

Abstract

We examined if students' understanding about the purpose and use of institutional assessment scores was affected by moving to remote testing due to COVID-19. Moreover, we examined if students' knowledge about the purpose of outcomes assessment related to their effort on these tests. If knowledge about accountability testing and effort were positively related, we could design interventions to increase knowledge and, in turn, increase effort. We gathered data on knowledge about institutional accountability testing and test-taking effort from students differing in year in school and whether tests were completed remotely or in person. Knowledge about assessment testing was high with negligible differences in knowledge across year in school and testing context. Knowledge related positively to test-taking effort. Testing context and year in school did not moderate this relation. In sum, students who better understood that outcomes assessment was used for accountability and improvement efforts expended more effort on these assessments.

**AUTHORS**

Sara J. Finney, Ph.D.
James Madison University

Dena A. Pastor, Ph.D.
James Madison University

Shanti Silver, BA
Kenyon College

Students' Understanding of Assessment for Institutional Accountability and Improvement: Relation with Test-Taking Effort and Remote Test Administration

Higher education institutions engage in outcomes assessment to respond to institutional accountability mandates (U.S. Department of Education, 2006) and to inform programming changes to improve student learning and development (Fulcher & Prendergast, 2021). The student learning and development outcomes that are assessed, reported, and used for improvement are often the outcomes of multi-faceted education experiences, such as general education programming (Mathers et al., 2018; Stone & Friedman, 2002), academic degree programs (Allen, 2004), quality enhancement plans (Miller et al., 2019; Smith & Finney, 2020), and student affairs programs (e.g., Kerr et al., 2020). Outcomes assessment data for institutional accountability and improvement purposes often is not associated with an individual course, but rather tied to several academic and/or student affairs learning experiences. Thus, the assessment of these outcomes often does not inform course grades, graduation, admission into a major, or other high-stakes outcomes for students. These institutional effectiveness assessments are often low stakes for students, meaning there are no personal consequences associated with their performance.

Studies have shown that students perform better when assessments are perceived as high stakes versus low stakes. Wise and DeMars (2005) summarized studies that compared test performance across examinee groups who were administered the same test but under high-stakes versus low-stakes conditions. Examinees in the low-stakes condition scored .59

CORRESPONDENCE

Email
finneysj@jmu.edu

SD lower than those in the high-stakes condition. During institutional effectiveness testing, students may not know if the assessments they are completing are low or high stakes. Their understanding of the stakes of the assessment may influence their test performance.

Moreover, studies have shown that students expend more effort when tests are perceived as high stakes versus low stakes. Sundre and Kistansas (2004) found that self-reported effort was lower when a test was described as low versus high stakes. Unlike high-stakes tests, where students tend to expend the effort necessary to reflect their ability, low-stakes tests tend to be associated with greater variability in effort, with some students expending a high degree of effort and others not. Expended effort positively covaries with test scores, indicating that test scores reflect expended effort to some extent (Cole et al., 2008; Eklöf et al., 2014; Myers & Finney, 2021). Moreover, effort covaries with students' perceived test importance (Finney et al., 2018; Penk & Richter, 2017; Rios, 2021), test emotions (Finney, Perkins, & Satkus, 2020; Finney, Satkus, & Perkins, 2020; Penk & Schipolowski, 2015; Perkins et al., 2021; Satkus & Finney, 2021), personality (Barry & Finney, 2016; Barry et al., 2010; Freund & Holling, 2011; Kopp et al., 2011), and attitudes toward accountability testing (Zhao et al., 2020; Zilberberg et al., 2014).

Students' Knowledge of the Purpose and Use of Outcomes Assessment Data

We were interested in examining if knowledge about the purpose and use of assessment scores also related to expended effort. If knowledge about assessment for institutional effectiveness purposes and test-taking effort were positively related, we could design interventions to increase knowledge and, in turn, possibly increase effort. Interventions to influence knowledge may be easier to create than interventions to influence test importance, test emotions, or attitudes toward accountability testing, and, of course, trying to change personality is futile. With that said, we were unsure if understanding the purpose and use of institutional effectiveness test data would relate positively or negatively with effort. It may be that having an accurate understanding of the low-stakes nature of the test leads to lower expended effort.

A student's understanding of the purpose and use of test scores for institutional effectiveness may be influenced by a variety of things, including what information is shared with students, how it is shared, and how receptive students are to the information. For instance, the information itself may be high quality, but the delivery of the information may be poor. Likewise, the information and delivery may be high quality, but students may not be engaged in receiving the information (e.g., do not read or listen to information provided). When our institutional accountability testing moved online due to the COVID-19 pandemic, it provided us with an opportunity to assess students' understanding of institutional-level assessment efforts and to assess if this understanding was impacted by the modality of the testing (in-person proctored testing prior to the pandemic versus remote, unproctored testing during the pandemic).

We were also able to examine if understanding of institutional assessment differed across student groups, specifically incoming first-year students and more advanced students (students who had completed 45-70 credit hours; typically 1.5 years at the institution). Students earlier in their college career may differ from more advanced students with respect to their understanding of institutional accountability testing and its impact on them personally. Numerous studies have found that older students exhibit lower effort on low-stakes tests than younger students (e.g., Finney et al., 2016; Rios & Guo, 2020; Thelk et al., 2009). It is unknown if the difference in effort across student age groups is due to older students understanding the low-stakes nature of the tests better than younger students. Thus, we examined if incoming first-year students differed from more advanced students in their understanding of institutional-level assessment efforts and if this difference related to differences in test-taking motivation.

Students' understanding of the stakes of the assessment may influence their test performance.

Although our institution shares a great deal of information about the purpose and use of institutional effectiveness testing data throughout a students' college career (described in the Methods section), we were unsure of the level of students' understanding. Anecdotally, we heard from some faculty that students do not understand why they are being tested, who sees the scores, or how scores are being used. Some faculty claimed this lack of knowledge resulted in low motivation to perform well on the tests. Other faculty claimed the opposite: more advanced students, unlike incoming students, do understand the purpose of testing and, in particular, understand there are no personal consequences for poor performance. These faculty would often attribute minimal increases in outcome scores to the decreased effort of more advanced students who better understood the low-stakes nature of the tests. Although we understood the logic of both claims, there were no data to support either hypothesis.

Moreover, studies examining college students' understanding of institutional accountability testing is limited. There are concerns that misunderstandings abound. "Many of the criticisms we hear about educational assessments appear to be based on misconceptions. Some of them are due to persons simply misunderstanding the meaning of test scores and their implications for instructional improvement and school accountability" (Goodman & Hambleton, 2005, p. 107). What do students understand regarding accountability testing? One study examined college students' understanding of federal K-12 accountability testing (Zilberberg et al., 2012). Performance was poor. For incoming students, item-level performance ranged from 20% of students answering correctly to 74% of students answering correctly, with less than 50% of students answering six of the nine items correctly. More advanced students had similar levels of misunderstanding, with 50% or less answering six of the nine items correctly. Paradoxically, both incoming and more advanced students indicated a moderate level of confidence in their answers. It is unknown what information was shared with these students about the purpose and use of the K-12 institutional accountability test scores. Thus, prior to conducting the current study, we did not hypothesize expected levels of understanding or how this understanding would relate to modality of testing, year in school, or expended effort on the test. Instead, this was an exploratory study to provide initial insight into students' understanding of the purpose and use of higher education institutional accountability assessment scores.

Despite widespread use of institutional effectiveness testing, students' understanding of its purpose and use remains unclear, with anecdotal reports suggesting misconceptions and varied levels of motivation.

Purpose of the Current Study

Despite the widespread use of testing for institutional accountability and improvement, little is known about students' understanding of institutional accountability testing, and even less is known about how this understanding relates to students' test-taking behavior. The purpose of our study was to examine students' understanding of the purpose of institutional accountability assessment scores at our higher education institution. We examined this understanding for both incoming students at the start of the fall semester and more advanced students with over a year at the institution. Moreover, given the move to remote testing due to COVID, we examined if different modalities of testing were associated with students' understanding of testing for institutional accountability and improvement. We also examined if students' knowledge about the purpose of outcomes assessment related to their effort on these tests. We were unsure if understanding the purpose of institutional accountability test data would relate positively or negatively with effort and if this relation would be moderated by student group or testing modality. No matter the results, there would be implications for testing practices.

Methods

Information Sharing and Testing Procedures

For more than 30 years, James Madison University has used Assessment Days to collect longitudinal data on student learning outcomes. Our model ensures that all incoming students are tested twice: once in the fall semester as incoming students and again in the spring semester after accumulating 45-70 credit hours (Pastor et al., 2019). Although a student completes only four instruments each Assessment Day, 25 different assessments are typically administered, thereby allowing for examination of learning gains on a variety of outcomes.

The goal of Assessment Days is to collect data for interpretation at the program level, not the individual student level. Thus, assessment results are high stakes for the educational programs being assessed but low stakes for students. Individual student scores are not reported, nor do the scores have any individual implications. In other words, students are simply required to attend Assessment Days but are not required to receive passing scores on the assessments.

Our study used data from Fall 2019 (FA19), Spring 2020 (SP20), Fall 2020 (FA20), and Spring 2021 (SP21) Assessment Days. The four administrations differed in the year in school of students being tested, with FA19 and FA20 testing incoming first-year students and SP20 and SP21 testing more advanced students. The administrations also differed in format, with FA19 and SP20 occurring before COVID restrictions and FA20 and SP21 occurring during COVID restrictions. The differences in procedures associated with the two administrative formats are described below.

In-person Proctored Testing prior to COVID

The FA19 and SP20 administrations were typical of the Assessment Day experience at our institution. The FA19 Assessment Day took place during first-year orientation, just prior to the start of fall classes. The SP20 Assessment Day took place on a Tuesday in mid-February. As the name suggests, both Assessments Days took place on a single day. All classes were cancelled until 4:00 p.m. in SP20 to allow students, faculty, and staff to participate in assessment activities.

For both FA19 and SP20 Assessment Days, students were randomly assigned one of three two-hour sessions where they completed three to four assessments aligned with learning objectives in general education and other wide-reaching university initiatives. The majority of testing took place in classrooms where students provided their responses on Scantrons (i.e., optical answer sheets); only 20% of students were tested in computer labs where responses were collected through computer-based testing platforms. All testing sessions were facilitated by trained proctors to ensure standardized conditions.

Students were informed about Assessment Day through the undergraduate catalog, multiple emails, and alerts via university social media outlets. A link to the Assessment Day website was provided in all communications and a video about the purpose of Assessment Day was shown to students just prior to testing. University policy for nonattendance on Assessment Day is a hold placed on the student's record which blocks the student from registering for next semester's classes. Once the student completes makeup testing, the hold is removed.

Remote, Unproctored Testing during COVID

COVID necessitated changes to Assessment Day procedures in FA20 and SP21 (Pastor & Love, 2020). In both FA20 and SP21, students were asked to complete the assessments remotely, without a proctor, and during a specific testing window. In FA20, the testing window spanned from about a week before the start of classes to a little over three weeks after the start of classes. Taking advantage of classes being cancelled for Spring Assessment Day, students were asked to complete their requirement on Assessment Day or the day afterward in SP21.

Thus, FA20 and SP21 Assessment Days differed from typical administrations in that all students were allowed to complete their assessments remotely, without a proctor, on a computer-based testing platform, and at any time they pleased during the testing window. Although students were told holds would be placed if they failed to complete the requirement by the deadline, to minimize disruption in academic progress, no holds were placed in FA20 and SP21.

Similar to the typical in-person administration, in FA20 and SP21 students were informed about Assessment Day through the undergraduate catalog, multiple emails, and alerts via university social media outlets. A link to the Assessment Day website was provided in all communications and a video about the purpose of Assessment Day was shown to all students just prior to the start of remote testing.

Assessment Days collect program-level data, not individual student scores, making them low-stakes for students.

Participants

Table 1 illustrates how the final sample for each Assessment Day administration was obtained. We begin with the number of students required to participate in each administration and end with the sample sizes used in the current study. Students were randomly assigned to assessments and only a subset of students were assigned to complete the assessments used in the present study. Because previous research indicates that students who attend make-up testing sessions differ from those who attend Assessment Day (Swerdzewski et al., 2009), we only considered students who completed the assessments on Assessment Day in FA19 and SP20 or by the deadlines in FA20 and SP21. We further limited the data to only those who provided research consent, were > 18 years of age, and provided valid, non-missing responses to all items. Because preliminary analysis indicated problems with streamlining responses (providing the same response to all items on a given scale), we also deleted students streamlining on any scale¹. Finally, for students tested during more than one administration, we randomly selected which administration's data to retain to reduce dependencies in the data, resulting in the final sample sizes in the far right column of Table 1. A total of 7,513 students comprised the final sample, with 577 (8%), 2,660 (35%), 1,901 (25%), and 2,375 (32%) tested during the FA19, FA20, SP20, and SP21 administrations, respectively. The majority (63%) of students in the final sample self-identified as female and 77% self-identified as White with all other races/ethnicities each represented by 7% or less of the students. These demographics align with those of the institution.

The results suggest that, despite the differences in administration format, Assessment Day scores were relatively stable over time and across cohorts.

Table 1
Process to Arrive at Final Sample Sizes by Assessment Day Administration

Administration	Required to participate	Assigned to complete the study assessments	Completed on Assessment Day or by deadline	≥ 18, provided research consent & valid, non-missing data	Did not provide streamlined responses	Only one record retained for those testing more than once (Final N)
FA19	4466	893	846	781	744	577
FA20	4462	3875	3381	2852	2700	2660
SP20	3797	2962	2600	1951	1906	1901
SP21	3524	3480	3142	2720	2578	2375
Total	16249	11210	9969	8304	7928	7513

Note. Students in a column are a subset of those students to the left of the column. For instance, of the 4,466 students in FA19 who were required to participate in Assessment Day, only 893 were assigned to complete the knowledge of institutional accountability measure (focus of this study); of those 893 students, 846 completed the measure on Assessment Day or by the deadline.

¹ The effort and importance items are responded to on a scale with values of 1 = *Strongly disagree*, 2 = *Disagree*, 3 = *Neither agree nor disagree*, 4 = *Agree*, and 5 = *Strongly agree*. With the exception of a response of 3 to all items, the same response to all items on either subscale is nonsensical given the presence of reverse-scored items. Students who provided responses of all 1s, 2s, 4s, or 5s to items on either subscale were deleted. All Assessment Day knowledge items are responded to on a scale with values of 1=*True* and 2=*False*. Because providing the same response to all items does not align with a reasonable response pattern, students doing so were considered unmotivated and deleted from the data. Interestingly, of the 376 students identified as streamliners, 87% completed the assessments under remote, unproctored conditions in either FA20 or SP21.

Measures

Knowledge about the Purpose and Use of Outcomes Assessment Data

We have never assessed the extent to which students understand the purpose and use of outcomes assessment data. Thus, we created a 12-item dichotomously-scored measure to assess their understanding. We purposefully avoided jargon related to institutional accountability and improvement, such as “value-added”, “accountability”, “accreditation”. Instead, we constructed items to describe these purposes without using unknown terms. Because the items were not created to measure a unidimensional construct, internal consistency reliability was not computed.

Test-Taking Motivation

The Student Opinion Scale (SOS) (Thelk et al., 2009) is a 10-item measure consisting of five items reflecting students’ perceived importance of the assessments they completed (e.g., “Doing well on these tests was important to me.”) and five items reflecting expended effort (e.g., “I gave my best effort on these tests.”). Students indicated their agreement with each statement using a 5-point scale (1 = Strongly disagree to 5 = Strongly agree). The SOS has been employed in at least nine countries, 33 universities, and 55 studies (Sessoms & Finney, 2015). A two-factor structure of scores and longitudinal invariance across one and a half years has been supported. Reliability estimates were adequate in the current study: .78 for perceived test importance and .83 for expended effort.

Results

Knowledge of Institutional Accountability Testing

Students performed incredibly well on the knowledge of institutional accountability testing measure. Across the four samples, students answered 89% to 93% of the items correctly (see Table 2) with more advanced students (SP20, SP21) performing negligibly better, on average, than incoming students. Likewise, there were trivial differences in average knowledge scores for students tested in person (FA19, SP20) versus remotely (FA20, SP21). In sum, all four samples performed well and were not practically different from one another in average scores.

Assessment data is only valuable when students understand its purpose and use.

Table 2
Average Knowledge about Institutional Accountability Testing and Test-taking Motivation

Measure	FA19 (N=577)		SP20 (N=1901)		FA20 (N=2660)		SP21 (N=2375)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Knowledge Score	10.84	1.00	11.22	0.98	10.71	1.19	11.30	1.16
Knowledge %	90.35	8.37	93.52	8.14	89.24	9.91	91.93	9.63
Expended Effort	3.77	0.65	3.83	0.64	3.74	0.67	3.61	0.66
Perceived Test Importance	3.26	0.72	2.88	0.80	3.33	0.69	3.07	0.75

Note. Knowledge Score can range from 0 to 12 with higher scores indicating more knowledge. Knowledge % is the Knowledge Score converted to a percent correct scale and can range from 0 to 100% correct. Expended Effort and Perceived Test Importance can range from 1 to 5 with higher scores indicating higher effort and importance.

Students’ self-reports of expended effort on the tests was moderately high across the four samples with no noteworthy differences based on year in school, and average scores being somewhat higher for students tested in person (FA19, SP20) relative to those tested remotely (FA20, SP21). In alignment with previous research, perceived test importance was moderate and

lower than expended effort for all samples, with higher average scores for incoming students (FA19, FA20) relative to more advanced students (SP20, SP21). Perceived test importance was similar for students tested in person (FA19, SP20) and remotely (FA20, SP21).

To investigate understanding of specific aspects of institutional accountability testing, we examined performance on each of the 12 items. In particular, we were interested in differences in performance on items that did and did not reference personal consequences to the student. In Table 3, we have italicized items that reference personal consequences to students. For example, the first item assesses if students understand that scores from Assessment Day tests are not factored into grade point average. For all but one italicized item, students indicated their understanding of the low-stakes nature of these tests to them personally (i.e., no personal consequences for poor performance). Specifically, almost all students understood that their performance would not appear on their transcript, impact their grade point average, be used to determine future coursework, or affect their academic record. Of note, many incoming students (FA19, FA20) mistakenly believed that faculty could see students' individual performance on the tests. More advanced students (SP20, SP21) performed better on this item; however, this item in general was the most difficult across the four samples. Other than that item and another asking students whether the state requires all state universities to assess student learning, students performed well on the remaining 10 items and differences in performance across testing context (in person versus remote) and year in school were trivial.

Relations between Knowledge and Test-Taking Motivation

We examined the relation between knowledge and test-taking motivation via two approaches. First, we examined the bivariate linear relation between total knowledge score and both effort and perceived test importance (see Table 4). Across testing modality and student age, the relations were not practically different. Effort related positively to knowledge; students who better understood the purpose and use of institutional assessment data reported expending more effort. Knowledge was negligibly related to perceived test importance; the amount a student understood the purpose and use of institutional assessment data did not relate to their perceived value of the test. As expected based on research, effort and importance were positively related.

Next, we examined the relation between effort and knowledge by estimating the average expended effort for students who answered each item correctly versus incorrectly (see Table 5). For items that referenced personal consequences (italicized), we were interested in whether students who understood the personal low-stakes nature of the test expended less effort than those who did not. The opposite occurred — students who correctly understood the lack of personal consequences of these tests had higher average effort. In fact, for all 12 items, students who answered correctly had higher effort with effect sizes ranging from 0.04 to 1.03 SDs.

Discussion

In general, students understood the purpose and use of outcomes assessment testing, with negligible differences in knowledge across testing context and year in school. Knowledge did not relate to perceived importance of these tests but it did relate positively to test-taking effort. Testing context and year in school did not moderate these relations. In sum, students understand the low-stakes nature of outcomes assessment to them personally and increased understanding was not associated with lower expended effort. Instead, students who better understood that outcomes assessment was used for accountability and improvement efforts expended more effort on these assessments.

Implications for Remote Testing

Fortunately, our transition to remote testing was not accompanied by misunderstanding the purpose or use of institutional accountability testing. Instead, knowledge about institutional accountability testing was similarly high for students tested in-person versus remotely. Likewise, knowledge related to perceived test importance and expended test-taking effort similarly for students tested in-person versus remotely. Thus,

These results suggest that remote testing can be a viable option for institutions to continue assessment efforts during times of disruption or beyond, without sacrificing student understanding or motivation.

Table 3
Percentage of Students Answering each Knowledge Item Correctly

Item	Correct Answer	Percentage of Students Answering Item Correctly			
		FA19	SP20	FA20	SP21
<i>Scores from the Assessment Day tests I just completed will be factored into my Grade Point Average (GPA).</i>	False	99	100	99	99
<i>Scores from the Assessment Day tests I completed will be used to determine which courses I enroll in next semester.</i>	False	97	99	91	98
Scores from the Assessment Day tests I just completed will be used to evaluate the quality of James Madison University.	True	93	97	87	90
<i>Faculty can see my individual scores on the tests I completed today.</i>	False	44	73	54	68
James Madison University students are assessed in the Fall as entering students and again after earning 45 to 70 credits.	True	97	95	91	92
I was supposed to prepare for Assessment Day by studying.	False	99	99	95	97
Faculty use results from Assessment Day to make improvements to James Madison University programs.	True	98	98	96	96
<i>My scores on the Assessment Day tests will appear on my transcript.</i>	False	99	98	97	98
Students are expected to have mastered all the concepts assessed during Assessment Day.	False	93	98	88	92
Students are expected to put forth their best effort on the Assessment Day tests.	True	98	99	97	97
<i>My performance on Assessment Day tests does not impact my academic record.</i>	True	95	97	92	93
The state of Virginia requires all state universities to assess student learning.	True	73	71	83	81

Note. The percentage of students who answered the item correctly is often called “difficulty” by assessment experts. Items that reference personal consequences to students are italicized.

Table 4
Relations between Knowledge and Test-Taking Motivation

Variable	Expended Effort	Perceived Test Importance	Knowledge Score
In-person Proctored Testing			
Expended Effort	1.00	0.42	0.16
Perceived Test Importance	0.34	1.00	0.03
Knowledge Score	0.16	-0.01	1.00
Remote Unproctored Testing			
Expended Effort	1.00	0.48	0.15
Perceived Test Importance	0.50	1.00	0.05
Knowledge Score	0.22	0.04	1.00

Note. Correlations above the diagonal are based on the incoming student sample. Correlations below the diagonal are based on the more advanced student sample.

potential arguments to avoid remote testing due to student confusion regarding testing were not supported. These are encouraging results, particularly for assessment programs quickly transitioning to similar testing modalities.

Although students' knowledge did not differ across testing contexts, there was a small difference in expended effort, with slightly lower effort associated with the remote administration. The extent to which lower effort during remote testing affects test performance was recently considered by Alahmadi and DeMars (2022) and is worthy of continued study.

Implications for Increasing Test-Taking Effort

Across the four student samples, expended test-taking effort averaged between 3.61 and 3.83 on a 5-point scale, where higher scores indicate higher expended effort. Given these averages and variability about them, there is an opportunity to increase expended effort in institutional accountability testing contexts. Because knowledge of institutional accountability testing was positively related to expended effort, professionals may suggest increasing knowledge as a possible way to increase effort. Our results and this suggestion align with previous recommendations to explain the purpose of accountability testing, given that examinees stated they would have expended more effort if they had known this information (Zilberberg et al., 2009). Moreover, there is evidence that students do not have positive attitudes about testing, nor do these attitudes improve over time (Paris et al., 1991; Zilberberg et al., 2013; Zilberberg et al., 2014). Of importance is the negative relation between understanding the purpose of accountability tests and disillusionment toward these tests (Zilberberg et al., 2013; Zilberberg et al., 2014). If students are aware of the tests' purpose, they may be less disillusioned and might expend effort. Although we did not examine disillusionment, we did find that understanding the purpose of these tests was related to expended effort. With that said, we have two caveats regarding the suggestion of increasing knowledge of institutional accountability testing as a possible way to increase effort.

First, we cannot claim increased knowledge causes increased effort. We can simply state that those students who expended more effort during institutional accountability testing also tended to understand better the purpose of institutional accountability testing. It may be that conscientiousness influenced both variables; students higher in conscientiousness better focused on information explaining institutional accountability testing (thus, they understand it) and they responsibly put forth more effort on tests. Yet, even if increased knowledge does not directly translate into better examinee behavior, informing students as to the purpose

Encouraging results for remote testing in institutional accountability contexts, with an opportunity to increase test-taking effort through increased knowledge of testing purposes.

Table 5
Average Expended Effort for Students with Incorrect and Correct Answers on each Knowledge Item (N = 7513)

Item	Incorrect Answer		Correct Answer		Mean Diff	d
	N	M (SD)	N	M (SD)		
Scores from the Assessment Day tests I just completed will be factored into my Grade Point Average (GPA).	60	3.38 (0.71)	7453	3.73 (0.66)	0.35	0.50
Scores from the Assessment Day tests I completed will be used to determine which courses I enroll in next semester.	310	3.65 (0.60)	7203	3.73 (0.67)	0.07	0.12
Scores from the Assessment Day tests I just completed will be used to evaluate the quality of James Madison University.	679	3.45 (0.71)	6834	3.75 (0.65)	0.30	0.43
Faculty can see my individual scores on the tests I completed today.	2818	3.69 (0.65)	4695	3.74 (0.67)	0.05	0.07
James Madison University students are assessed in the Fall as entering students and again after earning 45 to 70 credits.	529	3.51 (0.70)	6984	3.74 (0.66)	0.22	0.33
I was supposed to prepare for Assessment Day by studying.	250	3.46 (0.68)	7263	3.73 (0.66)	0.28	0.41
Faculty use results from Assessment Day to make improvements to James Madison University programs.	257	3.27 (0.69)	7256	3.74 (0.66)	0.47	0.70
My scores on the Assessment Day tests will appear on my transcript.	158	3.43 (0.75)	7355	3.73 (0.66)	0.30	0.42
Students are expected to have mastered all the concepts assessed during Assessment Day.	584	3.54 (0.72)	6929	3.74 (0.66)	0.20	0.29
Students are expected to put forth their best effort on the Assessment Day tests.	169	3.07 (0.64)	7344	3.74 (0.66)	0.66	1.03
My performance on Assessment Day tests does not impact my academic record.	467	3.54 (0.66)	7046	3.73 (0.66)	0.19	0.29
The state of Virginia requires all state universities to assess student learning.	1600	3.70 (0.67)	5913	3.73 (0.66)	0.02	0.04

Note. Mean Diff = $M_{Correct} - M_{Incorrect}$, $d = \frac{M_{Correct} - M_{Incorrect}}{\sqrt{\frac{SD_{Correct}^2 + SD_{Incorrect}^2}{2}}}$.

and use of institutional accountability testing is ethical practice. By sharing this information, we are being transparent and respectful to students who are providing the data we use to improve our educational programs and meet accountability requirements.

Second, although increasing student knowledge about institutional accountability testing is a worthy endeavor, it is likely to have little impact on increasing effort if knowledge is already high. At our institution, the vast majority of students understood the purpose and use of this testing. That is, for 10 of the 12 knowledge items, over 90% of the students answered the item correctly. Thus, proposing an intervention to raise knowledge about institutional accountability assessment would have limited impact. We have great comfort knowing our students hear and understand our messaging about outcomes assessment, which includes information shared via email, university social media alerts, on the Assessment Day website, and a video about the purpose of Assessment Day featuring student actors shown just prior to testing. Our results suggest this messaging is working. With that said, we do need to explain more clearly that: 1) faculty see only aggregated results, not results from individual students, and 2) the state requires all universities to gather accountability data. For other institutions where students' understanding of institutional accountability testing is limited, increasing knowledge is not only an ethical obligation but may be an effective and cheap strategy to increase effort. However, future studies are needed to investigate the strength of the relation between knowledge and effort when knowledge is more variable than found in our context.

In closing, we return to the two hypotheses presented by faculty regarding test-taking effort and knowledge of institutional accountability testing. Recall, some faculty believed that students did not understand the purpose of the testing and this lack of knowledge resulted in low motivation to perform well on the tests. Other faculty believed that more advanced students did understand that there are no personal consequences for poor test performance and thus did not expend test-taking effort. Our study identified flaws in both arguments. Incoming and advanced students have a good understanding of institutional accountability testing and its low-stakes nature; this is not an area of concern. Moreover, understanding the low-stakes nature of these tests did not result in decreased effort during the testing process. These results provide support for our current strategies to educate students about the testing and arguments to keep the testing low-stakes in nature.

Increasing student knowledge about institutional accountability testing is a worthy endeavor, but it is likely to have little impact on increasing effort if knowledge is already high. Our results suggest this messaging is working.

References

- Alahmadi, S., & DeMars, C. E. (2022). Large-scale assessment during a pandemic: Results from James Madison University's remote assessment day. *Research & Practice in Assessment*, 17, 5-15. <https://www.rpajournal.com/dev/wp-content/uploads/2022/02/Large-Scale-Assessment-During-a-Pandemic-RPA.pdf>
- Allen, M. J. (2004). *Assessing academic programs in higher education*. Jossey-Bass.
- Barry, C. L., & Finney, S. J. (2016). Modeling change in effort across a low stakes testing session: A latent growth curve modeling approach. *Applied Measurement in Education*, 29, 46-64. <https://doi.org/10.1080/08957347.2015.1102914>
- 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. <https://doi.org/10.1080/15305058.2010.508569>
- Cole, J., Bergin, D., & Whittaker, T. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology*, 33, 609-624. <https://doi.org/10.1016/j.cedpsych.2007.10.002>
- Eklöf, H., Pavešič, B. J., & Grønmo, L. S. (2014). A cross-national comparison of reported effort and mathematics performance in TIMSS. *Applied Measurement in Education*, 27, 31-45. <https://doi.org/10.1080/08957347.2013.853070>
- Finney, S. J., Myers, A. J., & Mathers, C. E. (2018). Test instructions do not moderate the indirect effect of perceived test importance on test performance in low-stakes testing contexts. *International Journal of Testing*, 18, 297-322. <https://doi.org/10.1080/15305058.2017.1396466>
- Finney, S. J., Perkins, B. A., & Satkus, P. (2020). Examining the simultaneous change in emotions during a test: Relations with expended effort and test performance. *International Journal of Testing*, 20, 274-298. <https://doi.org/10.1080/15305058.2020.1786834>
- Finney, S. J., Satkus, P., & Perkins, B. A. (2020). The effect of perceived test importance and examinee emotions on expended effort during a low-stakes test: A longitudinal panel model. *Educational Assessment*, 25, 159-177. <https://doi.org/10.1080/10627197.2020.1756254>
- Finney, S. J., Sundre, D. L., Swain, M. S., & Williams, L. M. (2016). The validity of value-added estimates from low-stakes testing contexts: The impact of change in test-taking motivation and test consequences. *Educational Assessment*, 21, 60-87. <https://doi.org/10.1080/10627197.2015.1127753>
- Freund, P. A., & Holling, H. (2011). Who wants to take an intelligence test? Personality and achievement motivation in the context of ability testing. *Personality and Individual Differences*, 50, 723-728. <https://doi.org/10.1016/j.paid.2010.12.025>
- Fulcher, K. H., & Prendergast, C. O. (2021). *Improving student learning at scale: A how-to guide for higher education*. Stylus.
- Goodman, D., & Hambleton, R. K. (2005). Some misconceptions about large-scale educational assessments. In R. P. Phelps (Ed.), *Defending standardized testing* (pp. 91-110). Lawrence Erlbaum Associates.
- Kerr, K. G., Edwards, K. E., Tweedy, J., Lichterman, H. L., & Knerr, A. R. (2020). *The curricular approach to student affairs: A revolutionary shift for learning beyond the classroom*. Stylus.
- Kopp, J. P., Zinn, T. E., Finney, S. J., & Jurich, D. P. (2011). The development and evaluation of the Academic Entitlement Questionnaire. *Measurement and Evaluation in Counseling and Development*, 44, 105-129. <https://doi.org/10.1177/0748175611400292>
- Mathers, C. E., Finney, S. J., & Hathcoat, J. (2018). Student learning in higher education: A longitudinal analysis and faculty discussion. *Assessment and Evaluation in Higher Education*, 43, 1211-1227. <https://doi.org/10.1080/02602938.2018.1443202>
- Miller, A. R., Miller, K. E., Bailey, S., Fletcher, M., France-Harris, A., Klein, S., & Vickery, R. P. (2019). Partnering academics and community engagement: A Quality Enhancement Plan for a diverse and non-traditional university. *Journal of Community Engagement and Scholarship*, 12, 1, 17-32. <https://digitalcommons.northgeorgia.edu/jces/vol12/iss1/5>
- Myers, A. J., & Finney, S. J. (2021). Does it matter if examinee motivation is measured before or after a low-stakes test? A moderated mediation analysis. *Educational Assessment*, 26, 1-19. <https://doi.org/10.1080/10627197.2019.1645591>
- Paris, S. G., Turner, J. C., Lawton, T. A., & Roth, J. L. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher*, 20(5), 12-20. <https://doi.org/10.3102/0013189X020005012>

- Pastor, D. A., Foelber, K. J., Jacovidis, J. N., Fulcher, K. H., Sauder, D. C., & Love, P. D. (2019). University-wide Assessment Days: The James Madison University model. *The Association for Institutional Research (AIR) Professional File, Article 144*, 1-13. https://www.airweb.org/docs/default-source/documents-for-pages/reports-and-publications/professional-file/apf-144-2019-spring_university-wide-assessment-days-the-james-madison-university-model.pdf
- Pastor, D. A., & Love, P. (2020, Fall). University-wide Assessment during Covid-19: An opportunity for innovation. *Intersection: A Journal at the Intersection of Assessment and Learning*, 2(1). <https://aalhe.scholasticahq.com/article/17617.pdf>
- Penk, C., & Richter, D. (2017). Change in test-taking motivation and its relationship to test performance in low-stakes assessments. *Educational Assessment, Evaluation and Accountability*, 29, 55-79. <https://doi.org/10.1007/s11092-016-9248-7>
- Penk, C., & Schipolowski, S. (2015). Is it all about value? Bringing back the expectancy component to the assessment of test-taking motivation. *Learning and Individual Differences*, 42, 27-35. <https://doi.org/10.1016/j.lindif.2015.08.002>
- Perkins, B. A., Pastor, D. A., & Finney, S. J. (2021) Between- versus within-examinee variability in test-taking effort and test emotions during a low-stakes test. *Applied Measurement in Education*, 34, 285-300. <https://doi.org/10.1080/08957347.2021.1987905>
- Rios, J. (2021). Improving test-taking effort in low-stakes group-based educational testing: A meta-analysis of interventions. *Applied Measurement in Education*, 34, 85-106. <https://doi.org/10.1080/08957347.2021.1890741>
- Rios, J., & Guo, H. (2020). Can culture be a salient predictor of test-taking engagement? An analysis of differential non-effortful responding on an international college-level assessment of critical thinking. *Applied Measurement in Education*, 33, 263-279. <https://doi.org/10.1080/08957347.2020.1789141>
- Satkus, P., & Finney, S. J. (2021). Antecedents of examinee motivation during low-stakes tests: Examining the variability in effects across different research designs. *Assessment and Evaluation in Higher Education*, 46, 1065-1079. <https://doi.org/10.1080/02602938.2020.1846680>
- Sessoms, J. C., & Finney, S. J. (2015). Measuring and modeling change in examinee effort on low-stakes tests across testing occasions. *International Journal of Testing*, 15, 356-388. <http://dx.doi.org/10.1080/15305058.2015.1034866>
- Smith, K. L., & Finney, S. J. (2020). Elevating program theory and implementation fidelity in higher education: Modeling the process via an ethical reasoning curriculum. *Research & Practice in Assessment*, 15, 5-17. <https://files.eric.ed.gov/fulltext/EJ1293385.pdf>
- Stone, J., & Friedman, S. (2002). A case study in the integration of assessment and general education: Lessons learned from a complex process. *Assessment & Evaluation in Higher Education*, 27:2, 199-211. <https://doi.org/10.1080/02602930220128760>
- Sundre, D. L., & Kistantas, 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. [https://doi.org/10.1016/S0361-476X\(02\)00063-2](https://doi.org/10.1016/S0361-476X(02)00063-2)
- Swerdzewski, P. J., Harmes, J. C., & Finney, S. J. (2009). Skipping the test: Using empirical evidence to inform policy related to students who avoid taking low-stakes assessments in college. *Journal of General Education*, 58, 167-195. <https://doi.org/10.1353/jge.0.0043>
- Thelk, A., Sundre, D. L., Horst, S. J., & Finney, S. J. (2009). Motivation matters: Using the Student Opinion Scale (SOS) to make valid inferences about student performance. *Journal of General Education*, 58, 129-15. [doi:10.1353/jge.0.0047](https://doi.org/10.1353/jge.0.0047).
- 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. http://dx.doi.org/10.1207/s15326977ea1001_1
- Zhao, A., Brown, G., & Meissel, K. (2020). Manipulating the consequences of tests: How Shanghai teens react to different consequences. *Educational Research and Evaluation*, 26, 221-251. <https://doi.org/10.1080/13803611.2021.1963938>
- Zilberberg, A., Anderson, R. A., Finney, S. J., & Marsh, K. R. (2013). American college students' attitudes toward institutional accountability testing: Developing measures. *Educational Assessment*, 18, 208-234. <https://doi.org/10.1080/10627197.2013.817153>

- Zilberberg, A., Anderson, R. A., Swerdzewski, P. J., Finney, S. J., & Marsh, K. R. (2012). Growing up with No Child Left Behind: An initial assessment of the understanding of college students' knowledge of accountability testing. *Research & Practice in Assessment*, 7, 12-25. <http://www.rpajournal.com/dev/wp-content/uploads/2012/07/A1.pdf>
- Zilberberg, A., Brown, A. R., Harmes, J. C., & Anderson, R. D. (2009). How can we increase student motivation during low-stakes testing? Understanding the student perspective. In D. M. McInerney, G. Brown, & G. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 255-277). Information Age.
- Zilberberg, A., Finney, S. J., Marsh, K. R., & Anderson, R. A. (2014). The role of students' attitudes and test-taking motivation on the validity of college institutional accountability tests: A path analytic model. *International Journal of Testing*, 14, 360-384. <https://doi.org/10.1080/15305058.2014.928301>