

The Relationship between Effort and Performance on a Low-Stakes Exam across High School
and College Freshmen

Zachary G. C. Kornhauser & Jessalynn K. James

Abstract

Low-stakes assessments are commonly used in educational institutions as measures of student learning, but suffer from some threats to validity. Low-student motivation is one issue facing the validity of these assessments, as research has found student motivation to impact performance on low-stakes exams. The current study aimed to determine (a) if effort on a low-stakes exam is consistent across both high school and college freshmen, (b) if the relationship between effort and performance is consistent across high school and college freshmen, and (c) if effort on a low-stakes exam is dependent on the type of task students are presented with. Results indicate that high school students expended more effort than did college students on a low-stakes assessment, and that effort was higher for essay-based items than multiple-choice items for both high school and college students.

Objectives:

- 1) Determine whether effort on a low-stakes exam is consistent across high school and college freshmen.
- 2) Determine whether the relationship between effort and performance is consistent across high school and college freshmen.
- 3) Determine whether effort differs depending on the type of task students are presented with.

Perspective

The movement towards higher education reform has led to an increased usage of low-stakes assessments (Cole & Osterlind, 2008; Cole, 2007). Low-stakes assessments are exams that do not have consequences for the individual test-taker but may influence institutions or schools administering the exams. Low-stakes assessments have been utilized at the higher-education level to hold public institutions accountable for the funding they receive (Wise & DeMars, 2005). At the K-12 level, the No Child Left Behind legislation (U.S. Department of Education, 2002) has enacted sanctions for schools that are consistent low performers, including public school choice, staff replacement, and restructuring for schools that receive Title I aid (Dee & Jacob, 2010).

Researchers often operate under the assumption that examinees give their best effort when taking an exam, although that supposition cannot necessarily be taken as fact. For high-stakes exams, one might reasonably assume that examinees will put forth their best effort, as the examinee typically has a large investment in the exam (Wise, Wise, & Bhola, 2006). A persistent

concern when considering the validity of low-stakes exams, however, is student motivation; if students do not have consequences tied to their performance, they may not be highly motivated to put forth their best effort (Wise & DeMars, 2010). Expectancy-value models of academic motivation are used to explain students' motivation when approaching a given task. According to these models, students' motivation on a task depends on their beliefs regarding their ability to perform well on a task in addition to how much they value a task (Pintrich & Schunk, 2002). Furthermore, students also might be motivated by the importance of a task and the cost of completing a task (Wigfield & Eccles, 2000). Applied to low-stakes assessments, these models would predict that students might not be motivated to put forth their best effort, as there are not any consequences tied to their performance (Wise & DeMars, 2005).

Previous research has examined the relationship between motivation and test performance. In the most comprehensive review on the topic, Wise and DeMars (2005) reviewed 25 studies that examined the difference between motivated and unmotivated examinees, and found that all but one of the studies yielded a testing advantage to motivated examinees. The average advantage conferred to motivated examinees' was .59 standard deviations.

Evidence suggests that factors influence the relationship between students' motivation and performance, including the type and the length of items examinees are presented with. For example, DeMars (2000) found that increasing exam stakes affected performance on constructed-response items more strongly than it did on multiple-choice items. However, DeMars did not include a measure of motivation on the low-stakes exam used in that study. Along the same lines, Sundre (1999) found that the effect size for having a consequence tied to performance, as compared to no consequence, was far greater for essay-based tests (1.59) than for multiple-choice exams (.79). Wise et al. (2006) found that item length (i.e., how much

reading the examinee needs to do in order to answer the item) and where the item appears in the test affects exam scores. Other researchers have similarly found that motivation on multiple-choice tests depends on the mental taxation required to answer items and the location of the items in a test (Wolf, Smith, & Birnbaum, 1995).

Liu, Bridgeman, and Adler (2012) examined students' performance on a low-stakes assessment across different motivational conditions. The authors found more value-added learning on multiple-choice questions than on essay questions. They reason, like other researchers, that it takes more effort to construct an essay than it does to answer multiple-choice questions.

Examinee gender has also been found to influence motivation. For example, evidence indicates that females exhibit more solution-based behavior on exams than do males (Wise & DeMars, 2010), and also females report higher levels of effort than do males (Cole, 2008; Eklof, 2007). The most comprehensive review on gender differences in test-taking motivation was conducted by DeMars, Bashkov, and Socha (2013). The authors reviewed the extant research on gender effects in motivation and noted that females generally exert more effort on low-stakes assessment than do males. The authors provide explanations for why females exhibit greater levels of motivation on low-stakes assessments, including differences in personality between the genders as well as differences in approach to learning.

Given the impact of student motivation on performance, researchers have attempted to use a number of strategies in order to increase student motivation. These methods have included offering financial incentives for performance on low-stakes exams (O'Neil, Abedi, Miyoshi, & Mastergeorge, 2005; Baumert & Demrich, 2001), raising the stakes of assessments (Wolf &

Smith, 1996; Sundre, 1999), and filtering out the responses of examinees who did not put forth effort on assessments (Sundre & Wise, 2003; Swerdzewski, Harnes, & Finney, 2011).

While the above research establishes the relationship between motivation and performance as well as some factors influencing examinee motivation, there is still much that is not known. As Liu and colleagues note, “In most value-added calculations, it is assumed that the levels of motivation remain somewhat equal between the benchmark class (e.g., freshmen or sophomores) and the comparison class (e.g., juniors or seniors). However, students in lower classes may be more motivated than their upper-class peers for a multitude of reasons such as being intimidated by tests or being less busy” (p. 359).

As mentioned, Wise and DeMars (2005) reviewed a large body of literature that examined the effect that low-motivation has on examinee performance. The studies that the authors reviewed included research on students as young as the second grade, who were paid based on their test performance, to older job applicants applying to work at a laundry shop. No research reviewed by the authors, however, included a sample of both high school and college students who were taking a similar low-stakes assessment. This lack of research is understandable, since typically the same types of assessments are not used at both the high school and college levels. However, since low-stakes assessments pervade both K-12 and higher education, understanding how educational level might affect motivation is critical. Therefore, the first aim of this study is to determine if effort on a low-stakes exam is consistent across high school and college freshmen. Similar to this first aim, the second aim of the study is to investigate whether the *relationship* between effort and task performance is similar across a sample of high school and college freshmen.

As discussed earlier, the length and type of items students encounter can impact their motivation. However, with the exception of Liu et al. (2012), research has generally not investigated whether there might be differences in motivation across an essay and multiple-choice based exam that assesses broad higher order thinking skills such as critical thinking. Further, this question warrants investigation at both the high school and college levels. Therefore, the third aim of this study is to determine if effort on a low-stakes assessment depends on the type of task that an examinee is presented with.

Methods

Sample

Participants included undergraduate freshmen at four-year colleges and universities using the Collegiate Learning Assessment (CLA+) and high school freshmen at secondary schools using the College and Work Readiness Assessment (CWRA+). Both the college and high schools students in this sample took the exam in the fall of 2013. More information on the sample can be found in the table below:

Table 1

Student sample summary across CLA+ and CWRA+

Demographic characteristic	College freshmen (CLA+)		High school freshmen (CWRA+)	
	N	%	N	%
All students	18,178	100%	8,291	100%
Gender				
Male	7,092	39%	4040	49%
Female	10,783	59%	4101	49%
Decline to state	303	2%	150	2%

Primary language

English	14,832	82%	7,454	90%
Other	3346	18%	837	10%

Race/ethnicity

American Indian / Alaska Native / indigenous	173	1%	219	3%
Asian (including Indian subcontinent and Philippines)	1,604	9%	753	9%
Native Hawaiian or other Pacific Islander	71	0%	65	1%
African-American / Black (including African and Caribbean), non-Hispanic	2,452	14%	775	9%
Hispanic or Latino	3,051	17%	653	8%
White (including Middle Eastern), non-Hispanic	9,553	53%	5,137	62%
Other	620	3%	418	5%
Decline to state	654	4%	217	3%

Instrument

The assessment used in this study, the CLA+, along with its high-school equivalent, the CWRA+, is a test of critical thinking and written communication. The CLA+ was created to focus on these broad abilities because they cut across academic majors and are frequently mentioned in the mission statements of many institutions of higher education (Klein, Benjamin, Shavelson, & Bolus, 2007). The CWRA+ was developed as an offshoot of the CLA+, with prompts and scenarios designed specifically for high-school aged students. Aside from differences in prompt topics and Lexiles to ensure the age-appropriateness of the materials, all components of the exams are identical across high school and college; students are presented

with the same opening instructions and test format regardless of whether they are in high school or college. Also, all forms of the assessment are equated to adjust for difficulty and allow for comparisons between high school students as well as comparisons across test forms.

Both assessments consist of two major components: a performance task (PT) and a series of selected-response questions (SRQs). The PT presents students with a real-world situation that requires a purposeful written response. Students are instructed to support their response by using information provided in a document library. Students have 60 minutes to complete this constructed-response task. In the second part of the examination, students are asked to answer 25 SRQs. These questions measure scientific and quantitative reasoning, critical reading and evaluation, and the ability to critique an argument. Students have 30 minutes to complete this section of the assessment.

The CLA+ and CWRA+ also include a brief post-assessment survey, which includes questions designed to measure students' effort and engagement during the exam. Students are asked how much effort they put into each of the two sections, with five response options listed on a Likert scale ranging from "no effort at all" to "my best effort."

Participating schools are responsible for selecting and recruiting representative samples of students, and schools are encouraged to use stratified sampling strategies.

Results

Analyses were run to examine the amount of effort that high school and college freshmen expended on a low-stakes exam. In regard to the amount of time spent on the tasks (in minutes), high school students spent significantly more time on both the PT $t(27080) = 33.32, p < .001$ and SRQ $t(27080) = 27.84, p < .001$ sections of the exam. In addition, high school students reported higher levels of effort than college students on both the PT $t(27297) = 29.24, p < .001$ and the

SRQs $t(27297) = 15.58, p < .001$. It should also be noted that both college and high school students reported lower effort on the multiple-choice SRQs than they did on the essay-based PT, which requires participants to answer an open-ended essay question. A table detailing these differences can be found below:

Table 2: *Time spent and effort reported by college and high school students*

	Time on PT	Time on SR	Effort on PT	Effort on SR
High School	37.47 (14.71)	22.19 (6.25)	4.08 (.83)	3.55 (.89)
College	31.13 (14.50)	19.72 (7.04)	3.74 (.91)	3.36 (.96)

Pearson correlations were conducted to determine if the relationship between effort and performance differed between high school and college freshmen. Results of these correlations found that, for high school students, scores on the PT positively correlated with amount of self-reported effort on the PT $r(8348) = .34, p < .01$. Further, scores on the SRQ positively correlated with amount of self-reported effort on the SRQ $r(8442) = .22, p < .01$. For college students, scores on the PT positively correlated with self-reported effort on the PT $r(18305) = .34, p < .01$. Also, scores on the SRQ positively correlated with amount of self-reported effort on the SRQ $r(18335) = .28, p < .01$. Results of these correlations can be found in Table 3.

Table 3: *Correlations between PT/SRQ Score and Effort*

	PT Score and Effort	SRQ Score and Effort
High School	.34**	.22**
College	.34**	.28**

Note: ** $P < .01$

A final set of analyses was conducted to determine if effort on a low-stakes exam is dependent on the type of task examinees are presented with. Mean levels of effort on the PT ($M=3.85$; $SD=.91$) were actually higher than those on the SRQs ($M=3.43$; $SD=.94$), and these differences were found to be statistically significant $t(28594) = 79.79, p<.001$.

In terms of the relationship between demographic characteristics and effort, for college students, there was a small but significant positive correlation between parental education level and self-reported effort on the PT $r(18708)=.08, p<.01$ and SRQs $r(18708)=.03, p<.01$. For high school students, the correlation, while still small, was higher for both the PT $r(8591)=.15, p<.01$ and the SRQs $r(8591)=.10, p<.01$. Results of these correlations can be found in Table 4. In addition, gender differences were found on effort expended on the PT $t(2893) = 5.22, p<.01$ with females expending more effort than males ($M=3.82, SD=.90$). Also, females ($M=3.46, SD=.94$) expended more effort than males ($M=3.39, SD=.94$), $t(2893) = 5.97, p<.01$ on the SRQs.

Table 4: *Correlations between Effort and Parent Education Levels*

	PT Effort and Parent Education	SRQ Effort and Parent Education
High School	.15**	.10**
College	.08**	.03**

Note: ** $P<.01$

To determine whether there is an interaction between gender and effort, a series of two-way ANOVAs were conducted. In the first ANOVA, gender and effort on the PT were used as factors, while PT score was used as an outcome. Results revealed a main effect of gender $F(2, 27875) 6.56, p<.01$, with females scoring higher than males. There was also a main effect of

effort on PT score $F(4, 27875) 178.01, p < .01$, with higher levels of reported effort leading to higher scores. In the second ANOVA, gender and effort on SRQ scores were used as factors, and SRQ score was used as the outcome. Results revealed no significant main effect of gender on SRQ score, although there was a main effect of self-reported effort on SRQ score $F(4, 27956) 149.05, p < .01$, with students who reported higher levels of effort scoring higher on the assessment. A small interaction was uncovered $F(8, 27956) 2.26, p < .01$ indicating that, when reporting low-levels of effort, males scored higher than females. When reporting higher levels of effort, however, females outperformed males. The effect size for this interaction, however, was very small $\eta^2 = .001$.

Discussion

The results of this study yielded a number of interesting findings. First, as Liu and colleagues (2012) hypothesized, high school students put forth greater effort than did college students on both the PT and SRQs. It is not clear exactly why high school students exerted more effort, although it is possible, as Liu et al. surmise, that these students may be more intimidated by these assessments or have fewer demands on their time. A further possibility is that low-stakes exams may be more novel for high school students than for college students. In accordance with this hypothesis, high school students experience a steeper drop-off in self-reported motivation from the PT to the SRQ sections than did college students. This may be evidence that the initial appeal of the exam attenuated over time for high school students.

A second finding of note is that the relationship between effort and task performance appears to be similar for high school and college students on the PT, but does differ somewhat on the SRQs. Although the difference is small, it does provide important information about the

relationship between motivation and performance, and suggests that this relationship is not necessarily stable across age groups. Further research should endeavor to examine the relationship between motivation and performance across participants of a number of different ages and class levels in order to more fully understand the relationship between motivation and performance.

A more surprising finding is that, in contrast to previous research (e.g. DeMars, 2000; Sundre, 1999) self-reported effort for the essay-based PT was higher than for the multiple-choice SRQs. The discrepancy with previous research may be explained by the fact that examinees took the PT first, so their interest may have waned by the time they reached the SRQs. Nevertheless, future research should endeavor to counterbalance the order of the tasks to determine if essay-based tasks, do, in fact, lead to lower levels of motivation than do multiple choice-based tasks.

In addition, small gender differences were found in this study between effort expended on the PT and the SRQ sections. The differences are not surprising and are consistent with what has been reported in previous research (e.g., DeMars et al., 2013), although it does provide further evidence that females do expend more effort than males on low-stakes exams.

Despite the contributions this study offers to the literature on motivation and performance, there are a number of limitations worth noting. First, there is a potential sampling issue in that although the college sample is nationally representative, the high school sample is not. Students from participating high schools tend to come from higher socioeconomic backgrounds and have parents who have attained higher levels of education than the national average. The increased motivation observed at the high school level may very well represent a difference in the samples collected as part of this study.

A further limitation of this study is that students' scores are "pre-filtered" before they are included in the sample. More specifically, participants whose responses do not address the prompt on the PT or answer less than half of the SRQs are not included in this sample. These requirements for inclusion likely remove the least motivated students from the sample. Therefore, the results of the study may be different if these requirements for inclusion were not applied.

Finally, the current study only assessed motivation through one question. More sophisticated measures of motivation, such as the Student Opinion Scale (Sundre, 2007), have been used to more reliably assess students' level of motivation, and using these scales may have provided a more true indicator of motivational level.

Despite these limitations, this study has both theoretical and practical applications. From a theoretical standpoint, the study adds information to the literature on the relationship between motivation and exam performance and whether that relationship is consistent cross-sectionally. From a practical standpoint, information from this current study may help guide future attempts at motivation filtering, as it elucidates that high school students put forth more effort on low-stakes exams than do college students. Finally, the study raises questions about the true level of motivation that examinees may exhibit on essay-based tests, and if that motivation is dependent upon the order of questions in an exam. 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, it is imperative to understand the issue of examinee motivation on low-stakes examinations.

References

- 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.
- Cole, J.S. & Osterlind, S.J. (2008). Investigating differences between low- and high- stakes test performance on a general education exam. *Journal of General Education, 57*, 119-130.
- Cole, J. S. (2007) *Motivation to do well on low-stakes tests* (Doctoral dissertation). Retrieved from Network Digital Library of Theses and Dissertations.
- Dee, T.S. & Jacob, B.A. (2010). The impact of No Child Left Behind on teachers, students, and schools. Brookings papers on economic activity, Fall 2010.
- DeMars, C. E. (2000). Test stakes and item format interactions. *Applied Measurement in Education, 13*(1), 55-77.
- DeMars, C. E., Bashkov, B. M., & Socha, A. B. (2013). The Role of Gender in Test-Taking Motivation under Low-Stakes Conditions. *Research & Practice in Assessment, 8*69-82.
- Eklof, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing, 7*, 311-326.
- Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. *Evaluation Review, 31*(5), 415-439.
- Liu, O., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Educational Researcher, 41*(9), 352-362.
- O'Neil, H. F., Abedi, J., Miyoshi, J., & Mastergeorge, A. (2005). Monetary Incentives for Low-Stakes Tests. *Educational Assessment, 10*(3), 185-208. doi:10.1207/s15326977ea1003_3

- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- 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.
- Sundre, D. L., & Wise, S. L. (2003). *'Motivation filtering': An exploration of the impact of low examinee motivation on the psychometric quality of tests.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.
- Sundre, D. L. (2007). *The student opinion scale (SOS): A measure of examinee motivation.* Harrisonburg, VA: The Center for Assessment and Research Studies.
- Swerdzewski, P. J., Harmes, J. C., & Finney, S. J. (2011). Two Approaches for Identifying Low-Motivated Students in a Low-Stakes Assessment Context. *Applied Measurement In Education, 24*(2), 162-188.
- U.S. Department of Education (2002) *No Child Left Behind: A desktop reference.* Washington, D.C.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology, 25*(1), 68-81.
- Wise, S. L. (2004) An investigation of the differential effort received by items on a low-stakes computer-based test. *Applied Measurement in Education, 19*(2), 95-114
- 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, V. L., Wise, S. L., & Bhola, D. S. (2006). The Generalizability of Motivation Filtering in Improving Test Score Validity. *Educational Assessment, 11*(1), 65-83.

Wolf, L. F., & Smith, J. K. (1996). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*(3), 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*(4), 341-351.