International Journal of Teaching and Learning in Higher Education 2022, Volume 33, Number 2, 119-134
http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Psychometric Properties of the Student Course Engagement Questionnaire (SCEQ): Measuring Engagement or Missing the Mark?

Lindsay Masland and Shawn Bergman
Appalachian State University

Jordan Ellis
East Carolina University

Despite widespread use of the Student Course Engagement Questionnaire (SCEQ; Handelsman, Briggs, Sullivan, & Towler, 2005) for the purposes of measuring college student engagement in scholarship of teaching and learning (SoTL) contexts, few studies have been published regarding its psychometric validity. The current study examined the structural and criterion validity of models proposed in the original study and in a subsequent validation study. The results indicated that neither model provided adequate fit to the current data. Concerns regarding structural, criterion, and content validity are discussed, and recommendations regarding the use of the SCEQ are provided in the context of current psychometric evidence and relevant theories of student engagement.

A glance at the agenda of nearly any university faculty meeting in recent years will reveal that at least a portion of the meeting is to be spent discussing metrics. Percentage of students retained, credit hours generated per instructor, time to graduation, and average student GPA are often topics of interest and debate. The metrics are internally scrutinized and publicly displayed, sometimes to the chagrin of the faculty members responsible for producing such numbers. Although the utility of this sort of assessment can be debated, it seems unlikely that the *accountability* frequently tied to such assessments will depart the higher education landscape any time soon. As such, colleges and universities are finding themselves in the position of needing to demonstrate that students are successful in their schools.

Student success can be measured in a variety of ways, but a common target is student retention. That is, if students do not return to their institutions year after year, then they cannot graduate, and the return on any financial investments will be extremely low. The research on college student retention is extensive (see Kuh, Kinzie, Buckley, Bridges, & Hayek, 2007, for a review), but it can be distilled into three general predictors of retention and success: 1) student background factors such as socioeconomic status, availability of financial aid, and quality of high school preparation, 2) institutional features such as student/learning-centered institutional missions and the availability of student support services, and 3) student engagement such as student/faculty interaction and time invested in collegiate activities. In contrast to the less-malleable factors of student background and institutional characteristics, student engagement, which is defined as the "quality of effort and involvement in productive learning activities" (Kuh, 2009, p. 6) has been viewed as a target worthy of augmentation (Pascarella, 2001). As such, the National Survey of Student Engagement (NSSE, 2008) was developed to measure and support efforts to improve student engagement. Specifically, the NSSE has been conceptualized as providing an institutional perspective of engagement that divides the construct into five sub-domains: academic rigor, active and collaborative learning, interactions between students and faculty, campus support, and quality of academic experience (Kahn, 2014).

Although an institutional perspective of student engagement is certainly informative, some have criticized that the NSSE directs researchers to conceptualize the construct in limited ways. That is, the NSSE focuses on *macro*-engagement, or the extent to which a student is engaged in his or her post-secondary institution *in general*, whereas the most malleable type of student engagement is likely to be *micro*-engagement, or the engagement that students experience in the context of individual academic courses (Taylor et al., 2011; Kuh, 2001; Pascarella, 2001). Although students bring certain unchangeable background characteristics into the classroom, research indicates that qualities of the classroom environment affect both the beliefs that students hold about themselves as learners, as well as the types of academic skills that students utilize in service of classroom success (Zumbrunn, McKim, Buhs, & Hawley, 2014). Therefore, it is likely that the sum total of the various levels of malleable micro-engagement displayed in each of a student's courses contributes to the overall institutional (i.e., macro-) engagement that is so predictive of student retention and success. If instructors are interested in contributing to student retention efforts via increased student engagement, they should attend to the levels of micro-engagement that students display in their own classrooms. Furthermore, because quality instructional practices lead to increased student engagement, which leads to academic success, it is important for instructors to have well-validated measures of the micro-engagement construct for the purposes of measuring and improving classroom engagement.

The (Intended) Current Study

At this point, it bears noting that the original intention of the authors of this manuscript was unrelated to the psychometric properties of measures of student engagement. In contrast, the original plan was to estimate a longitudinal model of changes in student engagement in psychology classes over the course of a semester. Through the use of confirmatory factor, exploratory factor, and regression analysis, we planned to examine how changes in engagement predict relevant outcome variables, including academic performance and course satisfaction. When we designed the original study, we did what we expect most researchers do: we completed a literature review of the constructs of interest and based on the results, we selected appropriate measures of those constructs. To measure the construct of student engagement, the Student Course Engagement Questionnaire (SCEQ; Handelsman et al., 2005) was chosen. Our review indicated that many researchers had utilized the SCEQ to measure engagement in the past. In fact, in addition to being cited over 700 times in the literature, the SCEQ was also cited as a psychometrically-valid scale in several publications regarding best practices in the measurement of student engagement (e.g., Lovelace & Brickman, 2005; Mandernach, 2015; Zabel & Heger, 2015). However, we fell prey to the incorrect assumption that the wide use of a scale is indicative of the psychometric validity of that scale.

After collecting semester-long data from nearly 800 students across more than 40 courses, we sat down to analyze them. Before launching into the estimation of our longitudinal models, we sought to confirm the structure of the scale that we used to estimate student engagement. Modeling experts will recognize that verifying the purported four-factor structure of the SCEQ through confirmatory factor analysis is a necessary preliminary step (Byrne, 2013; Crockett, 2012; Kline, 2015). To our surprise, the scale structure was not replicated. We were able to complete the intended analyses on a modified data set, but because of our psychometric concerns regarding the SCEQ, we questioned the validity and generalizability of our results. Thus, the previous study has been relegated to someone's file drawer.

The (Actual) Current Study

Although our initial inclination was to move on from this failed study, after further reflection, we recognized that others might benefit from the insights we gained from our work. However, this study is unique in that we have no a priori hypotheses to present. This was not designed as a psychometric validation study, and it lacks many of the elements one might expect in a conventional examination of construct validity. Nonetheless, in service of the goal of increased rigor in higher education assessment, we feel obligated to share these results. As such, this article will present a critique of several aspects of the SCEQ's construct validity through confirmatory and exploratory factor analyses and regressions, as well as recommendations specific to the use of the SCEQ in future Scholarship of Teaching and Learning (SoTL) with college student populations.

Method

Participants and Procedures

Participants were 792 undergraduate students enrolled in 41 separate introductory psychology classes taught at a state university. Data from the classes were collected over the course of three consecutive academic semesters. The majority of the student participants were White (88.9%; Black = 4.7%; Hispanic = 2%; Asian = 1%), female (67.0%) and freshmen (69.3%; sophomores = 18.8%; juniors = 7.9%; seniors = 3.6%). The average age of the participants was 18.75 years (SD = 1.14). Student engagement was assessed during the middle of the semester, and outcome data was collected during the last two weeks of classes. Final grades and learning management system activity were collected from each course instructor after grades were submitted. All of the assessments were collected using an online survey tool.

Measures

Student Engagement. Student micro-engagement within each course was assessed using the 23-item Student Course Engagement Questionnaire (Handelsman et al., 2005). The SCEQ asks students to rate the extent to which the behaviors, thoughts, and feelings described by each item relate to their own experience. Items are rated on a 5-point scale ranging from "not at all characteristic of me" to "very characteristic of me". See Table 1 for the complete list of items.

To create the SCEQ, Handelsman and colleagues (2005) asked undergraduate students and faculty members to generate items reflective of student engagement. It bears noting that the SCEQ was not created to align with any of the existing theories of student engagement, but rather to reflect student and faculty perceptions, as revealed through an inductive process.

The resulting 27 items were then administered to a sample of 266 undergraduates from several psychology, politics, and math classes, and exploratory factor analysis yielded a four-factor solution. Four of the original 27 items were dropped due to low factor loadings (i.e., they were removed due to insignificant relationships with the factors), resulting in a 23-item measure characterized by the following factors: skills engagement (Cronbach's α = .82), emotional engagement (Cronbach's α = .82), participation/interaction engagement (Cronbach's α = .79), and performance engagement (Cronbach's α = .75).

In addition to acceptable internal reliability estimates, Handelsman and colleagues (2005) examined

Table 1
CFA Results- Handelsman et al. (2005) Model

Item	Skills	Emotional	Participation	Performance
1. Making sure to study on a regular basis.	.72			
2. Putting forth effort.	.85			
3. Doing all the homework problems.	.72			
4. Staying up on the readings.	.67			
5. Looking over the class notes between classes to make sure I understand the material.	.66			
6. Being organized.	.66			
7. Taking good notes in class.	.70			
8. Listening carefully in class.	.78			
9. Coming to class every day.	64			
10. Finding ways to make the course material relevant to my life.	119.74(-.36)	.82	103.45(-.34)	
11. Applying the course material to my life.	116.22(-.34)	.80	113.79(-.38)	
12. Finding ways to make the course material interesting to me.		.79		
13. Thinking about the course between class meetings.		.71	92.34(.30)	
14. Really desiring to learn the material.		.81	113.05(.34)	
15. Raising my hand in class.		123.52(-.34)	.81	
16. Asking questions when I don't understand the instructor.			.80	
17. Having fun in class.		173.61(.40)	.78	
18. Participating actively in small-group discussions.			.80	
19. Going to the professor's office hours to review assignments or tests or to ask questions.			.50	
20. Helping fellow students.			.70	
21. Getting a good grade.				.89
22. Doing well on the tests.				.80
23. Being confident that I can learn and do well in the class.	138.14(.36)	258.44(.43)	191.09(.38)	.87

Notes: Factor loadings in bold are the standardized CFA factor loadings from the data in the current study. Numbers in italics are the modification indices (i.e., approximate decrease in model chi-square if parameter were to be estimated in a subsequent model) and the estimated standardized expected parameter change from the CFA model that indicated potential cross-loading items.

correlations among the factors, which ranged from .23 to .44, and determined there was adequate discriminant validity. Initial convergent validity was also demonstrated via expected correlations with other relevant measures of engagement and motivation, including self-reported course-specific and general motivation, incremental self-theory, and goal orientation. As evidence of criterion validity, regression analyses indicated that performance, participation/interaction, and skills engagement all demonstrated significant relationships with midterm exam grades. In contrast, performance engagement was the only factor that predicted significant variance in homework assignment grades. Finally, the participation/interaction factor was the only significant predictor of final exam grades. As noted in previous research, these results provide evidence that the SCEQ meets minimum standards of reliability and validity (e.g., Lovelace & Brickman, 2005; Mandernach, 2005; Zabel & Heger, 2015).

Several years later, the SCEQ was revisited by a second group of authors for the purpose of contributing to the validation of the scale through a paper published in the *Journal on Excellence in College Teaching* (Taylor et al., 2011). Taylor and colleagues (2011) sought to explore whether the SCEQ would maintain the same factor structure when administered to students in a large-scale introductory marketing class, a different course type than those examined in Handelsman and colleagues' study. Their confirmatory factor analysis seeking to verify the original model demonstrated poor fit. As such, Taylor and colleagues (2011) sought to devise a better-fitting model via exploratory factor analysis. The model that demonstrated the best fit was a five-factor solution that replicated three of the factors from the original SCEQ model (Handelsman et al., 2005) and that split the fourth factor into two factors (see Table 2). Despite conceptual replication of three of the factors from the Handelsman et al. (2005) model, empirical replication was not achieved, as some items cross-loaded on multiple factors and were removed, whereas other items loaded on factors in ways not specified by the original model. Additionally, the fourth and fifth factors identified by Taylor and colleagues reconceptualized skill engagement as two separate factors: attention directed inside the classroom and attention directed outside of the classroom. This finding has interesting implications regarding which items should be used to create which factors, as well the total number of factors to be used when assessing classroom engagement via the SCEQ.

Table 2
CFA Results- Taylor et al. (2011) Model

Item	Skills-In	Skills-Out	Emotional	Participation	Performance
1. Making sure to study on a regular basis.	-	-	-	-	-
2. Putting forth effort.	17.90(1.23)	.78		47.97(-.41)	52.92(.42)
3. Doing all the homework problems.	.74				
4. Staying up on the readings.	-	-	-	-	-
5. Looking over the class notes between classes to make sure I understand the material.	17.96(-.79)	.50			
6. Being organized.	-	-	-	-	-
7. Taking good notes in class.	.72				
8. Listening carefully in class.	.81	15.76(.50)			
9. Coming to class every day.	.67				
10. Finding ways to make the course material relevant to my life.			.91		
11. Applying the course material to my life.			.88		
12. Finding ways to make the course material interesting to me.	-	-	-	-	-
13. Thinking about the course between class meetings.	-	-	-	-	-
14. Really desiring to learn the material.	-	-	-	-	-
15. Raising my hand in class.	-	-	-	-	-
16. Asking questions when I don't understand the instructor.				.77	
17. Having fun in class.				.76	
18. Participating actively in small-group discussions.				.73	
19. Going to the professor's office hours to review assignments or tests or to ask questions.				.47	
20. Helping fellow students.				.71	
21. Getting a good grade.					.98
22. Doing well on the tests.					.83
23. Being confident that I can learn and do well in the class.	-	-	-	-	-

Notes: Factor loadings in bold are the standardized CFA factor loadings from the data in the current study. Numbers in italics are the modification indices (i.e., approximate decrease in model chi-square if parameter were to be estimated in a subsequent model) and the estimated standardized expected parameter change from the CFA model that indicated potential cross-loading items. "-" represents items not included in the final Taylor et al. (2011) model.

Student Outcome Variables. To examine criterion validity for the scale in the current study, student performance, online activity, and course satisfaction were assessed. Student performance was assessed by course instructor reports of each student's final grade percentage. Student online activity was assessed by acquiring the number of times the student logged into their course website via the university's online learning management system throughout the semester. The course instructor provided the frequency of online visits at the end of the semester. Student satisfaction was assessed via the following item: "Overall, I rate this course as excellent", to which students indicated their level of agreement on a 5-point rating scale (1 "strongly disagree" to 5 "strongly agree").

Data Analysis Plan

Several steps were used to test the fit of the previously proposed models. First, confirmatory factor analyses through MPlus version 8.0 (Muthén & Muthén, 1998-2017) with the weighted least squares with means and variance adjusted estimator (WLSMV; Brown, 2015) were utilized to determine the fit of the present data to Handelsman et al.'s four-factor and Taylor et al.'s five-factor model. Criteria for good model fit were determined by the following fit indexes: Comparative Fit Index (CFI), Tucker-Lewis Fit Index (TLI), root mean square residual (RMSEA), and weighted root mean square residual (WRMR). Values at or above .95 on the CFI and TLI have been advised as the cutoff that represents a well-fitting model (Hu & Bentler, 1999). Hu and Bentler (1999) suggested that a RMSEA value of 0.06 represents a good model fit, RMSEA values above 0.10 indicate a poor fitting model, and values of .08 indicating adequate fit (Browne & Cudeck, 1993; MacCallum, Browne, & Sugawara, 1996; Steiger, 1989). A cut-off value below 1.0 for WRMR is generally regarded as indicating good model fit (DiStefano, Liu, Jiang, & Shi, 2018).

After the fit of the previously proposed models was determined, a series of exploratory and confirmatory analyses were conducted in MPlus version 8.0 to determine the model that best fit the current data. While we acknowledge that this type of exploratory approach will always lead to a model with superior fit to the study's data, we sought to determine the extent to which this exploratory approach, similar to the ones used in previous empirical investigations of the SCEQ, would lead to a similar factor structure of the measure. First, an exploratory factory analysis using a WLSMV estimator with promax rotation was used to determine the number of latent factors that were present in the data. These exploratory models did not restrict the item loadings and allowed all 23 items to load on all of the factors in the model. An examination of the scree plot and the fit indices for each of the factor solutions was used to determine the number of factors that provided the most parsimonious fit to the data while still reflecting the theoretical conceptualization of a classroom engagement model. Once we discovered the number of latent factors that best fit the data, a confirmatory factor analysis approach carried out in Mplus 8.0 using a WLSMV estimator was utilized as a scale reduction technique. The preliminary confirmatory factor analysis (CFA) model used those items that had EFA factor pattern coefficients above .30 on only one factor. Items from the EFA that had factor pattern coefficients above .30 on multiple factors were not included in the CFA. During the CFA iterative process, items that cross-loaded were dropped one-by-one in accordance with their loadings, and the CFA model was re-estimated until a model with acceptable fit was found.

After the model of best fit using the same fit indices and criteria employed for the Handelsman and Taylor models was determined from this exploration, the internal reliability of the factors from the three models was estimated using Cronbach's alpha (Cronbach's alpha is a measure of the intercorrelation among items, and it is used to assess the extent to which a group of items "hangs together"). Zero-order correlations and associated 95% confidence intervals were then calculated between each factor and the student outcomes. These correlation coefficients were first used to compare the differential criterion-related validity between the Handelsman et al. (2005), Taylor et al. (2011), and newly-created model of best fit.

Finally, the criterion-related validity of the three models was explored through the use of multiple regression paired with a relative weights analysis (RWA; Johnson, 2000). This set of analyses first included all of a model's factors as predictors of an outcome in multiple regression analyses to determine the joint prediction of the factors on a specific outcome. These regression analyses were carried out in Mplus 8.0, which allowed the outcomes to correlate and accounted for the correlated errors in the linear equations. Next, an RWA was conducted to determine the relative contribution that each factor made to the prediction of each outcome. This determination is not easily achieved through the use of conventional processes such as examining the zero-order correlations, standardized the regression weights, or squared semi-partial correlations when the multicollinearity between predictors in a regression model is relatively high (Johnson & Lebreton, 2004; Lebreton, Ployhart, & Ladd, 2004) Given that previous studies had shown a robust correlation among SCEQ factors, which would introduce multicollinearity into the regression models when the

Table 3
EFA Total Variance Explained

Factor	Total	% of Variance	Cumulative %
1	9.21	40.02%	40.02%
2	2.20	9.57%	49.60%
3	1.83	7.96%	57.55%
4	1.58	6.88%	64.43%
5	1.05	4.56%	68.99%
6	0.71	3.10%	72.10%
7	0.68	2.95%	75.05%
8	0.60	2.60%	77.65%
9	0.57	2.48%	80.13%
10	0.52	2.24%	82.37%
11	0.49	2.12%	84.49%
12	0.44	1.93%	86.42%
13	0.42	1.82%	88.24%
14	0.38	1.65%	89.89%
15	0.37	1.60%	91.50%
16	0.33	1.42%	92.92%
17	0.29	1.27%	94.19%
18	0.28	1.22%	95.41%
19	0.28	1.21%	96.62%
20	0.24	1.06%	97.68%
21	0.21	0.90%	98.58%
22	0.19	0.83%	99.41%
23	0.14	0.59%	100.00%

Notes: Extraction Method: Weighted least square mean and variance adjusted estimator

Table 4
Exploratory Factor Analysis Fit Indices

Solution	χ^2	*df*	CFI	TLI	RMSEA	RMSEA 90 CI	WMRM
2-factor	4350.70	208	0.80	0.76	0.16	.15-.16	3.21
3-factor	2624.29	187	0.88	0.84	0.13	.12-.13	2.18
4-factor	1070.11	167	0.96	0.93	0.08	.08-.09	1.18
5-factor (selected)	556.22	148	0.98	0.97	0.06	.05-.06	0.76
6-factor	358.11	130	0.99	0.98	0.05	.04-.05	0.58
7-factor	268.30	113	0.99	0.98	0.04	.04-.05	0.47
8-factor	184.99	97	1.00	0.99	0.03	.03-.04	0.38
9-factor	152.93	82	1.00	0.99	0.03	.03-.04	0.32

Note: EFA with WLSMV estimator and Promax rotation. EFA model was selected was the most parsimonious model that provided "good fit" to the data (see Browne & Cudeck, 1993; MacCallum et al., 1996; Steiger, 1989) as determined by fit indices cutoffs. The 5-factor EFA solution is the first model that provided good fit to the data. When comparing the 5-factor EFA solution to the 4-factor solution, the RMSEA TLI, and WMRM fit indices suggested that the 5-factor solution was a better fit to the data compared with the 4-factor solution. When comparing the 5-factor solution to the 6-factor solution, the solutions produced comparable fit indices (e.g., the 90% RMSEA CIs overlapped). Thus, the more parsimonious 5-factor solution was used as the starting point for the iterative CFA process.

resulting factors were used as predictors, RWA was employed as a way to further explore the relative contribution of each factor in the multiple regression models. RWA computes an estimate of the proportionate contribution that each predictor makes to R^2 by considering both a predictor's independent relationship with an outcome and the joint relationship a predictor has with an outcome when considered with the other predictors in a regression analysis. The results of a RWA can be expressed as both a relative weight (RW) for each predictor, which sum to the overall model R^2, and a relative importance (RI) score, which is the percentage of the R^2 value accounted for by a predictor, summing to 100%. The RWA was conducted in R using the "iopsych" package and allowed the outcomes to correlate (Goebl, Jones, & Beatty, 2016). The results of the correlation, regression, and RWA were used to determine the extent to which each factor was both a statistically and practically significant predictor of a specific outcome.

Results

CFAs of the Proposed Structure of the SCEQ

CFA using all 23 items included in Handelsman et al.'s (2005) original model for the SCEQ showed the current data were a poor fit to the hypothesized four-factor model (χ^2 (224) = 2635.86, $p < .001$, CFI = .88, TLI = .87, RMSEA = .12, RMSEA 90% CI = .11 - .12; WRMR = 2.60). An examination of the model's CFA factor loadings and modification indices revealed a number of potential cross-loading items (see Table 1). Confirmatory factor analysis of the 15 items included in Taylor et al.'s (2011) five-factor model showed the data were an adequate fit (χ^2 (80) = 489.50, $p < .001$, CFI = .97, TLI = .96, RMSEA = .08; RMSEA 90% CI = .07 - .09; WRMR = 1.38). An examination of the model's CFA factor loadings and modification indices, however, revealed a number of potential cross-loading items (see Table 2), and the internal consistency of the out-of-class skills factor was unacceptable (Cronbach's alpha =.50).

Although Taylor et al.'s (2011) model indicated an "adequate" fit and a better fit than the Handelsman et al. (2005) model, fit indices suggested a more appropriate model could be derived. To determine the model that best fit the data in the present study, an exploratory factor analysis (EFA) was conducted that allowed all 23 items in the model to load on models with one to ten factors. An examination of the scree plot (see Table 3 for initially extracted eigenvalues) and fit indices for each of the proposed EFA models indicated that a five-factor solution provided the most parsimonious and best fit to the data (see Table 4). An iterative CFA process for scale reduction was conducted until an acceptable fit was achieved (see Tables 5 and 6). The study's final model included 16-items that loaded on five-factors (see Table 4) and had a good to adequate fit to the data, χ^2 (94) = 401.93, $p < .001$, CFI = .98, TLI = .98, RMSEA = .06; RMSEA 90% CI = .06 - .07; WRMR = 1.13. All five of the factors in the study's final model were found to have acceptable internal reliability estimates (see Table 7).

The study's final 16-item, five-factor model resembled Taylor et al.'s (2011) model, although more items were retained within the participation and skills domains (see Tables 2 and 4). Similar to Taylor et al.'s (2011) model, the final model indicated the divergence of the skills component of engagement into two factors; however, the present model retained a three-item out-of-class skills factor. Individual items retained in both models tended to load within the same theoretical domains, with the exception of "putting forth effort", which loaded on the out-of-class skills factor in Taylor et al.'s (2011) model and on the in-class skills factor in the current model. Interestingly, "doing all of the homework problems" loaded on the *in*-class skills factor in the final model, as it had in Taylor et al.'s (2011) model. The study's participation factor included three items, compared to Taylor et al.'s (2011) five items with only two overlapping items (i.e., "asking questions when I don't understand the instructor" and "participating actively in small-group discussions"). Both models indicated identical two-item emotional and performance engagement factors.

Table 5
Confirmatory Factor Analysis Fit Indices for Each Iteration Towards the Final Model

Iteration #	χ^2	*df*	CFI	TLI	RMSEA	RMSEA 90 CI	WMRM
1	1821.11	179	.92	.90	0.11	.10-.11	2.20
2	1507.86	160	.93	.91	0.10	.10-.11	2.06
3	1108.44	142	.95	.94	0.09	.09-.10	1.82
4	873.52	125	.96	.95	0.09	.08-.09	1.64
5	665.03	109	.97	.96	0.08	.07-.09	1.45
6- Final Model	401.93	94	.98	.98	0.06	.06-.07	1.13

Note: Items were removed between iterations due to cross-loadings. Iterations continued until a "good fit" (see Browne & Cudeck, 1993; MacCallum et al., 1996; Steiger, 1989) determined by fit indices cutoffs was achieved.

Table 6

Iterative Item Removal Process to Derive Current Study's Final Model Factor Loadings

Item	Study Model Final Loadings	When Item was Dropped from Final Model	Notes
Skills			
1. Making sure to study on a regular basis.	.85 (1)		
2. Putting forth effort.	.86 (2)		
3. Doing all the homework problems.	.74 (2)		
4. Staying up on the readings.	.78 (1)		
5. Looking over the class notes between classes to make sure I understand the material.	.74 (1)		
6. Being organized.	.69 (2)		
7. Taking good notes in class.	.72 (2)		
8. Listening carefully in class.	.78 (2)		
9. Coming to class every day.	.65 (2)		
Emotional			
10. Finding ways to make the course material relevant to my life.	.91		
11. Applying the course material to my life.	.89		
12. Finding ways to make the course material interesting to me.	-	CFA #3 (Cross-loaded)	Skills-Out: 142.43 (.33) Skills-In: 182.73 (.36) Participation: 111.15(.28) Performance: 103.74(.27)
13. Thinking about the course between class meetings.	-	EFA (Cross-loaded)	Skills-Out (.30) Emotional (.33)
14. Really desiring to learn the material.	-	CFA #2 (Cross-loaded)	Skills-Out: 111.40(.32) Skills-In: 158.00(.38) Participation: 235.51(.44) Performance: 125.05(.29)
Participation			
15. Raising my hand in class.	.87		
16. Asking questions when I don't understand the instructor.	.83		
17. Having fun in class.	-	CFA #5 (Cross-loaded)	Skills-Out: 146.64 (.28) Skills-In: 211.86 (.31) Emotion: 196.15(.35) Performance: 138.55(.28)
18. Participating actively in small-group discussions.	.85		
19. Going to the professor's office hours to review assignments or tests or to ask questions.	-	EFA (Cross-loaded)	Skills-Out (.42) Participation (.39)
20. Helping fellow students.	-	CFA #4 (Cross-loaded)	Skills-Out: 128.49 (.28) Skills-In: 124.76(.25)
Performance			
21. Getting a good grade.	.99		
22. Doing well on the tests.	.82		
23. Being confident that I can learn and do well in the class.	-	CFA #1 (Cross-loaded)	Skills-Out: 135.58(.30) Skills-In: 136.12(.38) Emotion: 262.91(.42) Participation: 193.49(.36)

Notes: Study Model Final Loadings are standardized loadings from the final CFA. (1) indicates the out-of-class skills factor. (2) indicates the in-class skills factor. "-" represents items not included in the final CFA model. The standardized factor loadings above .30 are presented for the two items dropped during the EFA. Modification indices (i.e., approximate decrease in model chi-square if parameter were to be estimated in a subsequent model) and the estimated standardized expected parameter change from the CFA model that indicated potential cross-loading items are shown for items removed during the iterative CFA process.

2022, Volume 33, Number 2, 119-134
ISSN 1812-9129

Table 7
Zero-Order Correlations Between Factors in Each Engagement Model

Factor	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	
Handelsman et al. Model																
1. Skills	3.58	0.69	(.87)													
2. Emotional	3.52	0.72	0.58	(.83)												
3. Participation	3.06	0.75	0.49	0.53	(.83)											
4. Performance	3.91	0.68	0.48	0.38	0.37	(.82)										
Taylor et al. Model																
5. Out-of-Class Skills	3.50	0.77	0.85	0.55	0.46	0.40	(.50)									
6. In-Class Skills	3.83	0.74	0.89	0.52	0.44	0.48	0.64	(.77)								
7. Emotional	3.58	0.88	0.43	0.88	0.36	0.29	0.41	0.38	(.85)							
8. Participation	3.10	0.73	0.50	0.54	0.99	0.38	0.48	0.44	0.37	(.78)						
9. Performance	3.93	0.73	0.44	0.29	0.30	0.95	0.35	0.44	0.22	0.31	(.83)					
Study Model																
10. Out-of-Class Skills	3.10	0.88	0.82	0.50	0.43	0.27	0.84	0.52	0.35	0.44	0.24	(.79)				
11. In-Class Skills	3.83	0.71	0.94	0.54	0.44	0.53	0.71	0.96	0.40	0.45	0.48	0.57	(.84)			
12. Emotional	3.58	0.88	0.43	0.88	0.36	0.29	0.41	0.38	1.00	0.37	0.22	0.35	0.40	(.85)		
13. Participation	3.08	0.93	0.40	0.43	0.92	0.32	0.38	0.39	0.28	0.86	0.26	0.32	0.38	0.28	(.85)	
14. Performance	3.93	0.73	0.44	0.29	0.30	0.95	0.35	0.44	0.22	0.31	1.00	0.24	0.48	0.22	0.26	(.83)

Note: Correlations greater than .07 in absolute magnitude were significant at $p<.05$, correlations greater than .10 in absolute magnitude were significant at $p<.01$, and correlations greater than .13 in absolute magnitude were significant at $p<.001$. Internal reliability estimates presented on the diagonal. Correlations in bold represent correlations among factors in the same model.

 2022, Volume 33, Number 2, 119-134
ISSN 1812-9129

Factor to Outcomes Relationships Across Models

To test the discriminant validity of the factors in the three models, zero-order correlations, regression analyses, and RWA were considered (see Tables 8 and 9). Examination of the zero-order correlations revealed a number of overlapping 95% confidence intervals across domains (see Table 8). Thus, regression and RWA were employed to better understand the relationships between factors and outcomes (see Table 9).

Skills Engagement. To investigate the predictive validity of skills engagement, the results of the correlations, regression, and RWA were examined. Results indicated that the single skills factor in the Handelsman et al. (2005) model was a statistically significant, but only marginally important predictor of final grade. Equivalently, in the Handelsman et al. (2005) model, the combined skills engagement scale failed to be a statistically significant predictor of online activity in the regression analysis, but was found to have a statistically significant zero-order correlation[1]. A slightly different pattern of results emerged when looking at the regression results for the two models that separated skills engagement into in- and out-of-class skills engagement. As shown in Table 9, in-class skill engagement was a relatively stronger predictor of final grades in both the Taylor et al. and current study models. These results are consistent with the zero-order correlations showing in-class engagement having a stronger relationship with final grade than out-of-class skill engagement (see Table 8). These results suggest there could be value in separating in- versus out-of-class skill engagement.

Further support of the potential value of separating in- versus out-of-class skill engagement comes from the results of the other regression models. Specifically, the in-class skills factor was found to be a statistically and practically significant predictor for course satisfaction, whereas the out-of-class skills factor was not. However, out-of-class skills engagement was a statistically significant predictor of online activity in both regression models (see Table 9). Interestingly, out-of-class skills engagement was found to be a statistically significant and *negative* predictor of final grade in both regression models (see Table 9). These results were unexpected, given that the out-of-class skills engagement factor in Taylor's model had a modest, but statically significant, positive zero-order correlation with final grades but a non-significant zero-order relationship in the study's model (see Table 8). An examination of the RWA revealed that the out-of-class skills engagement factor was a relatively unimportant predictor of final grades in both models, explaining less than 1% of the total variance in final grade when considered alongside the other factors (RW_{Taylor}=.008; RW_{Study}=.009) and making a relatively unimportant contribution to model R^2 (RI_{Taylor}=.03; RI_{Study}=.04; see Table 9). Despite this surprising finding, the apparent differential prediction of in-class and out-of-class skills across outcomes provides some support for the discriminant validity of these engagement factors.

Emotional Engagement. The results of the correlations, regression, and RWA were examined to determine the predictive validity of emotional engagement when considered alongside the other engagement factors. An examination of the results from the regression models revealed that emotional engagement was the dominant predictor of course satisfaction in the Handelsman et al. (2005) model (see Table 9). A similar pattern of results emerged when looking at the Taylor et al. (2011) and current study models in which emotional engagement was a statistically significant predictor that made the relatively strongest contribution to the models' R^2 (see Table 9). Interestingly, the emotional engagement factor in the Handelsman model was a slightly stronger predictor of course satisfaction than in either the Taylor or study models (see Tables 8 and 9). However, the emotional engagement factor did emerge as the strongest predictor of course satisfaction in all three models and did not meaningfully contribute to the prediction of final grade or online activity (see Table 9). These results provide evidence of the discriminant validity of the emotional engagement factor.

Participation Engagement. When the predictive validity of participation engagement was examined in the regression analyses, results indicated that participation engagement was not a positive, statistically significant predictor of any outcomes (see Table 9). Specifically, despite having statistically significant zero-order correlations with course satisfaction and online activity (see Table 8), participation engagement was not a significant contributor to the prediction of these outcomes in the regression models. In fact, the regression results indicated that participation engagement had a statistically significant and negative relationship with final grade (see Table 9). These results were unexpected, given that the participation factor had a non-significant zero-order correlation with final grade in the Handelsman, Taylor, and study models (see Table 8). Further investigation of the results from the

---

[1] When treating online activity as a count variable and re-running the Handelsman model multiple regression analysis, skills engagement was found to be a statistically significant predictor (p=.042) but was not a statistically significant predictor in the linear multiple regression (p=.068). However, the zero-order correlations and RWA indicate that skills engagement in the Handelsman model was a relatively weak predictor of online activity regardless of its p-value.

Table 8

Zero-Order Correlations Between Factors and Outcome Variables

Factor	Course Satisfaction	Final Grade	Online Activity
Handelsman et al. (2005) Model			
Skills	0.31 (0.25, 0.37)	0.25 (0.18, 0.31)	0.16 (0.09, 0.23)
Emotional	0.41 (0.35, 0.47)	0.11 (0.04, 0.18)	0.12 (0.05, 0.18)
Participation	0.23 (0.17, 0.30)	0.04 (-0.04, 0.11)	0.14 (0.07, 0.21)
Performance	0.25 (0.19, 0.32)	0.47 (0.41, 0.52)	0.02 (-0.05, 0.09)
Taylor et al. (2011) Model			
Out-of-Class Skills	0.28 (0.21, 0.34)	0.13 (0.06, 0.20)	0.18 (0.11, 0.24)
In-Class Skills	0.31 (0.24, 0.37)	0.31 (0.25, 0.38)	0.08 (0.01, 0.15)
Emotional	0.32 (0.26, 0.38)	0.09 (0.02, 0.15)	0.09 (0.02, 0.16)
Participation	0.25 (0.18, 0.31)	0.05 (-0.02, 0.12)	0.14 (0.07, 0.21)
Performance	0.20 (0.14, 0.27)	0.50 (0.47, 0.55)	0.04 (-0.03, 0.11)
Study Model			
Out-of-Class Skills	0.20 (0.13, 0.27)	0.04 (-0.03, 0.11)	0.20 (0.13, 0.27)
In-Class Skills	0.33 (0.26, 0.39)	0.33 (0.27, 0.39)	0.11 (0.04, 0.18)
Emotional	0.32 (0.26, 0.38)	0.09 (0.02, 0.15)	0.09 (0.02, 0.16)
Participation	0.18 (0.11, 0.24)	0.05 (-0.02, 0.12)	0.14 (0.07, 0.21)
Performance	0.20 (0.14, 0.27)	0.50 (0.47, 0.55)	0.04 (-0.03, 0.11)

Note: 95% confidence interval is indicated in parentheses. Result presented above are Pearson correlation coefficients. The bi-variate relationships presented in this table were also examined using Spearman rho and Kendall Tau and by conducting a series of single regression models that treated each model predictor as continuous and course satisfaction as ordinal and web activity as count. The results from these sets of nonparametric analysis did not change the overall pattern of results. Additionally, the conclusions reached using the parametric analyses were the same that would have been reached using the nonparametric equivalent.

RWA indicated that participation engagement was an extremely weak and practically unimportant factor in the three regression models, contributing to less than 1% of the total prediction ($RW_{Handelsman}$=.007; RW_{Taylor}=.007; RW_{Study}=.004) and making a relatively feeble contribution to the regression models' R^2 ($RI_{Handelsman}$=.03; Ri_{Taylor}=.02; RI_{Study}=.02). Together, these results call into question discriminant validity of the participation engagement factor.

Performance Engagement. The results of the regression analyses revealed that performance engagement was a statistically significant, yet modest predictor of course satisfaction in the Handelsman et al. (2005) model, but it did not contribute to course satisfaction in the other two models (see Table 9)[2]. Evidence of discriminant validity for the performance engagement factor was found in the strength of the results of the correlation (see Table 8) and regression analyses (see Table 9) predicting final grades. Performance engagement was found to be the strongest and most dominant predictor of final grade in all three models.

Results Summary

The results of the CFAs indicated that neither the four-factor Handelsman et al. (2005) nor the five-factor Taylor and colleagues (2011) SCEQ models provided good fit to the current data. While the Taylor and colleagues model provided adequate fit, the internal reliability of the out-of-class engagement scale was unacceptably low. Furthermore, Taylor and colleagues reported poor reliability and validity for the out-of-class engagement scale in their 2011 paper, stating, "if, in fact, future research supports the two factors identified, more (and perhaps different) scale items will be necessary to overcome this limitation" (p. 46). An exploratory CFA process in the current study produced a five-factor study model that provided good fit to the data with acceptable internal reliability estimates for all engagement subscales, providing a more internally consistent out-of-class engagement subscale.

[2] When treating online activity as a count variable and re-running the Handelsman model multiple regression analysis, performance engagement was not found to be a statistically significant predictor (p=.109) but was a statistically significant predictor in the linear multiple regression (p=.036). However, as the zero-order correlations and RWA indicate, skills engagement in the Handelsman model was only a modest predictor of course satisfaction, regardless of its p-value.

Table 9
Regression and Relative Importance Analysis Results

Engagement Factor	Course Satisfaction			Final Grade			Online Activity		
Handelsman et al. (2005) Model	β	RW	RI	β	RW	RI	β	RW	RI
Skills	0.08	0.04	0.21	0.11**	0.03	0.13	0.14	0.02	0.37
Emotional	0.33*	0.10	0.57	-0.05	0.01	0.03	0.07	0.01	0.22
Participation	-0.01	0.02	0.10	-0.14**	0.01	0.04	0.10	0.02	0.30
Performance	0.09*	0.02	0.12	0.45***	0.16	0.80	-0.12	0.01	0.10
	R^2 = 0.18***			R^2 = 0.21***			R^2 = 0.05*		
Taylor et al. (2011) Model									
Out-of-Class Skills	0.05	0.03	0.16	-0.11*	0.01	0.03	0.21**	0.03	0.51
In-Class Skills	0.14**	0.04	0.24	0.22***	0.05	0.21	-0.10	0.00	0.08
Emotional	0.20***	0.06	0.38	0.00	0.00	0.01	0.04	0.01	0.14
Participation	0.07	0.02	0.14	-0.13**	0.01	0.03	0.10	0.01	0.25
Performance	0.07	0.01	0.08	0.44***	0.17	0.72	-0.04	0.00	0.02
	R^2 = 0.15***			R^2 = 0.24***			R^2 = 0.05*		
Study Model									
Out-of-Class Skills	-0.02	0.01	0.08	-0.15***	0.01	0.04	0.21**	0.04	0.60
In-Class Skills	0.21***	0.05	0.32	0.23***	0.06	0.23	-0.05	0.00	0.06
Emotional	0.22***	0.07	0.43	-0.02	0.00	0.01	0.04	0.01	0.13
Participation	0.03	0.01	0.07	-0.08*	0.00	0.02	0.11	0.01	0.19
Performance	0.06	0.01	0.09	0.41***	0.17	0.69	-0.03	0.00	0.01
	R^2 = 0.15***			R^2 = 0.24***			R^2 = 0.06*		

Note: β = standardized regression coefficient; RW = raw relative weight for each predictor, which sum to the overall model R^2; RI = relative importance score, which is the percentage of the R^2 value accounted for by the predictor. Rounding error may result in these values not summing to unity. **p* < .05, ** *p* < .01, *** *p* < .001.

Additionally, the results of the zero-order correlations, regressions, and RWA provided support for discriminant validity of the Taylor et al. (2011) and current study's five-factor models. Most notably, the separation of in-class versus out-of-class skills engagement provided predictive value, as in-class engagement added to the prediction of course satisfaction, whereas the single skills engagement factor did not. Additionally, the out-of-class engagement factor predicted out-of-class online activity, whereas the single skills engagement factor also failed to predict this outcome. These findings suggest that, at least empirically, the skills factor might be better conceptualized as two separate factors. The emotional engagement factor was largely replicated and showed evidence of discriminant validity (i.e., emotional engagement predicted course satisfaction across all models, but it did not explain variance in grades or online activity). Similarly, the performance engagement factor replicated the item structure of previous models, predicted outcome variables in accordance with expectations, and showed evidence of discriminant validity. In contrast, although the participation factor did replicate in the current study, item-loading patterns varied considerably across the three models. Additionally, the participation factor had poor predictive validity (i.e., did not substantially predict any student outcomes), which calls into question its utility as a measure of student engagement.

Discussion

Structural and Criterion Validity of the SCEQ

Our first set of analyses concerned the replicability of the factor structures isolated in previous work concerning the SCEQ. The present findings indicate that the four-factor structure proposed by Handelsman et al. (2005) was a poor fit to our data. Additionally, although the five-factor structure of the Taylor et al. (2011) model provided better fit to the current data set than did the Handelsman et al. (2005) model, the fit, modification indices, and poor internal consistency of a subscale suggested that a better factor model could be derived. As such, we utilized exploratory and confirmatory factor analyses to determine the best-fitting model for our data. Results indicated that a five-factor model that loosely replicated the five factors obtained by Taylor et al. (2011) had the best fit. This is an interesting finding, given that no additional researchers have utilized a five-factor structure when assessing student engagement via the SCEQ. It is possible that engagement is best understood as a four-factor construct in some classrooms and as a five-factor in others, but because most researchers have not confirmed the factor structure of the SCEQ in the course of their work, this possibility cannot be thoroughly evaluated at this time. Instead, our review of the literature indicates that researchers utilizing the SCEQ in a variety of classroom types (e.g., politics, marketing, psychology, STEM) typically assess engagement using the original four factors of skill, emotional, participation/interaction, and performance engagement (Lovelace & Brickman, 2005; Mandernach, 2015; Zabel & Heger, 2015). Given that such a four-factor structure has yet to replicate, this is a concern. In other words, perhaps an important takeaway is not that the current model has a factor structure more closely aligned with the rarely-cited Taylor et al. model, but that both the current and the Taylor et al. models call into question the four-factor model that is most frequently cited as meeting minimal standards for reliability and validity.

In addition to understanding the *number* of factors represented by the scale, further work is needed to understand the *composition* of each of the factors. First, two of the five factors that were retained in the final model were specified by only two items per factor. This is because several items with sufficient factor loadings in the Handelsman et al. study failed to load on a factor in the current study model. For example, "going to the professor's office hours" did not display strong loadings on any of the factors. Given the reliability and efficiency of email as a method of student-instructor communication, it may be that office hour attendance is no longer an important indicator of student engagement. Additionally, items such as "doing all the homework problems" and "participating actively in small-group discussions" may not be uniformly applicable across all classroom contexts. Flipped classrooms might not assign homework problems, and small seminar classes may not break into smaller discussion groups. As such, scores on these items might contribute irrelevant variance to the overall factor scores and might explain why some of the factors were only identified by two indicators, which can lead to empirical under-identification in CFA models (Brown, 2015).

An additional concern involves the composition of the fourth and fifth factors implicated in both the current study and the Taylor et al. (2011) study. The performance, participation/interaction, and emotional engagement factors were loosely replicated across all three studies, give or take a few items per factor. However, in both the current study and the Taylor et al. (2011) study, the best-fitting solution split the skills engagement factor from the Handelsman et al. (2005) paper into two separate factors. Taylor et al. (2011) conceptualized these two factors as in-class skills and out-of-class skills. This conceptualization is somewhat puzzling, given that the item "doing all the homework problems" loaded on the *in*-class factor in both studies. In the current study, the psychology courses in which the subjects participated did not involve the completion of homework exercises during class time. Although such information is not available for the Taylor et al. (2011) study, it can be assumed that *home*work for their introductory marketing classes was not completed *in class*. As such, the label of "in-class skills" for that factor should be reconsidered. As shown in Table 6, the items that loaded onto the current study's "in-class skills" factor involve both specific academic skills (e.g., taking good notes, doing homework problems), as well as general academic conscientiousness (e.g., putting forth effort, being organized). In contrast, the "out-of-class skills" factor includes three studying behaviors: studying regularly, staying up on readings, and looking at notes between class periods. With the exception of the performance engagement factor, the "in-class skills" factor is the strongest predictor of both expected and actual final grades. Therefore, more work is needed to clarify the nature of this factor, as it may be one of the most important factors for instructors to focus on in pursuit of increased learning in their classrooms.

A final concern related to the criterion validity of the SCEQ is the performance of the participation factor. Although the structure of this factor was replicated across all three studies, it failed to predict any of the outcome variables examined in the current work. Because engagement is frequently conceptualized as a mediator of the relationship between academic motivation and academic outcomes (Appleton, Christenson, & Furlong, 2008), we expected that participation engagement would have predicted at least one of the outcome variables in the current study. It is possible that participation engagement could share important relationships with criterion variables that we did not measure, but we question the utility of a factor that does not reliably predict relevant outcomes such as grade or course satisfaction.

Content Validity of the SCEQ

The performance engagement factor is one of the strongest correlates of grade-based outcome variables when compared to the other engagement factors, and this strong correlation is often presented as evidence of the predictive validity of the measure. In fact, in the current study, performance engagement was one of the strongest predictors of student outcomes across models. However, recall that the performance engagement factor is specified by the following two items in both the current model and the Taylor et al. (2011) model: 1) getting a good grade, and 2) doing well on the tests. Given that self- or instructor-reported grades are typically utilized as relevant outcome variables, it is not surprising to find that students who rate themselves on the SCEQ as getting good grades and as doing well on tests, do, in fact, get good grades and do well on tests. Researchers typically interpret high scores on the performance engagement subscale as indicative of extrinsic motivation and performance achievement goals (Handelsman et al., 2005), but the wording of these items does not permit that interpretation. That is, participants are asked to rate the extent to which each item is "very characteristic of me," so scores on these items are indicative of appraisals of academic performance, not a preoccupation with grades. Although the original validation study included "being confident that I can learn and do well in class" in the performance engagement subscale, of the three loading items, this item was the least correlated with the factor, and it did not load on the performance factor in the current work. Therefore, interpretations that equate scores on the performance subscale with extrinsic motivation are not advised.

Furthermore, it is interesting that the authors of the original study elected to include the performance engagement items in their measure at all. Traditional conceptualizations of student engagement seek to identify motivation states and academic behaviors that predict eventual academic success (Appleton et al., 2008). Academic success is viewed as the desired *outcome*, not as one of the ways in which engagement exerts its effects. Although it is true that many researchers espouse models of student engagement that consist of several factors, the most commonly cited factors are behavioral engagement (e.g., work ethic, participation, attendance), emotional engagement (e.g., interest in learning, sense of belongingness), and cognitive engagement (e.g., self-regulation, study strategies; see Appleton et al., 2008 for a review). Performance engagement is not included in most conceptualizations of student engagement, possibly due to its overlap with the constructs we are attempting to predict.

Additionally, the benefit of measuring students' performance engagement is unclear. If, for example, an instructor discovers that her students routinely rate their performance engagement as quite low in a course, that instructor might devise an intervention to increase students' ratings on the performance engagement items. It is certainly possible, however, to produce high performance engagement ratings in the absence of similarly high outcome variables (e.g., test grades, course scores). It is not beneficial to have students *believe* that they will perform well in a class, when they will not. In contrast, an intervention focused on other aspects of student engagement could be more worthwhile. Given demonstrated causal relationships between academically conscientious behavior and academic performance (Chamorro-Premuzic & Furnham, 2003; Wagerman & Funder, 2007), an instructor would be better served to attempt to increase students' organizational abilities, note-taking skills, or study strategies than to focus on the students' self-assessments of anticipated course performance. As such, in addition to splitting the single skills factor into two factors (i.e., in-class and out-of-class skills, but perhaps with different names), we recommend that future research utilizing the SCEQ should remove the performance engagement items from the scale.

Implications and Future Directions

Although the SCEQ was created over 15 years ago and although only a handful of studies have examined its psychometric properties, citation searches at the time of publication indicate that over 900 manuscripts have referenced the original measure. Some of these manuscripts have utilized the SCEQ as inspiration for the creation of new measures of engagement, but many of them have utilized the SCEQ as originally conceptualized to measure student engagement in their studies. Given that the original factor structure has not replicated across studies, and given the confusion regarding the nature and importance of the factors that the measure represents, this is of concern. For example, consider the 2011 study by Miller, Rycek, and Fritson that examined the effects of a variety of high-impact learning experiences (i.e., internships, undergraduate research, service learning, and learning communities) on student engagement. The four subscales of the original SCEQ were utilized to measure engagement, and results indicated that learning experience type predicted differences in skills and emotional engagement but not participation or performance engagement. Given concerns about the structure of the skills engagement subscale and the relevance of performance engagement subscale as measured by the SCEQ, the statistical conclusion validity of this work should be considered. That is, would the results replicate with a different measure of engagement? Is it accurate to conclude, as these researchers did, that undergraduate internships and research are more impactful than other learning activities when it comes to engagement, or is that an artifact of the engagement measure that was utilized?

Similar concerns exist regarding the statistical conclusion validity of papers published in other higher education journals, as well. For example, in a 2014 *Teaching of Psychology* article, Troisi demonstrated that student engagement serves as a mediator between participation on a student management team (SMT) and improved academic performance in a Psychology class. However, analyses indicated that this effect was only marginally statistically significant. Is this because SMT participation only marginally affects performance, or because the scale that was utilized to measure engagement (the SCEQ) insufficiently covers the criterion space? Similarly, Richmond, Berglund, Epelbaum, and Klein (2015) examined predictors of student ratings of instructors, and they concluded that student perceptions of professor-student rapport were the strongest predictor of student ratings. Although SCEQ-measured engagement did share a statistically significant relationship with the outcome variable, the improvement in variance explained over rapport was only 3%. Given what we know about the relationship between the classroom experience and student engagement, at least when engagement is defined as "the degree to which students devote cognitive resources to class material" (Richmond et al. 2015, p. 120), this finding is surprising. Is it really true that engagement does not matter, at least when compared to rapport? Or does this particular measure of student engagement deserve more pointed examination before such conclusions are made?

It is beyond the scope of this article to fully detail the threats to statistical conclusion validity in every study using the SCEQ. However, given the concerns first stated by Taylor and colleagues (2011) and amplified by our current work, we believe that the original four-factor conceptualization of the SCEQ may not be suitable for the assessment of student engagement in higher education classrooms. This is not to say that the SCEQ is a poor measure of student engagement but that more research is needed is to understand what it measures and what it predicts. Despite such concerns, measurement of course-specific micro-engagement in college courses remains a worthy task. Although instructors have limited control over many predictors of student success (e.g., institutional mission, availability of support services, student background; see Kuh et al., 2007), they do have control over what they do in their own classrooms. Given that the quality of instructional practices is strongly correlated with classroom engagement (Umbach & Wawrzynski, 2005), and that classroom engagement is predictive of retention and success, instructors should make efforts to teach using evidence-based practices. Assessment of the effects of these educational innovations should also be completed to determine the conditions under which such practices yield maximum success. In the absence of psychometrically valid measures of engagement, however, the assessment of meaningful student change can be obscured.

References

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with *applications, and programming*. New York, NY: Routledge.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research.* New York, NY: Guilford Press.

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.

Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge.

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic college student course engagement. *The Journal of Educational Research*, *98*(3), 184-192.

Crockett, S. A. (2012). A five-step guide to conducting SEM analysis in counseling determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130-149.

DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2018). Examination of the weighted root mean square residual: Evidence for trustworthiness? *Structural Equation Modeling, 25*(3), 453–466. https://doi.org/10.1080/10705511.2017.1390394

Goebl, A. P., Jones, J. A., and Beatty, A. S. (2016). *iopsych: Methods for industrial/organizational psychology.* (R package version 0.90) [Computer software] https://cran.r-project.org/web/packages/iopsych/

Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *The Journal of Educational Research*, *98*(3), 184-192.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance imperative to look within. Bloomington, IN: Indiana University Center for Postsecondary Research. indices in organizational research. *Organizational Research Methods*, *7*(3), 238-257. https://dx.doi.org/10.1177/1094428104266510

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research, 35*(1), 1-19.

Johnson, J. W., & Lebreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, *7*(3), 238-257. https://dx.doi.org/10.1177/1094428104266510

Kahn, P. E. (2014). Theorising student engagement in higher education. *British Educational Research Journal*, *40*(6), 1005-1018. doi:10.1002/berj.3121

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Kuh, G. D. (2001). Assessing what really matters to student learning: Inside the National Survey of Student Engagement. *Change, 3*3(3), 10-17. http://www.jstor.org/stable/40165768

Kuh, G. D. (2009). The national survey of student engagement: Conceptual and empirical foundations. *New Directions for Institutional Research*, 141, 5-20.

Kuh, G. D., Kinzie, J., Buckley, J., Bridges, B., & Hayek, J. C. (2007). Piecing together the student success puzzle: Research, propositions, and recommendations. ASHE Higher Education Report, 32(5). San Francisco: Jossey-Bass.

Lebreton, J. M., Ployhart, R. E., & Ladd, R. T. (2004). A Monte Carlo comparison of relative importance methodologies. *Organizational Research Methods*, *7*(3), 258-282. https://dx.doi.org/10.1177/1094428104266017

Lovelace, M., & Brickman, P. (2013). Best practices for measuring students' attitudes toward learning science. *Cell Biology Education*, *12*(4), 606-617.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and management teams on student classroom outcomes. *Teaching of Psychology*, *41*(2), 99-103.

Mandernach, B. J. (2015). Assessment of student engagement in higher education: A synthesis of literature and assessment tools. *International Journal of Learning, Teaching and Educational Research*, *12*(2), 1-14. www.aajhss.org/index.php/ijlter/article/viewFile/367/167

Miller, R. L., Rycek, R. F., & Fritson, K. (2011). The effects of high impact learning experiences on student engagement. *Procedia-Social and Behavioral Sciences*, *15*, 53-59.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus: Statistical analysis with latent variables: User's guide* (Version 8). Los Angeles, CA: Authors.

National Survey of Student Engagement. (2008). *Promoting engagement for all students: The imperative to look within.* Bloomington, IN: Indiana University Center for Postsecondary Research.

Pascarella, E. T. (2001). Identifying excellence in undergraduate education: Are we even close? *Change,* *33*(3), 19-23. http://dx.doi.org/10.1080/00091380109601796

Richmond, A. S., Berglund, M. B., Epelbaum, V. B., & Klein, E. M. (2015). a+(b1) school: Critical conceptual and methodological issues of the construct. *Science*, *42*(5), 661-684. http://dx.doi.org/10.1007/s11251-014-9310-0

Steiger, J. H. (1989). *Causal modeling: A supplementary module for SYSTAT and SYGRAPH*. Evanston, IL: SYSTAT.

Taylor, S. A., Goodwin, S. A., Melton, H., & Hunter, G. (2011). An analysis of the the student success puzzle: Research, propositions, and recommendations. *ASHE Higher Education Report, 3*2(5).

Troisi, J. D. (2014). Making the grade and staying engaged: The influence of student management teams on student classroom outcomes. *Teaching of Psychology*, *41*(2), 99-103.

Umbach, P. D., & Wawrzynski, M. R. (2005). Faculty do matter: The role of college *User's Guide* (Version 8). Los Angeles, CA: Authors.

Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality*, *41*(1), 221-229.

Zabel, K. L., & Heger, A. (2015). Student engagement toward coursework: Measures, considerations, and future directions. In R. S. Jhangiani, J. D. Troisi, B. Fleck, A. M. Legg, and H. D. Hussey, H. D. (Eds.), *A compendium of scales for use in the scholarship of teaching and learning* (87-101). Society of the Teaching of Psychology.

Zumbrunn, S., McKim, C., Buhs, E., & Hawley, L. R. (2014). Support, belonging, motivation, and engagement in the college classroom: a mixed method study. *Instructional Science*, *42*(5), 661-684. http://dx.doi.org/10.1007/s11251-014-9310-0

DR. JORDAN ELLIS is a Licensed Clinical Psychologist with the United States Air Force. Jordan received his doctorate in Clinical Health Psychology from East Carolina University and a Master of Arts in Clinical Psychology from Appalachian State University. Dr. Ellis's work focuses on examining the measurement properties of a variety of instruments and provides evidence-based behavioral medicine to improve the individual and group health of the men and women of the United States Air Force.

DR. LINDSAY C. MASLAND is the Associate Director for Faculty Professional Development and Associate Professor of Psychology at Appalachian State University. She serves as a Consulting Editor for the journals *Teaching of Psychology* and *Scholarship of Teaching and Learning in Psychology*. She is the Director of the Annual Conference on Teaching, which is offered by the Society for the Teaching of Psychology. Dr. Masland teaches courses in education, psychology, and statistics to future educators and has won university and national teaching awards, including the Appalachian State Excellence in Teaching Award and the Jane S. Halonen Teaching Excellence Award for Early Career Psychologists.

DR. SHAWN BERGMAN is the Kulynych/Cline Distinguished Professor of Psychology and the founder and co-director of the HR Science Research Team at Appalachian State University. Shawn is also the co-founder and Research Director at the Vela Institute, a non-profit organization dedicated using evaluation and data to increase educational outcomes. Dr. Bergman is a member of the Academy of Outstanding Teachers and Academy of Outstanding Graduate Mentors, has received multiple Psi Chi Excellence in Teaching Awards, earned one of Appalachian's Board of Governors Excellence in Teaching Awards, and has been inducted into Appalachian State Faculty Hall of Fame.