evaluation Review*, 31(5), 415-439.
- O'Neil, 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(2), 135-157.
- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Raven, J. C. (1962). *Advanced Progressive Matrices. Set II*. London: H. K. Lewis & Co.
- Schmitt, N., Oswald, F. L., Kim, B. H., Imus, A., Merritt, S., Friede, A., et al. (2007). The use of background and ability profiles to predict college student outcomes. *Journal of Applied Psychology*, 92, 165-179.
- Silva, E. (2008). *Measuring skills for the 21st century*. Washington, DC: Education Sector.
- Sundre, D. L. (1999, April). *Does examinee motivation moderate the relationship between test consequences and test performance?* Paper presented at the annual meeting of the American Education Research Association, Montreal.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, 25(1), 68-81.
- Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1), 1-17.
- Wise, S. L., & DeMars, C. E. (2010). Examinee non-effort and the validity of program assessment results. *Educational Assessment*, 15, 27-41
- Wise, S. L., & Smith, L. F. (2012). A model of examinee test-taking effort. In J. Bovaird (Ed.) *Contemporary issues in high stakes testing*. Washington, DC: American Psychological Association.
- Wolf, L. F., & Smith, J. K. (1996). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, 8(3), 227-242.
- Zahner, D., Ramsaran, L. M., & Steedle, J. T. (2012) *Comparing alternatives in the predictors of college success*. Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, Canada.