Abstract

Motivational theory is often used to develop strategies for boosting student effort on assessments, particularly in low stakes situations. Increasing students' cognitive engagement on such assessments may also impact student effort. However, before such interventions can be evaluated, a sound measure of cognitive engagement must be identified. This study examines the factor structure of a scale (CE-S) modified to measure students' cognitive engagement specifically on assessment tests. A 2-factor model of cognitive engagement supports the interpretation of two subscale scores. The relationship between these subscale scores and scores on measures of motivation and goal orientation further supports two separate subscales of cognitive engagement. Future research and implications for use of the CE-S in assessment practice is discussed.

Authors
Whitney Smiley, M.A.
James Madison University

Robin Anderson, Psy.D.
James Madison University

Email
smileywf@dukes.jmu.edu

MEASURING STUDENTS' COGNITIVE ENGAGEMENT ON ASSESSMENT TESTS: A CONFIRMATORY FACTOR ANALYSIS OF THE SHORT FORM OF THE COGNITIVE ENGAGEMENT SCALE

As with K-12 institutions, higher education institutions are feeling the pressure from the state governing bodies to provide evidence that learning is occurring, in return for the hard-earned tax dollars the states dispense to colleges and universities. In response, many higher education institutions are designing methods to assess student learning and development as evidence of the effectiveness of their academic programs. These assessments are typically viewed as low-stakes for the students because there are no consequences regardless of how they perform. However, if institutions want to demonstrate what students are learning to stakeholders, students must be motivated to put forth effort on the test (Wise & DeMars, 2005). It often falls to assessment specialists to ensure that assessment data are collected in a meaningful way, especially in low-stakes situations.

While students may receive no direct consequences from their performance on such assessments, these tests often represent a high stakes situation for the institution. Failure to provide evidence that programs are effective could result in serious consequences at the hands of accrediting organizations and state governing bodies. It is of little surprise that the low-stakes nature for students on such assessments would make institutions skeptical about using findings inferred from low-stakes assessment data. Research findings indicating that low motivation hinders the validity of inferences made from student scores (Wise & Demars, 2005), further support such institutional concerns. Concerns regarding the impact of low motivation on assessment results have prompted assessment practitioners to employ motivational theory in an attempt to find ways to encourage students to put forth effort. However, relying on motivational theory alone may exclude other factors that play a role in student effort on low-stakes assessment.

One factor that is less understood is the role that cognitive engagement plays in student effort. Newmann, Wehlarge, and Lamborn's (1992) definition of cognitive engagement, "the student's psychological investment in and effort directed toward learning, understanding, or mastering the knowledge, skills, or crafts that academic work is intended to promote" (p.12), is specific to academic work situations and is therefore relevant for assessment contexts. For example, students may put forth more effort on assessments that they find more cognitively engaging. Thus, assessment specialists may be able to improve student effort by utilizing more cognitively engaging assessments. We expect that if students are more engaged the costs associated with taking the test (i.e. effort, time, etc.) will be reduced and students should get more out of the test, boosting the value they place on the assessment. As Wigfield and Eccles (2000) pointed out, value is a tradeoff between what students get out of the test and the costs associated with taking the test. This increased engagement and the resulting boost in value placed on the assessment may result in increased effort. However, these are empirical questions and current assessment practices have largely ignored cognitive engagement as an area of research.

"...assessment specialists may be able to improve student effort by utilizing more cognitively engaging assessments."

Cognitive Engagement

School and government policies have been put in place to require students to attend schools; however, engagement in academic settings is tough to mandate. Newmann et al. (1992) point out that disengaged students can disrupt the classroom, skip classes, fail to complete assignments, etc. However, the more typical disengaged student can come to class every day, complete all of their work, behave well, and yet have neither excitement nor commitment to the material. They may in turn lack mastery of the material. Of course, while attendance can be regulated, engagement cannot. In situations where attendance is regulated but engagement is lacking, students may become bored and uninvolved throughout the school day; in many cases, they might as well be absent (Newmann et al., 1992). Because of this, it is important to study cognitive engagement so that policy and practices can be developed to reduce the likelihood of such cognitive absences. This is especially important in low-stakes assessment testing situations where students are mandated to attend but cannot be mandated to engage. If students are not engaged while taking the test, institutions will have assessment results, but what inferences can we draw from these results?

The construct of cognitive engagement can be talked about in a myriad of ways. Appleton, Christenson, and Furlong (2008) reviewed several definitions of cognitive engagement and were able to classify the definitions into eight types: engagement, engagement in schoolwork, academic engagement, school engagement, student engagement, student engagement in academic work, student engagement in/with school, and participation identification. Measuring cognitive engagement during assessments would fall under the student engagement with academic work subtype.

Cognitive engagement in academic work has been defined by Marks (2000) as, "A psychological process involving the attention, interest, investment, and effort students expend in the work of learning" (pp. 154-155). Newmann et al. (1992) defined cognitive engagement in academic work as, "The student's psychological investment in and effort directed toward learning, understanding, or mastering the knowledge, skills, or crafts that academic work is intended to promote" (p.12). Both of these definitions involve psychological investment and effort. The Newmann et al. definition is the more specific one stating that the construct involves engagement for the purpose of mastering knowledge, skills, or crafts; whereas, Marks' definition does not address the issue of purpose for engagement. The definition used by the current study more closely aligns with Newmann et al.'s definition. We are most interested in students' psychological investment directed toward a specific academic event (assessment testing). Students may complete academic work and perform well without being engaged in mastery of material. In fact, a significant body of research indicates that "students invest much of their energy in performing rituals, procedures, and routines without developing substantive understanding" (Newmann et al., 1992, p. 12). Our understanding of cognitive engagement can be furthered by distinguishing among behaviors as on a continuum between deep and shallow engagement (Greene & Miller, 1996). Students who exhibit behaviors that allow them to master academic work are seen to have deep cognitive engagement, while students who exhibit behaviors such as rote memorization and rituals they perceive will help them do well without developing mastery of the material are demonstrating shallow engagement. In the context of assessment testing, deeply engaged students will come in and make sure they read each answer carefully and try to formulate thoughtful answers while students who simply come in and provide vague, unrelated, or not well thought out answers, exhibit behaviors associated with shallow engagement.

To further understand how cognitive engagement may impact student performance, one must understand how cognitive engagement differs from other related constructs. For example, it is important to distinguish between cognitive engagement and motivation. Effort is incorporated into both of the above definitions of cognitive engagement. Motivation scales often include items designed to assess effort as a subscale of motivation (e.g. the Student Opinion Survey; Sundre, 1999). However, engagement implies more than motivation, although motivation is necessary for cognitive engagement. Motivation is more of a general trait; that is, one can be a motivated person without being engaged in a specific task (Appleton, Christenson, Kim, & Reschly, 2006; Newmann et al., 1992). However, cognitive engagement is context dependent. This can be shown in the research of Marks (2000) who found that students in his sample reported higher cognitive engagement behavior in their mathematics courses than in their social studies courses. Marks concluded that this difference shows that cognitive engagement can change across contexts, or in this case, educational experiences.

Another construct that might be confused with cognitive engagement is goal orientation. Goal orientation refers to the reason a person engages in an academic task. Initially, research was focused on two types of goal orientation: performance and mastery (Dweck, 1986; Nicholls, 1984). Performance goals involve competence relative to others whereas mastery goals are seen as competence related to task mastery. However, over time goal orientation has grown to include five different orientation types including the original two, as well as work-avoidance, performance-avoidance, and mastery-avoidance. These avoidance items are used to distinguish between people who want to perform well on a task, versus people who want to avoid performing badly at a task (Baranik, Barron, & Finney, 2010). In Newmann et al.'s (1992) definition of cognitive engagement, they make it clear that the goal of an engaged student is mastery of knowledge, which is a factor in goal orientation.

Consistent with Newmann et al.'s (1992) definition of cognitive engagement, Meece, Blumenfeld, and Hoyle (1988) found a significant relationship between goal orientation and engagement patterns. They found a strong positive correlation between the task mastery subscale of their goal orientation measure on the Science Activity Questionnaire (i.e. a child's goal to learn something new and understand his or her work, or learn as much as possible) and active cognitive engagement. Also, scores on the ego/social scale as well as the work-avoidant scale on the same measure correlated positively with superficial cognitive engagement. This research shows that while these constructs are highly correlated, they are also likely two separate constructs. The difference between these constructs is also contextual. Goal orientation refers to a general orientation toward learning (Meece et al., 1988) whereas cognitive engagement in academic tasks refers to a specific task and can change across tasks.

Problems with Measuring Cognitive Engagement

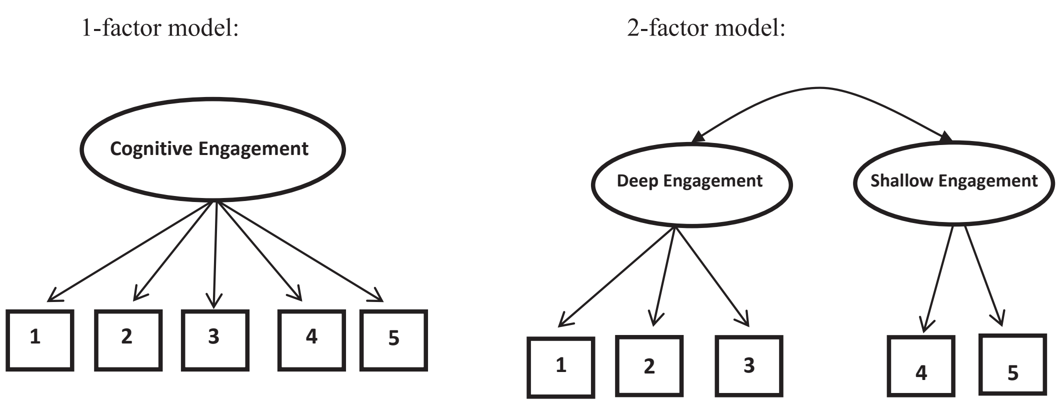
As expressed above, cognitive engagement is an important construct to measure within the context of assessment practice because higher cognitive engagement could result in more effort exerted from students on low-stakes assessment tests. As Newmann et al. (1992) point out, simply attending an environment (assessment day, classroom, or computer lab) and completing necessary work (assessment tests) are not good indicators of cognitive engagement. Rather, engagement is a construct that is used to describe internal behaviors such as effort to learn and quality of understanding. In order to make valid inferences regarding students' level of cognitive engagement across different tasks, researchers must have a measure of cognitive engagement that produces reliable scores and demonstrates evidence for the validity of the inferences made from those scores. Currently, many of the instruments used to measure cognitive engagement are focused on a specific discipline and cannot be used across a variety of tasks. For example, the Science Activity Questionnaire (SAQ) is designed to assess engagement in the context of science activities. The Attitudes towards Mathematics Survey developed by Miller, Greene, Montalvo, Ravindran, and Nichols (1996) assesses academic engagement in mathematics courses. Items on these two scales may not be suitable for tasks outside of the science and/or mathematics classroom.

Current Study

In the current study, faculty members wanted to examine cognitive engagement within the context of a large-scale arts, humanities, and literature assessment situation. The original items on the Attitudes Towards Mathematics Survey (Greene & Miller, 1996) were modified to address, specifically, student engagement on a low-stakes general education fine arts and humanities assessment instrument. Some of the original cognitive engagement items had to be excluded because they were irrelevant to the assessment context. Any time test users shorten a scale (American Educational Research Association [AERA], American Psychological Association [APA], & The National Council on Measurement in Education [NCME], 1999; Smith, McCarthy, & Anderson, 2000) or change the context of the questions (Baranik et al., 2010), the test users should re-examine the reliability of scores and the validity of inferences made from those scores. One such re-examination would be to test whether the factor structure of the original scale applies to the adapted measure.

The dimensionality of the scale can affect scoring, which in turn impacts inferences from findings. In order to determine whether student scores should be interpreted as an overall cognitive engagement factor, or as two separate factors, (deep and shallow) as Greene and Miller (1996) suggested, the dimensionality of the adapted scale was examined using a confirmatory factor analysis (CFA). Researchers hypothesized that because the context of the new cognitive engagement scale was more specific (pertaining to one 45 minute testing session instead of an entire course) the items would be more closely related and represent a unidimensional model. A one and two factor CFA was run to test this hypothesis. For a priori hypothesis models, see Figure 1. Researchers examined global and local fit indices

Figure 1



to determine which model best represents the data. In addition, researchers established the internal consistency (Cronbach's alpha) of the instrument based on the factor structure as recommended by Cortina (1993). Finally, researchers examined the relatedness of this scale to constructs that have shown to be correlated to cognitive engagement, specifically goal orientation and motivation. The development of a sound measure of cognitive engagement for students in large-scale assessment situations could assist faculty and assessment specialist in examining empirical questions such as, "Which assessment tests are most engaging for participants?"

Participants and Procedure

Assessment specialists gathered responses to the short form of the cognitive engagement instrument (CE-S) from students participating in university-wide assessment day activities at a mid-sized, mid-Atlantic university. All incoming freshmen and students with 45-70 earned credits are required to participate in the university's assessment activities. First-year, incoming students complete assessments in the fall on the last day of freshmen orientation. Students with 45-70 earned credits complete assessments in the spring. Students are assigned to testing rooms according to the last two digits of their university identification number. Using this method, the assessment specialists were able to randomly assign students to complete a specific battery of assessment instruments based on their room assignment. Participants included 243 students who completed the assessment activities during the fall of 2010 as incoming freshman or in the spring of 2011 after having earned 45-70 credits. The assessment specialists assigned the students in this study to an assessment battery that included the university's fine arts and humanities assessment tests.

"If students are not engaged while taking the test, institutions will have assessment results, but what inferences can we draw from these results?"

Instruments

In addition to completing the university's open-ended, constructed-response fine arts and humanities general education assessments, each participant completed a series of student development instruments. Among these instruments were scales designed to measure participants' overall goal-orientation, as well as their motivation and cognitive engagement associated with the fine arts and humanities assessment.

Cognitive Engagement – Short form. The CE-S was adapted from a cognitive engagement scale written by Greene and Miller (1996). Five items were adapted from the scale and reworded to specifically refer to the specific large-scale assessment context. Participants are asked to respond to each question using a 1 to 5 scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). Three of the questions are used to measure meaningful cognitive engagement while two questions were used to measure shallow cognitive engagement. Greene and Miller found a Cronbach's alpha of .90 for their longer version of the meaningful engagement subscale and .81 for their longer version of the shallow engagement subscale. The current study examines the internal consistency of the shorter CE-S scale (For the CE-S items, see Appendix).

Student Opinion Scale. The Student Opinion Scale (SOS; Sundre, 1999) is a 10-item questionnaire used to measure examinee motivation. This scale is frequently used to help faculty understand motivation during low-stakes testing situations. Participants are asked to respond to each question using a 1 to 5 scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). The questionnaire contains two subscales measuring importance and effort. Thelk, Sundre, and Horst (2009) used Cronbach's alpha as a measure of reliability and found the subscales to have an alpha value ranging between .80 to .89 for Importance and .83 to .87 for Effort. In the current study, internal reliability was found to be slightly lower for both the effort (α=.74) and importance (α=.77) subscales.

Achievement Goal Questionnaire. The Achievement Goal Questionnaire contains 12 goal orientation items (Finney, Pieper, & Barron, 2004), plus four work avoidance items (Pieper, 2004), and and new mastery-avoidance items from Elliot and Murayama (2008). The AGQ consists of five subscales: mastery-approach, performance-avoidance, work avoidance, performance-approach, and mastery-avoidance that coincide with achievement goal theory mentioned previously. Cronbach's alpha for the subscales range from .65 to .89.

Results

Data Cleaning and Screening

Before running the models, the data were checked for outliers and normality. Data were screened for univariate and multivariate outliers. A graphical plot of the cognitive engagement scores was used to screen for univariate outliers. Researchers used a SPSS macro written by DeCarlo (1997) to screen for multivariate outliers. Analyses suggest that there are no outliers. Univariate normality was assessed by examining skewness and kurtosis. All of the skewness and kurtosis values fell below the recommended cutoffs of |2| for skewness and |7| for kurtosis (Bandalos & Finney, 2010; see Table 1). A histogram with an overlying normal curve was used to examine normality for each item. The responses appeared to depart from the normal curve, a possible function of the categorical nature of the data. Evidence of multivariate non-normality was also found using Mardia's normalized multivariate kurtosis; therefore, the researchers decided to use robust diagonally weighted least squares estimation methods

Table 1.

Item Correlations and Descriptive Statistics (N = 243)

Item	1	2	3	4	5
1	1.00				
2	.46	1.00			
3	.31	.39	1.00		
4	-.28	-.13	-.32	1.00	
5	-.39	-.31	-.26	.62	1.00
M	2.81	3.67	3.45	2.81	2.42
SD	1.04	0.88	0.98	1.04	0.84
Skew	0.14	-0.88	-0.66	0.45	1.03
Kurt	-0.91	0.62	-0.22	-0.66	1.26

"...simply attending an environment and completing necessary work are not good indicators of cognitive engagement."

Factor Analysis

The asymptotic covariance matrix used for the analyses was produced in PRELIS 2.71, and the confirmatory factor analyses were conducted using LISREL 8.72 (Jöreskog & Sörbom, 2005). A unidimensional model was fit to the data to obtain evidence that the CE-S items are measuring cognitive engagement as a single construct. A two-factor model was fit to the data to see if the items are measuring cognitive engagement as two separate factors as previously found by Miller et al. (1996). Hu and Bentler (1998, 1999) recommend reporting at least one absolute fit index and one incremental fit index in addition to X^2. Therefore, four global fit indices were examined to evaluate model fit: the X^2, the standardized root mean square residual (SRMR), the robust root mean square residual (RMSEA), and the robust comparative fit index (CFI). The X^2 test is an absolute fit index that is sensitive to sample size. Like the X^2, the SRMR and RMSEA are absolute fit indices, meaning that they assess how well the hypothesized model reproduces the sample asymptotic covariance matrix. It is recommended the SRMR and RMSEA values be .08 or less (Browne & Cudeck, 1993; Hu & Bentler, 1999). The CFI is an incremental fit index and, unlike the other indices, larger values indicate adequate model fit. Hu and Bentler (1998, 1999) recommend a cutoff of .95 or above.

Table 2 shows the fit indices for the one and two factor models. None of the fit indices for the one factor model are within the suggested cutoffs. However, all of the indices for the 2-factor model are within the recommendations set forth by previous research. Localized misfit in the 2-factor model was investigated by looking at the

Table 2.

Fit indices for the one and two factor models (N = 243)

Model	χ^2	*df*	SRMR	RMSEA	CFI
Two Factor	10.29	4	.05	.08	.97
One Factor	43.63	5	.10	.18	.82

* Note: RDWLS estimation used

Table 3.

Standardized Polychoric Residuals for the One and Two Factor model (N = 243)

Two Factor Model:

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	--				
Item 2	.95	--			
Item 3	-1.70	.80	--		
Item 4	-.17	1.78	-1.70	--	
Item 5	-.73	.33	.41	--	--

One factor Model:

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	--				
Item 2	3.54	--			
Item 3	-.02	1.85	--		
Item 4	1.56	3.07	-.08	--	
Item 5	1.18	1.54	1.71	7.71	--

Table 4.

Standardized Factor Pattern Coefficients, Correlations, and Cronbach's Alpha for the Two-Factor Model (N = 243)

Items	Deep	Shallow	Error Variance	R^2
1	.69		.53	.47
2	.64		.59	.41
3	.56		.69	.31
4		.68	.54	.46
5		.91	.17	.83
Deep	1.00			
Shallow	-.57	1.00		
Cronbach's α	.57	.71		

standardized polychoric residuals. The 1-factor model has several areas of local misfit that exceed the recommended cutoff of |3|; while the 2-factor model had no areas of misfit (see Table 3). Since the 2-factor model had appropriate values for both the fit indices as well as the standardized polychoric residuals, we championed this model. Reliability for the two subscales was also examined. While the deep subscale reliability (α=.56) is not acceptable for program-level inferences, it is higher than expected considering the number of items in the subscale. The two-item shallow subscale has an impressive reliability of .71, indicating it may be appropriate for program-level inferences (Nunnally, 1978). No ΔX^2 was reported as the fit indices for the one factor model clearly did not represent the data.

Having championed the 2-factor model, we looked at the parameter estimates (See Table 4) to understand how much of the variance in the item is accounted for by the latent factor (or how much variance was due to measurement error). The standardized coefficients ranged from .56 to .91 and were all significant at $p < .05$. Squaring these standardized estimates produced the R^2 for each item. R^2 values ranged from .31 to .83. These values indicate that items such as item 3 had low variance accounted for (31%) by the latent factor (deep cognitive engagement) and large amounts of unexplained variance. Item 5 on the other hand had a large amount of variability explained by the latent factor (83%). The standardized error variances ranged between .17 and .69 for all items. Finally, the factor intercorrlations were estimated (Table 4). The deep and shallow factors had a moderate negative correlation (-.57) suggesting that as deep engagement increases, shallow engagement decreases.

"In order to make valid inferences regarding students' level of cognitive engagement across different tasks, researchers must have a measure of cognitive engagement that produces reliable scores and demonstrates evidence for the validity of the inferences made from those scores."

Relationships with External Variables

Table 5 shows the correlations between the two subscales of the CE-S with the SOS total score and each subscale of the SOS and AGQ. The deep subscale is positively related to the SOS total score as well as each SOS subscale, suggesting that as deep engagement goes up so does both effort and importance. However, these correlations are only moderate in nature, suggesting that these two constructs are related but different. The deep subscale also has low to moderate correlations with the AGQ subscales. As expected based on previous literature, AGQ mastery performance subscale scores are related to the deep subscale of the CE-S. There was no significant correlation with the deep subscale of the CE-S to

Table 5.

Correlations with External Variables (N = 243)

	Deep	Shallow
SOS total	.54**	.25**
Effort	.45**	.13*
Importance	.43**	.29**
Mastery Approach	.29**	.14*
Performance Approach	.13*	.11
Mastery Avoidance	.22**	.17**
Performance Avoidance	.05	.13*
Work Avoidance	-.25**	-.13*

** Correlation is significant at the .01 level.

* Correlation is significant at the .05 level.

the performance avoidance subscale of the AGQ and only a slight negative correlation with the work avoidance subscale of the AGQ. While several of the correlations are statistically significant, correlations between the shallow subscale scores and the AGQ subscales were all small (less than r = .17).

Discussion

After examining both a 1-factor and a 2-factor solution, we have championed a 2-factor model of cognitive engagement as measured by the CE-S. This is consistent with Miller et al. (1996) and Meece et al. (1988). Therefore, in this case, shortening a parent questionnaire and changing the context to be more specific did not affect the factor structure of the instrument. Reliability of the subscale scores was higher than expected considering the small number of items composing the two subscales.

We also looked at external correlations which seem to support that cognitive engagement is related to other constructs in expected ways. The positive and moderate correlation between deep cognitive engagement and motivation shows that the two constructs are related, yet distinct from one another (Appleton et al., 2006; Nemann et al., 1992). Deep engagement is also positively related to mastery approach and not related to performance avoidance, which is consistent with Meece et al. (1988). Shallow engagement showed much smaller correlations with these variables, further supporting the two-factor model by showing that the deep and shallow are related to other variables in different, yet predicted ways.

Future Research

In the future, more work should be done to continue to develop the CE-S as a psychometrically sound instrument for cognitive engagement. As mentioned earlier, this work is important to both assessment and educational practices. The development of additional items designed to tap into the deep and shallow engagement factors may improve subscale score reliability. However, we do still want to make sure that we retain only a small amount of items to make sure that use of the cognitive engagement instrument is feasible and easy to add into existing assessment processes.

"Students who exhibit behaviors that allow them to master academic work are seen to have deep cognitive engagement, while students who exhibit behaviors such as rote memorization and rituals they perceive will help them to do well without developing mastery of the material are demonstrating shallow engagement."

In addition to adding items, this study should be replicated with a new sample of participants to examine the stability of the 2-factor model. Another future direction could be developing a cognitive engagement scale to examine cognitive engagement on selected-response assessments, as the CE-S was developed for use with construct-response assessments only. The ultimate goal of this instrument development process should be to develop a general cognitive engagement instrument that can be used flexibly across all assessments.

Once a sound instrument of cognitive engagement is fully developed, future research can examine empirical questions related to assessment practice. One example of an interesting question that could be relevant to an assessment specialist is, "which assessment produces higher cognitive engagement in different contexts, (open-ended vs. multiple choice, paper and pencil vs. computer based testing, etc.)"? Once a good measure is established, assessment specialists may also want to model the relationship among cognitive engagement, effort, and performance. Understanding the connectedness of these constructs may assist in the development of interventions designed to increase students' cognitive engagement on low-stakes assessments. Also of interest may be whether students are giving quality responses on constructed response tests, making sure rapid responding is diminished on multiple-choice assessments, and investigating whether participants are skipping fewer questions when compared to less cognitively engaging assessments.

Conclusion

Cognitive engagement currently is under-researched in applied assessment contexts. The study of this construct may provide unique information regarding students' effort and performance on assessment tests beyond that currently understood through motivation theory alone. Considering the factor structure and reliabilities of the CE-S scale, this scale appears to have potential as a psychometrically sound measure of deep and shallow cognitive engagement. The addition of a few quality items would likely increase the utility of the measure. The establishment of such as method would allow assessment practitioners to test empirically multiple hypotheses regarding the role of cognitive engagement in assessment practice.

References

American Educational Research Association (AERA), American Psychological Association (APA), & The National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing.* Washington, DC American Educational Research Association.

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, *45*, 369-386.

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology, 44*, 427-445.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 125-155). Florence, KY: Routledge Education.

Banta, T. W. (2008). Editor's notes: Trying to clothe the emperor. *Assessment Update*, 20(2), p. 3-4, 15-16.

Baranik, L. E., Barron, K. E., & Finney, S. J. (2010). Examining specific versus general measures of achievement goals. *Human Performance, 23*, 155-172.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98-104.

DeCarlo, L. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*, 292–307.

Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040-1048.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*, 613-628.

Finney, S. J., Pieper, S., L., & Barron, K. E. (2004). Examining the psychometric properties of the Achievement Goal Questionnaire in a general academic context. *Educational and Psychological Measurement, 64*, 365-382.

Greene, B. A., & Miller, R. B. (1996). Influences on achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Psychology, 21*, 181-192.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Jöreskog, K. G., & Sörbom, D. (2005). LISREL 8.72 [Computer software]. Lincolnwood, IL: Scientific Software.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37*, 153-184.

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80*, 514-523.

Miller, R. B., Greene, B. A., Montalvo, G. P., Ravindran, B., & Nichols, J. D. (1996). Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology, 21*(4), 388-422.

Newmann, F. M., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11-39). New York, NY: Teachers College Press.

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328-346.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Pieper, S. L. (2004). Refining and extending the 2 × 2 achievement goal framework: Another look at work-avoidance. Doctoral dissertation, James Madison University, 2004. *Dissertation Abstracts International, 64*, 4436.

Pike, G. R. (2006). Assessment measures: Value-added models and the collegiate learning assessment. *Assessment Update, 18*(4), 5-7.

Shermis, M.D. (2008). The collegiate learning assessment: A critical perspective. *Assessment Update*, *20*(2), 10-12.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*, 102-111.

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).

Thelk, A. D., Sundre, D. L., Horst, S. J., & Finney, S. J. (2009). Motivation matters: Using the Student Opinion Scale to make valid inferences about student performance. *The Journal of General Education, 58*, 129-151.

Terenzini, P. T. (2010). Assessment with open eyes: Pitfalls in studying student outcomes. In J. Fredericks Volkwein (Ed.), *New Directions in Institutional Research Special Issue: Assessing Student Outcomes*, (S1), 29-46.

Wigfield, A., & Eccles, J. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*, 68-81.

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, *10*(1), 1-17.

Appendix

Earlier in today's assessment session you completed two assessment tests designed to assess your performance on learning goals associated with JMU's General Education Cluster 2 (Fine Arts and Humanities). These assessments were the humanities test and the aesthetics test. The humanities test asked you to respond to two separate texts while the aesthetics test asked you to respond to a painting, musical work and play. Please consider these two particular assessments when responding to the following items.

1) When approaching the questions on the Cluster 2 assessments, I planned out or organized my response prior to writing my answer.
2) When preparing to answer the questions on the Cluster 2 assessments, I stopped to reflect on my experience with the works (text, video, music, painting) presented.
3) When experiencing the works (text, video, music, painting) presented in the Cluster 2 assessments, I considered issues related to culture when considering their meaning or significance.
4) When answering the questions on the Cluster 2 assessments, I considered how those reviewing the answers would want me to respond.
5) When answering the questions on the Cluster 2 assessments, I looked for clues of how to respond with the test itself.