University Student Engagement Inventory (USEI): Transcultural Validity Evidence Across Four Continents

Hugo Assunção[1], Su-Wei Lin[2], Pou-Seong Sit[3], Kwok-Cheung Cheung[3],
Heidi Harju-Luukkainen[4], Thomas Smith[5], Benvindo Maloa[6],
Juliana Álvares Duarte Bonini Campos[7], Ivana Stepanovic Ilic[8], Giovanna Esposito[9],
Freda Maria Francesca[9] and João Marôco[1]*

[1] William James Center for Research, ISPA - University Institute of Psychological, Social and Life Sciences, Lisbon, Portugal,
[2] Department of Education, National University of Tainan, Tainan, Taiwan, [3] Faculty of Education, University of Macau, Macau,
China, [4] Faculty of Education and Arts, Nord University, Bodø, Norway, [5] Department of Educational Technology, Research
and Assessment, Northern Illinois University, DeKalb, IL, United States, [6] Universidade Pedagógica de Maputo, Maputo,
Mozambique, [7] School of Pharmaceutical Sciences, São Paulo State University, São Paulo, Brazil, [8] Department of
Psychology, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia, [9] Department of Humanistic Studies, University
of Naples Federico II, Naples, Italy

Academic engagement describes students' involvement in academic learning and achievement. This paper reports the psychometric properties of the University Student Engagement Inventory (USEI) with a sample of 3992 university students from nine different countries and regions from Europe, North and South America, Africa, and Asia. The USEI operationalizes a trifactorial conceptualization of academic engagement (behavioral, emotional, and cognitive). Construct validity was assessed by means of confirmatory factor analysis and reliability was assessed using Cronbach's alpha and McDonald's omega coefficients. Weak measurement invariance was observed for country/region, while strong measurement invariance was observed for gender and area of graduation. The USEI scores showed predictive validity for dropout intention, self-rated academic performance, and course approval rate while divergent validity with student burnout scores was also evident. Overall, the results indicate that the USEI can produce reliable and valid data on academic engagement of university students across the world.

Keywords: student engagement, transcultural invariance, measurment, confirmatory factor analysis, university

INTRODUCTION

The concept of engagement emerged in professional and occupational contexts, but has recently been expanded to the educational context as well (Kuh, 2009; Vasalampi et al., 2009; Bresó et al., 2011; Reschly and Christenson, 2012). Student engagement is viewed as a malleable, developing, and multidimensional construct that evolves over time. It can be affected by interventions that enhance positive performance and prevent potential dropout (Appleton et al., 2008). Engaged students invest more in their performance, participate more in school activities, and tend to develop mechanisms to help them persevere and self-regulate their learning processes (Raykov, 2001; Klem and Connell, 2004). Academic engagement is both the cause and consequence of having positive academic and social outcomes (Klem and Connell, 2004; Wonglorsaichon et al., 2014), leading to more satisfaction and self-efficacy (Elmore and Huebner, 2010; Coetzee and Oosthuizen, 2012),

and lower incidence of achievement problems and dropout (Fredricks et al., 2004; Gilardi and Guglielmetti, 2011; Reschly and Christenson, 2012).

An early conceptualization of engagement comes from Maslach and Leiter's (1997) work on the burnout construct. These authors define burnout as the erosion of engagement (Maslach and Leiter, 1997). The burnout syndrome is considered to have three dimensions: emotional exhaustion, depersonalization, and personal accomplishment (Maslach et al., 1996), later generalized to exhaustion, cynicism, and professional efficacy (Schaufeli et al., 1996). Thus, in earlier works engagement was conceptualized as the opposite of burnout and defined as the attribution of meaning and importance to work with feelings of energy, commitment, and accomplishment. When engagement fades, energy turns into exhaustion, involvement turns into cynicism, and efficacy turns into ineffectiveness, leading workers into burnout. In this perspective, people exist in a burnout-engagement continuum in relation to their work (Maslach and Leiter, 1997). However, this conceptualization has a major drawback: people with low levels of burnout are not necessarily engaged in their work. Responding to this critique, a new conceptualization of engagement was proposed by Schaufeli et al. (2002) where three dimensions were considered (vigor, dedication, and absorption), and where engagement was defined as vigor (energy and resilience), absorption (concentration and immersion), and dedication (involvement and enthusiasm). In this view, burnout and engagement, although negatively correlated, are not conceptual opposites. While vigor is the conceptual opposite of exhaustion (activation continuum) and dedication is the opposite of depersonalization/cynicism (identification continuum), absorption and inefficacy are not conceptual opposites (Schaufeli et al., 2002). Absorption is characterized by being "fully concentrated and happily engrossed in one's work, whereby time passes quickly, and one feels carried away by one's job." Based on these nomological considerations, Schaufeli and Bakker (2004) proposed the Utrecht Work Engagement Scale (UWES) to measure engagement. Several authors have since proposed other models that combine behavioral and psychological dimensions (Audas and Douglas Willms, 2001); behavioral, emotional, and cognitive dimensions (Fredricks et al., 2004; Hart et al., 2011); and even a fourth dimension such as academic engagement or agency (Appleton et al., 2008; Reeve and Tseng, 2011; Sinatra et al., 2015). Proposals for the construct dimensionality have ranged from two to eight (learning strategies, academic integration, institutional emphasis, co-curricular activity, diverse interactions, effort, overall relationships, and workload; Lanasa et al., 2009) and higher dimensional models also have been proposed (Martin, 2007).

In this paper, we follow the conceptualization described in Maroco et al. (2016) that expands on the Nystrand and Gamoran (1989) definition of students' engagement with the North American model (Nystrand and Gamoran, 1989; Fredricks et al., 2004; Maroco et al., 2016). This model has received considerable attention and extensive empirical examination (Janosz et al., 2008; Mo et al., 2008; Archambault et al., 2009; Vasalampi et al., 2009; Bresó et al., 2011; Wang et al., 2011; Tuominen-Soini and Salmela-Aro, 2014;

Wang and Fredricks, 2014; Alrashidi et al., 2016; Salmela-Aro and Upadyaya, 2017). Based on this model, Maroco et al. (2016) devised the University Student Engagement Inventory (USEI) which includes behavioral, cognitive, and emotional dimensions of academic engagement with university students. The behavioral dimension is related to positive normative class behaviors (e.g., respecting the social and institutional rules). The cognitive dimension refers to students' thoughts, perceptions, and strategies related to the acquisition of knowledge or development of competencies to academic activities (e.g., learning approaches). The emotional dimension refers to positive and negative feelings and emotions related to the learning process, class activities, peers, and teachers (Sheppard, 2011; Carter et al., 2012; Maroco et al., 2016). Based on the nomology of the first order engagement constructs, their theoretical closedness as well as the moderate to strong inter-construct correlations, Maroco et al. (2016) proposed a second order factor termed "Engagement." This second order construct provides an overall measure of the student engagement that unifies the construct (three dimensions, one overall measure), useful for both education psychologists and educators.

Other engagement scales, such as the UWES, have suffered from several criticisms ranging from the construct definitions and dimensionality to its applicability to university students (Lanasa et al., 2009; Wefald and Downey, 2009; Fiorini et al., 2014; Kulikowski, 2017). The USEI was created to measure student engagement in the university context as opposed to the organizational context (Wefald and Downey, 2009; García-Ros et al., 2017) or the elementary student's context (Fredricks et al., 2011).

Content-related validity evidence based on response processes of the behavioral, cognitive, and emotional as dimensions of academic engagement was evaluated with a focus group of psychologists and university students in the original proposal of Maroco et al. (2016). The USEI has been shown to present appropriate validity, reliability, and measurement invariance across gender and the area of graduation using confirmatory factor analysis (CFA) (Sinval et al., 2018). Although measurement invariance was found across genders and area of studies, no studies so far have analyzed the USEI's measurement invariance across countries. In this paper, we expect to replicate previous findings by analyzing the USEI's factorial validity, internal consistency reliability, and convergent and discriminant validity evidence (H1). We also expect the USEI to present measurement invariance across genders, areas of study, and different countries/regions (H2). Finally, we expect that the USEI presents evidence of criterion predictive validity with academic relevant variables such as students' dropout intention, academic performance, course expectations, course approval rate, and student burnout scores (H3).

MATERIALS AND METHODS

Participants

Minimum sample size for CFA was determined by Monte-Carlo simulation as suggested by Brown (2015) with criteria defined by Muthén and Muthén (2017): (a) Bias of parameters

estimates <10%; (b) 95% confidence intervals coverage >91%; and (c) percentage of significant coefficients (power) ≥80%. *Mplus* software (v. 8; Muthén and Muthén, 2017) was used for simulations with the second-order CFA model using factor loadings from the original USEI study (Maroco et al., 2016). A total of 1000 replications employing sample sizes of 100, 200, and 300 were simulated. A minimum sample size of 200 was shown to be enough to attain bias <1% for both parameters and parameters' standard errors; 99% confidence interval coverage >95%, and minimum power of 90%. However, to ensure that the study sample (which was non-probabilistic) would capture a large amount of the normative population variance we set the sample size at a minimum of 300 students per country/region (i.e., 20 participants per item of the model as suggested by Marôco, 2014).

We collected a sample of 4479 university students (ages ranging from 16 to 70 years; *M* = 23.2; SD = 5.6; Mdn = 21) from Portugal (1067), Brazil (424), Mozambique (413), United Kingdom (314), United States (316), Finland (356), Serbia (409), Macau SAR and Taiwan (762), and Italy (418).

The typical participant was female (60%), pursuing a bachelor's degree (74%) in human and social sciences (51%) in a public (88%) university (80%), living with their family (54%), which financed their studies (56%) (see **Table 1** for further details).

Measuring Instruments
University Student Engagement Inventory
The USEI (Maroco et al., 2016) was used to measure student engagement. In the USEI, student engagement is conceptualized as a second-order factor construct that is reflected as behavioral, emotional, and cognitive dimensions. Behavioral engagement is defined as students' participation in classroom tasks, student conduct, and participation in school-related extracurricular activities. Cognitive engagement is defined as the students' investment and willingness to exert the necessary efforts for the comprehension and mastering of complex ideas and difficult skills. Emotional engagement is defined as attention to teachers' instructions, perception of school belonging, and beliefs about the value of schooling. The USEI consists of 15 self-report items, each associated with Likert-type response options ranging from "1-never" to "5-always." Each of the three first-order factors is composed of five items. The USEI has previously been assessed for factorial validity and reliability (Maroco et al., 2016) and measurement invariance across genders and areas of study (Sinval et al., 2018) but only for Portuguese speaking students. In this study, we used five versions of the scale: Portuguese (for Portugal, Brazil, and Mozambique), English (for the United Kingdom, the United States, and Finland), Serbian (Serbia), Italian (Italia), and simplified Chinese (Macau SAR and Taiwan; see **Supplementary Data Sheet 1**). The Portuguese and English versions used were the original ones of Maroco et al. (2016). The Serbian, Italian, and simplified Chinese were translated by authors from Maroco et al. (2016) and checked for cross-cultural equivalence.

Maslach Burnout Inventory – Student Survey
The Maslach Burnout Inventory – Student Survey (MBI-SSi; Maroco et al., 2014) was used to measure student burnout. Student burnout is conceptualized as a second-order construct reflected on the first-order exhaustion, cynicism, and inefficacy dimensions. The MBI-SSi consists of 15 self-report items rated with a 7-point Likert frequency scale from "0-Never" to "6-Every day." In its original formulation (Schaufeli et al., 2002),

TABLE 1 | Demographic variables by country.

Country	PT	BR	MZ	UK	USA	FIN	SER	TW&MO	IT
Age (mean)	22.9	23.3	26.3	22.6	21.9	26.2	22.0	22.3	23.3
Age (median)	21	22	25	21	20.5	24	22	21	21
Age (SD)	6.7	5.3	6.8	5.3	4.3	7.8	2.2	5.4	6.0
Women (%)	65.3	43.4	62.2	47.1	34.8	62.1	83.9	66.1	77.5
Public school (%)	90	85	97	86	81	94	96	80	90
Human and social sciences (%)	33.4	39.2	71.9	49.0	37.0	56.5	53.3	66.5	30.4
Exact sciences (%)	29.6	38.7	11.4	36.0	50.0	32.3	20.8	22.0	0.0
Biological sciences (%)	9.7	9.2	12.1	8.9	8.2	6.7	4.6	6.3	69.6
Health sciences (%)	27.4	13.0	4.6	6.1	4.7	4.5	15.9	5.1	0.0
1st year (%)	17.7	13.6	39.5	26.3	20.1	15.6	21.3	21.9	21.3
2nd year (%)	18.6	18.1	20.1	20.5	29.4	15.6	21.8	18.4	19.7
3rd year (%)	21.9	20.3	8.7	14.1	36.3	20.8	23.3	28.4	22.2
4th year (%)	15.6	24.1	26.6	22.9	4.3	17.3	28.7	19.1	19.4
5th year (%)	7.0	19.8	3.9	4.7	3.0	28.3	1.3	5.2	8.5
6th year (%)	4.1	1.4	0	4.0	1.3	0.9	1.8	3.7	2.6
7th year (%)	7.3	1.7	0.7	2.0	1.7	0.9	0.8	0.9	2.8
8th year (%)	6.8	0.7	0.5	2.4	2.6	0.3	0	1.6	2.7
9th year (%)	0.6	0.2	0	3.0	1.3	0.3	1.0	0.5	0.7
10th year (%)	0.4	0	0	0	0	0	0	0.3	0.2

Note: PT - Portugal, BR - Brazil, MZ - Mozambique, UK - The United Kingdom, USA - The United States of America, FIN - Finland, SER - Servia, TW&MO - Taiwan and Macao SAR, IT - Italy.

the Efficacy dimension has its items positively worded while Emotional Exhaustion and Cynicism are composed of negatively worded items. Here we use a version of the MBI-SS (MBI-SSi; Maroco et al., 2014) where the items in the Efficacy dimension were negatively worded to give rise to the Inefficacy (INEF) dimension. Four versions of the scale were used in this study: Portuguese (Portugal, Brazil, Mozambique), English (the United Kingdom, United States, and Finland), Serbian (Serbia), and simplified Chinese (Macau and Taiwan).

Demographic and Academic-Related Questions

The demographic variables assessed were gender, age, region, household, and financial support. The self-reported academic variables were the name of the degree, area of degree (human and social sciences, exact sciences, biological sciences, and health sciences), type of degree (bachelor's, master, doctorate), type of school (public/private university), year of school, time of classes, order of preference for the course, self-reported academic performance, dropout intention, total number of classes, and number of failed classes. The class approval rate was calculated by subtracting from one the ratio of the number of failed classes with the number of total classes the student has attended. Five versions of the demographic and academic-related questions were used in this study: Portuguese (Portugal, Brazil, Mozambique), English (the United Kingdom, United States, and Finland), Serbian (Serbia), Italian (Italia), and simplified Chinese (Macau SAR and Taiwan).

Procedures

An online questionnaire containing two scales measuring student engagement using USEI (Maroco et al., 2016) and student burnout using the MBI (MBI-SSi; Maroco et al., 2014) was created using the Qualtrics platform. The order of appearance of the two scales was randomized between participants. At the end of the questionnaire, participants answered a series of demographic and academic-related questions. The survey was designed to take 15 min to complete. The content, objectives, duration, risks, data policy, ethics approval, and contacts were provided at the start of the questionnaire. Informed consent was required to participate as well as confirmation of enrollment in a higher education institution. To move forward in the questionnaire all answers were mandatory. Only completed questionnaires with no missing data were considered for data analysis. At the end of the survey, participants were asked to voluntarily leave a comment about the survey and to provide their e-mail to receive the results of the study if they wanted to. Faculty members and student associations were contacted in each country/region and invited to distribute the survey via e-mail and online social media.

Data Analysis

Descriptive Statistics and Item Sensitivity

Descriptive statistics were obtained using the *skimr* package (v. 1.0.5; McNamara et al., 2018) and the *psych* package (v. 1.8.12; Revelle and Revelle, 2015) for the *R* statistical system (v. 3.5.3; R Core Team, 2013). The minimum, maximum, average, standard deviation, skewness, and kurtosis were calculated, and histograms

were created for each item. Absolute skewness and kurtosis values above 7 and 3, respectively, were considered indicative of strong deviations from normality (Finney and DiStefano, 2013) and low item psychometric sensitivity (Marôco, 2014).

Confirmatory Factor Analysis

Confirmatory factor analysis was conducted with the *lavaan* package (v. 0.6.4; Rosseel, 2012) to evaluate the psychometric properties of the data gathered with the USEI and MBI. CFA was conducted to verify whether the first- and second-order factor structure presented an adequate fit for the sample data. We used the following goodness-of-fit indices: χ^2 (Chi-square statistic), comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). The fit of the model was considered acceptable when CFI and TLI values were >0.90 and RMSEA and SRMR values were <0.06 and <0.08, respectively (Hu and Bentler, 1999; Marôco, 2014).

Although the USEI items are ordinal, because not all response categories were present in all the nine participant countries/regions, it was not possible to use WLSMV estimation to test threshold invariance. However, when the categorical items have at least five categories and a normal-shaped distribution, as it was observed for our sample, Pearson correlations estimate well the associations between variables (Bentler, 1988; Marôco, 2014). Thus, CFA and analysis of invariance by means of multigroup CFA were carried out using robust maximum-likelihood (MLR) estimation implemented in *lavaan* to account for the small deviations from normality and overestimation of fit indices. No measurement errors of items were correlated for both the USEI and MBI measurement models.

Evidence of Convergent and Discriminant Validity Evidence

To analyze the convergent and discriminant validity evidence, the average variance extracted (AVE; Fornell and Larcker, 1981) and the heterotrait–monotrait (HTMT; Henseler et al., 2015) correlations were calculated using the *semTools* package (v. 0.5.1; Jorgensen et al., 2018). Values of AVE ≥ 0.5 were considered acceptable indicators of convergent validity evidence. For two factors x and y, when AVE_x and $AVE_y \geq r^2_{xy}$ (Fornell and Larcker criterion), or when HTMT correlation values are <0.7, the two factors show evidence of discriminant validity.

Evidence of Reliability

Evidence of reliability was assessed using internal consistency measures with the "SemTools" R package (v. 0.5.1; Jorgensen et al., 2018): the Cronbach's alpha coefficient (α; Cronbach, 1951), the coefficient omega (ω; McDonald, 2013), and the hierarchical omega coefficient (ω_h; Green and Yang, 2009; McDonald, 2013; Kelley, 2016) for each factor. Alpha and omega values ≥ 0.7 were satisfactory indicators of internal consistency (Marôco, 2014).

Evidence of Measurement Invariance

Measurement invariance was tested for country/region, gender, and area of studies. We created a set of comparisons within a

group of seven nested models based on the recommendations of Millsap and Yun-Tein (2004) and Wu and Estabrook (2016) for second-order models. A configural model was created, where factor loadings, item intercepts, regression coefficients (second-order structural loadings), first-order factor intercepts, and second-order factor means were freely estimated between groups. This model served as a baseline for further invariance testing. Four nested models were thereafter created where factor loadings, item intercepts, regression coefficients, factor intercepts, and means were sequentially fixed between groups. Fit indices of the nested models were assessed to probe for invariance. Invariance was assessed using the |ΔCFI| < 0.01 criteria (Cheung and Rensvold, 2002) and the |ΔRMSEA| < 0.01 criterion set by Rutkowski and Svetina (2014) were used. χ^2 difference tests were not used because the large sample sizes would result in statistical significance even when very little invariance was evident. When first-order factor loadings and regression coefficients were invariant between groups, but intercepts were not invariant, weak or metric invariance was assumed. Metric invariance means that the contribution of each item to the factor remains constant across different groups and, thus, relationships of the constructs to other variables can be compared validly among groups. When factor loadings and intercepts were invariant across groups, strong or scalar invariance was assumed. Scalar invariance enables comparisons between group means (Millsap and Yun-Tein, 2004). When factor loadings, intercepts, and second-order factor loadings were invariant across groups, full measurement invariance was assumed. Analysis of invariance may stop at this level because invariance between residuals is considered too restrictive (Marôco, 2014). To ensure equal contributions to the invariance analysis of all eight countries/regions and obtain model convergence, a random sample of 313 students from each participant country/region was drawn from the original sample. To ensure the equal contribution of all areas of study and achieve convergence in invariance analysis between areas of study, a random sample of 335 students from each area was selected from the original sample.

Evidence of Criterion and Concurrent-Related Validity

To assess criterion validity, dropout intention, self-rated academic performance, course approval rate, and student burnout scores were simultaneously regressed on student engagement. Evidence of criterion predictive validity was obtained with MLR or probit regression (for ordinal outcomes) using the *lavaan* package (v. 0.6.4; Rosseel, 2012).

Student Engagement Scores

Student engagement scores were estimated using the *lavaan* package (v. 0.6.4; Rosseel, 2012) under the weak (metric) invariance assumption among countries/regions. Engagement, behavioral, emotional, and cognitive factors' scores were estimated, and the following statistics/plots were generated for each dimension: sample size, mean, standard deviation, quartiles, and histogram.

RESULTS

Items' Distributional Properties

Summary measures, including skewness (sk) and kurtosis (ku), as well as the histogram for each of the USEI items are presented in **Table 2**. No USEI item showed absolute value of ku and sk indicative of strong deviations from the normal distribution or lack of psychometric sensitivity.

Factorial Validity Evidence

The USEI first-order three-factor model presented an acceptable fit to the data [χ^2(84) = 751.528, CFI = 0.936, TLI = 0.923, RMSEA = 0.052, SRMR = 0.040). With the addition of a second-order latent variable (**Figure 1**) goodness of fit indices remained the same. The regression (structural) coefficients for the academic engagement second-order factor model were high for behavioral engagement (γ = 0.85; p < 0.001) and emotional engagement (γ = 0.74; p < 0.001) and medium for cognitive engagement (γ = 0.64; p < 0.001).

Convergent and Discriminant Validity Evidence

The AVE was acceptable for EE (0.56) and CE (0.49) but low for BE (0.34). Convergent validity evidence was acceptable for the EE and CE factors and poor for the BE factor. The AVE_{EE} was greater than $r^2_{EE.CE}$ (0.25) and $r^2_{EE.BE}$ (0.42). The AVE_{CE} was greater than $r^2_{CE.EE}$ (0.25) and $r^2_{CE.BE}$ (0.32). The AVE_{BE} was greater than $r^2_{BE.CE}$ (0.31), but not greater than $r^2_{BE.EE}$ (0.42) (**Table 3**). All HTMT inter-construct correlations were below the recommended threshold of 0.70 ($HTMT_{BE.EE}$ = 0.63, $HTMT_{BE.CE}$ = 0.55, and $HTMT_{EE.CE}$ = 0.50). These results altogether show acceptable evidence of convergent- and discriminant-related validity of the USEI dimensions.

Reliability Evidence

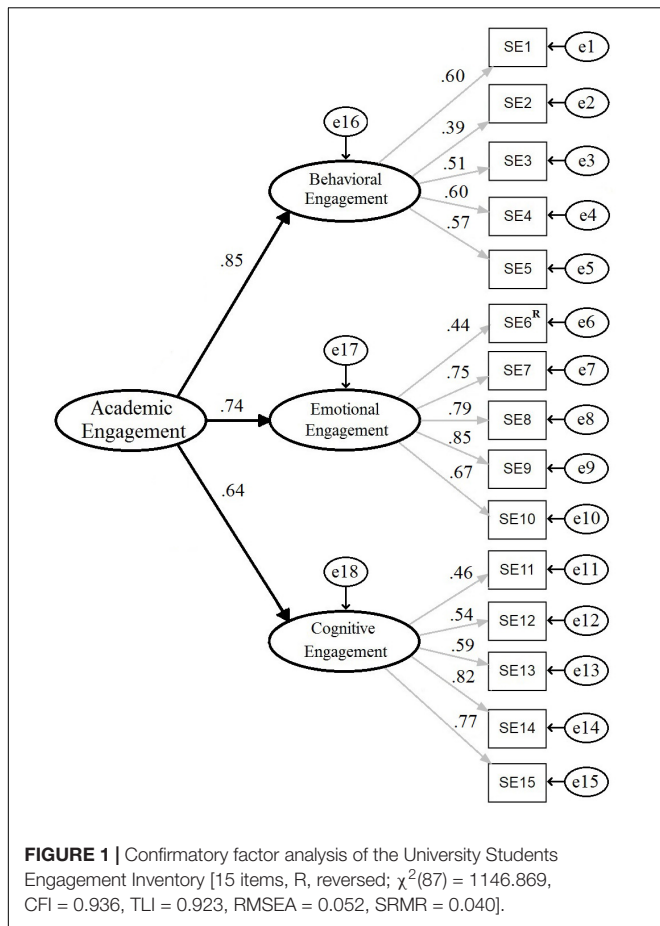The α values were >0.70 for all factors and >0.8 for the total scale (**Table 4**). The hierarchical omega statistic for the total scale was high (ω_h = 0.88), which gives support to a second-order factor as observed elsewhere (Maroco et al., 2016; Sinval et al., 2018). This result provides evidence of acceptable internal consistency reliability.

Evidence of Measurement Invariance
Invariance by Country/Region

To detect whether the second-order latent USEI model holds in different countries/regions, a group of nested models for the nine participating countries/regions was created. **Table 5** lists goodness of fit measures for all models (factor loadings, item intercepts, regression coefficients, factor intercepts, and means). Using the Cheung and Rensvold (2002) ΔCFI criterion (|ΔCFI| < 0.01) and the Rutkowski and Svetina (2014) ΔRMSEA criterion (|ΔRMSEA| < 0.01), metric invariance was found between all countries. Following the lack of global scalar invariance, an analysis of invariance was conducted for pairs of countries/participants. Scalar invariance was found between

TABLE 2 | Distributional properties of USEI's items (R, reversed).

Item	M	SD	Min	p25	p50	p75	Max	sk	ku	Histogram
1	3.71	0.85	1	3	4	4	5	−0.71	0.56	
2	4.32	0.79	1	2	3	4	5	−1.25	1.88	
3	3.76	1.03	1	3	4	4	5	−0.71	−0.07	
4	2.99	1.21	1	3	3	4	5	0.01	−0.95	
5	4.01	1.01	1	4	4	5	5	−1.05	0.77	
6[R]	3.19	1.18	1	3	4	4	5	−0.19	−0.78	
7	3.02	1.01	1	3	4	5	5	−0.03	−0.52	
8	3.4	1.08	1	4	4	5	5	−0.36	−0.53	
9	3.39	1.02	1	3	4	5	5	−0.26	−0.45	
10	3.02	1.05	1	2	3	4	5	−0.06	−0.59	
11	3.51	1.11	1	4	4	5	5	−0.4	−0.53	
12	3.28	1.03	1	2	3	4	5	−0.13	−0.51	
13	4.14	0.92	1	2	3	4	5	−1	0.7	
14	3.83	0.93	1	3	4	4	5	−0.59	0.02	
15	3.84	0.98	1	3	3	4	5	−0.62	−0.13	

FIGURE 1 | Confirmatory factor analysis of the University Students Engagement Inventory [15 items, R, reversed; $\chi^2(87)$ = 1146.869, CFI = 0.936, TLI = 0.923, RMSEA = 0.052, SRMR = 0.040].

Portugal and Brazil and between the United Kingdom and the United States.

Information regarding each model's goodness of fit [χ^2(df), CFI, TLI, RMSEA, SRMR] and model fitness comparison (Δdf, $\Delta\chi^2$, ΔCFI, ΔRMSEA) can be found in **Table 5**.

Information regarding the ΔCFI for each pair of countries can be found in **Table 6**.

Measurement Invariance by Gender

To detect whether the USEI invariance holds across genders, a group of nested models with indications of equivalence was created. **Table 7** lists goodness of fit measures for all models (factor loadings, item intercepts, regression coefficients, factor intercepts, and means). Using the Cheung and Rensvold (2002) ΔCFI criterion (|ΔCFI| < 0.01) and the Rutkowski and Svetina (2014) ΔRMSEA criterion (|ΔRMSEA| < 0.01), scalar measurement invariance was found for gender. Information regarding each model's goodness of fit [χ^2(df), CFI, TLI, RMSEA, SRMR] and model's goodness of fit comparison (Δdf, $\Delta\chi^2$, ΔCFI, ΔRMSEA) can be found in **Table 7**.

TABLE 3 | Average variance extracted (main diagonal), explained variance (R^2; lower triangular matrix), and HTMT correlations (upper triangular matrix).

USEI dimension[a]	BE	EE	CE
BE	0.34	0.63	0.55
EE	0.17	0.56	0.50
CE	0.09	0.06	0.49

[a]BE, behavioral engagement; EE, emotional engagement; CE, cognitive engagement.

TABLE 4 | Internal consistency reliability of USEI dimensions.

USEI dimension[a]	α	ω	ω hierarchical
BE	0.71	0.66	0.65
EE	0.85	0.83	0.83
CE	0.81	0.79	0.81
Total	0.87	0.88	0.88

[a]BE, behavioral engagement; EE, emotional engagement; CE, cognitive engagement.

TABLE 5 | USEI model comparison for country/region invariance.

Model	df	χ^2	CFI	TLI	RMSEA	SRMR	Δdf	$\Delta\chi^{2\dagger}$	ΔCFI	ΔRMSEA
Configural	783	1666	0.92	0.904	0.064	0.057				
Loadings	879	1846	0.912	0.905	0.063	0.069	96	181***	−0.008	−0.001
Intercepts	975	2932	0.824	0.829	0.085	0.084	96	1160***	−0.088	0.022
Regressions	991	2952	0.823	0.831	0.084	0.087	16	23	−0.001	−0.001
Means	1015	3390	0.786	0.8	0.091	0.108	24	440***	−0.037	0.007

*** $p < 0.001$. $\dagger$Scaled $\Delta\chi^2$ test using the Satorra–Bentler correction.

TABLE 6 | ΔCFI (models with fixed loadings and free intercepts vs. model with fixed loadings plus fixed intercepts) for each pair of countries/regions.

ΔCFI	BR	MZ	UK	USA	FIN	SER	M&T	IT
Portugal	−0.005*	−0.049	−0.043	−0.042	−0.022	−0.026	−0.044	−0.106
Brazil		−0.031	−0.022	−0.021	−0.012	−0.021	−0.035	−0.079
Mozambique	−0.031		−0.062	−0.06	−0.065	−0.081	−0.064	−0.097
United Kingdom	−0.022	−0.062		−0.001*	−0.012	−0.037	−0.052	−0.114
United States	−0.021	−0.06	−0.001*		−0.015	−0.033	−0.04	−0.126
Finland	−0.012	−0.065	−0.012	−0.015		−0.017	−0.042	−0.071
Serbia	−0.021	−0.081	−0.037	−0.033	−0.017		−0.027	−0.108
Macau and Taiwan	−0.035	−0.064	−0.052	−0.04	−0.042	−0.027		−0.131
Italy	−0.079	−0.097	−0.114	−0.126	−0.071	−0.108	−0.131	

*|ΔCFI| < 0.01. Column names represent the two letter code for the countries found on the first column.

TABLE 7 | USEI model comparison for gender invariance.

Model	df	χ^2_{scaled}	CFI	TLI	RMSEA	SRMR	Δdf	$\Delta\chi^{2\dagger}$	ΔCFI	ΔRMSEA
Configural	174	1224	0.939	0.927	0.056	0.041				
Loadings	186	1242	0.939	0.931	0.055	0.041	12	15	0	−0.002
Intercepts	198	1454	0.928	0.924	0.058	0.045	12	237***	−0.011	0.003
Regressions	200	1463	0.928	0.924	0.057	0.046	2	9*	0	0
Means	203	1488	0.927	0.924	0.057	0.047	3	25***	−0.001	0

* $p < 0.05$; *** $p < 0.001$. $\dagger$Scaled $\Delta\chi^2$ test using the Satorra–Bentler correction.

Measurement Invariance by Area of Study

To detect whether the second-order latent model invariance holds across different areas of study, a group of nested models for the four areas of study (Social Sciences, Exact Sciences, Biological Sciences, and Health Sciences) was created. **Table 8** lists the goodness of indicators for all models (factor loadings, item intercepts, regression coefficients, factor intercepts, and means). Using the Rutkowski and Svetina (2014) |ΔRMSEA| < 0.01 criterion, strong measurement invariance was achieved among the four areas of study. Information regarding each model's goodness of fit [χ^2(df), CFI, TLI, RMSEA, SRMR] and model's goodness of fit comparison (Δdf, $\Delta\chi^2$, ΔCFI, ΔRMSEA) can be found in **Table 8**.

MBI Factorial Validity and Internal Consistency Evidence

The first-order three-factor MBI-SSi model presented an adequate fit to the data [χ^2(87) = 2573.694, CFI = 0.911, TLI = 0.892, RMSEA = 0.084, and SRMR = 0.056]. With the addition of a second-order latent variable, goodness of fit indices remained the same. The regression coefficients for the burnout second-order factor model were high for exhaustion (γ = 0.80; $p < 0.001$), for cynicism (γ = 0.86; $p < 0.001$), and inefficacy (γ = 0.90; $p < 0.001$). The α and ω values were >0.85 for all factors and >0.90 for the total scale. The hierarchical omega for the total scale was high (ω_h = 0.943). These results provide evidence of adequate internal consistency reliability.

Criterion Validity Evidence

The USEI showed predictive criterion-related validity with dropout intention (β = −0.407, R^2 = 0.165, $p < 0.001$), self-reported academic performance (β = 0.533, R^2 = 0.284, $p < 0.001$), course expectations (β = 0.528, R^2 = 0.279, $p < 0.001$), and course approval rate (β = 0.244, R^2 = 0.059, $p < 0.001$). Evidence for divergent validity with the students' burnout was also observed (r = −0.69, $p < 0.001$).

Student Engagement Scores

Table 9 contains the information of the engagement global and dimensions scores by country/region. Mozambique had

TABLE 8 | USEI model comparison for area invariance.

Model	df	χ^2	CFI	TLI	RMSEA	SRMR	Δdf	$\Delta\chi^{2\dagger}$	ΔCFI	ΔRMSEA
Configural	348	982	0.924	0.908	0.063	0.052				
Loadings	384	1027	0.923	0.915	0.061	0.057	36	46.375	−0.001	−0.003
Intercepts	420	1259	0.900	0.900	0.066	0.062	36	247***	−0.023	0.005
Regressions	426	1262	0.900	0.902	0.065	0.063	6	4.734	0	−0.001
Means	435	1358	0.890	0.894	0.068	0.072	9	100***	−0.01	0.003

***p < 0.001. †Scaled $\Delta\chi^2$ test using the Satorra–Bentler correction.

TABLE 9 | USEI scores (1–5) by country/region.

Country/region	Factor	n	Mean	SD	Min	p25	p50	p75	Max	Histogram
Portugal	Engagement	1067	3	0.58	0.39	2.62	3.02	3.42	4.44	
	- Behavioral	1067	2.97	0.75	−0.47	2.51	3.02	3.51	4.69	
	- Emotional	1067	3.02	0.71	0.88	2.55	3.06	3.56	4.62	
	- Cognitive	1067	3.03	0.7	0.1	2.6	3.08	3.52	4.31	
Brazil	Engagement	424	2.82	0.69	0.78	2.39	2.84	3.28	4.44	
	- Behavioral	424	2.75	0.88	−0.05	2.22	2.82	3.38	4.69	
	- Emotional	424	2.8	0.85	0.83	2.21	2.81	3.49	4.62	
	- Cognitive	424	2.94	0.81	0.51	2.38	3.03	3.63	4.31	
Mozambique	Engagement	413	3.37	0.58	0.39	3.04	3.44	3.81	4.44	
	- Behavioral	413	3.38	0.75	−0.47	3.02	3.45	3.89	4.69	
	- Emotional	413	3.54	0.66	0.91	3.15	3.6	4.02	4.62	
	- Cognitive	413	3.27	0.77	0.1	2.81	3.37	3.89	4.31	
United Kingdom	Engagement	314	3.05	0.58	1.09	2.67	3.07	3.39	4.44	
	- Behavioral	314	3.08	0.78	0.23	2.61	3.14	3.59	4.69	
	- Emotional	314	3.05	0.73	0.96	2.59	3.13	3.55	4.62	
	- Cognitive	314	3.02	0.68	0.47	2.61	3.04	3.56	4.31	
United States	Engagement	316	3.04	0.63	0.71	2.71	3.13	3.48	4.44	
	- Behavioral	316	3.15	0.82	0.23	2.73	3.26	3.71	4.69	
	- Emotional	316	2.95	0.76	0.82	2.46	3.04	3.53	4.62	
	- Cognitive	316	2.98	0.76	0.48	2.46	3.06	3.56	4.31	
Finland	Engagement	356	3	0.59	0.39	2.66	3.07	3.4	4.33	
	- Behavioral	356	3.05	0.76	−0.47	2.71	3.17	3.51	4.56	
	- Emotional	356	3.01	0.72	0.91	2.57	3.13	3.56	4.6	
	- Cognitive	356	2.89	0.71	0.1	2.46	2.98	3.35	4.29	
Serbia	Engagement	409	2.94	0.65	1.01	2.51	3.03	3.4	4.44	
	- Behavioral	409	2.84	0.86	0.27	2.25	2.96	3.45	4.69	
	- Emotional	409	2.93	0.79	1.03	2.38	3.01	3.53	4.62	
	- Cognitive	409	3.12	0.75	0.24	2.72	3.25	3.67	4.31	
Taiwan and Macau	Engagement	762	2.83	0.58	0.41	2.43	2.84	3.18	4.44	
	- Behavioral	762	2.93	0.75	−0.46	2.46	2.97	3.41	4.69	
	- Emotional	762	2.68	0.71	0.82	2.16	2.7	3.12	4.62	
	- Cognitive	762	2.76	0.71	0.19	2.29	2.79	3.21	4.31	
Italy	Engagement	411	3.11	0.5	1.07	2.84	3.17	3.48	4.35	
	- Behavioral	411	3	0.7	0.25	2.6	3.06	3.44	4.56	
	- Emotional	411	3.25	0.59	1.05	2.87	3.31	3.66	4.56	
	- Cognitive	411	3.15	0.59	0.93	2.79	3.21	3.53	4.3	

mean engagement of 3.37, Italy had a mean of 3.11, the United Kingdom had a mean of 3.05, the United States had a mean of 3.04, Portugal and Finland had means of 3.00, and Serbia had a mean of 2.94. Lowest mean engagement values were observed in Taiwan and Macau (2.83) and Brazil (2.82). However, the reader should refrain from comparing the means between all countries/participants since no scalar invariance was observed (not even partial scalar invariance; data not shown) and thus two mean values that are equal in value can have distinct interpretations. Therefore, comparing mean scores is only valid for the pairs of countries which displayed scalar invariance.

DISCUSSION

Engagement in university life has proven to be a determinant for learning, academic success, reduce dropout, and promote individual and social well-being (Klem and Connell, 2004; Wonglorsaichon et al., 2014). The measurement of engagement has emerged from the organizational and workplace framework (Schaufeli et al., 2002), but its importance in other activities, like studying, has led to the expansion of the construct and the development of measurement instruments for the school and university context (see, e.g., Appleton et al., 2008; Reeve and Tseng, 2011; Sinatra et al., 2015). In this paper, we report the psychometric properties of engagement data collected with the USEI (Maroco et al., 2016) in higher education systems from nine countries and regions from four continents.

Item sensitivity analysis revealed that the psychometric sensitivity for the 15 items composing the USEI was adequate (**Table 1**). Further CFA showed that the USEI presented adequate evidence of factorial validity, with goodness-of-fit indices indicating a very good fit of the second-order factorial engagement structure to the data from the nine participant countries/regions. Engagement, as a second-order construct presented high loading values for the first-order behavioral engagement and emotional factors and some-how lower, but still medium for the cognitive factor (**Figure 1**). Reliability, as evaluated by internal consistency measures, was quite high for the emotional and cognitive factors and medium for the behavioral factor (**Table 3**). The convergent validity evidence was satisfactory for the cognitive and emotional factors, but low for the behavioral factor. The discriminant validity evidence was appropriate for the emotional and cognitive factors according to the Fornell–Larcker criterion and appropriate for all factors according to the HTMT criterion. These results show that although the three first-order factors of engagement (Cognitive, Emotional, and Behavioral) are strongly correlated, they do measure specific factors of engagement (**Table 4**). Taken together, these results indicate that the USEI presents adequate internal structure validity with data from higher education systems in countries/regions as diverse as the United States, Taiwan and Macau SAR, Finland, Brazil, Servia, Portugal, Italy, and Mozambique. The three-factor scores of the USEI are valid and reliable measures that can be combined to form a reliable total score of academic engagement. These results are in accordance with previous findings of Portuguese students (Costa et al., 2014; Maroco et al., 2016; Sinval et al., 2018) and with our first hypothesis (H1) with students from nine different countries and regions.

With regards to measurement invariance, strong measurement invariance was found for gender and the four areas (Social Sciences, Exact Sciences, Biological Sciences, and Health Sciences). With regards to measurement invariance between countries/regions, we found evidence of strong measurement invariance between Portugal and Brazil and between the United Kingdom and the United States. The remaining combination of countries achieved only weak measurement invariance (**Table 7**). These results indicate that

the USEI's mean scores can be directly compared between genders and between areas of study within countries/regions, but not across all accessed countries/regions. This result partially confirms our second hypothesis (H2).

Because weak measurement invariance between participating countries/regions was found, it is possible to compare regression models of USEI scores on criterion variables between different countries/participants. We, therefore, investigated the USEI evidence of predictive criterion validity. The USEI can significantly predict dropout intention, academic performance, course approval rate, and course expectations as well as burnout scores (see, e.g., Maslach and Leiter, 1997). Most strikingly, USEI scores shared almost half of their variance with the burnout scores, and can explain a quarter of the variability of subjective academic performance and dropout intention. These results indicate that the USEI scores are significantly related to other aspects of academic life and can be used to make reasonable predictions about students' academic success and intention to drop out, therefore confirming our third hypothesis (H3).

The USEI generated data with adequate psychometric characteristics that make it an instrument that produces valid and reliable scores to access student engagement in the university context and its behavioral, emotional, and cognitive dimensions. Other engagement scales that measure engagement, such as the UWES (Schaufeli and Bakker, 2004), have suffered several criticisms ranging from the construct definitions and dimensionality to their application to university students (Lanasa et al., 2009; Wefald and Downey, 2009; Campbell and Cabrera, 2011; Mills et al., 2012). Our results support the adequacy of the USEI to measure student engagement in the university context as opposed to the organizational context (Wefald and Downey, 2009; Upadaya and Salmela-Aro, 2012) or the elementary student's context (Fredricks et al., 2011). Although psychometric analysis showed adequate psychometric qualities of data gathered with the USEI on a diversity of higher education systems, there is still room for improvement. One issue with the conceptualization of student engagement as behavioral, emotional, and cognitive factors is that the behavioral aspect of student engagement dominates the variance attributed to the USEI's global score. The high structural coefficient from the second-order engagement to the behavioral first-order factor contrasts with its reduced internal consistency reliability and AVE. When analyzing the behavioral factor item-by-item we found that item 2 (I follow the school's rules) and item 3 (I usually do my homework on time) somehow produces low factor loadings ($0.4 < \lambda < 0.5$), which explain the reduced internal consistency and AVE of this factor. Item 2 also suffered from a ceiling effect, having the highest absolute skewness of all items on the scale. Item 3 refers to homework and may not have the same meaning across different courses and education systems as expressed by some students that commented on the appropriateness of this item for their university experience. The high structural coefficient value of engagement on the behavioral factor shows that this factor can be more important for the global score than the emotional and cognitive factors. If this is the case, conceptualizing sub-types of behavioral engagement could prove to be useful. Future research can assess if the behavioral factor benefits from additional items

or item rephrasing to better specify all the different behaviors associated with academic engagement.

The emotional and cognitive factors can also be improved, as items 6 and 11 have low factor loadings ($0.4 < \lambda < 0.5$). Item 6 has previously been identified as a problematic item in this scale as it is the only reverse-coded item (Sinval et al., 2018). It can be positively worded as "I feel very accomplished at this school" because reversed items may have reduced sensitivity (as demonstrated by Maroco et al., 2014) for the efficacy dimension of the MBI-SS. Item 11 ("When I read a book, I question myself to make sure I understand the subject I'm reading about") refers to reading a book and may not have the same relevance across different courses or education systems as many students may use a diverse set of media and often read specific chapters of books. A possible solution to this problem is to modify item 11 so that it is not specific to reading a book (e.g., "When I study, I question myself to make sure I understand the subject I'm studying about"). These hypotheses can lead the way to future research and improvement of the USEI.

Limitations and Future Research

Because of the cross-sectional nature of the current study causation should not be inferred from the data. It is important to avoid causal interpretations of the results, as that would require that longitudinal and experimental methods be used. Therefore, causal association between the USEI scores and criterion variables are to be taken with caution. Future studies could consider studying these variables using longitudinal and/or experimental methods.

A second limitation of this study is the self-selection and self-report nature of the data collected that may create a self-selection bias and a social desirability bias (e.g., in the course approval rate or the self-rated academic performance). Future research could improve on these methods by gathering data in a more systematic manner from the official students' records and by using objective criterion variables. Further research should also look at the student engagement predictors and consequences paying special attention to students' academic performance, health, and well-being.

CONCLUSION

This study shows that the USEI with students from nine different countries/regions from Europe, North and South America, Africa, and Asia can be used to collect valid and reliable data on student engagement. It shows the stability of the second-order factor structure observed previously with Portuguese students (Costa et al., 2014; Maroco et al., 2016; Sinval et al., 2018). Metric (weak) invariance was found between countries and scalar (strong) invariance was found between Portugal and Brazil and between the United Kingdom and the United States. Furthermore, the USEI shows strong measurement invariance between genders and between areas of study. USEI scores can confidently be used to predict students' academic performance and other academic-related variables.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The study was properly validated by the ISPA-IU's Ethics Commission (Process: I/017/02/2019) and the Northern Illinois University International Review Board (decision 1504/2014; FWAA00004025). Informed consent was obtained from all individual participants in the study. All procedures performed in studies involving human participants were in accordance with the ethical standards of the Institutional Ethics Research Committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

AUTHOR CONTRIBUTIONS

HA collected the data, did the data analysis, and wrote the first draft of the manuscript. S-WL, P-SS, K-CC, HH-L, BM, II, TS, JC, GE, and FF collected the data and revised the manuscript critically for important intellectual content. JM planned and coordinated the project, did the data analysis, and co-wrote the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02796/full#supplementary-material

REFERENCES

Alrashidi, O., Phan, H. P., and Ngu, B. H. (2016). Academic engagement: an overview of its definitions, dimensions, and major conceptualisations. *Int. Educ. Stud.* 9, 41–52. doi: 10.5539/ies.v9n12p41

Appleton, J. J., Christenson, S. L., and Furlong, M. J. (2008). Student engagement with school: critical conceptual and methodological issues of the construct. *Psychol. Sch.* 45, 369–386. doi: 10.1002/pits.20303

Archambault, I., Janosz, M., Morizot, J., and Pagani, L. (2009). Adolescent behavioral, affective, and cognitive engagement in school: relationship to dropout. *J. Sch. Health* 79, 408–415. doi: 10.1111/j.1746-1561.2009.00428.x

Audas, R., and Douglas Willms, J. (2001). *Applied Research Branch Strategic Policy Human Resources Development Canada Engagement and Dropping Out of School: A Life-Course Perspective.* Available at: http://www.hrdc-drhc.gc.ca/arb (accessed July 2, 2019).

Bentler, P. M. (1988). "Causal modeling via structural equation systems," in *Handbook of Multivariate Experimental Psychology*, ed. R. Cattell, (Boston, MA: Springer), 317–335. doi: 10.1007/978-1-4613-0893-5_9

Bresó, E., Schaufeli, W. B., and Salanova, M. (2011). Can a self-efficacy-based intervention decrease burnout, increase engagement, and enhance performance? A quasi-experimental study. *High. Educ.* 61, 339–355. doi: 10.1007/s10734-010-9334-6

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research.* New York, NY: Guilford Press.

Carter, C. P., Reschly, A. L., Lovelace, M. D., Appleton, J. J., and Thompson, D. (2012). Measuring student engagement among elementary students: Pilot of the student engagement instrument-elementary version. *Sch. Psychol. Q.* 27, 61–73. doi: 10.1037/a0029229

Campbell, C. M., and Cabrera, A.F. (2011). How sound is NSSE?: investigating the psychometric properties of NSSE at a public, research-extensive institution. *Rev. High. Ed.* 35, 77–103.

Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model. Multidiscip. J.* 9, 233–255. doi: 10.1207/S15328007SEM0902_5

Coetzee, M., and Oosthuizen, R. M. (2012). Students' sense of coherence, study engagement and self-efficacy in relation to their study and employability satisfaction. *J. Psychol. Africa* 22, 315–322. doi: 10.1080/14330237.2012.10820536

Costa, A. R., Araújo, A. M., and Almeida, L. S. (2014). *Envolvimento Académico de Estudantes de Engenharia: Contributos para a Validação Interna e Externa de uma Escala de Avaliação.* Available at: http://www.revistaepsi.com (accessed July 2, 2019).

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334. doi: 10.1007/BF02310555

Elmore, G. M., and Huebner, E. S. (2010). Adolescents' satisfaction with school experiences: Relationships with demographics, attachment relationships, and school engagement behavior. *Psychol. Sch.* 47, 525–537. doi: 10.1002/pits.20488

Finney, S. J., and DiStefano, C. (2013). "Nonnormal and categorical data in structural equation modeling," in *A Second Course in Structural Equation Modeling*, 2nd Edn, eds G. R. Hancock, and R. O. Mueller, (Charlotte, NC: Information Age).

Fiorini, S., Shepard, L., and Ouimet, J. (2014). "Using NSSE to Understand Student Success: A Multi-Year Analysis," in *Proceedings of the 10th National Symposium on Student Retention*, ed. O. K. Norman, (Louisville: The University of Oklahoma), 460–473.

Fornell, C., and Larcker, D. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *J. Mark. Res.* 18, 382–388. doi: 10.1177/002224378101800313

Fredricks, J. A., Blumenfeld, P. C., and Paris, A. H. (2004). School engagement: potential of the concept. State of the evidence. *Rev. Educ. Res.* 74, 59–109. doi: 10.3102/00346543074001059

Fredricks, J., Mccolskey, W., Meli, J., Montrosse, B., Mordica, J., and Mooney, K. (2011). *Measuring Student Engagement in Upper Elementary Through High School: a Description of 21 Instruments.* Available at: http://ies.ed.gov/ncee/edlabs (accessed July 12, 2019).

García-Ros, R., Tomás, J. M., and Fernández, I. (2017). The schoolwork engagement inventory: factorial structure, measurement invariance by gender and educational level, and convergent validity in secondary education (12-18 Years) assessment higher education view project promotion of sustainable development and road safety for cyclists view project. *Artic. J. Psychoeduc. Assess.* 36:073428291668923. doi: 10.1177/0734282916689235

Gilardi, S., and Guglielmetti, C. (2011). University life of non-traditional students: engagement styles and impact on attrition. *J. Higher Educ.* 82, 33–53. doi: 10.1080/00221546.2011.11779084

Green, S. B., and Yang, Y. (2009). Commentary on coefficient alpha: a cautionary tale. *Psychometrika* 74, 121–135. doi: 10.1007/s11336-008-9098-4

Hart, S. R., Stewart, K., and Jimerson, S. R. (2011). The student engagement in schools questionnaire (SESQ) and the teacher engagement report form-new (TERF-N): examining the preliminary evidence. *Contemp. Sch. Psychol.* 15, 67–79.

Henseler, J., Ringle, C. M., and Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J. Acad. Mark. Sci.* 43, 115–135. doi: 10.1007/s11747-014-0403-8

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscip. J.* 6, 1–55. doi: 10.1080/10705519909540118

Janosz, M., Archambault, I., Morizot, J., and Pagani, L. S. (2008). school engagement trajectories and their differential predictive relations to dropout. *J. Soc. Issues* 64, 21–40. doi: 10.1111/j.1540-4560.2008.00546.x

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., and Rosseel, Y. (2018). *semTools: Useful tools for Structural Equation Modeling (R package version 0.4–15.930) [Computer Software].*

Kelley, K. (2016). Confidence intervals for population reliability coefficients: evaluation of methods, recommendations, and software for composite measures. *Psychol. Methods* 21, 69–92. doi: 10.1037/a0040086.supp

Klem, A., and Connell, J. P. (2004). *Relationships Matter: Linking Teacher Support to Student Engagement and Achievement.* Hoboken, NJ: Wiley Online Library.

Kuh, G. D. (2009). The national survey of student engagement: conceptual and empirical foundations. *New Dir. Institutional Res.* 2009, 5–20. doi: 10.1002/ir.283

Kulikowski, K. (2017). Do we all agree on how to measure work engagement? Factorial validity of Utrecht Work Engagement Scale as a standard measurement tool – A literature review. *Int. J. Occup. Med. Environ. Health* 30, 161–175. doi: 10.13075/ijomeh.1896.00947

Lanasa, S. M., Cabrera, A. F., and Trangsrud, H. (2009). The construct validity of student engagement: a confirmatory factor analysis approach. *Res. High. Educ.* 50, 315–332. doi: 10.1007/s11162-009-9123-1

Marôco, J. (2014). *Análise de Equações Estruturais: Fundamentos Teóricos, Software & Aplicações*, 2nd Edn. Pêro Pinheiro: ReportNumber.

Maroco, J., Maroco, A. L., Bonini Campos, J. A. D., and Fredricks, J. A. (2016). University student's engagement: Development of the University Student Engagement Inventory (USEI). *Psicol. Reflex. e Crit.* 29:21. doi: 10.1186/s41155-016-0042-8

Maroco, J., Maroco, A. L., and Campos, J. A. D. B. (2014). Student's Academic efficacy or inefficacy? an example on how to evaluate the psychometric properties of a measuring instrument and evaluate the effects of item wording. *Open J. Stat.* 04, 484–493. doi: 10.4236/ojs.2014.46046

Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *Br. J. Educ. Psychol.* 77, 413–440. doi: 10.1348/000709906X118036

Maslach, C., Jackson, S. E., and Leiter, M. P. (1996). *The Maslach Burnout Inventory Manual.* Hove: Psychology Press, 191–298. doi: 10.1002/job.4030020205

Maslach, C., and Leiter, M. P. (1997). *The Truth About Burnout: How Organizations Cause Personal Stress and What to do About It.* Hoboken, NJ: Wiley.

McDonald, R. P. (2013). *Test Theory.* Hove: Psychology Press, doi: 10.4324/9781410601087

McNamara, A., Arino de la Rubia, E., Zhu, H., Ellis, S., and Quinn, M. (2018). *skimr: Compact and flexible summaries of data. R Package. version, 1.*

Mills, M. J., Culbertson, S. S., and Fullagar, C. J. (2012). Conceptualizing and measuring engagement: an analysis of the utrecht work engagement scale. *J. Happiness Stud.* 13, 519–545.

Millsap, R. E., and Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behav. Res.* 39, 479–515. doi: 10.1207/S15327906MBR3903_4

Mo, Y., Singh, K., and Caskey, M. M. (2008). Parents' relationships and involvement: effects on students' school engagement and performance. *Res. Middle Level Educ.* 31, 1–11. doi: 10.1080/19404476.2008.11462053

Muthén, L. K., and Muthén, B. O. (2017). *MPlus: Statistical Analysis With Latent Variables User's Guide.* Available at: www.StatModel.com (accessed July 15, 2019).

Nystrand, M., and Gamoran, A. (1989). *Instructional Discourse and Student Engagement.* Madison, WI: National Center on Effective Secondary Schools, 43.

Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *Br. J. Math. Stat. Psychol.* 54, 315–323. doi: 10.1348/000711001159582

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* Vienna: R Core Team.

Reeve, J., and Tseng, C. M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemp. Educ. Psychol.* 36, 257–267. doi: 10.1016/j.cedpsych.2011.05.002

Reschly, A. L., and Christenson, S. L. (2012). "Jingle, jangle, and conceptual haziness: evolution and future directions of the engagement construct," in *Handbook of Research on Student Engagement*, ed. L. C. Sandra, (Boston, MA: Springer), 3–19. doi: 10.1007/978-1-4614-2018-7_1

Revelle, W., and Revelle, M. W. (2015). *Package 'psych.'*. Available at: https://personality-project.org/r/psych (accessed July 11, 2019).

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Rutkowski, L., and Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educ. Psychol. Meas.* 74, 31–57. doi: 10.1177/0013164413498257

Salmela-Aro, K., and Upadaya, K. (2012). The schoolwork engagement inventory. *Eur. J. Psychol. Assess.* 28, 60–67. doi: 10.1027/1015-5759/a000091

Salmela-Aro, K., and Upadyaya, K. (2017). Co-development of educational aspirations and academic burnout from adolescence to adulthood in Finland. *Res. Hum. Dev.* 14, 106–121. doi: 10.1080/15427609.2017.1305809

Schaufeli, W. B., Martínez, I. M., Pinto, A. M., Salanova, M., and Bakker, A. B. (2002). Burnout and engagement in university students: a cross-national study. *J. Cross. Cult. Psychol.* 33, 464–481. doi: 10.1177/0022022102033005003

Schaufeli, W., and Bakker, A. (2004). *UWES UTRECHT WORK ENGAGEMENT SCALE Preliminary Manual*. Available at: https://www.wilmarschaufeli.nl/publications/Schaufeli/TestManuals/Test_manual_UWES_English.pdf (accessed July 2, 2019).

Schaufeli, W., Leiter, M., Maslach, C., and Jackson, S. (1996). *MBI-general Survey*. Available at: https://scholar.google.pt/scholar?hl=en&as_sdt=0%2C5&q=Schaufeli%2C+W.+B.%2C+Leiter%2C+M.+P.%2C+Maslach%2C+C.%2C+%26+Jackson%2C+S.+E.+%281996%29.&btnG= (accessed July 2, 2019).

Sheppard, S. L. (2011). School engagement: a "Danse Macabre"? *J. Philos. Educ.* 45, 111–123. doi: 10.1111/j.1467-9752.2010.00782.x

Sinatra, G. M., Heddy, B. C., and Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educ. Psychol.* 50, 1–13. doi: 10.1080/00461520.2014.1002924

Sinval, J., Casanova, J. R., Marôco, J., and Almeida, L. S. (2018). University student engagement inventory (USEI): Psychometric properties. *Curr. Psychol.* 2, 1–13. doi: 10.1007/s12144-018-0082-6

Tuominen-Soini, H., and Salmela-Aro, K. (2014). Schoolwork engagement and burnout among Finnish high school students and young adults: profiles, progressions, and educational outcomes. *Dev. Psychol.* 50, 649–662. doi: 10.1037/a0033898

Vasalampi, K., Katariina, S.-A., and Jari-Erik, N. (2009). Adolescents' self-concordance, school engagement, and burnout predict their educational trajectories. *Eur. Psychol.* 14, 332–341.

Wang, M.-T., and Fredricks, J. A. (2014). The reciprocal links between school engagement, youth problem behaviors, and school dropout during adolescence. *Child Dev.* 85, 722–737. doi: 10.1111/cdev.12138

Wang, M.-T., Willett, J. B., and Eccles, J. S. (2011). The assessment of school engagement: examining dimensionality and measurement invariance by gender and race/ethnicity. *J. Sch. Psychol.* 49, 465–480. doi: 10.1016/j.jsp.2011.04.001

Wefald, A. J., and Downey, R. G. (2009). Construct dimensionality of engagement and its relation with satisfaction. *J. Psychol.* 143, 91–112. doi: 10.3200/JRLP.143.1.91-112

Wonglorsaichon, B., Wongwanich, S., and Wiratchai, N. (2014). *The Influence of Students School Engagement on Learning Achievement: A Structural Equation Modeling Analysis*. Amsterdam: Elsevier.

Wu, H., and Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika* 81, 1014–1045. doi: 10.1007/s11336-016-9506-0