The Student Engagement in Schools Questionnaire (SESQ) and the Teacher Engagement Report Form-New (TERF-N): Examining the Preliminary Evidence

Shelley R. Hart, Kaitlyn Stewart & Shane R. Jimerson,
University of California Santa Barbara

Student engagement in school is an important construct that has been associated with student success. For the current study, researchers examined the psychometrics of the Student Engagement in Schools Questionnaire (SESQ) and the Teacher Engagement Report Form (TERF-N) of student engagement. The results revealed that both the SESQ and the TERF-N have good internal consistency. The exploratory factor analysis results for the SESQ demonstrated alignment with the theoretically driven development (five factors: Affective Engagement-Liking for Learning, Affective Engagement-Liking for School, Behavioral Engagement-Effort & Persistence, Behavioral Engagement-Extracurricular, and Cognitive Engagement) whereas the results for the TERF-N were more complicated. The items did not load as conceptualized in a 3-factor model, but instead loaded on one, General Engagement factor. Finally, while it may be that teachers viewed a student's level of engagement as a global construct, the correlations between the measures indicated that they might be used to provide helpful, convergent information obtained from a variety of sources regarding a student's levels of engagement. Future directions and implications for school psychologists are discussed.

Engagement is a growth-producing activity through which an individual allocates attention in active response to the environment (Csikszentmihalyi, 1990). Engagement related to school activity (or student engagement) has become an important concept related to multiple educational outcomes (e.g., achievement, attendance, behavior, dropout/completion; e.g., Finn, 1989; Jimerson, Campos, & Greif, 2003; Jimerson, Renshaw, Stewart, Hart, & O'Malley, 2009). Student engagement has been identified as a primary variable in understanding dropout, particularly as a gradual process operating in a student's life and influencing that final decision to withdraw (Jimerson et al., 2009). Numerous studies have linked student engagement with improved academic performance and it has repeatedly demonstrated to be a robust predictor of achievement and behavior in the schools (Appleton, Christenson, & Furlong, 2008; Shernoff & Schmidt, 2008). It has also been correlated with both health compromising (e.g., substance abuse, depression, suicidality, aggression, early sexual activity) and health promoting (e.g., exercise, nutrition, safe sex activities) behaviors (Carter, McGee, Taylor, & Williams, 2007).

As a result of its demonstrated relationships with a variety of outcomes, it is postulated that an understanding of student engagement might help educators prevent deleterious outcomes and promote positive ones for at-risk students. Student engagement is a construct that resonates with most consumers of education, including students and parents (Appleton, Christenson, & Furlong, 2008) and presents an attractive focus for researchers and educators, in that compared to other predictors of academic success that are static (e.g., socioeconomic status [SES], ethnicity), it is believed to be a malleable characteristic and therefore a more appropriate focus for interventions (e.g., Christenson, Sinclair, Lehr, & Godber, 2001). In addition, both the individual *and* the environment shape a student's level of engagement, thus, there are many factors in the school environment (e.g. interpersonal relationships, recognition) that may enhance it (Fredricks, Blumenfeld, & Paris, 2004). Indeed, researchers have shown that effective interventions to promote student engagement and motivation also enhance the probability of high school completion (Reschly, Appleton, & Christenson, 2007). For these reasons it can be viewed as an asset associated with positive student outcomes (Furlong et al., 2003).

Correspondence may be sent to Shelley R. Hart, UCSB, GGSE, CCSP, Santa Barbara, CA 93106-9490 or e-mail: shart@education.ucsb.edu or jimerson@education.ucsb.edu

DEFINING AND MEASURING STUDENT ENGAGEMENT IN SCHOOL

Despite its apparent utility, student engagement remains a nebulous construct with researchers using ambiguous or inconsistent definitions resulting in equally nebulous measures. Several recent reviews have focused on defining this meta-construct and setting the stage for future scholarship (see Appleton et al., 2008; Fredricks et al., 2004; Jimerson et al., 2003). These scholars (i.e., Appleton et al., 2008; Jimerson et al., 2003) suggest that student engagement in school is multi-dimensional and appears to overlap with several similar constructs (e.g., school connectedness, school bonding). The proposed definition includes both *indicators* (i.e., affective, behavioral, and cognitive) and *facilitators* (i.e., both personal and contextual factors that influence engagement) of engagement (Appleton et al., 2008). Each component is vital to a complete understanding of student engagement. Appleton and colleagues (2008) have suggested that indicators are proposed to "…convey a student's degree or level of connection with learning"; while facilitators are "…factors [that] influence the strength of the connection" (p. 382).

The current study is focused primarily on the indicators of student engagement, and therefore, each indicator will be discussed further. *Affective* engagement refers to a student's feelings toward his school, learning, teachers, and peers (e.g., the student has positive feelings toward his teachers; Jimerson et al., 2003). The terms psychological and emotional engagement have also been used in the current literature to describe this construct (Appleton, Christenson, Kim, & Reschly, 2006; Reschly et al., 2007). *Behavioral* engagement includes observable student actions or participation while at school and is investigated through a student's positive conduct, effort, and participation (e.g., participation in extracurricular activities, attendance, and work habits; (Fredricks et al., 2004). Historically, research has been focused primarily on this aspect of student engagement. *Cognitive* engagement includes a student's perceptions and beliefs associated to school and learning (e.g., I will do well in this class if I try). It refers to the cognitive processing a student brings to academic tasks as well as the amount and type of strategies a student utilizes (Walker, Greene, & Mansell, 2006).

Some researchers propose the notion of academic engagement as a fourth indicator of student engagement (e.g., Reschly & Christenson, 2006). Academic engagement has been defined as time spent in academic learning. We contend that academic engagement can be better explained as an aspect of one of the three more commonly identified indicators (e.g., time-on-task is more accurately described as a behavioral indicator) or as an *outcome* of student engagement (e.g., Grade Point Average [GPA]).

Whereas there seems to be a general consensus that three indicators of engagement exist, there still remain differences in precisely *how* these indicators are defined and measured. For example, Jimerson et al. (2003) locate motivation within the affective engagement indicator, while Fredricks et al. (2004) define this construct as a cognitive indicator of engagement, and Patrick, Ryan, and Kaplan (2007) describe it as a cognitive precursor to engagement. Therefore, an obvious challenge remains for researchers of student engagement in parsing out the characteristics of each component.

PURPOSE OF THE PRESENT STUDY

It appears to be the multidimensional nature of student engagement that has created confusion in the field. For example, researchers may focus on only one component (unidimensional approach) or mix elements of several components (mixed approach), nonetheless operationalizing it as "student engagement." We argue that part of the reason for this confusion is the lack of a comprehensive measure to examine the meta-construct of student engagement. Thus, a psychometrically sound, universal measure of student engagement would advance scholarship in this area.

In order to fill this need for a comprehensive instrument, researchers from more than 19 countries collaborated to develop such a measure and to study student engagement internationally (Lam & Jimerson, 2008). *The Student Engagement in Schools Questionnaire* (SESQ) was the product of this collaboration. As a self-report measure, the SESQ obtains information from the student's perspective related to both the indicators and facilitators of engagement. Optimally, in an assessment of any construct, information would be collected from a variety of sources, in a variety of contexts, through a variety of methods, and over a period of time. To further this aim and provide a complement to the self-report SESQ and to evaluate the three indicators of engagement from the teacher perspective, the *Teacher Engagement*

Report Form - New (TERF-N)[1] was also examined.

The main purpose of this study is to establish the psychometric properties of the SESQ and TERF-N. Specifically, reliability and validity evidence will be evaluated through internal consistency estimates, exploratory factor analysis and correlations between measures.

METHOD

Participants

The present study utilized a sample drawn from one junior high and one high school located in the central coast area of California. For analyses of the SESQ, a sample of N = 428 seventh- through ninth-grade students was obtained. There were very few eighth-grade students included in the sample (5%), while ninth-graders composed the majority of the sample (59%), followed by seventh-graders (36%). Fifty-four percent of the sample was male, 42% Hispanic, 25% African American, 6% White (non-Hispanic), and 2% other. Due to the return rate of the TERF-N by the teachers (N = 4), for these analyses, a subsample (N = 129 seventh-grade students; 48% male) of the larger sample was utilized. The classrooms are considered to be representative of the schools, as well as the community because the demographics of the classrooms from which the same was drawn are similar to the both the schools and the communities where they are situated.

Procedure

Participation was requested through direct contact with school administrators. Two schools agreed to participate. Next the teachers at the two schools were contacted to determine their interest in and availability for the project. Ten of the teachers contacted agreed to participate. Finally, researchers obtained consent from the students and the survey was completed during one class period in spring 2008 and spring 2009. The researchers, school psychology graduate students from the University of California, Santa Barbara, presented the surveys to students, provided directions, fielded questions, and collected completed surveys. Students were provided with an alternative to opt-out of the data collection procedure. Teachers completed their ratings of students, while students were completing the self-report.

Measures

Student Engagement in Schools Questionnaire (SESQ). Scholars from more than 19 countries collaborated in the development of the SESQ (see Lam & Jimerson, 2008, for a description of this process and the international scholars who participated). The SESQ is a 109-item paper-and-pencil, Likert-type, self-report questionnaire focused on the comprehensive assessment of the construct of student engagement. After agreeing on the definition of student engagement, scholars developed a questionnaire to encompass this construct. Items were drawn from existing research, increasing the content validity of the measure (see Lam & Jimerson, 2008, for a detailed description of this process and the resulting measure). The SESQ contains four composites (i.e., Student Engagement in the Schools, Motivational Beliefs, Social-Relatedness Contexts, Student Outcomes) within which are 13 domains and 15 sub-domains. Students respond according to a Likert-type scale of 1-5 (e.g., *1 = never*, *5 = always*) and typically require approximately 35-minutes to complete. Due to the sampling restrictions associated with a survey of 109-items (i.e., a very large sample would be required for evaluation of the entire survey), for purposes of this study the items representing only the indicators of engagement (i.e., Affective, Behavioral, and Cognitive) are examined in the exploratory factor analysis; therefore, only the composite of Student Engagement in Schools (ENG; 33 items) was explored. However, reliability estimates are examined for the entire survey. As proposed by researchers, the SESQ-ENG is composed of five factors (Affective Engagement: Liking for Learning; Affective Engagement: Liking for School; Behavioral Engagement: Effort and Persistence; Behavioral Engagement: Extracurricular Activities; and Cognitive Engagement; Lam & Jimerson, 2008).

[1]An original teacher report (*Teacher Engagement Report Form-Original* [TERF-O]; Lam & Jimerson, 2008) was developed as part of the international collaboration. However, we anticipated that there were indicators of engagement that were not measured by the original 6 questions requested; therefore we developed the TERF-N to expressly access teachers' impressions of all three indicators (i.e., affective, behavioral, and cognitive engagement) for each student.

Teacher Engagement Report Form (TERF-N). The TERF-N is a 10-item, paper-and-pencil chart, where the teacher fills in 10 boxes, one for each item, per student. Each item is completed using a Likert-type scale of 1-5 (e.g., *1 = strongly disagree*, *5 = strongly agree*). The TERF-N requires approximately 45 minutes to complete for 30 students. The questionnaire items address aspects of affective, behavioral, and cognitive engagement.

Data Analyses

Analyses for this study were selected for scale development and validation. Establishing the reliability of a measure is a crucial first step in scale development. Therefore, the analyses for each measure began with internal consistency estimates to examine reliability. Next, as these measures have not been analyzed prior, exploratory factor analyses (EFA) were conducted on each to examine the latent factor structure of each measure. Finally, correlations were conducted to examine the relationship of the scales and to examine external validity. Internal consistency estimates and correlations were conducted utilizing the SPSS package (version 16.0; SPSS, 2007), and the EFA's were conducted using Mplus software (version 5.21; Muthén & Muthén, 2009).

RESULTS

Preliminary Analysis

Preliminary assumption testing was conducted to check for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multi-collinearity, with no serious violations noted. Skewness and kurtosis for each item within both the SESQ-ENG and the TERF-N were evaluated to be within acceptable (±2) limits.

Internal Consistency

Cronbach's coefficient alpha (α) is the most common coefficient of reliability (Cronbach & Shavelson, 2004). Alpha is the ratio of the variance of the true scores to the observed scores; therefore, the higher the reliability, the closer the true scores will be to the observed scores (Gliner & Morgan, 2000). This measure of internal consistency is used to demonstrate how well a set of items measures a unidimensional latent construct (e.g., affective engagement). For this reason, separate coefficient analyses were run for each domain of the SESQ and TERF-N. The literature demonstrates a range of acceptable alpha levels from $.60 \leq \alpha \leq .90$ (Gliner & Morgan, 2000). Acceptable coefficients for this study were set at $\alpha \geq .70$.

Table 1 lists the internal consistency estimates for both the domains of the SESQ and the overall TERF-N. In general, both measures demonstrate good reliability. Estimates for the SESQ range from $.65 \leq \alpha \leq .95$. Only one domain (Attributions) did not demonstrate the acceptable level of $\alpha \geq .70$. The data for the TERF-N indicates good internal consistency ($\alpha = .83$) between the 10 items.

Table 1 *Cronbach's Alpha (α) Estimates for the Domains of the Student Engagement in Schools Questionnaire (SESQ) and the Overall Teacher Engagement Report Form (TERF-N)*

Domain	α	Domain	α
SESQ: Affective Engagement	.88[a]	*SESQ*: Teacher Support	.83[a]
SESQ: Behavioral Engagement	.85[a]	*SESQ*: Peer Support	.84[a]
SESQ: Cognitive Engagement	.93[a]	*SESQ*: Peer Aggression	.84[a]
SESQ: Goal Orientations	.85[a]	*SESQ*: Peer Victimization	.78[a]
SESQ: Attributions	.65	*SESQ*: Parental Support	.82[a]
SESQ: Learning Self-Efficacy	.84[a]		
SESQ: Motivating Instructional Contexts	.95[a]	*TERF-N*	.83[a]

Note. [a] Domain meets or exceeds the acceptable $\alpha \geq .70$.

Factor Structure

Establishing the internal (factor) structure of a measure is an important step in psychometrics. Factor analysis is a common way to do this. This process allows researchers to a) clarify the number of factors within a set of items, b) reveal the associations among the factors, and c) link items to factors (Furr & Bacharach, 2008). It accomplishes these tasks through a number of statistical techniques that aim to simplify complex data sets (Kahn, 2006; Kline, 2002). As described by Brown (2006), factors (or latent, unobservable, variables/constructs) account for the variation and covariation among a set of observed indicators (or items). That is, the indicators are intercorrelated because they are influenced by a common construct/factor/latent variable and if that latent variable were partialed out, the correlations between the indicators would be zero. As such, factor analysis provides for a more parsimonious understanding of a construct, as there are fewer factors than indicators. Exploratory factor analysis (EFA) is the first step in establishing the factors underlying a measure and in conducting data reduction. Therefore, EFA is the focus of the current study.

EFA requires several decisions as a result of a variety of rotation and estimation options. For any analysis with two or more factors, there exist an infinite number of equally good-fitting solutions, with each solution represented by a different factor-loading matrix (Brown, 2006). This means that any multiple-factor EFA model does not have a unique solution and researchers must make decisions about the solution interpreted among those infinite equally fitting solutions (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Pursuit of the property of simple structure is the criterion most typically used for the selection of solutions (Thurstone, 1947). Simple structure is defined as a solution where a) each factor is defined by a subset of indicators that load highly (primary loading) on the factor[2], and b) each indicator has a high loading on one factor and a trivial (i.e., close to zero) loading on any remaining factors (i.e., secondary loadings). In order to obtain simple structure and increase interpretability, rotation of the factors in multidimensional space is recommended.

Two basic types of rotation exist: orthogonal and oblique (Brown, 2006). In an orthogonal rotation, the factors are constrained to be uncorrelated (i.e., the axes of the factors remain at 90° angles), whereas with an oblique rotation, the factors are allowed to intercorrelate (i.e., the axes are allowed to be more or less than 90° angles). An oblique rotation is believed to provide a more realistic representation of the interrelated nature of the factors (or underlying constructs; Fabrigar et al., 1999), provides solutions that are more likely to generalize to CFA, and if the factors are truly not correlated, an oblique rotation will provide a solution virtually the same as an orthogonal rotation (Brown, 2006). Therefore, oblique rotation was selected in this study. Several oblique rotation methods have demonstrated satisfactory solutions, without one approach clearly dominating research (Fabrigar et al., 1999; Kline, 2002); therefore, Geomin rotation (Yates, 1987) was selected, as it is the default rotation used in the Mplus software for EFA.

Additionally, an estimation procedure (i.e., finding parameter values of a model that best fit the data) needs to be indicated. Maximum Likelihood (ML) estimation was chosen due to its ability to provide goodness-of-fit statistics. Goodness-of-fit statistics provide an overall summary of the model's ability to reproduce the observed covariance matrix. There are three different types of fit statistics: absolute, parsimony corrected, and comparative/incremental. Within these types exist a variety of indices; at least one index should be considered from each type when evaluating the overall fit of the model (Brown, 2006). Absolute fit statistics assess the magnitude of discrepancy between the sample and fitted or predicted variance-covariance matrices (Hu & Bentler, 1999) and include the chi-square (χ^2) statistic and the standardized root mean square residual (SRMR). Parsimony correction fit statistics incorporate a penalty for poor model parsimony (i.e., more than needed freely estimated model parameters). The most commonly used statistic from this category is the root mean square error of approximation (RMSEA). Finally, comparative fit statistics assess the fit of the model specified with a null model -- typically one where the covariances among all indicators are fixed to zero (variances are not constrained; Brown, 2006), with the most popular being the comparative fit index (CFI). The best fitting model would have a non-significant

[2]Factor loadings are indicated by lambda (λ) and are defined as completely standardized estimates of the regression slopes predicting the indicators by the latent variable, or factor (Brown, 2006).

chi-square (χ^2) statistic (although this is very sensitive to sample size; Brown, 2006), a Comparative Fit Index of > .95 (CFI), a Root Mean Square Error of Approximation of < .06 (RMSEA), a Standardized Root Mean-Squared Residual of < .08 (SRMR; Hu & Bentler, 1999).

Recommendations for the number of factors to extract, an optimal sample size, and factor loading cut-offs vary in the literature. One common method for determining the number of factors to retain is Kaiser's criterion (retain factors with eigenvalues over 1.00); however, after reviewing numerous studies that indicated this procedure tended to overestimate the number of factors, Fabrigar et al. (1999) stated knowledge of "…no study of this rule that shows it to work well" (p. 278). Additionally, it has been suggested that this criterion is only appropriate with principal components analysis (PCA; Kahn, 2006). Another method utilizing eigenvalues is Cattell's "scree test" (Cattell, 1966). The scree plot is examined to identify the last substantial drop in magnitude of the eigenvalues. A non-horizontal line can be drawn through the point where the eigenvalues "flatten out," and the number of factors before that flattening out are retained. This criterion has been criticized due to its subjectivity (Fabrigar et al., 1999). Both methods are provided in this study as they are still frequently referenced in the literature.

Sample size is another issue debated in the literature. The ratio of item-to-subject for the SESQ was 428:33 (or 13:1), while recommendations typically state at least a 5:1 ratio (Kahn, 2006). The sample size for the TERF-N was smaller (N=87, still resulting in a 8:1 ratio). A recent study exploring best practices in EFA (Costello & Osbourne, 2005) suggests that 60-70% of the time, the SESQ ratio would likely result in the correct factor structure, while 40-60% of the time, the TERF-N ratio would result in the correct factor structure. However, it has also been discussed that the importance in sample size is related to the communalities (not simply the size of the population). An item's communality is the variance explained in that item by the common factors. For communalities in the moderate range (such as demonstrated by the SESQ and TERF-N), a sample size of at least 200 is recommended (Fabrigar et al., 1999).

While EFA is a data-driven approach (i.e., the number of factors is not specified a priori), it is important to note that in order for an EFA to be useful it must make sense. Theory may play a large part in the determination of the number of factors extracted. Additionally, a balance between parsimony and plausibility must be struck (i.e., the fewest number of factors used to account for the correlations among the variables; Fabrigar et al., 1999). After the number of factors was determined, factor-loading cut-offs of $\lambda \geq .32$ were used in the current study per recommendations in the literature (Tabachnick & Fidell, 2007).

Student Engagement in the Schools Questionnaire-Engagement Composite (SESQ-ENG)

Goodness-of-fit statistics for the Engagement items of the SESQ (SESQ-ENG) EFA are located in Table 2. Eigenvalues and the scree plot can be found in Figure 1. According to the Kaiser criterion, a six-factor model is suggested; while the SESQ scree plot indicated a two-factor model and the chi-square statistic remains significant. Again, these methods for determining the number of factors to extract are problematic. The CFI indicated a six-factor model, while the RMSEA and the SRMR might indicate either the five- or six-factor model. The improvements in fit for the CFI and the RMSEA are incremental from the five- to the six-factor model; therefore, in the interest of parsimony and theory (a five-factor model was hypothesized) the five-factor model was evaluated.

Table 3 displays the factor loadings and communalities of the SESQ-ENG items. The factors were all well-determined (i.e., at least three items loaded on each factor) and relatively simple structure (i.e., each item loaded significantly on only one factor) was established. Results indicated that items loaded significantly on hypothesized and theoretically determined factors. The cumulative percentage of variance explained by the five factors was 61.45% (a breakdown by factor can be found in Table 4). The correlations between the factors indicated medium to large correlations (i.e., between .37 - .66), but not too large to indicate multicollinearity; therefore, it appears that each factor represents a unique construct within the concept of student engagement (Table 5 displays the factor correlations). Two items cross-loaded (i.e., item loaded on more than one factor at a significant level). These two items loaded on both factors one (Affective - Liking for Learning) and factor three (Behavioral - Effort and Persistence). However, these items loaded more strongly on the theoretically-driven factor. The majority of the correlations were significant at $p < .01$. There is a wide range of values from trivial-to-large ($r = .01$ - $.74$). The average inter-item correlation is in the moderate range ($r = .34$).

Table 2 Goodness-of-Fit Statistics for the SESQ-ENG EFA

Model	χ^2	DF	CFI	RMSEA	SRMR
1	3400.96***	495	.61	.12 (.11-.12)	.10
2	2124.26***	463	.78	.09 (.09-.10)	.06
3	1488.41***	432	.86	.08 (.07-.08)	.05
4	1160.08***	402	.90	.07 (.06-.07)	.04
5	900.98***	373	.93	.06 (.05-.06)[a]	.03[a]
6	739.17***	345	.95[a]	.05 (.05-.06)[a]	.03[a]

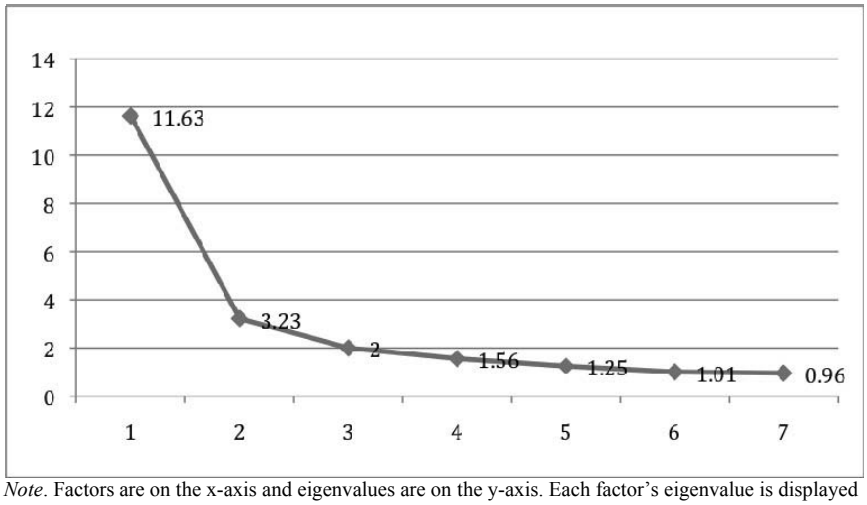
Note. DF=Degrees of Freedom. CFI = Comparative Fit Index. RMSEA = Root Mean Square of Error of Approximation. SRMR = Standard Root Mean-Squared Residual. AIC = Aikake Information Criterion. BIC = Bayesian Information Criterion.
*** $p < .001$.
[a] Model meets standards for that index established in the literature.

Table 3 *Factor Loadings for the SESQ-ENG Items*

Factor	Item	λ	Communality
	I am very interested in learning. (a1)***	.44	.57
		(.38)	
1	I think what we are learning in school is interesting. (a3)	.85	.74
Affective:	I like what I am learning in school. (a5)	.79	.70
Liking for	I enjoy learning new things in class. (a7)	.52	.61
Learning	I think learning is boring. (a9R)***	.39	.41
		(.33)	
2	I like my school. (a11)	.92	.79
Affective:	I am proud to be at this school. (a13)	.87	.78
Liking for	Most mornings, I look forward to going to school. (a15)	.34	.36
School	I am happy to be at this school. (a17)	.77	.77
	I try hard to do well in school. (a2)	.45	.50
	In class, I work as hard as I can. (a4)	.45	.53
3	When I'm in class, I participate in class activities. (a6)	.32	.40
Behav.:	I pay attention in class. (a8)	.63	.55
Effort &	When I'm in class, I just act like I'm working. (a10R)	.66	.45
Persist.	In school, I do just enough to get by. (a12R)	.62	.41
	When I'm in class, my mind wanders. (a14R)	.42	.19
	If I have trouble understanding a problem, I go over it again until I understand it. (a16)	.50	.36
	When I run into a difficult homework problem, I keep working at it until I think I've solved it. (a18)	.52	.52
4	I am an active participant of school activities such as sport day and school picnic. (a20)	.72	.59
Behav.:	I volunteer to help with school activities such as sport day and parent day. (a22)	.82	.71
Extracurr	I take an active role in extracurricular activities in my school. (a24)	.59	.59
	When I study, I try to understand the material better by relating it to things I already know. (b1)	.59	.47
	When I study, I figure out how the information might be useful in the real world. (b2)	.58	.38
	When learning new information, I try to put the ideas in my own words. (b3)	.62	.52
	When I study, I try to connect what I am learning with my own experiences. (b4)	.76	.57
5	I make up my own examples to help me understand the important concepts I learn from school. (b5)	.62	.51
Cognitive	When learning things for school, I try to see how they fit together with other things I already know. (b6)	.70	.53
	When learning things for school, I often try to associate them with what I learnt in other classes about the same or similar things. (b7)	.78	.64
	I try to see the similarities and differences between things I am learning for school and things I know already. (b8)	.79	.66
	I try to understand how the things I learn in school fit together with each other. (b9)	.73	.62
	I try to match what I already know with things I am trying to learn for school. (b10)	.81	.67
	I try to think through topics and decide what I'm supposed to learn from them, rather than studying topics by just reading them over. (b11)	.64	.50
	When studying, I try to combine different pieces of information from course material in new ways. (b12)	.70	.65

Note.*** Items that cross-loaded. Both items loaded on Factors 1 & 3. Values on the conceptualized factors were higher (values in parentheses are the factor loadings on the second factor). The average communality for the SESQ = .55.

Figure 1. *Scree Plot of the Eigenvalues for the SESQ*



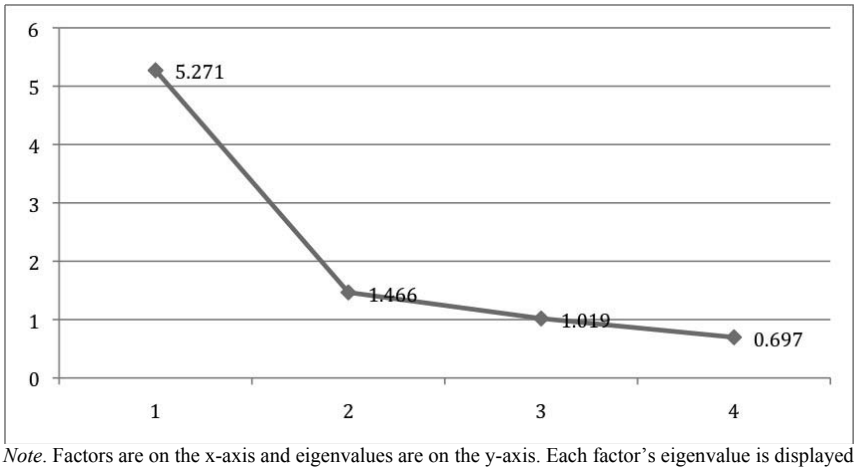
Note. Factors are on the x-axis and eigenvalues are on the y-axis. Each factor's eigenvalue is displayed in the chart (e.g., factor 1 had an eigenvalue of 11.63).

Table 4 *Variance Explained in the SESQ-ENG*

Factor	Percent of variance explained
Cognitive Engagement	36.82%
Behavioral Engagement (Effort & Persistence)	9.81%
Affective Engagement (Liking for Learning)	6.21%
Affective Engagement (Liking for School)	4.73%
Behavioral Engagement (Extracurricular Activities)	3.88%

Note. Cumulative percent of variance explained = 61.45%

Figure 2. *Scree Plot of Eigenvalues for the TERF-N*



Note. Factors are on the x-axis and eigenvalues are on the y-axis. Each factor's eigenvalue is displayed in the chart (e.g., factor 1 had an eigenvalue of 5.271).

Table 5 *Factor Correlations for the SESQ-ENG*

	AFF: Lkg Lrg	AFF: Lkg Sch	BEH: Eff&P	BEH: ECA
AFF: Lkg Lrg				
AFF: Lkg Sch	.51**			
BEH: Eff&P	.66**	.41**		
BEH: ECA	.38**	.37**	.42**	
COG	.44**	.38**	.47**	.45**

Note. Correlation is significant at the $p < .01$ level.

Teacher Engagement Report Form - New (TERF-N). The scree plot with eigenvalues is displayed in Figure 2. Utilizing Kaiser's criterion, a three-factor model would be optimal; while Cattell's scree test would indicate a one-factor model. The fit indices (Table 6) provide divergent indications of the best-fitting model, although it seems the four-factor model is the best fit. An ultra-Heywood case[3] was evident in the four-factor model rendering this model un-interpretable. The five-, three-, and two-factor models resulted in several undetermined factors and complex structure. Additionally, in each of these models, most of the items loaded on one factor. Therefore, the one-factor model is interpreted.

Table 7 lists the factor loadings for the factor (titled General Engagement). With the exception of two items, all items significantly loaded on the factor. The percentage of variance explained by the factor was 52.88%. The correlations for the TERF-N items range from trivial ($r = -.07$) to large ($r = .87$). With the exception of item 6, the majority of correlations are significant. The average inter-item correlation is moderate ($r = .43$).

Correlations between the SESQ and the TERF-N

After establishing internal consistency and structure of the measures, establishing that these measures are correlated is important. Bivariate correlations between the SESQ and the TERF-N were conducted. This analysis demonstrated positive, significant, moderate correlations between the TERF-N (total score) and the SESQ for the Affective ($r = .33$), Behavioral ($r = .35$), and Total ($r = .33$) engagement composites, each statistically significant at $p < .01$. The Cognitive engagement composite correlation was not significant ($r = .06$).

Table 6 *Goodness-of-Fit Statistics for the TERF-N EFA*

Model	χ^2	df	CFI	RMSEA	SRMR
1	169.58***	35	.85	.17 (.15-.20)	.09
2	96.28***	26	.92	.15 (.11-.18)	.04
3	59.65***	18	.95[a]	.13 (.10-.17)	.03
4	19.16[a]	11	.992[a]	.08 (.00-.13)	.02
5	7.70[a]	5	.997[a]	.07 (.00-.15)	.01[a]

Note. DF=Degrees of Freedom. CFI = Comparative Fit Index. RMSEA = Root Mean Square of Error of Approximation. SRMR = Standard Root Mean-Squared Residual. AIC = Aikake Information Criterion. BIC = Bayesian Information Criterion.
*** $p < .001$.
[a] Model meets standards for that index established in the literature.

[3]When parameter estimates have out-of-range values, they are referred to as Heywood cases. In the case of the TERF-N, the communality parameter estimate was more than one, which is referred to as an ultra-Heywood case and, which renders the solution invalid and un-interpretable (Brown, 2006).

Table 7 *Factor Loadings for the TERF-N Items*

ITEM & HYPOTHESIZED DOMAIN	General Engagement λ	Communality
AFFECTIVE		
Seems interested in school	.89	.92
Gets along with peers	.56	.51
Seems to care about grades	.94	.90
BEHAVIORAL		
Has good attendance	.50	.97
Participates in class discussions/activities	.74	.70
Is referred for out-of-class disciplinary procedures	-.29	.54
Is respectful to staff	.60	.62
COGNITIVE		
Persists on more challenging tasks	.23	.87
Demonstrates appropriate effort for task	.91	.89
Is self-motivated	.92	.91

Note. The average communality for the TERF-N = .60.

DISCUSSION

The primary purpose of this study was to begin to examine the psychometric properties of the Student Engagement in Schools Questionnaire (SESQ) and the Teacher Engagement Report Form (TERF-N). Analyses included; a) internal consistency estimates to examine reliability, b) exploratory factor analysis to examine the latent factor structure of each measure, and c) correlations to examine the relationship of the scales and begin to examine external validity.

Internal Consistency

Both measures yielded good internal consistency ($\alpha \geq .70$) with the exception of the Attributions domain ($\alpha = .65$) of the SESQ. Examination of the Attributions domain indicates that it may be problematic due to the structure of the questions, in particular the response format appeared to create confusion (as evidenced by questions from students during the administration and also by the number of responses that were not plausible). Based on careful review of responses to this item, it is recommended that this portion of the questionnaire be eliminated or altered to avoid this challenge.

Factor Structure

The factor structure of the SESQ Engagement items (i.e., the items measuring the affective, behavioral, and cognitive indicators of engagement) retained its conceptualized structure through exploratory factor analysis. As hypothesized, a five-factor model consisting of Affective (i.e., Liking for Learning and Liking for School), Behavioral (i.e., Effort & Persistence, Extracurricular Activities), and Cognitive Engagement factors was demonstrated. Two items (a1 & a9R) from the Affective: Liking for Learning factor cross-loaded on the Behavioral: Effort & Persistence factor. While these items loaded more strongly on the conceptualized factors, they demonstrated the weakest loadings on the factor. These items were also among the lowest communalities of the factor. Next steps for the SESQ will include reducing the length of the survey. These two cross-loading items will likely be candidates for elimination.

The EFA for the TERF-N produced interesting results. While a three-factor structure was hypoth-

esized, a one-factor model of General Engagement was the most viable, interpretable model. It may be that a student's levels of engagement are viewed through less differentiated eyes by a teacher (i.e., teachers see students who are engaged and those who are not rather than students who are affectively engaged and not behaviorally engaged) in comparison to students self-reporting (i.e., students may feel more variation than is actively displayed). Another consideration is that the teacher relies on easily observable characteristics and creates an overall, rather stable impression of a student, while the student relies on internal characteristics and is able to see more variation in his attitudes and beliefs over time and situation. A final explanation regarding the unidimensionality of the TERF-N may be related to the size of the questionnaire (i.e., 10 items).

Two items did not load significantly on the General Engagement factor (i.e., "Is referred for out-of-class disciplinary procedures" and "Persists on more challenging tasks"). These items also had very low communalities. While these items did not load as expected on the factor, it is believed by the researchers that these items provide important information to the evaluation of student engagement. According to the conceptualization of the construct, out-of-class discipline and persistence in academic tasks are considered related to a student's level of engagement at school. Additionally, the sample size of the TERF-N study was relatively small. It will be important to further evaluate the TERF-N on larger sample sizes before making any decisions about eliminating or changing items.

External Validity

Finally, moderate, significant correlations between the SESQ Affective and Behavioral Engagement composites and the overall TERF-N scores were evident. These correlations imply a relationship between how teachers and students view engagement; however, the moderate size of the correlations suggests that important information is gathered from both informants. The Cognitive Engagement composite score was not statistically significant and close to zero, implying no relationship between the two. One possible reason for this might be that cognitive engagement is an internal characteristic of an individual. Behavioral engagement is typically easily observable by the classroom teacher in a way that cognitive engagement is not. These results suggest that information regarding a student's cognitive engagement in school is best recognized by the individual themselves. Following further exploration, it may be that the cognitive dimension of the TERF-N is not feasible.

Limitations and Future Directions

The interpretation of the results of this exploratory study warrant caution, related to the limitations of the study. In particular, the small size of the TERF-N sample and the geographically limited sample of both analyses are notable limitations. Future studies can be conducted to further examine the results of the current study by utilizing a larger sample size, including more classroom teachers, with a variety of grade levels. Second, the length of the SESQ is problematic. The self-report form, consisting of 109 items, may cause fatigue in the students. Future efforts will include creating a psychometrically sound brief version of the SESQ. Additionally, a computer-based version of both forms would be beneficial. With a computer-based questionnaire students and teachers could complete the form outside of class time and over more than one session if necessary. A computer-based questionnaire would also increase accessibility and widen the population from which to sample.

Additionally, researchers should focus on gathering more reliability and validity data. Information regarding outcomes (e.g., grades, attendance, dropout) should be collected in order to demonstrate correlations between the SESQ and TERF-N and these outcomes. Ideally, longitudinal data will be gathered to demonstrate the importance of these measures. Also, measures accessing both similar and different constructs (e.g., Psychological Sense of School Membership, Goodenow, 1993; Children's Depression Inventory; Kovacs, 1992) should be utilized in order to confirm convergent and divergent validity. Internal consistency and confirmatory factor analyses should continue to be investigated in order to provide statistical bases for the measures. Future studies should gather gender and ethnicity information in more reliable ways (e.g., from student records), and analyze gender and ethnicity differences in student engagement.

Finally, the goal of assessment is to drive interventions. The SESQ and TERF-N will provide important information to the understanding of a student's experience at school. Once measures are created and tested, the next logical step is to use the information from those measures to inform interventions; future research should focus on creating and evaluating school-based engagement programs for prevention and intervention. There are many opportunities for research in student engagement in the schools.

CONCLUSIONS

Student engagement is a complex construct that continues to be important in promoting positive outcomes for students. Challenges remain in the conceptualization and measurement of this construct. One question that persists is: Should engagement be measured by its apparent components or should all types of engagement be lumped together as one overarching construct? Researchers should focus on the areas of examining and measuring engagement in order to further refine measures (e.g., short forms), and create interventions directly related to an individual student's needs. Further efforts are warranted in the development of a more comprehensive perspective of assessing student engagement, for example, using self-, teacher-, and parent-reports. Information from the multiple sources may provide a better understanding of students. Additionally, student engagement data collected on the large scale can provide administrators with additional school climate information, and may direct interventions at the universal, school-wide level.

Ongoing efforts related to the conceptualization and measurement of student engagement also need to seek out helpful information about how the construct relates directly to positive student outcomes. The current models of student engagement assume that there are multiple components, thus, outcomes should be present and measureable in, thus, defined components.

It is important for school psychologists to be aware of the literature and the ongoing research efforts in the area of student engagement in the schools. Interventions suggested by researchers (e.g., Reschly et al., 2007) can be targeted for specific students who are at-risk, or used for prevention efforts at the school-wide level. School psychologists can provide a context that is consultation- and collaboration-friendly; they can use their knowledge to provide psychoeducation to teachers on the importance of engagement in the classroom, in addition to strategies to enhance student engagement in school.

This study contributes to continued efforts by school staff, school psychologists, and educational researchers to investigate student engagement. Through collaboration among practitioners and academics, student engagement interventions may be part of the key to promoting school completion and academic outcomes. It is hoped that further work in this area will result in better measures of student engagement in school and increased positive outcomes for all students.

- - -

Shelley R. Hart, PhD., is a recent graduate of the University of California, Santa Barbara.

Kaitlyn Stewart, M.Ed., is a recent graduate of the University of California, Santa Barbara.

Shane R. Jimerson, PhD., is a professor at the University of California, Santa Barbara.

REFERENCES

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. doi: 10.1002/pits.20303

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. doi: 10.1016/j.jsp.2006.04.002

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Carter, M., McGee, R., Taylor, B., & Williams, S. (2007). Health outcomes in adolescence: Associations with family, friends and school engagement. *Journal of Adolescence, 30*, 51-62. doi: 10.1016/j.adolescence.2005.04.002

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Christenson, S.L., Sinclair, M.F., Lehr, C.A., & Godber, Y. (2001). Promoting successful school completion: Critical conceptual and methodological guidelines. *School Psychology Quarterly, 16*, 468-484. doi: 10.1521/scpq.16.4.468.19898

Costello, A.B., & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.

Cronbach, L.J., & Shavelson, R.J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*, 391-418. doi: 10.1177/0013164404266386

Csikszentmihalyi, M., (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299. doi: 10.1037/1082-989X.4.3.272

Finn, J.D. (1989). Withdrawing from school. *Review of Educational Research, 59*, 117-142.

Fredricks, J.A., Blumenfeld, P.C., & Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research 74*(1), 59-109. doi: 10.3102/00346543074001059

Furr, R.M., & Bacarach, V.R. (2008). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage.

Furlong, M.J., Whipple, A.D., St. Jean, G., Simental, J., Soliz, A., & Punthuna, S. (2003). Multiple contexts of school engagement: Moving toward a unifying framework for educational research and practice. *The California School Psychologist, 8*, 99-114.

Gliner, J.A., & Morgan, G.A. (2000). *Research methods in applied settings: An integrated approach to design and analysis*. Mahweh, NJ: Lawrence Erlbaum & Associates.

Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools, 30*, 79-90. doi: 10.1002/1520-6807(199301)30:1<79::AID-PITS231030011 3>3.0.CO;2-X

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. doi: 10.1080/10705519909540118

Jimerson, S.R., Campos, E., & Greif, J.L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *The California School Psychologist, 8*, 7-27.

Jimerson, S.R., Renshaw, T.L., Stewart, K., Hart, S., & O'Malley, M. (2009). Promoting school completion through understanding school failure: A multi-factorial model of dropping out as a developmental process. *Romanian Journal of School Psychology, 2*, 12-29.

Kahn, J.H. (2006). Factor analysis in counseling psychology research, training, and practice: Principles, advances, and applications. *The Counseling Psychologist, 34*, 684-718. doi: 10.1177/0011000006286347

Kline, P. (2002). *An easy guide to factor analysis*. New York: Routledge.

Kovacs, M. (1992). *The Children's Depression Inventory*. New York: Multi-Health Systems.

Lam, S.F., & Jimerson, S.R. (2008). *Exploring student engagement in schools internationally: Consultation paper*. Chicago, IL: International School Psychologist Association.

Muthén, L.K., & Muthén, B.O. (2009). *Mplus Statistical Analysis with Latent Variables: User's Guide* (5th Ed.). Los Angeles, CA: Author.

Patrick, H., Ryan, A.M., & Kaplan, A. (2007). Early adolescents' percpetions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology, 99*, 83-98. doi: 10.1037/0022-0663.99.1.83

Reschly, A.L., Appleton, J.J., & Christenson, S.L. (2007). Student engagement at school and with learning: Theory and interventions. *NASP Communiqué, 35*(8).

Reschly, A.L., & Christenson, S.L. (2006). Prediction of dropout among students with mild disabilities: A case for inclusion of student engagement variables. *Remedial and Special Education, 27*, 276-292. doi: 10.1177/07419 325060270050301

Shernoff, D.J., & Schmidt, J.A. (2008). Further evidence of an engagement-achievement paradox among U.S. high school students. *Journal of Youth and Adolescence, 37*(5), 564-580. doi: 10.1007/s10964-007-9241-z

SPSS, Inc. (2007). *SPSS Base 16.0 for Mac User's Guide*. Chicago, IL: Author.

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics*. Boston, MA: Allyn & Bacon.

Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.

Walker, C.O., Greene, B. A., & Mansell, R.A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences, 16*, 1-12. doi: 10.1016/j.lindif.2005.06.004

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. Albany, NY: State University of New York Press.