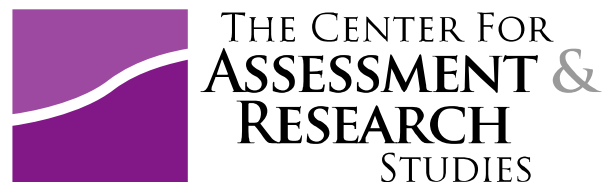


The Student Opinion Scale (SOS)

A measure of examinee
motivation

Test Manual

Donna L. Sundre
2007



MSC 6806
Harrisonburg, VA 22807
540.568.6706 Phone
540.568.7878 Fax
assessment@jmu.edu
www.jmu.edu/assessment

The Student Opinion Scale (SOS) Examinee Motivation Instrument

Table of Contents

SECTION 1. THE NATURE OF THE INSTRUMENT.....3

1.1. INTENDED USE..... 3

SECTION 2. TEST DEVELOPMENT.....3

2.1. THEORETICAL FRAMEWORK..... 3

2.2. ITEM POOL AND SCALE DEVELOPMENT PROCESS..... 4

2.3. SCORING ITEMS 4

SECTION 3. ADMINISTRATIVE PROCEDURES 5

3.1. TRAINING AND PROCEDURES..... 5

3.1.1. *Paper and Pencil Administrations* 5

3.1.2. *Computer-Administered Testing*..... 5

3.2. EXTENT OF EXCHANGEABILITY..... 5

SECTION 4. TECHNICAL INFORMATION 6

4.1. SCORE INTERPRETATION 6

4.2. EVIDENCE OF RELIABILITY..... 6

4.2.1. *SOS Reliabilities for General Education assessment use (four-year colleges)*..... 6

4.2.2. *SOS Reliabilities for high-stakes use (four-year college)*..... 7

4.2.3. *SOS Reliabilities for General Education assessment use (Community College)*..... 7

4.3. EVIDENCE OF VALIDITY 7

4.3.1. *The Substantive Stage*..... 7

SECTION 5. WORK TO BE CONDUCTED TO ENHANCE TEST USE AND INTERPRETATION 14

SECTION 6. ADDITIONAL INFORMATION..... 14

SECTION 7. REFERENCES 14

The SOS Test Manual – 2007

Section 1. The Nature of the Instrument

Purpose of the assessment:
To evaluate examinee
motivation in a testing context.

Standardized tests play an important role in providing interested parties with information concerning student achievement and growth. When test results are analyzed, score interpretations are often questioned on several fronts: it is possible that a lack of motivation to perform well on these tests may produce scores that are spuriously low. When performances are lower than expected, many stakeholders, generally relying on anecdotal evidence, insist that these low results are due to lack of motivation. The true meaning of test scores remains in question. Are the scores lower than the students' true achievement level, or is the observed performance an accurate indication of achievement? In either case, the lack of validity information quiets the conversation and perhaps stifles needed reform. Rarely is a measure of motivation available to help inform interpretations. Lack of motivation may present a potential threat to appropriate interpretation of score meaning. Knowing how large a threat it represents would be very useful. Researchers have been interested in trying to gauge examinee motivation in a variety of testing conditions to explore the presence and magnitude of this potential source of score bias.

The Student Opinion Scale (SOS) is one such means for gauging examinee motivation. The SOS is a self-report tool that has been used in various testing contexts (e.g., consequential and non-consequential). The measure has garnered empirical support for internal consistency and validity of score inferences through more than 12 years of use in research and practical applications.

1.1. Intended Use

The purpose of the SOS is to measure the test-taking motivation of examinees. This measure provides users (e.g., faculty and researchers) with information about student motivation during a testing situation.

The SOS should be administered *after* students have completed the achievement tests, as a post test. The instrument can be administered after either a battery of tests or a single test. If the instrument is used after a battery of tests, the item wording should be modified to read “these tests...” (see Administration Procedures for more details).

The SOS is not intended to be used to make decisions about individual students given its primary function as a self-report of motivation processes. There have been a few internal assessment-related activities at James Madison University (Harrisonburg, VA) that have used SOS scores to identify students with low motivation, in order to filter unmotivated examinees out of a particular analysis. (e.g., Lau, 2006; Harnes, Swerdzewski, & Zeng, 2006; Sundre & Wise, 2003). However, whether this measure functions adequately as a motivation filter is still under investigation.¹

The best use of this instrument is to describe examinee motivation, which can be especially useful in a low-stakes test administration. Low-stakes tests present no personal consequences to the examinee, although the results from such tests may be consequential to the institution administering the exam. The interpretation of achievement test results may be improved with SOS subscale scores and other descriptive statistics (e.g., standard deviation, correlations). Reporting SOS scores along with achievement test results will help guide the audience toward a more robust interpretation of the test scores. If low test scores and SOS scores are observed, users of the test data may question the validity of the scores. However, if SOS scores are high, test score users can more confidently interpret the test scores, as they are more likely to be reflective of true student achievement.

SOS scores are intended to be reported in aggregate form and not for individual students.

Section 2. Test Development

2.1. Theoretical Framework

Examinee motivation draws from the expectancy-value model of achievement motivation theory (Eccles et al., 1983; Pintrich, 1989; Pintrich & DeGroot, 1990). Expectancy-value theorists (e.g., Pintrich, 1989) maintain that motivation to perform well can be defined by three factors: expectancy for success, value of the task, and affect. A person's expectancy for success is integrated in their estimation of ability to be successful on the task (i.e., competence). The value of the task involves a person's perception that this task is important, interesting, or useful. The third component, affect, captures how a person feels about the task; specifically, it involves dispositions such as anxiety.

¹ If motivation filtering is a desired use, the instrument user is directed to also explore research on response time effort (see Wise & Kong, 2005).

The SOS finds its theoretical foundation in the work of:
Pintrich (1989)
Wolf and Smith (1995)
Sundre (1999)
Sundre and Finney (2002)
[among other sources]

There are three important differences to note in the application of the expectancy-value theory in developing test-taking motivation as a construct. The first distinction is in how value is interpreted in this context. For our purposes, value is defined as how important doing well on the test is to the student. The second distinction is in how effort was conceptualized as a component of value in the expectancy-value model and how effort was extracted and made to be more distinct in Wolf and Smith’s (1995) model. Effort is defined as how much mental effort or taxation was involved in answering a test item (Wolf, Smith, & Birnbaum, 1995). The final distinction is that no affective measures were included in scale development. Thus, operationalization of examinee motivation took the form of the importance and effort components.

Following the substantive work, Wolf and Smith (1993) created the Motivation Questionnaire, the predecessor to the Student Opinion Scale (Sundre, 1999). This scale consisted of eight items (five measuring importance and three measuring effort) and was considered unidimensional. Sundre (1999) revised the original scale by adding two effort items and modifying the wording of others in an effort to further delineate and strengthen the representation and conceptualization of examinee motivation. Subsequent research using the scale found two orthogonal factors that measured Importance and Effort; however, fit was optimal following the removal of one item (Sundre & Finney, 2002). While such research garnered further support for the two-factor conceptualization of examinee motivation as measured by the SOS, further study needs to be undertaken to evaluate the performance of the instrument under various conditions and with different samples.

2.2. Item Pool and Scale Development Process

The SOS is a revised version of an earlier eight-item Likert-type instrument (i.e., Motivation Questionnaire; Wolf & Smith, 1993). Researchers had conducted factor analyses that resulted in two orthogonal dimensions that appeared to be assessing the self-reported Importance of the Task completed and the amount of Effort the examinee reported expending. The Effort factor had only three items contributing to it, and Sundre (1999) supplemented the scale to include two additional items designed to assess the Effort construct.

Additional samples were administered the 10-item scale, and the two-factor structure was consistently evidenced. Sundre & Finney (2002) completed a confirmatory factor analysis that provided strong support for the two-factor model with all designed items contributing uniquely to the appropriate factor.

The items comprising the final SOS scale are presented in Table 1. Items are measured on a 1 to 5 scale, where 1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree and 5=Strongly Agree. Negatively worded items should be reverse scored; these are denoted by an asterisk (*).

Table 1
Student Opinion Scale items

- | | |
|-----|--|
| 1. | Doing well on this test was important to me. |
| 2. | I engaged in good effort throughout this test. |
| 3. | I am not curious about how I did on this test relative to others. * |
| 4. | I am not concerned about the score I receive on this test. * |
| 5. | This was an important test to me. |
| 6. | I gave my best effort on this test. |
| 7. | While taking this test, I could have worked harder on it. * |
| 8. | I would like to know how well I did on this test. |
| 9. | I did not give this test my full attention while completing it. * |
| 10. | While taking this test, I was able to persist to completion of the task. |

* Denotes items that are reversed prior to scoring.

2.3. Scoring Items

The SOS is comprised of two subscales: Importance and Effort (see Table 2). Separate scores should be reported for each subscale. It is not recommended that a total score be presented, given strong evidence that examinee motivation is not a unidimensional construct. Reporting a total score would conceal this essential feature of the construct and would therefore be highly misleading, and possibly ambiguous, to the instrument users. The range of possible scores for each subscale is 5 to 25.

The SOS item pool:

Was developed from a study by Wolf and Smith (1993), with additional items added by Sundre (1999).

A confirmatory factor analysis demonstrates an acceptable two-factor solution.

Scoring subscales of the

SOS:

- Effort

- Importance

Evidence does not support the practice of reporting a total score across the two subscales.

Table 2
Test Blueprint for SOS

| Subscale | # Items | Items |
|--|---------|----------------|
| Importance Definition: How important doing well on the test is to the student (the consequence of the test for the student). | 5 | 1, 3, 4, 5, 8 |
| Effort Definition: The perceived degree of work or mental taxation put forth in completing the test. | 5 | 2, 6, 7, 9, 10 |

Administering the SOS...

Is comparable to the administration of any other multiple-choice test; however, the SOS should always be the last test administered in a series of tests.

Section 3. Administrative Procedures

3.1. Training and Procedures

Proctors who administer the SOS require minimal training. Test instructions are straightforward and adhere to traditional testing precepts common among multiple-choice tests. Proctors should familiarize themselves with the instrument and the instrument instructions prior to the testing session to ensure they can expeditiously respond to students' questions during the testing session. As with all assessments, testing sessions that include an administration of the SOS should be standardized across all administrations to decrease threats of construct irrelevant variance in test scores. Assessment session temperature, climate, and lighting should be at acceptable levels to ensure students are able to successfully complete the instrument without environmental distractions.

The SOS is unique among instruments in that it is designed to accompany instruments that measure other constructs. For example, previous administrations of the SOS have paired the instrument with measures of quantitative and scientific reasoning, fine arts, American history, global history, social sciences, and a variety of developmental instruments aimed at motivation, perceived cohesion, and self-efficacy, among other traits. Recalling that the SOS is a measure of a student's motivation toward a given instrument or set of instruments, in the testing session the instrument should be administered at the end of the test or set of tests for which motivation data is desired.

3.1.1. Paper and Pencil Administrations

If administered via paper and pencil, participants receive a reproduction of the instrument and an answer sheet, such as a Scantron form. If using pre-printed Scantron response sheets, students should receive explicit instructions detailing how to code identifying information (e.g., student ID, name) and responses. Instructions should always include how much time a student has to complete the test.

Sample Instructions:

Pass out tests and answer sheets; then read the following:

"Please think about the test or tests that you just completed. Mark the answer that best represents how you feel about statements 1 through 10 below. You have five minutes to complete this instrument."

(Allow the entire five minutes. Remind at 1 minute.)

3.1.2. Computer-Administered Testing

Administering the SOS via computer requires minimal changes to the process for administering the SOS via paper and pencil. In addition to the instructions stated above, students should be briefed in how to operate the testing software (e.g., enter identifying information, select a response). Proctors should be familiar with error recovery procedures in case a testing workstation ceases to function properly.

Administration of the SOS has been successfully conducted in two ways: (1) students are allotted a specific amount of time to complete the SOS, and all students begin the SOS at the same time; or (2) instruments are freely accessible to students during the assessment, and a given student begins the SOS when he or she has completed the other instrument(s) for which motivation data is desired.

3.2. Extent of Exchangeability

Computer-administered testing could be problematic for test takers who are unfamiliar with using a computer. In addition, technical problems may occur with computer-administered testing that do not occur with pencil-and-paper testing. Specific issues related to administration methods should be researched in the future.

To date, there has been no evidence to indicate a lack of exchangeability of the SOS among delivery formats. As with all assessments, students should be familiar with how to operate the testing equipment (e.g. computer, Scantron response form). Level of exchangeability among delivery options shall be researched further.

Section 4. Technical Information

Technical information for the SOS includes:

- Score interpretation
- Evidence of reliability
- Evidence of validity

4.1. Score Interpretation

SOS items are scored on a scale of 1 to 5. The highest score that someone can receive on either subscale is 25. A score of 25 on the Effort subscale means that the student perceived that they fully engaged in effortful behavior on the test (or tests). A score of 25 on the Importance subscale means that the student thought it important to do well on the test or tests. Low scores obviously mean that the student put forth less effort or perceived doing well to be of less importance. It would be inappropriate to say of students who scored low on one or both of the subscales that they were “not motivated”. Similarly, it would be inappropriate to say of students who scored high that they were “motivated”.

It is also essential to report scores for each subscale and to avoid adding the subscale scores together to form a total score. To do so would intimate that the construct is unidimensional, when in fact research has not supported this (see section 4.3.2 – Internal validation).

4.2. Evidence of Reliability

In addition to having used the scale for its intended purpose, accurate score interpretation of the results requires an understanding of the reliability of the scale scores.

Cronbach’s coefficient alpha is a measure of internal consistency reliability. Internal consistency tells us how strongly the items are relating to one another and how well they are working to represent the construct for which they were written to represent. Cronbach’s alpha has an upper bound of 1 (indicating high internal consistency) and a lower bound of 0. While coefficients with a value of .70 or higher have traditionally been considered adequate for scale use, reliabilities above .80 are desirable (Nunnally, 1978, as cited in Lance, Butts, & Michels, 2006). Internal consistency should be reported for each administration of the SOS, as reliability is dependent on the characteristics of the sample and may fluctuate. Next, Cronbach’s alpha values for the SOS subscales that have been observed for various samples in various contexts are provided to demonstrate what level of reliability can be expected.

4.2.1. SOS Reliabilities for General Education assessment use (four-year colleges)

At James Madison University, first-year and second-year students complete the SOS in reference to a required battery of general education assessments in a low-stakes context. Table 3 presents the SOS subscale means, standard deviations, and Cronbach’s alpha values that have been observed for several administration sessions.

Table 3
Cronbach’s alpha, raw scores, and percent scores for SOS subscales from low-stakes use at a mid-Atlantic four-year liberal-arts school (James Madison University)

| Setting | Subscale | Reliability (α) | Raw Score Average | Raw Score SD |
|--|------------|--------------------------|-------------------|--------------|
| Fall 2005 General Education – first-year students (N = 3111) | Importance | .80 | 14.74 | 4.06 |
| | Effort | .84 | 17.20 | 4.04 |
| Fall 2006 General Education – first-year students (N = 3343) | Importance | .80 | 14.94 | 3.98 |
| | Effort | .83 | 17.62 | 3.96 |
| Spring 2005 General Education -- sophomores (N = 1965) | Importance | .83 | 13.84 | 4.36 |
| | Effort | .85 | 16.97 | 3.97 |
| Spring 2006 General Education -- sophomores (N = 2210) | Importance | .82 | 13.37 | 4.28 |
| | Effort | .86 | 17.08 | 4.06 |

Table 4

Cronbach's alpha, raw scores, and percent scores for SOS subscales from low-stakes use at a mid-Western four-year liberal-arts school (Truam State University)

| Setting | Subscale | Reliability (α) | Raw Score Average | Raw Score SD |
|---|------------|--------------------------|-------------------|--------------|
| Fall 2003 and Spring 2004 General Education – seniors (N = 1002) | Importance | .84 | 17.44 | 4.13 |
| | Effort | .85 | 17.91 | 3.88 |

4.2.2. SOS Reliabilities for high-stakes use (four-year college)

Graduating Social Work majors at James Madison University completed the SOS in reference to their motivation on their compulsory comprehensive examination for their major, a very high-stakes assessment. Table 5 presents the SOS data aggregated over a seven-year period spanning from 1998-2004 to allow reporting of a larger sample.

Table 5

Cronbach's alpha, raw scores, and percent scores for SOS subscales from high-stakes use at four-year school

| Setting | Subscale | Reliability (α) | Raw Score Average | Raw Score SD |
|---|------------|--------------------------|-------------------|--------------|
| Fall 2005 Social Work Seniors (N=14) | Importance | .46 ^a | 22.21 | 2.49 |
| | Effort | .83 | 21.29 | 2.89 |
| Social Work Seniors Fall 1998 – Spring 2004 (N=232) | Importance | .59* | 21.64 | 2.74 |
| | Effort | .76 | 21.67 | 2.92 |

^a Low variability in Importance subscale items contributed to resulting low reliability values.

4.2.3. SOS Reliabilities for General Education assessment use (Community College)

Table 6

Cronbach's alpha, raw scores, and percent scores for SOS subscales from high-stakes use at a two-year college

| Setting | Subscale | Reliability (α) | Raw Score Average | Raw Score SD |
|--|------------|--------------------------|-------------------|--------------|
| 2005 Assessment testing for graduating students at a single institution (N=332) | Importance | .89 | 17.29 | 4.41 |
| | Effort | .87 | 18.29 | 3.93 |
| 2005 Assessment testing for graduating students at all institutions in state (N=2045; includes above sample) | Importance | .86 | 17.43 | 4.36 |
| | Effort | .86 | 17.99 | 4.02 |

4.3. Evidence of Validity

According to Benson (1998), there are three critical stages in the validation process of an instrument: the substantive stage, internal validation, and external validation. The substantive stage involves the foundational stage of defining the construct in the theoretical domain and asking critical questions of whether the construct, as defined and operationalized, adequately represents theory. The second stage of this process is the structural stage wherein exploratory and confirmatory factor analysis studies are conducted to test the theoretical assumptions of the construct as operationalized by the instrument. The third and final stage is the external stage where the construct being measured is tested against known correlates or constructs for which there is an established theoretical relationship.

4.3.1. The Substantive Stage

The SOS is comprised of two subscales, one that gauges *how important* the exam was to the individual, and another that measures *effort put forth* by the student. Each contributes different information about task motivation. The Importance factor is related to the subjective task value element of the Expectancy-Value Theory (see section 2.1). It is generally accepted that importance scores in low-stakes testing will be lower than those generated in high-stakes administrations, as the student would perceive the test to have lower importance when there were no personal consequences associated with it. While the Importance subscale does not offer information about levels of particular types of subjective task value (attainment, utility, or intrinsic) held by the examinee, it does provide a gauge of student regard for task value in the general sense. The Effort subscale gives an indication of how much effort and persistence the student put toward completion of the exam. Effort and test performance are positively correlated; however, students may suffer lower scores due to reduced participation, or may endorse lower effort as a result of lack of confidence about participation. A limitation of the Effort score is that it does not provide information about which of these two scenarios has been enacted. Effort and Importance are distinct despite their positive correlation. Certain individuals may choose to put forth high effort when the test has little relevance for the individual, while others will choose not to exert much cognitive energy because of the low stakes.

Originally termed The Motivation Scale, the SOS was developed by Wolf (Wolf & Smith, 1995), who employed eight items to measure a single construct. Later research detected a two-factor structure, with three items measuring Effort and the remaining five items measuring Importance (Sundre, 2000). Sundre (1999) added two items to the original scale, both intended to gauge effort, and reworded some of the existing items to further develop the Effort and Importance subscales. The current version resulted (see Table 1).

Using the earliest version of the instrument, Sundre (1996) compared results from data obtained through a university-wide assessment day to the low- and high-stakes consequence group results from Wolf and Smith (1995). The purpose of this comparison was to determine whether assessment day testing at the researcher's institution (required for incoming students, as well as second-year students with 45-70 completed credits) yielded motivation scores that more closely matched to Wolf and Smith's low-consequences group or high-consequences group. Results of independent sample *t*-tests showed that first-year students had higher motivation scores than sophomores. The freshman means for Sundre's low-stakes group were much lower than Wolf and Smith's high-stakes group means; the same trend was found with Sundre's sophomore scores. However, when freshmen and sophomore scores were compared to Wolf and Smith's low-consequences group, no significant differences were found. Sundre concluded that students at her university participating in assessment-day activities correctly perceived the tests to be of low consequences. Furthermore, examinee motivation was no different than that observed for a course-embedded test that did not count.

Based on the results of numerous studies, Sundre and Moore (2002) reported coefficient alphas in the .80s for both SOS subscales. Administering the instrument to freshmen and juniors at a four-year college, they found that while effort scores are only slightly higher for the incoming students, the juniors attach lower importance to low-stakes tests than freshmen do. Sundre and Moore also indicated that when the SOS is given under high stakes, students endorse similar levels of motivation across the two subscales, whereas in low-stakes testing, effort may be high even if their perception of the test's importance is low.

4.3.2. Internal Validation

Using low-stakes assessment-day data, Sundre (1997) conducted a factor analysis of the original eight-item SOS, and a two-factor structure was supported. This structure was distinct from the one-factor model presented by Wolf and Smith (1995). The study went on to examine whether the factor structure could be confirmed across student cohort groups. The results of the analyses confirmed that the two-factor model was indeed preserved over different grade levels.

Two confirmatory studies have been conducted to test the proposed factor structures, as well as to test for invariance of the SOS across groups of undergraduate students. Sundre and Finney (2002) tested a two-factor model and one-factor model for groups of freshmen and sophomores. The researchers initially did not find support for the two-factor model fit. Items 1 and 5 were identified as being the most problematic. Two alternative models (i.e., model 1: a two-factor without item 1 and model 2: a two-factor without item 5) were tested on an independent sample and they found that the two-factor without item 1 fit somewhat better than the two-factor without item 5. The researchers suggested that since items 1 and 5 were redundant in wording, item 1 could be removed. Sundre and Finney (2002) also investigated measurement invariance across freshmen and sophomores and found that measure was invariant across the groups, but that items 2 and 10 (Effort) were relating differently to the factor for each group. They offered that there could exist a theoretical underpinning to explain why the construct would change from one sample to the other. This needs to be explored further.

Thelk (2004) investigated the factor structure under a consequential testing condition and found that while neither the two-factor nor the one-factor model exhibited overall fit, the two factors fit better than the one factor. Items 3, 4, and 8 were identified as being problematic in terms of large local misfit (i.e., standardized residuals $> |3|$). Thelk suggested that under consequential conditions, there may be a ceiling effect for these items. A ceiling effect causes the variability to be low, which results in low correlations that cannot be reproduced well.

Thelk (2006) researched whether the 2-factor structure could be replicated with data collected from two-year college (community college) students taking tests in a setting similar to the assessment day environment at James Madison University. The two-factor structure was upheld for this population, suggesting that the two groups may be similar in terms

of test-taking motivation. In addition, it appears that the Expectancy-Value theory of motivation may be viable for the community-college student population.

Throughout the confirmatory studies, Importance and Effort have maintained a moderate correlation ($r = .410$ or higher), reinforcing the notion that the two constructs are related yet not redundant.

4.3.3. External Validation

Various studies have been performed with regard to external correlates. Wise (2006) correlated the SOS with an achievement tests (i.e., Information Literacy final exam), response time effort, and SAT Total. Effort was moderately correlated with ILT final score ($r = .460$) and response time effort ($r = .433$). Effort and SAT Total were not significantly related, which is theoretically sound, given the observation that general ability has little to do with Effort. Importance was less correlated with ILT final score ($r = .258$) and response time effort ($r = .293$).

In general, a ceiling effect can be observed with SOS subscale scores: in high-stakes administrations, as scores go up, variance (and therefore reliability) goes down (see Table 5). This phenomenon acknowledges that the SOS has the greatest utility in low-stakes applications, since this is when motivation varies the most, i.e. students may choose to enact or withhold effort with low risk of personal loss. Table 7 contains effect size calculations of the difference in subscale scores when differences in testing stakes are considered.

Table 7
Effect sizes (Cohen’s *d*) for differences between high stakes (Social Work seniors) and low stakes (Spring 2006 assessment day examinees) administrations for SOS subscales

| Importance | Effort |
|------------|--------|
| 2.30 | 1.30 |

To determine whether motivation and self-regulation accounted for more test-score variance in low-stakes testing as opposed to high-stakes testing, Sundre and Kitsantas (2003) gave undergraduate students two versions of a test (multiple-choice and essay) paired with low and high test consequences. The researchers speculated that motivation and self-regulation would be better predictors of final test score in the low-stakes condition, since motivation is uniformly higher in high-stakes conditions. Self-regulation refers to behavioral and metacognitive variables that promote learning for the individual. Similar previous studies had found that self-regulation strategy usage adequately predicted academic achievement, as measured by multiple-choice tests, in both consequential and non-consequential test conditions. Self-regulation predicted essay scores only in the consequential test condition. However, test-taking motivation emerged as a significant predictor of performance in non-consequential conditions, but not in the consequential conditions as had been expected. In addition, test taking motivation was found to explain variance very well in non-consequential essay performances. This research reinforces that motivation accounts for differences in test scores in low-stakes situations.

4.3.4. Norm Groups

To determine how the students at your institution performed in relation to college students at other institutions, refer to Tables 8-11. The percentile ranks associated with each SOS raw score are presented for the total group and by gender. Note that Table 11, which presents information for the high-stakes application, is not broken down by gender. This major is highly populated by females; the sample size for males is too small to allow for comparisons.

Table 8
Percentile ranks for SOS scores for freshmen at a mid-Atlantic 4-year institution

| Score | Total Group | | Males | | Females | |
|-------|-------------|--------|------------|--------|------------|--------|
| | Importance | Effort | Importance | Effort | Importance | Effort |
| 25 | 99.60 | 97.50 | 99.58 | 97.41 | 99.63 | 97.56 |
| 24 | 98.91 | 93.25 | 99.06 | 92.98 | 98.87 | 93.60 |
| 23 | 97.94 | 89.83 | 98.30 | 89.77 | 97.78 | 90.15 |
| 22 | 96.52 | 86.11 | 96.94 | 86.62 | 96.41 | 86.21 |
| 21 | 94.51 | 81.56 | 94.82 | 82.89 | 94.53 | 81.34 |
| 20 | 91.10 | 73.44 | 91.71 | 76.25 | 91.04 | 72.53 |
| 19 | 85.60 | 62.34 | 86.48 | 66.02 | 85.48 | 61.23 |
| 18 | 78.42 | 51.82 | 79.08 | 55.51 | 78.51 | 50.76 |
| 17 | 70.13 | 42.06 | 71.02 | 45.57 | 70.36 | 40.89 |
| 16 | 59.84 | 33.10 | 61.40 | 36.29 | 59.70 | 31.88 |
| 15 | 47.65 | 25.23 | 50.61 | 27.95 | 46.55 | 24.20 |
| 14 | 36.85 | 18.10 | 40.76 | 20.17 | 35.16 | 17.38 |
| 13 | 28.36 | 12.07 | 32.38 | 14.00 | 26.44 | 11.25 |
| 12 | 21.16 | 8.05 | 25.02 | 10.27 | 19.37 | 6.99 |
| 11 | 15.57 | 5.38 | 19.09 | 7.49 | 14.01 | 4.48 |
| 10 | 11.19 | 3.48 | 14.04 | 5.00 | 9.82 | 2.90 |
| 9 | 8.03 | 2.14 | 10.27 | 3.02 | 6.92 | 1.82 |
| 8 | 5.94 | 1.26 | 8.15 | 1.65 | 4.80 | 1.11 |
| 7 | 4.26 | 0.78 | 6.31 | 0.99 | 3.23 | 0.69 |
| 6 | 2.98 | 0.46 | 4.62 | 0.61 | 2.19 | 0.42 |
| 5 | 1.26 | 0.18 | 1.93 | 0.24 | 0.94 | 0.17 |

Note: Data used were collected during Fall 2006 assessment day.

Table 9
Percentile ranks for SOS scores for sophomores at a Mid-Atlantic 4-year institution

| Score | Total Group | | Males | | Females | |
|-------|-------------|--------|------------|--------|------------|--------|
| | Importance | Effort | Importance | Effort | Importance | Effort |
| 25 | 99.66 | 98.26 | 99.34 | 97.76 | 99.82 | 98.61 |
| 24 | 99.21 | 95.36 | 98.42 | 94.33 | 99.61 | 96.12 |
| 23 | 98.76 | 92.53 | 97.76 | 91.70 | 99.32 | 93.17 |
| 22 | 97.94 | 89.12 | 96.51 | 88.41 | 98.79 | 89.76 |
| 21 | 96.47 | 85.18 | 94.33 | 84.72 | 97.72 | 85.88 |
| 20 | 94.21 | 77.44 | 91.70 | 78.52 | 95.73 | 77.42 |
| 19 | 90.86 | 66.49 | 88.80 | 68.64 | 92.14 | 65.86 |
| 18 | 86.20 | 55.84 | 84.39 | 58.43 | 87.48 | 54.87 |
| 17 | 80.45 | 46.15 | 78.06 | 48.81 | 82.18 | 45.06 |
| 16 | 73.64 | 37.58 | 71.67 | 39.72 | 75.18 | 36.74 |
| 15 | 64.25 | 29.16 | 63.18 | 31.29 | 65.33 | 28.38 |
| 14 | 53.67 | 21.65 | 53.36 | 23.85 | 54.30 | 20.91 |
| 13 | 44.28 | 15.20 | 45.32 | 16.86 | 44.17 | 14.72 |
| 12 | 35.86 | 10.27 | 37.75 | 11.26 | 35.24 | 10.03 |
| 11 | 28.71 | 7.35 | 30.37 | 7.91 | 28.17 | 7.29 |
| 10 | 21.97 | 5.23 | 23.32 | 5.60 | 21.48 | 5.19 |
| 9 | 15.79 | 3.51 | 17.06 | 4.15 | 15.29 | 3.27 |
| 8 | 11.70 | 2.40 | 12.91 | 3.03 | 11.24 | 2.13 |
| 7 | 9.14 | 1.70 | 10.47 | 2.24 | 8.61 | 1.46 |
| 6 | 7.01 | 1.18 | 8.17 | | 6.54 | 0.85 |
| 5 | 3.03 | 0.52 | 3.49 | 0.92 | 2.84 | 0.32 |

Note: Data used were collected during Spring 2006 assessment day.

Table 10
Percentile ranks for SOS scores for mid-Atlantic 2-year institution (community college)

| Score | Total Group | | Males | | Females | |
|-------|-------------|--------|------------|--------|------------|--------|
| | Importance | Effort | Importance | Effort | Importance | Effort |
| 25 | 97.44 | 97.59 | 96.97 | 96.21 | 97.75 | 98.50 |
| 24 | 93.67 | 91.42 | 93.56 | 89.39 | 93.75 | 92.75 |
| 23 | 90.21 | 85.09 | 90.15 | 84.09 | 90.25 | 85.75 |
| 22 | 86.75 | 79.82 | 85.61 | 79.17 | 87.50 | 80.25 |
| 21 | 82.23 | 75.00 | 80.68 | 73.86 | 83.25 | 75.75 |
| 20 | 72.59 | 66.11 | 72.35 | 64.77 | 72.75 | 67.00 |
| 19 | 61.45 | 54.82 | 61.36 | 53.79 | 61.50 | 55.50 |
| 18 | 52.56 | 45.18 | 50.76 | 43.94 | 53.75 | 46.00 |
| 17 | 44.43 | 37.50 | 40.91 | 35.98 | 46.75 | 38.50 |
| 16 | 36.90 | 31.17 | 32.95 | 28.79 | 39.50 | 32.75 |
| 15 | 28.77 | 22.14 | 26.14 | 20.83 | 30.50 | 23.00 |
| 14 | 19.73 | 13.25 | 17.05 | 13.64 | 21.50 | 13.00 |
| 13 | 13.70 | 8.43 | 10.61 | 8.33 | 15.75 | 8.50 |
| 12 | 10.84 | 5.87 | 7.95 | 5.30 | 12.75 | 6.25 |
| 11 | 8.43 | 3.46 | -- | 3.41 | 9.50 | 3.50 |
| 10 | 6.17 | 1.96 | 6.06 | 2.65 | 6.25 | 1.50 |
| 9 | 4.67 | 1.05 | 4.92 | 1.52 | 4.50 | 0.75 |
| 8 | 3.77 | -- | 4.17 | -- | 3.50 | -- |
| 7 | 3.01 | 0.45 | 3.41 | -- | 2.75 | -- |
| 6 | -- | -- | -- | -- | -- | -- |
| 5 | 1.36 | 0.15 | 1.52 | 0.38 | 1.25 | 0.25 |

-- indicates that this frequency was not present in the data set.

Table 11

Percentile ranks for SOS Social Work program seniors at 4-year mid-Atlantic institution (high-stakes application)

| | Total Group | |
|--------------|--------------------|---------------|
| Score | Importance | Effort |
| 25 | 92.24 | 90.95 |
| 24 | 79.53 | 75.43 |
| 23 | 67.03 | 62.28 |
| 22 | 51.29 | 48.92 |
| 21 | 34.91 | 36.42 |
| 20 | 21.34 | 25.65 |
| 19 | 12.28 | 15.95 |
| 18 | 7.11 | 9.05 |
| 17 | 5.17 | 5.39 |
| 16 | 4.31 | 3.02 |
| 15 | 3.66 | 1.94 |
| 14 | 2.80 | -- |
| 13 | 1.72 | -- |
| 12 | 1.08 | 1.51 |
| 11 | -- | 1.08 |
| 10 | 0.43 | 0.65 |
| 9 | -- | -- |
| 8 | -- | -- |
| 7 | -- | -- |
| 6 | -- | 0.22 |
| 5 | -- | -- |

-- indicates that this frequency was not present in the data set.

Section 5. Work to be Conducted to Enhance Test Use and Interpretation

Further Studies of the SOS Should Include:

Further exploration of the scale with regard to the third phase of Benson's (1998) strong program of construct validation, the external stage.

The validation of this instrument follows Benson's (1998) strong program of construct validation. As discussed above, extensive work has been done to provide evidence for the first of three stages, the substantive stage. This evidence takes the form of research in the test motivation literature about the construct of interest, which subsequently informed the current 10-item instrument in use today.

The second of the three phases of Benson's (1998) strong program of construct validation is the internal stage. During this stage evidence is amassed to address the internal structure of the instrument, especially the interpretive factor structure of the SOS. At present, there is convincing evidence to support the hypothesized two-factor solution for the instrument.

The third phase in Benson's strong program of construct validation requires the placement of the SOS within the larger context of instruments theoretically hypothesized to correlate with the SOS. Potential instruments that should be related to the SOS are test anxiety, test avoidance, and the citizenship one feels toward a test. One would theorize that test anxiety would be moderately negatively related to the effort subscale of the SOS as students who are anxious about engaging in testing would likely not put forth a great deal of effort because this effort is cognitively dissuaded by the student's somatic reaction to test-taking. Similarly, test avoidance should be theoretically negatively related with both the Effort and Importance SOS factors, given that students who dislike and actively avoid tests are likely to place little value in the tests (SOS Importance) and expend little energy on the test (SOS Effort). A student's level of test-taking citizenship, or the inherent worth and responsibility a student feels toward the taking of a test, should moderately or highly correlate with the SOS Effort and Importance subscales. Future studies should root the SOS in a theoretical nomological net composed of these and other constructs, thus allowing a clearer understanding of how the test-taking motivation and other similar (and dissimilar) constructs interrelate.

Section 6. Additional Information

For more information, please contact Donna L. Sundre at sundredl@jmu.edu or call: (540) 568-3483

To review and download additional information from the Center for Assessment and Research Studies, go to: <http://www.jmu.edu/assessment> -- Look under the Resources and Measures sections.

Section 7. References

- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, 17, 10-17.
- Eccles, J. F., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and Achievement Motives* (pp. 75-146). San Francisco, CA: W.H. Freeman.
- Harnes, J. C., Swerdzewski, P. J., & Zeng, X. (2006). Cluster Two: The Fine Arts (Annual report of general education assessment results). James Madison University Center for Assessment and Research Studies.
- Lau, A. R. & Finney, S. J. (2006, May). *Measuring diversity orientation: An examination of the Miville-Guzman Universality-Diversity Scale Short Form*. Paper presented at the meeting of the Association for Psychological Sciences, New York.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria. *Organizational Research Methods*, 9(2), 202-220.

- Sundre, D. L. (1999). *Does examinee motivation moderate the relationship between test consequences and test performance?* (Report No. TM029964). Harrisonburg, Virginia: James Madison University. (ERIC Documentation Reproduction Service No. ED432588).
- Sundre, D. L. & Finney S. J. (2002). *Enhancing the validity and value of learning assessment: Furthering the development of a motivation scale*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Sundre, D. L., & Kitsantas, A. (2003). 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.
- Sundre, D. L., & Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation. *Assessment Update, 14*(1), 8-9.
- Sundre, D. L., & Wise, S. L. (2003, April). 'Motivation Filtering': *An exploration of the impact of low examinee motivation on the psychometric quality of tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Thelk, A. D. (2004). *Factor structure of the Student Opinion Scale under high-stakes conditions*. Unpublished manuscript.
- Thelk, A. D. (2006). *Examinee Awareness of Performance Expectations and Its Effects on Motivation and Test Scores*. Unpublished doctoral dissertation.
- 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. & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*(2), 162-183.
- Wise, V. L. (personal communication, March 23, 2004).
- Wolf, L. F. & Smith, J. K. (1993, April). *The effects of motivation and anxiety on test performance*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta.
- Wolf, L. F. & Smith, J. K. (1995). The consequences 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). Consequences of performance, test motivation, and mentally taxing items. *Applied Measurement in Education, 8*(4), 341-351.