Journal of Psychoeducational Assessment
http://jpa.sagepub.com/

Methodological Measurement Fruitfulness of Exploratory Structural Equation Modeling (ESEM): New Approaches to Key Substantive Issues in Motivation and Engagement

Herbert W. Marsh, Gregory Arief D. Liem, Andrew J. Martin, Alexandre J. S. Morin and Benjamin Nagengast

Journal of Psychoeducational Assessment 2011 29: 322 originally published online 28 June 2011
DOI: 10.1177/0734282911406657

Journal of Psychoeducational Assessment
29(4) 322–346

DOI: 10.1177/0734282911406657
http://jpa.sagepub.com

Methodological Measurement Fruitfulness of Exploratory Structural Equation Modeling (ESEM): New Approaches to Key Substantive Issues in Motivation and Engagement

Herbert W. Marsh[1,2], Gregory Arief D. Liem[1], Andrew J. Martin[1], Alexandre J. S. Morin[1], and Benjamin Nagengast[3]

Abstract

The most popular measures of multidimensional constructs typically fail to meet standards of good measurement: goodness of fit, measurement invariance, lack of differential item functioning, and well-differentiated factors that are not so highly correlated as to detract from their discriminant validity. Part of the problem, the authors argue, is undue reliance on overly restrictive independent cluster models of confirmatory factor analysis (ICM-CFA) in which each item loads on one, and only one, factor. Here the authors demonstrate exploratory structural equation modeling (ESEM), an integration of the best aspects of CFA and traditional exploratory factor analyses (EFA). On the basis of responses to the 11-factor Motivation and Engagement Scale (n = 7,420, M_{age} = 14.22), we demonstrate that ESEM fits the data much better and results in substantially more differentiated (less correlated) factors than corresponding CFA models. Guided by a 13-model taxonomy of ESEM full-measurement (mean structure) invariance, the authors then demonstrate invariance of factor loadings, item intercepts, item uniquenesses, and factor variances-covariances, across gender and over time. ESEM has broad applicability to other areas of research that cannot be appropriately addressed with either traditional EFA or CFA and should become a standard tool for use in psychometric tests of psychological assessment instruments.

Keywords

Factor analysis, Motivation and Engagement; measurement invariance; differential item functioning

Exploratory and confirmatory factor analysis (EFA and CFA) are widely recommended and used in the development and refinement of psychoeducational assessment instruments (e.g., Floyd &

[1]University of Sydney, Australia
[2]King Saud University, Saudi Arabia
[3]University of Oxford, UK

Corresponding Author:
Herbert M. Marsh, Department of Education, University of Oxford, Oxford OX2 6PY, UK
Email: herb.marsh@education.ox.ac.uk

Widaman, 1995; Reise, Waller, & Comrey, 2000). Historically, psychoeducational researchers have relied on EFAs to identify key constructs. However, assumptions of measurement invariance (in relation to multiple groups, time, and covariates) underpinning psychoeducational research cannot be appropriately evaluated with traditional EFAs. Moreover, in many instances, item-level CFAs failed to provide clear support for instruments that apparently were well established in EFA research (Marsh et al., 2010, 2009). Common problems are typically associated with goodness of fit, differentiation of factors, measurement invariance across groups or time, and differential item functioning (Marsh et al., 2010, 2009). We argue that part of the problem is linked to the overly restrictive independent cluster models of CFA (ICM-CFA) in which items are required to load on one, and only one, factor, with nontarget loadings constrained to be zero. Although there are many methodological and strategic advantages to ICM-CFAs, they are sometimes inappropriate and many strategies used to compensate for this inappropriateness tend to be dubious, misleading, or simply wrong (see further discussion by Marsh et al., 2010, 2009). Furthermore, requiring nontarget loadings to be zero typically leads to inflated factor correlations that might lead to biased estimates in structural equation models (SEMs) incorporating other variables (Asparouhov & Muthén, 2009; Marsh et al., 2010, 2009). This seems to be the case for motivation and engagement measures especially if a large number of latent factors and items are involved in the analysis (Martin, 2007, 2009). Here we outline a new approach called exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009; Marsh et al., 2010, 2009)—an integration of EFA, CFA, and SEM—which has the potential to resolve this dilemma and has wide applicability to all disciplines that are based on the measurement of latent constructs.

We propose that ESEM solutions based on responses to most multifactor psychological instruments will result in a substantially better fit to the data and substantially less correlated factors than corresponding ICM-CFA solutions, while still retaining much of the power and flexibility normally associated with CFA. The present study aims to demonstrate the fruitfulness of ESEM to substantively important issues in motivation and engagement research, with potential applicability to other psychoeducational research based on multidimensional constructs. Within the ESEM framework, the researcher has access to typical CFA/SEM parameters and statistical advances: standard errors; goodness-of-fit statistics; comparisons of competing models through tests of statistical significance and fit indices; inclusion of correlated residuals; inclusion of both CFA and EFA factors based on the same, different, or overlapping sets of items; estimation of method effects and bifactor models; multiple-indicators multiple-causes models (MIMIC); multiple-group and longitudinal invariance analyses; growth modeling (for illustrations of a diverse range of ESEM models, see Asparouhov & Muthén, 2009; Marsh et al., 2010, 2009). Guided by a taxonomy of full-measurement invariance models (Marsh et al., 2010, 2009), we apply ESEM approaches to testing factorial/measurement invariance (factor loadings, factor variances and covariances, uniquenesses, intercepts, and true latent means) of the Motivation and Engagement Scale (MES; Martin, 2007, 2009) across multiple groups (gender) and over time (i.e., in a longitudinal study).

Multidimensional Motivation and Engagement

Martin (2007, 2009) developed a theoretical integrative multidimensional framework for understanding cognitive and behavioral components of academic motivation and engagement. This 11-factor framework (Figure 1) was developed in response to calls (e.g., Pintrich, 2003) for the development of more integrative, yet instrumental, approaches to motivation and engagement research and theorizing.

Of particular relevance to the present investigation, research into the validity of the MES also involved the examination of factor correlations using a conservative ICM-CFA approach. Martin

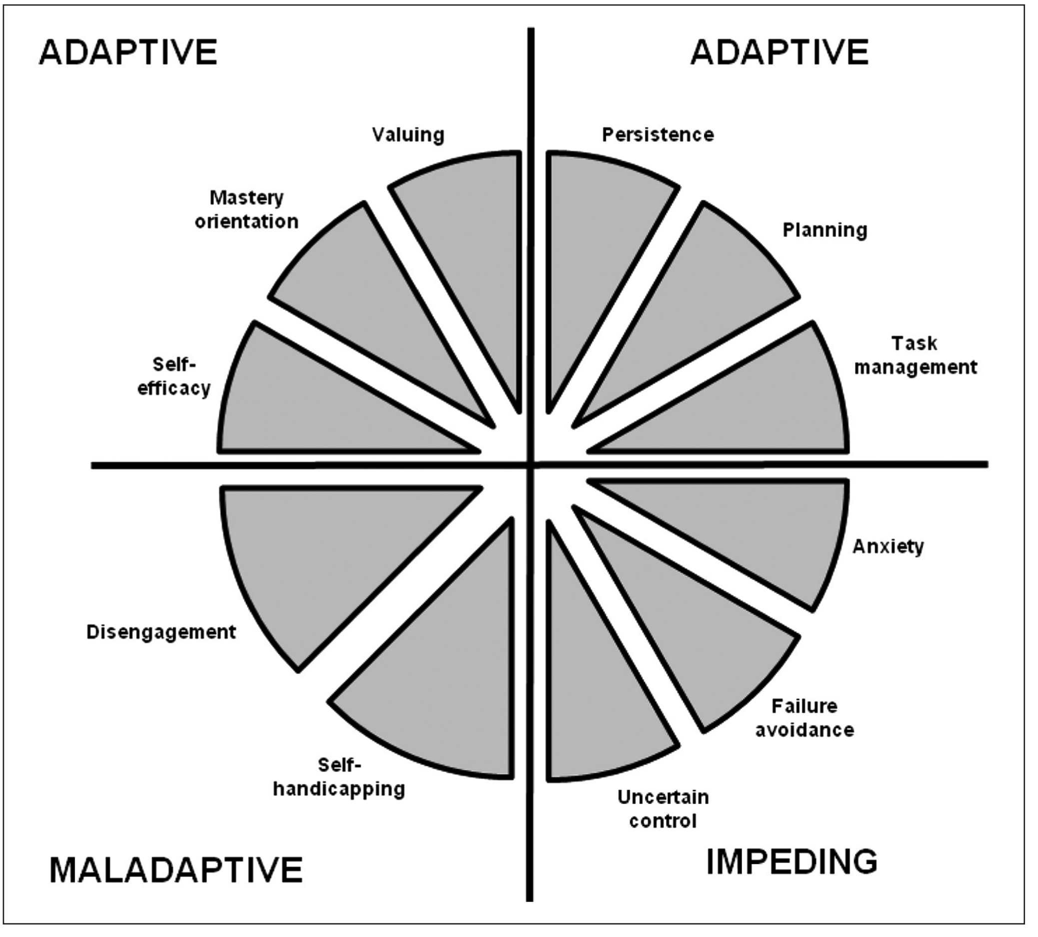

Figure 1. Motivation and engagement wheel
Source: Reproduced with permission from Martin, A. J. (2010). *Building Classroom Success: Eliminating Academic Fear and Failure*. London: Continuum.

(2007, 2009) reported that the correlations among factors were relatively high. Hence, though there is sufficient evidence for the multidimensionality and distinctiveness of the MES motivation and engagement factors, the MES factors based on ICM-CFAs typically show elevated correlations (Martin, 2007, 2009). Recent simulations studies designed in part to compare ESEM and ICM-CFA solutions showed that ICM-CFA solutions tended to result in inflated factor correlations (Asparouhov & Muthén, 2009; Marsh et al., 2009; 2010) and, thus, that the observed elevated correlations between the MES factors may in fact be inflated due to the highly restrictive ICM-CFA assumptions. This situation could lead to serious multicollinearity problems and suppression effects, obscuring the ability of individual MES factors in predicting outcome variables.

Furthermore, key substantive issues in psychoeducational research involve examining mean differences in gender or age in motivation and engagement factors (Halpern, 2006). MES studies have shown that girls typically have more adaptive levels of motivation and engagement than boys do and that student motivation and engagement typically decline after the middle school transition (Martin, 2007, 2009). In addressing these substantive issues, tests of the invariance of MES factors have commonly focused on the invariance of factor loadings, item uniquenesses,

Table 1. Taxonomy of Multiple Group Tests of Invariance Testable With ESEM

Model	Parameters constrained to be invariant
1	None (configural invariance)
2	FL [1] (weak factorial/measurement invariance)
3	FL Uniq [1, 2]
4	FL, FVCV [1, 2]
5	FL, Inter [1, 2] (strong factorial/measurement invariance)
6	FL, Uniq, FVCV [1, 2, 3, 4]
7	FL, Uniq, Inter [1, 2, 3, 5] (strict factorial/measurement invariance)
8	FL, FVCV, Inter [1, 2, 4, 5]
9	FL, Uniq, FVCV, Inter [1-8]
10	FL, Inter, FMn [1, 2, 5] (latent mean invariance)
11	FL, Uniq, Inter, FMn [1, 2, 3, 5, 7, 10] (manifest mean invariance)
12	FL, FVCV, Inter, FMn [1, 2, 4, 5, 6, 8, 10]
13	FL, Uniq, FVCV, Inter, FMn [1-12] (complete factorial invariance)

Note: FL = factor loadings; FVCV = factor variance-covariances; Inter = item intercepts; Uniq = item uniquenesses; FMn = factor means. Models with latent factor means freely estimated constrain intercepts to be invariant across groups, whereas models where intercepts are free imply that mean differences are a function of intercept differences. Values in brackets values represent nesting relations in which the estimated parameters of the less general model are a subset of the parameters estimated in the more general model under which it is nested. All models are nested under Model 1 (with no invariance constraints), whereas Model 13 (complete invariance) is nested under all other models. Table 1 was adapted from Marsh et al. (2009)

and variances-covariances (see Martin, 2007, 2009) but not of item intercepts (strong and strict factorial/measurement invariance) and latent factor means, and thus have precluded tests of differential item functioning (Marsh et al., 2010, 2009). Evidence for the invariance of the MES item intercepts is important if we wish to ascertain that mean differences based on a latent construct (e.g., self-belief) are not an artifact of differential item functioning. In addition, though all of these tests are available in the context of traditional ICM-CFAs as well as new ESEM approaches presented here, they were not available for EFA before the advent of ESEM. Hence, guided by the taxonomy of factorial/measurement invariance models proposed by Marsh et al. (2010, 2009) and presented in Table 1, we systematically test the invariance of all MES parameters across gender and over time.

The Present Investigation

Substantively important questions in relation to motivation and engagement research, as assessed by the MES instrument (with broad applicability to other psychoeducational assessment research), are addressed in the present study with ESEM. The present investigation is a substantive-methodological synergy that explores the utility of ESEM in testing (a) the invariance of the factor structure (factor loadings, variances-covariances), (b) the invariance of other measurement parameters (item uniquenesses, item intercepts, and latent factor means), and (c) its potential for reducing otherwise high correlations that can pose problems of multicollinearity and threats to support for discriminant validity. These issues cannot be appropriately addressed with either traditional EFA or CFA approaches and so this study offers methodological insights to those interested in measurement and substantive insights in relation to diverse psychological assessment instruments.

Method

Participants

The total sample comprised 7,420 students (M_{age} = 14.22, SD_{age} = 1.56; 45% girls). The responses, collected as part of the MES archive, are based on a broad cross-section of independent and government schools located in urban and regional areas in different Australian states. Teachers administered the MES (Martin, 2007) to students during class/pastoral care. After the rating scale was explained, students completed the instrument on their own and returned the completed form to teachers at the end of the period. Portions of the data have been reported elsewhere with a more substantial construct validity focus (see Martin, 2007, 2009). Of the total sample, 1,866 (38.8% girls) students were part of studies in which test–retest data were collected approximately 1 year later. The average age of the participants was 13.86 years (*SD* = 1.28) at Time 1.

Motivation and Engagement Scale

MES version designed for high school students (MES-HS) comprises 11 motivation and engagement scales, each assessed by 4 items on a 7-point response scale: 1 (*strongly disagree*) to 7 (*strongly agree*). The 11-factor structure is supported by many CFA studies (e.g., Martin, 2007, 2009) as well as in the present investigation. A summary of the 11 scales, Cronbach's alphas estimates of reliability, and sample items are provided in Appendix.

The ESEM Approach

Here we briefly summarize selected ESEM features relevant to the present investigation. All analyses in the present investigation were done with Mplus 5.2 (Muthén & Muthén, 2008), using robust maximum likelihood estimator (MLR) with standard errors and tests of fit that are robust in relation to nonnormality of observations and the use of categorical variables when there are at least four or more response categories (e.g., Beauducel & Herzberg, 2006; DiStefano, 2002; Dolan, 1994; Muthén & Kaplan, 1985; Rhemtulla, Brosseau-Liard, & Savalei, 2010).

In a basic version of the ESEM model, all parameters can be identified with the maximum likelihood (ML) estimation method or robust alternatives. However, when more than one factor ($m > 1$) is posited, further constraints are required to achieve an identified solution (Asparouhov & Muthén, 2009; Marsh et al., 2009). In the first step, an unconstrained factor structure (Jöreskog & Sörbom, 1979) is estimated—potentially restricted to be equal across different groups or time points. Given the need to estimate all loadings, a total of m^2 constraints have to be built into the model to achieve identification. These constraints are generally implemented by specifying the factor variance-covariance matrix as an identity matrix and constraining factor loadings in the right upper corner of the factor-loading matrix to 0 (for the ith factor, i-1 factor loadings are restricted to 0). In the second step, this initial, unrotated solution is rotated using a wide set of orthogonal and oblique rotation procedures (see Asparouhov & Muthén, 2009; Sass & Schmitt, 2010) again restraining parameters of the rotated solution to be equal across groups or time points when appropriate. Regarding the ESEM mean structure, the identification is similar to regular CFA: All items intercepts are freely estimated, and all latent factor means are constrained to 0. However, due to the rotational issues, the alternative CFA method of constraining one intercept per factor to 0 to freely estimate the latent means is unavailable in ESEM. All of these constraints are built as default in the Mplus estimation process so that the analyst does not need to worry about them. In addition, Mplus uses multiple random starting values in the estimation process to protect against nonconvergence and local minimums in the rotation algorithms. For a more detailed presentation of the identification and estimation issues in ESEM, the interested readers are referred to Asparouhov and Muthén, Marsh et al. (2010, 2009), and Sass and Schmitt. Although a wide variety of orthogonal and oblique rotation procedures are available, the choice of the most appropriate procedure is to some extent still an open research area. Following Marsh et al. (2010, 2009) we used an oblique geomin rotation (the default in Mplus) with an epsilon value of 0.5 and the full-information MLR estimator to correct for small amounts of missing data (due to the nature of data, the amount of missing data was very small—less than 1%). More detailed descriptions of alternative rotation procedures are available elsewhere (Asparouhov & Muthén, 2009; Marsh et al., 2010, 2009; Sass & Schmitt, 2010)

Test of factorial and measurement invariance. Of particular substantive importance for psychoeducational assessment research are mean-level differences across multiple groups (e.g., male vs. female, age groups, single-sex vs. coeducational schools) or over time (i.e., observing the same group of participants at multiple occasions, perhaps before and after an intervention). Tests of whether the underlying factor structure is the same in the different groups or for multiple occasions have typically been ignored in such studies. In particular, an important assumption in such comparisons is the invariance of item intercepts and problems associated with differential item functioning. For example, if gender or longitudinal differences vary substantially for different items used to infer this construct in a manner that is unrelated to respondents' true levels on the latent constructs, then the observed differences might be idiosyncratic to the particular items used. From this perspective, it is important to be able to evaluate the full measurement invariance of responses.

The evaluation of model invariance over different groups (e.g., gender) or over time for the same group is widely applied in SEM studies (Jöreskog & Sörbom, 1979; Meredith & Teresi, 2006). Indeed, such tests of invariance might be seen as a fundamental advantage of CFA/SEM over EFA. Although related multiple-group methods have been proposed in EFA settings (e.g., Meredith, 1964), they mainly focus on the similarity of factor patterns. However, ESEM model are easily extended to multiple-group test of the invariance. Marsh et al. (2010, 2009) operationalized a taxonomy of 13 partially nested models varying from the least restrictive model of configural invariance with no invariance constraints to a model of complete invariance positing strict invariance as well as the invariance of the latent means and of the factor variance-covariance matrix (Table 1). This taxonomy begins with a model with no invariance of any parameters or *configural invariance* (Model 1). The initial focus is on the invariance of the factor loadings (Model 2)—sometimes referred to as weak measurement invariance or pattern invariance—which requires that factor loadings be invariant over groups or over time. *Strong measurement invariance* (Model 5) requires that the indicator intercepts and factor loadings are invariant over groups or time and justifies comparison of latent means. *Strict measurement invariance* (Model 7) requires in invariance of item uniquenesses (addition to invariant factor loadings and intercepts) and justifies the comparison of manifest means over groups or time. Although we have only emphasized a few models, other models provide an important basis of comparison for evaluating these key models. Although the order of the models follows a logical progression, the order in which the models are used is not critical. However, the nesting relations between models is important in that it is particularly relevant to compare each of the key models with alternative models that are nested under it (see Table 1).

Goodness of fit. In applied CFA/SEM research, there is a predominant focus on indices that are sample size independent (e.g., Marsh, Balla, & Hau, 1996; Marsh, Hau, & Grayson, 2005; Marsh, Hau, & Wen, 2004) such as the root mean square error of approximation (RMSEA), the Tucker-Lewis Index (TLI), and the Comparative Fit Index (CFI)—as well as the χ^2 test statistic and an evaluation of parameter estimates. The TLI and CFI vary along a 0-to-1 continuum, and values greater than 0.90 and 0.95 typically reflect acceptable and excellent fit to the data. RMSEA values of less than 0.05 and 0.08 reflect a close fit and a reasonable fit to the data, respectively. However, for purposes of model comparison, comparison of the relative fit of models imposing more or fewer invariance constraints is more important than the absolute level of fit for any one model—so long as the fit of the best-fitting model is acceptable. Cheung and Rensvold (2001) and Chen (2007) suggested that if the decrease in fit for the more parsimonious model is less than 0.01 for incremental-fit indices like the CFI, then there is reasonable support for the more parsimonious model. Chen suggests that when the RMSEA increases by less than 0.015 there is support for the more constrained model. For indices that incorporate a penalty of lack of parsimony, it is also possible for a more restrictive model to result in a better fit than a less restrictive model. Although we rely on these guidelines in the present investigation, it is important to emphasize that these are only rough guidelines (Marsh et al., 2005) and that more research is needed on their appropriateness for ESEM studies for which the number of estimated parameters is typically substantially greater than the typical ICM-CFA study. In the meantime, we suggest that applied researchers use an eclectic approach based on a subjective integration of a variety of different indices, including the chi square, detailed evaluations of the actual parameter estimates in relation to theory, a priori predictions, common sense, and a comparison of viable alternative models specifically designed to evaluate goodness of fit in relation to key issues. This is consistent with the approach we used here.

Results

MES Factor Structures and Correlations: ESEM Versus CFA

The critical starting point for the present investigation is the hypothesis that the ESEM model provides a better fit to responses to the MES items than a traditional ICM-CFA model and that it reduces the size of the typically large factor correlations. The ICM-CFA solution provides an acceptable fit to the data (CFI = 0.935, TLI = 0.928, RMSEA = 0.033; see TG1CFA in Table 2). However, the corresponding ESEM solution fits the data even better (CFI = 0.977, TLI = 0.958, RMSEA = 0.025; see TG1ESEM in Table 2). This indicates that the highly restrictive ICM-CFA model in which all nontarget loadings are constrained to be zero is, apparently, overly restrictive even though it is also much more parsimonious.

We then proceeded with an evaluation of parameter estimates in the ESEM and ICM-CFA solutions. In terms of the 44 target factor loadings (the values indicated in bold face in Table 3), the sizes of most loadings are substantial. When both target and nontarget factor loadings are considered together, the ICM-CFA and ESEM solutions resulted in a very similar pattern with a profile similarity index = 0.925 (PSI, the correlation between the 484 ESEM factor loadings and the corresponding CFA values), suggesting that the ESEM and ICM-CFA factor loadings were highly related.

An evaluation of the factor correlations among the MES factors demonstrates a critical advantage of the ESEM approach over the ICM-CFA approach. Although patterns of correlations between the ESEM and CFA solutions are similar (PSI = 0.803), the ICM-CFA factor correlations (–.70 to .77, |M| = 0.40, SD = 0.22) tend to be systematically larger than the ESEM factor correlations (–.33 to .38; |M| = 0.17, SD = 0.11). Thus, for example, the negative correlation between valuing of school and disengagement is r = –.70 based on the ICM-CFA solution but r = –.33 for the ESEM solution. Similarly, the correlation between planning and study management is r = .77 for the ICM-CFA solution but r = .38 for the ESEM solution. These show that the 11 motivation and engagement factors are substantially more distinct in the ESEM solution than in the CFA solution. In summary, the ESEM model is clearly better in terms of goodness of fit. In particular, even though the ICM-CFA model fits the data reasonably well, the requirement that nontarget loadings are constrained to be zero is overly restrictive and results in systematically inflated correlations among the factors. Hence, the superiority of the ESEM model has practical implications in relation to the usefulness of the factors.

Table 2. Summary of Goodness-of-Fit Statistics for All Models

Model	χ^2/df	NFParm	CFI	TLI	RMSEA	AIC	BIC	corBIC	Description
Total group (TG) models									
TG1 CFA	7609.974/847	187	0.935	0.928	0.033	106,064	106,192	106,133	Total group CFA—11F corr
TG1 ESEM	2913.351/517	517	0.977	0.958	0.025	1,055,246	1,058,818	1,057,175	Total group ESEM—11F corr
Multiple (two) group invariance (MGI—gender)									
MGI1	3476.910/1034	1034	0.977	0.955	0.025	1,050,228	1,057,370	1,054,084	IN=none
MGI2	3889.021/1397	671	0.976	0.968	0.022	1,050,085	1,054,719	1,052,587	IN=FL
MGI3	4566.331/1441	627	0.970	0.961	0.024	1,050,896	1,055,220	1,053,231	IN=FL, Uniq
MGI4	4014.696/1463	605	0.976	0.969	0.022	1,050,113	1,054,292	1,052,370	IN=FL, FVCV
MGI5	4137.989/1430	638	0.974	0.966	0.023	1,050,302	1,054,709	1,052,682	IN=FL, Inter
MGI6	4707.130/1507	561	0.970	0.962	0.024	1,050,959	1,054,827	1,053,047	IN=FL, Uniq, FVCV
MGI7	4809.795/1474	594	0.968	0.959	0.025	1,051,101	1,055,197	1,053,313	IN=FL, Uniq, Inter
MGI8	4273.261/1496	572	0.974	0.967	0.022	1,050,332	1,054,283	1,052,466	IN=FL, FVCV, Inter
MGI9	4948.282/1540	528	0.968	0.960	0.024	1,051,167	1,054,807	1,053,132	IN=FL, FVCV, Inter, Uniq
MGI10	4464.223/1441	627	0.971	0.962	0.024	1,050,680	1,055,010	1,053,018	IN=FL, Inter, FMn
MGI11	5126.102/1485	583	0.966	0.956	0.026	1,051,478	1,055,498	1,053,649	IN=FL, Uniq, Inter, FMn
MGI12	4585.490/1506	562	0.971	0.963	0.024	1,050,711	1,054,586	1,052,804	IN=FL, FVCV, Inter, FMns
MGI13	5267.524/1551	517	0.965	0.957	0.025	1,051,542	1,055,107	1,053,467	IN=FL, FVCV, Inter, Uniq, FMn
Longitudinal (two) wave invariance (LGI—test–retest)									
LGI0	6028.961/2849	1155	0.951	0.934	0.024	519,661	526,050	522,381	IN=none, no CWCUs
LGI1	4717.932/2805	1199	0.970	0.960	0.019	518,169	524,802	520,993	IN=none
LGI2	5130.141/3168	836	0.970	0.963	0.018	518,037	522,661	520,005	IN=FL
LGI3	5263.982/3212	792	0.968	0.962	0.019	518,150	522,531	520,014	IN=FL, Uniq
LGI4	5233.379/3234	770	0.969	0.963	0.018	518,044	522,303	519,857	IN=FL, FVCV
LGI5	5235.473/3201	803	0.969	0.962	0.018	518,084	522,525	519,974	IN=FL, Inter

(continued)

Table 2. (continued)

Model	χ^2/df	NFParm	CFI	TLI	RMSEA	AIC	BIC	corBIC	Description
LGI6	5360.394/3278	726	0.968	0.962	0.018	518,146	522,162	519,855	IN=FL, Uniq, FVCV
LGI7	5366.634/3245	759	0.967	0.961	0.019	518,196	522,394	519,983	IN=FL, Uniq, Inter
LGI8	5338.006/3267	737	0.968	0.963	0.018	518,091	522,167	519,826	IN=FL, FVCV, Inter
LGI9	5461.255/3311	693	0.967	0.962	0.019	518,192	522,026	519,824	IN=FL, FVCV, Inter, Uniq
LGI10	5361.592/3212	792	0.967	0.960	0.019	518,209	522,590	520,074	IN=FL, Inter, FMn
LGI11	5493.346/3256	748	0.965	0.959	0.019	518,322	522,459	520,083	IN=FL, Uniq, Inter, FMn
LGI12	5463.366/3278	726	0.966	0.961	0.019	518,216	522,232	519,926	IN=FL, FVCV, Inter, FMns
LGI13	5587.409/3322	682	0.965	0.960	0.019	518,318	522,090	519,924	IN=FL, FVCV, Inter, Uniq, FMn

Note: χ^2/DF = chi square/degrees of freedom ratio; CFI= comparative fit index; TLI = Tucker–Lewis Index; NFParm = number of free parameters; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; corBIC = sample-size-adjusted BIC; RMSEA= root mean square error of approximation. For multiple-group invariance models, the "IN=" means the sets of paramaters constrained to be invariant across the multiple groups: FL= factor loadings; FVCV = factor variance-covariances; Inter = item intercepts; Uniq = item uniquenesses; FMn = factor means; CWCUs = cross-wave correlated uniquenesses.

Table 3. ESEM (Exploratory Structural Equation Modeling) and CFA (Confirmatory Factor Analysis) Solutions: 11 Motivation and Engagement Factors Based on Responses to 44 MES Items

Factors	ESEMfactor loadings											CFAfactor loadings
Items	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	
S-efficacy (F1)												
Eff1	**0.56**	0.07	0.11	0.04	0.04	0.03	−0.02	−0.05	−0.00	0.01	−0.06	**0.71**
BEff2	**0.51**	0.04	0.05	0.04	−0.02	0.17	−0.05	−0.07	0.00	−0.00	−0.06	**0.67**
Eff3	**0.45**	0.08	0.12	0.01	0.10	0.06	0.06	0.00	−0.01	−0.02	−0.01	**0.61**
Eff4	**0.61**	0.13	0.03	0.04	0.04	0.05	−0.00	−0.01	−0.02	−0.03	−0.05	**0.74**
Value (F2)												
Val1	−0.04	**0.54**	0.11	0.06	0.04	0.07	−0.03	−0.01	0.01	0.04	−0.01	**0.57**
Val2	0.17	**0.38**	0.16	0.03	0.03	0.02	0.07	−0.01	0.01	−0.02	−0.23	**0.74**
Val3	0.05	**0.74**	−0.01	0.03	0.01	0.02	−0.02	0.01	−0.01	0.01	−0.05	**0.68**
Val4	0.21	**0.45**	0.14	0.02	0.03	0.07	0.05	0.00	−0.03	−0.07	−0.09	**0.77**
Mastery (F3)												
Mas1	0.20	−0.05	**0.59**	0.04	0.05	0.07	0.06	−0.00	−0.01	−0.01	−0.08	**0.64**
Mas2	0.10	0.05	**0.48**	0.04	0.08	0.03	0.10	0.02	−0.01	−0.03	−0.09	**0.70**
Mas3	0.06	0.23	**0.59**	0.03	0.02	0.04	−0.02	−0.01	−0.02	−0.02	0.01	**0.77**
Mas4	0.06	0.17	**0.65**	0.09	0.02	0.04	−0.04	0.02	0.02	0.01	−0.03	**0.79**
Planning (F4)												
Pln1	0.09	0.06	0.07	**0.26**	0.29	0.15	−0.01	−0.07	0.02	−0.06	−0.06	**0.66**
Pln2	0.01	0.05	0.07	**0.73**	0.01	0.06	0.00	0.01	0.01	−0.03	−0.02	**0.76**
Pln3	0.05	0.04	0.01	**0.76**	0.06	0.07	0.01	0.00	−0.01	0.00	−0.03	**0.80**
Pln4	−0.01	0.05	−0.01	**0.32**	0.23	0.11	−0.03	0.01	0.05	−0.02	−0.06	**0.55**
Management (F5)												
Mng1	0.06	0.07	0.09	−0.00	**0.63**	0.07	0.01	−0.03	−0.00	−0.03	−0.06	**0.71**
Mng2	0.05	0.02	0.04	0.27	**0.47**	0.08	0.02	−0.02	0.02	−0.03	−0.07	**0.73**
Mng3	0.09	0.04	0.02	0.15	**0.70**	0.05	0.02	−0.00	0.00	−0.04	−0.04	**0.86**
Mng4	0.08	0.07	0.08	0.13	**0.49**	0.11	0.01	0.02	−0.01	−0.02	−0.03	**0.70**
Persistence (F6)												
Per1	0.07	0.02	0.09	−0.00	0.02	**0.53**	0.05	−0.03	−0.01	0.03	−0.03	**0.60**
Per2	0.02	0.04	0.04	0.08	0.09	**0.60**	0.04	−0.01	−0.01	−0.03	−0.11	**0.72**
Per3	0.08	0.13	0.06	0.20	0.02	**0.50**	0.00	−0.04	0.01	−0.04	−0.00	**0.75**
Per4	0.14	0.12	0.01	0.11	0.02	**0.62**	−0.03	−0.00	0.01	−0.02	−0.05	**0.80**
Anxiety (F7)												
Anx1	0.02	−0.00	0.05	0.01	0.03	0.03	**0.77**	0.03	0.02	−0.00	−0.02	**0.74**
Anx2	0.02	0.05	0.05	0.01	0.02	−0.02	**0.56**	0.17	0.08	0.02	−0.05	**0.68**
Anx3	−0.04	0.03	0.01	0.02	−0.02	−0.05	**0.51**	0.14	0.05	0.07	0.09	**0.60**
Anx4	−0.03	0.01	0.03	0.00	0.02	0.03	**0.62**	0.07	0.10	0.02	0.04	**0.69**
Control (F8)												
Cnt1	−0.02	0.00	0.03	−0.03	0.00	−0.06	0.01	**0.58**	0.03	0.08	0.02	**0.62**
Cnt2	−0.10	−0.01	0.03	−0.03	−0.01	−0.01	0.14	**0.44**	0.17	0.08	0.08	**0.66**
Cnt3	−0.05	−0.02	0.00	0.00	−0.02	−0.03	0.04	**0.72**	0.01	0.04	0.03	**0.73**
Cnt4	−0.03	0.01	−0.02	−0.01	−0.00	−0.05	0.04	**0.63**	0.08	0.07	0.08	**0.76**
Avoidance (F9)												
Avd1	−0.03	−0.02	0.02	−0.01	0.01	−0.01	0.04	0.07	**0.74**	0.05	0.02	**0.78**
Avd2	−0.05	−0.01	−0.01	0.00	0.00	0.01	0.02	0.08	**0.80**	0.07	0.03	**0.83**
Avd3	0.08	−0.02	−0.00	−0.03	0.01	−0.09	0.17	0.07	**0.42**	0.04	0.13	**0.56**
Avd4	−0.00	0.02	0.01	0.02	0.00	−0.01	0.11	0.07	**0.53**	0.09	0.07	**0.65**
S-handicapping (F10)												
Sh1	−0.02	0.00	−0.01	−0.11	0.02	−0.06	−0.04	0.12	0.04	**0.47**	0.09	**0.61**
Sh2	−0.03	−0.05	−0.00	−0.04	−0.04	−0.05	−0.01	0.08	0.02	**0.66**	0.08	**0.76**

(continued)

Table 3.(continued)

Factors	ESEMfactor loadings											
Items	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	CFAfactor loadings
Sh3	–0.08	–0.03	–0.03	0.00	–0.03	–0.04	0.00	0.04	0.06	**0.69**	0.08	**0.77**
Sh4	–0.02	0.03	–0.02	–0.09	–0.04	–0.08	0.06	0.05	0.05	**0.61**	0.07	**0.72**
Disengagement (F11)												
Deg1	0.04	–0.09	–0.05	–0.06	–0.06	–0.14	0.02	0.08	0.01	0.13	**0.39**	**0.60**
Deg2	–0.05	–0.16	–0.06	–0.06	–0.05	–0.06	–0.06	0.03	0.00	0.03	**0.62**	**0.80**
Deg3	–0.09	–0.04	–0.02	–0.05	–0.00	–0.06	0.06	0.04	0.04	0.07	**0.48**	**0.61**
Deg4	–0.09	–0.09	–0.04	–0.04	–0.02	–0.05	0.01	0.01	0.01	0.04	**0.73**	**0.83**

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
ESEM factor correlations											
Efficacy F1)	1.00										
Value (F2)	0.35	1.00									
Mastery (F3)	0.36	0.37	1.00								
Planning (F4)	0.20	0.23	0.19	1.00							
Management (F5)	0.24	0.21	0.21	0.38	1.00						
Persistence (F6)	0.32	0.27	0.23	0.36	0.26	1.00					
Anxiety (F7)	0.01	0.04	0.11	0.01	0.05	0.01	1.00				
Control (F8)	–0.18	–0.03	0.01	–0.07	–0.05	–0.14	0.26	1.00			
Avoidance (F9)	–0.06	–0.03	0.00	–0.00	0.02	–0.04	0.23	0.25	1.00		
S-handicapping (F10)	–0.14	–0.07	–0.07	–0.15	–0.11	0.06	–0.19	0.20	0.26	1.00	
Disengagement (F11)	–0.27	–0.33	–0.21	–0.20	–0.18	0.04	–0.28	0.14	0.18	0.28	1.00
CFA factor correlations											
Efficacy (F1)	1.00										
Value (F2)	0.73	1.00									
Mastery (F3)	0.70	0.76	1.00								
Planning (F4)	0.53	0.54	0.52	1.00							
Management (F5)	0.58	0.57	0.55	0.77	1.00						
Persistence (F6)	0.67	0.64	0.58	0.71	0.63	1.00					
Anxiety (F7)	–0.00	0.10	0.17	0.04	0.09	0.03	1.00				
Control (F8)	–0.39	–0.22	–0.14	–0.24	–0.22	–0.33	0.48	1.00			
Avoidance (F9)	–0.19	–0.14	–0.07	–0.09	–0.07	–0.15	0.44	0.52	1.00		
S-handicapping (F10)	–0.40	–0.35	–0.29	–0.40	–0.37	–0.45	0.18	0.53	0.42	1.00	
Disengagement (F11)	–0.61	–0.70	–0.54	–0.52	–0.52	–0.61	0.05	0.41	0.29	0.56	1.00

Note: F1, F2, F3 . . . F11 indicate factor numbers, such as Factor 1, Factor 2, and Factor 3, and so on. The ESEM model was an exploratory factor analysis with 11 MES factors (see model ESEM model for the total group (TGESEM) in Model 2 for goodness-of-fit statistics). All parameter estimates are completely standardized. In order to conserve space, the CFA factor loadings are presented in condensed format such that only the target loading relating each item to its a priori factor is presented (as all nontarget loadings are zero). A priori target loadings are shown in bold face.

Invariance Across Gender

Preliminary support for the similarity of parameter estimates over gender comes from an evaluation of Model MGI1 with parameters freely estimated for girls and boys. Both the pattern and level of factor loadings are very similar as are correlations among the factors, though uniquenesses were marginally higher for boys ($M = 0.49$) than for girls ($M = 0.47$). The results for Model MGI1 (Table 2, also Table 4 in technical appendix) provide strong support for configural invariance. Hence, we

Table 4. Eleven Motivation and Engagement Factors Based on Responses to 44 MES (Motivation and Engagement Scale) Items for Boys and Girls

	Factor loadings																						Inter		Uniq	
	Girls											Boys														
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	G	B	G	B
Eff (F1)																										
Eff1	**0.63**	0.09	0.06	0.06	0.01	0.02	–0.03	–0.03	–0.00	0.02	–0.07	**0.51**	0.05	0.16	0.02	0.06	0.03	–0.01	–0.08	–0.00	0.00	–0.06	5.53	5.11	0.45	0.53
Eff2	**0.48**	0.04	0.08	0.03	–0.04	0.20	–0.09	–0.07	0.00	–0.01	–0.06	**0.54**	0.04	0.04	0.06	–0.01	0.14	–0.01	–0.07	–0.01	–0.00	–0.07	4.02	3.98	0.55	0.53
Eff3	**0.42**	0.10	0.12	0.04	0.08	0.06	0.07	0.00	–0.00	0.00	0.04	**0.46**	0.07	0.12	0.00	0.11	0.06	0.05	0.00	–0.01	–0.03	–0.03	4.67	4.47	0.66	0.60
Eff4	**0.61**	0.11	0.03	0.03	0.06	0.05	0.00	–0.03	–0.03	–0.03	–0.06	**0.61**	0.14	0.04	0.05	0.03	0.06	–0.00	–0.01	–0.02	–0.03	–0.04	4.81	4.60	0.43	0.44
Val (F2)																										
Val1	0.01	**0.53**	0.06	0.08	0.05	0.04	–0.05	0.00	0.02	0.03	–0.02	–0.07	**0.54**	0.14	0.05	0.04	0.10	–0.02	–0.02	–0.00	0.05	–0.01	4.26	3.80	0.63	0.59
Val2	0.13	**0.45**	0.17	0.03	0.01	0.00	0.09	0.00	–0.01	–0.05	–0.21	0.17	**0.35**	0.16	0.02	0.05	0.03	0.04	–0.02	0.02	0.01	–0.23	6.00	5.31	0.46	0.52
Val3	0.06	**0.70**	–0.01	0.02	–0.01	0.06	–0.05	0.02	0.00	0.03	–0.05	0.05	**0.76**	–0.01	0.04	0.03	0.00	0.00	0.01	–0.01	0.00	–0.05	4.11	3.70	0.43	0.35
Val4	0.18	**0.47**	0.14	0.03	0.03	0.08	0.06	–0.00	–0.03	–0.08	–0.09	0.23	**0.46**	0.14	0.02	0.03	0.07	0.03	0.00	–0.02	–0.07	–0.09	5.45	4.74	0.46	0.45
Mas (F3)																										
Mas1	0.12	–0.12	**0.60**	–0.00	0.05	0.08	0.03	0.04	0.03	–0.00	–0.09	0.04	–0.01	**0.59**	0.05	0.05	0.06	0.07	–0.04	–0.02	–0.02	–0.07	5.57	4.71	0.53	0.53
Mas2	0.26	0.05	**0.47**	0.01	0.07	0.01	0.10	0.05	–0.01	–0.05	–0.09	0.17	0.05	**0.50**	0.04	0.08	0.03	0.09	–0.00	–0.01	–0.02	–0.09	6.37	5.19	0.48	0.51
Mas3	0.08	0.22	**0.61**	0.05	0.02	0.02	–0.02	–0.03	–0.04	–0.01	0.01	0.12	0.24	**0.55**	0.02	0.02	0.05	–0.02	0.01	–0.01	–0.03	0.01	5.07	4.57	0.39	0.41
Mas4	0.02	0.18	**0.66**	0.10	0.01	0.05	–0.03	–0.01	0.02	0.02	–0.03	0.09	0.17	**0.62**	0.09	0.02	0.05	–0.04	0.04	0.03	0.01	–0.03	4.87	4.33	0.37	0.36
Pln (F4)																										
Pln1	0.03	0.05	0.06	**0.26**	0.30	0.13	0.00	–0.07	0.02	–0.06	–0.09	0.10	0.05	0.08	**0.26**	0.29	0.16	0.00	–0.08	0.01	–0.05	–0.04	3.10	3.08	0.56	0.55
Pln2	0.05	0.03	0.05	**0.79**	–0.02	0.05	–0.00	0.01	0.00	–0.01	–0.03	0.01	0.06	0.08	**0.70**	0.03	0.05	–0.00	0.00	0.01	–0.04	–0.02	3.01	2.71	0.30	0.37
Pln3	0.02	0.04	0.01	**0.74**	0.07	0.07	0.00	0.00	–0.02	–0.01	–0.03	0.06	0.04	0.02	**0.77**	0.05	0.07	0.01	–0.00	–0.01	0.01	–0.03	3.13	2.82	0.31	0.27
Pln4	0.08	0.07	–0.01	**0.30**	0.25	0.10	–0.01	0.01	0.05	–0.03	–0.04	–0.03	0.03	0.00	**0.34**	0.22	0.12	–0.02	0.01	0.03	–0.00	–0.08	1.88	1.90	0.70	0.69
Mng (F5)																										
Mng1	0.08	0.07	0.08	0.01	**0.63**	0.08	–0.02	–0.01	0.01	–0.03	–0.07	0.04	0.04	0.11	–0.01	**0.64**	0.06	0.02	–0.04	–0.01	–0.03	–0.06	3.93	3.69	0.44	0.45
Mng2	0.06	0.02	0.05	0.25	**0.51**	0.06	0.03	–0.03	–0.01	–0.05	–0.07	0.05	0.03	0.04	0.28	**0.45**	0.09	0.01	–0.01	0.03	–0.02	–0.07	2.57	2.48	0.45	0.48
Mng3	0.11	0.08	0.04	0.15	**0.70**	0.08	0.02	0.00	–0.01	–0.02	–0.02	0.08	0.11	0.01	0.15	**0.69**	0.04	0.01	–0.00	0.01	–0.03	–0.05	3.43	3.16	0.23	0.25
Mng4	0.06	0.07	0.12	0.14	**0.51**	0.08	0.02	–0.01	–0.01	–0.02	–0.02	0.10	0.08	0.05	0.13	**0.48**	0.12	–0.01	0.05	–0.00	–0.01	–0.03	3.35	3.10	0.52	0.53
Per (F6)																										
Per1	0.09	0.01	0.07	0.00	0.02	**0.54**	0.04	–0.01	–0.01	–0.04	–0.02	0.06	0.01	0.10	0.01	0.02	**0.53**	0.05	–0.05	–0.01	–0.03	–0.04	4.22	3.69	0.61	0.61
Per2	0.03	0.04	0.02	0.07	0.08	**0.62**	0.02	–0.01	0.01	–0.03	–0.12	0.10	0.03	0.05	0.08	0.10	**0.59**	0.05	–0.02	–0.01	–0.03	–0.10	3.48	3.17	0.43	0.45
Per3	0.07	0.14	0.05	0.21	0.04	**0.52**	–0.00	–0.04	0.02	–0.03	–0.02	0.10	0.12	0.06	0.20	0.01	**0.49**	0.00	–0.04	0.01	–0.04	0.01	3.72	3.34	0.43	0.48
Per4	0.14	0.12	0.04	0.13	0.00	**0.61**	–0.01	–0.01	–0.00	–0.03	–0.05	0.15	0.11	–0.00	0.10	0.03	**0.63**	–0.04	0.00	0.02	–0.01	–0.05	3.54	3.40	0.34	0.36

(continued)

Table 4. (continued)

	Factor loadings																									
	Girls											Boys											Inter		Uniq	
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	G	B	G	B
Anx (F7)																										
Anx1	−0.00	−0.01	0.05	−0.01	0.04	0.03	**0.78**	0.04	0.03	−0.03	−0.01	0.02	0.00	0.05	0.01	0.02	0.03	**0.77**	0.01	0.00	0.01	−0.03	2.68	2.33	0.34	0.38
Anx2	0.04	0.04	0.05	0.01	0.03	−0.04	**0.61**	0.17	0.09	0.02	−0.04	0.02	0.05	0.04	0.00	0.00	−0.01	**0.52**	0.17	0.09	0.01	−0.06	3.07	2.40	0.49	0.61
Anx3	−0.07	0.06	0.02	0.01	−0.04	−0.05	**0.51**	0.13	0.06	0.07	0.12	−0.02	0.01	−0.01	0.02	−0.01	−0.06	**0.51**	0.15	0.04	0.06	0.08	2.61	2.27	0.61	0.63
Anx4	−0.03	0.02	0.05	0.03	0.02	0.02	**0.67**	0.06	0.10	0.01	0.01	0.02	0.00	0.01	−0.02	0.02	0.05	**0.58**	0.07	0.10	0.03	0.06	2.08	1.89	0.47	0.58
Cnt (F8)																										
Cnt1	−0.04	0.01	0.01	−0.03	−0.02	−0.05	−0.00	**0.59**	0.02	0.10	0.01	−0.01	0.00	0.04	−0.03	0.01	−0.06	0.04	**0.56**	0.03	0.07	0.03	2.27	2.13	0.58	0.60
Cnt2	−0.11	−0.02	0.03	−0.01	−0.02	−0.01	0.13	**0.45**	0.18	0.08	0.06	−0.09	−0.01	0.02	−0.05	−0.01	−0.02	0.14	**0.43**	0.16	0.09	0.08	2.07	1.87	0.58	0.58
Cnt3	−0.04	−0.00	0.02	−0.01	−0.03	−0.03	0.02	**0.76**	0.01	0.02	0.02	−0.05	−0.02	−0.01	0.01	−0.02	−0.02	0.05	**0.69**	0.02	0.05	0.04	2.07	1.83	0.37	0.44
Cnt4	−0.11	0.00	−0.01	−0.01	0.02	−0.05	0.03	**0.64**	0.08	0.07	0.07	−0.10	0.01	−0.03	−0.01	−0.01	0.05	0.05	**0.64**	0.08	0.06	0.07	1.96	1.78	0.43	0.43

	Factor loadings																									
	Girls											Boys											Inter		Uniq	
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	G	B	G	B
Avd (F9)																										
Avd1	−0.03	−0.02	0.02	−0.02	0.01	−0.01	0.03	0.05	**0.76**	0.04	0.04	−0.02	−0.03	0.01	−0.00	0.01	−0.02	0.04	0.08	**0.72**	0.06	0.01	1.61	1.53	0.37	0.39
Avd2	−0.06	−0.01	−0.00	−0.00	0.01	−0.01	0.01	0.07	**0.82**	0.06	0.03	−0.04	−0.10	−0.02	0.01	−0.00	0.02	0.04	0.09	**0.78**	0.07	0.03	1.70	1.64	0.25	0.28
Avd3	0.10	−0.01	−0.03	0.00	−0.03	−0.09	0.17	0.09	**0.41**	0.05	0.15	0.06	−0.03	0.02	−0.05	0.03	−0.09	0.17	0.06	**0.42**	0.03	0.12	2.13	2.13	0.67	0.67
Avd4	0.03	0.03	−0.02	0.01	−0.03	0.01	0.11	0.09	**0.54**	0.09	0.05	−0.03	0.01	0.02	0.02	0.02	−0.02	0.11	0.05	**0.53**	0.09	0.09	1.79	1.69	0.59	0.60
Sh (F10)																										
Sh1	−0.00	−0.02	−0.02	−0.11	0.03	−0.04	−0.04	0.10	0.06	**0.48**	0.10	−0.02	0.10	−0.01	−0.10	0.01	−0.06	−0.04	0.13	0.03	**0.47**	0.08	1.66	1.60	0.62	0.63
Sh2	−0.03	−0.05	−0.01	−0.06	−0.05	−0.03	−0.02	0.07	0.02	**0.67**	0.11	−0.03	−0.05	−0.01	−0.03	−0.04	−0.06	−0.01	0.08	0.03	**0.65**	0.07	1.70	1.60	0.39	0.45
Sh3	−0.06	−0.04	−0.02	−0.01	−0.02	−0.08	−0.02	0.02	0.06	**0.71**	0.05	−0.08	−0.02	−0.04	0.01	−0.04	−0.01	0.02	0.05	0.07	**0.67**	0.10	1.75	1.64	0.38	0.40
Sh4	−0.02	0.05	−0.03	−0.11	−0.04	−0.10	0.06	0.07	0.03	**0.61**	0.08	−0.02	0.02	−0.01	0.08	−0.03	−0.06	0.07	0.04	0.06	**0.62**	0.07	1.77	1.66	0.46	0.48
Deg (F11)																										
Deg1	0.00	−0.08	−0.04	−0.06	−0.06	−0.12	0.01	0.07	−0.00	0.16	**0.40**	0.06	−0.09	−0.06	−0.06	−0.05	−0.15	0.03	0.10	0.01	0.12	**0.38**	1.59	1.55	0.61	0.64
Deg2	−0.05	−0.19	−0.03	−0.06	−0.05	−0.05	−0.06	0.01	0.00	0.02	**0.63**	−0.05	−0.14	−0.08	−0.05	−0.05	−0.06	−0.06	0.05	0.00	0.04	**0.61**	1.48	1.45	0.35	0.38
Deg3	−0.07	−0.04	−0.05	−0.08	0.02	−0.11	0.07	0.06	0.04	0.07	**0.44**	−0.09	−0.04	−0.02	−0.04	−0.02	−0.03	0.04	0.03	0.05	0.06	**0.50**	1.62	1.51	0.62	0.61
Deg4	−0.10	−0.07	−0.05	−0.04	−0.00	−0.04	−0.00	−0.01	0.01	0.02	**0.76**	−0.09	−0.09	−0.03	−0.04	−0.03	−0.05	0.01	0.02	0.00	0.06	**0.72**	1.57	1.48	0.25	0.27

(continued)

Table 4. (continued)

	Factor correlations																					
	Girls											Boys										
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
Eff (F1)	1.00											1.00										
Val (F2)	0.35	1.00										0.34	1.00									
Mas (F3)	0.36	0.33	1.00									0.37	0.39	1.00								
Pln (F4)	0.21	0.23	0.18	1.00								0.19	0.22	0.20	1.00							
Mng (F5)	0.23	0.18	0.20	0.36	1.00							0.24	0.22	0.22	0.39	1.00						
Per (F6)	0.32	0.26	0.20	0.35	0.25	1.00						0.31	0.27	0.24	0.36	0.27	1.00					
Anx (F7)	–0.01	0.05	0.11	0.02	0.05	–0.02	1.00					0.02	0.03	0.10	–0.01	0.04	0.02	1.00				
Cnt (F8)	–0.17	–0.02	0.03	–0.07	–0.05	–0.14	0.24	1.00				–0.18	–0.04	–0.02	–0.07	–0.05	–0.15	0.26	1.00			
Avd (F9)	–0.06	–0.02	–0.00	–0.01	–0.00	–0.04	0.22	0.24	1.00			–0.07	–0.03	0.01	0.00	0.03	–0.04	0.24	0.26	1.00		
Sh (F10)	–0.12	–0.08	–0.08	–0.18	–0.12	–0.21	0.03	0.18	0.24	1.00		–0.15	–0.07	–0.08	–0.14	–0.11	–0.18	0.08	0.27	0.21	1.00	
Deg (F11)	–0.26	–0.33	–0.21	–0.21	–0.16	–0.28	0.04	0.14	0.15	0.28	1.00	–0.27	–0.32	–0.21	–0.20	–0.20	–0.27	0.04	0.21	0.14	0.28	1.00

Note: F1, F2, F3 . . . F11 indicate factor numbers, such as Factor 1, Factor 2, and Factor 3, and so on. G = girls, B = boys; Eff = self-efficacy, Val = valuing of school, Mas = mastery orientation, Pln = planning, Mng = study management, Per = persistence, Anx = anxiety, Cnt = uncertain control, Avd = failure avoidance, Sh = self-handicapping, Deg = disengagement; Uniq = item uniquenesses; Inter = item intercepts.

now proceed to evaluate results from key models in the 13-model taxonomy of invariance (Table 1) to evaluate: How generalizable is the 11-MES factor structure over gender? Are there systematic gender differences in latent means and are the underlying assumptions met to justify interpretations of these results?

Weak factorial/measurement invariance (MGI1 vs. MGI2 in Table 2) tests whether the unstandardized factor loadings are the same for girls and boys. In support of the invariance of factor loadings over gender, fit indices that control for parsimony are better for the more parsimonious MGI2 than for MGI1 (TLI = 0.968 vs. 0.955, RMSEA = 0.022 vs. 0.025). Thus, the difference in CFI and RMSEA are less than the 0.01 and 0.015 cutoff values typically used to reject the more parsimonious model. In summary, these results provide reasonable support for weak factorial/measurement invariance of the ESEM factor structure over gender.

Strong measurement invariance requires that item intercepts as well as factor loadings are invariant over groups. The critical comparison here is between Models MGI2 and MGI5, though any pair of model differing only in the items intercepts being free or constrained to invariance can also be used to complement this comparison. Nonsupport for this model would imply differential item functioning (though it would still be possible to posit partial invariance; see Marsh et al., 2010, 2009), whereas strong support for this model would imply that differences between groups at the item level can be explained in terms of differences at the latent factor mean level. Fit indices for MGI5 are comparable to those of MGI2 (TLI = 0.966 vs. 0.968, CFI = 0.974 vs. 0.976, RMSEA = 0.023 vs. 0.022). The small changes in fit indices provide reasonable support for the more parsimonious model MGI5, demonstrating the strong measurement invariance across gender, the absence of differential item functioning, and a justification for the comparison of latent mean differences.

Strict measurement invariance (MGI5 vs. MGI7) requires that item uniqueness, item intercepts, and factor loadings are all invariant across groups. A lack of support for this model would suggest that measurement error differs in the two groups and, thus, precludes the comparison of manifest scale scores (i.e., simple unweighted averages or sums of responses to items designed to measure each factor that are the basis of typical ANOVAs). In support of strict invariance, fit indices for MGI7 and MGI5 are comparable (TLI = 0.959 vs. 0.966, CFI = 0.968 vs. 0.974, RMSEA = 0.025 vs. 0.023), and the differences are less than traditional cutoff values.

Factor variance-covariance invariance (MG2 vs. MG4) is typically not a focus in studies of measurement invariance, particularly ones that focus on single constructs. However, this is frequently an important focus of studies of the discriminant validity of multidimensional constructs that might subsequently be extended to include relations with other constructs. Although the comparison of correlations between the 11 MES factors for different groups, for example, gender, is common, these are typically based on manifest scores that do not control for measurement error (particularly if measurement error differs for the groups); thus, such a comparison makes implicit invariance assumptions that are rarely tested. Fit indices that control for parsimony are nearly the same for MGI4 compared to MGI2 and MGI1 (0.969 vs. 0.968 and 0.955 for TLI, 0.022 vs. 0.022 and 0.025 for RMSEA), whereas the differences in CFIs (0.976 vs. 0.976 and 0.977) are less than the 0.01 cutoff value.

Latent factor mean comparison across gender. Finally, we are now in a position to address the issue of the invariance of the factor means across the two groups. In the taxonomy the final four models (MGI10-MGI13 in Table 2) all constrained mean differences between girls and boys to be zero—in combination with the invariance of other parameters. Again, there are several models that could be used to test mean invariance across gender: MGI5 vs. MGI10, MGI7 vs. MGI11, MGI8 vs. MGI12, and MGI9 vs. MGI13. Evaluation of fit indices for each of these pairs of models shows a consistent pattern (ΔCFIs < 0.01 and ΔRMSEAs < 0.015), which suggests that the mean levels of the 11 MES factors do not differ substantially for girls and boys.

Table 5. ESEM Test–Retest Factor Solution: 11 Motivation and Engagement Factors Based on Responses to 44 MES Items Collected at Time 1 and Time 2

	Factor loadings																										
	Time 1											Time 2											Inter		Uniq		Corr
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	Time 1	Time 2	Time 1	Time 2	Uniq
Eff (F1)																											
eff1	**0.60**	0.02	0.11	0.08	0.03	0.01	–0.02	–0.04	0.00	0.03	–0.11	**0.49**	0.05	0.20	–0.00	0.03	0.02	–0.01	–0.12	–0.07	0.03	–0.12	5.54	5.28	0.45	0.48	0.10
eff2	**0.49**	0.10	0.10	0.03	–0.03	0.16	–0.05	–0.10	–0.10	–0.03	0.00	**0.55**	0.02	0.10	0.10	–0.00	0.16	–0.07	–0.05	–0.05	–0.00	–0.09	4.12	4.24	0.52	0.42	0.19
eff3	**0.45**	0.08	0.12	–0.00	0.13	0.06	0.06	0.01	–0.00	–0.03	0.04	**0.53**	0.17	0.04	0.02	0.10	0.10	0.08	–0.00	0.02	–0.08	0.01	4.63	4.55	0.62	0.47	0.10
eff4	**0.59**	0.15	0.04	0.06	0.05	0.06	–0.01	–0.02	–0.01	–0.05	–0.06	**0.61**	0.12	0.11	0.04	0.09	0.05	0.01	–0.01	–0.02	–0.04	–0.09	4.89	4.70	0.41	0.33	0.03
Val (F2)																											
Val1	0.01	**0.54**	0.06	0.06	0.02	0.07	–0.03	0.01	0.02	0.03	–0.04	–0.10	**0.58**	0.15	0.05	0.07	0.06	–0.03	0.04	–0.06	0.03	–0.05	4.25	3.77	0.61	0.52	0.21
Val2	0.19	**0.36**	0.10	0.03	0.04	0.00	0.08	0.00	0.01	0.01	–0.29	0.19	**0.35**	0.21	–0.03	0.06	0.05	0.07	0.03	–0.00	0.00	–0.22	6.00	5.11	0.49	0.46	0.19
Val3	0.09	**0.69**	–0.03	0.05	0.02	0.05	–0.01	0.02	–0.01	–0.00	–0.07	0.06	**0.75**	–0.05	0.00	0.02	0.05	0.01	–0.04	–0.00	0.02	–0.07	4.17	3.78	0.40	0.36	0.14
Val4	0.14	**0.47**	0.14	0.01	0.05	0.04	0.04	–0.01	–0.05	–0.04	–0.12	0.28	**0.44**	0.12	0.09	–0.02	0.03	0.04	0.02	0.00	–0.07	–0.13	5.63	4.80	0.51	0.40	0.05
Mas (F3)																											
Mas1	0.04	–0.09	**0.71**	0.05	0.02	0.04	–0.01	0.04	0.02	–0.01	–0.09	0.02	–0.01	**0.70**	0.02	0.06	0.09	0.02	–0.02	0.01	–0.01	–0.05	5.19	4.90	0.43	0.40	0.11
Mas2	0.20	0.05	**0.54**	0.04	0.05	–0.01	0.10	0.01	–0.02	–0.02	–0.09	0.15	0.03	**0.58**	0.01	0.06	0.08	0.07	0.03	–0.02	–0.01	–0.13	5.67	5.09	0.46	0.40	0.12
Mas3	0.12	0.30	**0.50**	0.03	0.02	0.05	0.01	–0.03	–0.02	–0.03	0.05	0.12	0.30	**0.50**	0.13	–0.01	–0.00	–0.10	0.01	–0.02	–0.04	0.01	5.08	4.60	0.45	0.40	0.06
Mas4	0.10	0.22	**0.55**	0.08	0.03	0.08	–0.02	0.01	0.02	0.00	–0.00	0.13	0.27	**0.49**	0.15	–0.02	0.06	–0.03	0.01	0.03	–0.01	0.02	4.77	4.32	0.42	0.39	0.15
Pln (F4)																											
Pln1	0.13	0.02	0.05	**0.29**	0.27	0.14	–0.03	–0.09	0.01	–0.06	–0.07	0.08	0.01	0.09	**0.30**	0.25	0.18	–0.05	–0.11	0.00	–0.02	–0.07	3.18	3.10	0.55	0.54	0.19
Pln2	0.05	0.07	0.05	**0.71**	0.00	0.07	0.01	–0.02	0.00	–0.00	–0.04	0.02	0.03	0.05	**0.74**	0.02	0.05	0.03	0.01	0.01	–0.01	–0.09	2.97	2.82	0.36	0.32	0.16
Pln3	0.05	0.03	0.04	**0.76**	0.05	0.06	0.03	–0.01	–0.02	–0.00	–0.02	0.05	0.07	0.05	**0.73**	0.08	0.07	0.03	–0.02	–0.01	–0.01	–0.03	3.12	2.95	0.29	0.27	–0.00
Pln4	–0.04	0.05	–0.01	**0.33**	0.24	0.09	–0.06	0.02	0.03	–0.09	–0.08	–0.05	0.04	0.00	**0.40**	0.24	0.13	–0.04	0.03	0.03	–0.04	–0.01	2.01	1.98	0.68	0.64	0.30
Mng (F5)																											
Mng1	0.09	–0.02	0.06	–0.02	**0.64**	0.10	0.01	–0.03	–0.00	–0.04	–0.09	0.03	0.01	0.15	0.00	**0.65**	0.07	0.02	–0.02	–0.03	–0.02	–0.09	3.95	3.82	0.44	0.40	0.15
Mng2	0.03	0.03	0.10	0.27	**0.47**	0.06	–0.02	–0.02	0.04	–0.07	–0.08	0.03	0.05	0.05	0.19	**0.57**	0.07	–0.02	–0.02	–0.02	–0.04	–0.06	2.61	2.64	0.45	0.42	0.20
Mng3	0.08	0.11	0.04	0.14	**0.72**	0.05	–0.00	0.01	–0.01	–0.00	–0.03	0.13	0.10	–0.02	0.12	**0.68**	0.09	0.01	0.01	0.03	–0.05	–0.03	3.40	3.29	0.23	0.28	–0.04
Mng4	0.12	0.11	0.05	0.18	**0.48**	0.08	0.03	–0.07	–0.00	–0.01	0.01	0.09	0.11	0.08	0.17	**0.51**	0.05	0.03	0.00	–0.01	–0.03	–0.06	3.33	3.31	0.49	0.46	0.14
Per (F6)																											
Per1	0.07	–0.03	0.12	–0.01	0.03	**0.54**	0.04	–0.04	–0.03	–0.02	–0.04	–0.00	0.03	0.18	–0.06	0.06	**0.61**	0.01	–0.03	–0.05	0.02	–0.02	4.08	3.90	0.60	0.50	0.11
Per2	0.03	0.03	0.02	0.10	0.06	**0.66**	0.02	0.01	–0.01	–0.04	–0.09	0.01	0.04	0.07	0.06	0.08	**0.67**	0.01	–0.01	0.01	–0.04	–0.12	3.46	3.39	0.40	0.34	0.05
Per3	0.07	0.14	0.07	0.18	0.04	**0.50**	0.01	–0.07	0.01	–0.04	0.01	0.12	0.07	0.04	0.29	–0.04	**0.48**	0.01	–0.01	0.00	–0.06	–0.05	3.71	3.69	0.48	0.43	0.11
Per4	0.14	0.15	0.00	0.09	0.05	**0.59**	–0.00	–0.02	0.00	–0.02	–0.05	0.19	0.15	–0.04	0.14	0.03	**0.58**	–0.00	–0.01	0.02	–0.03	–0.04	3.55	3.48	0.38	0.34	0.06

(continued)

Table 5. (continued)

	Factor loadings																						Inter		Uniq		Corr
	Time 1											Time 2															
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	Time 1	Time 2	Time 1	Time 2	Uniq
Anx (F7)																											
Anx1	0.01	0.00	0.05	0.00	0.02	0.05	**0.77**	0.06	0.03	−0.02	−0.03	0.02	0.02	0.04	0.02	0.01	0.02	**0.80**	0.05	0.05	−0.01	−0.01	2.44	2.39	0.35	0.30	0.10
Anx2	0.01	0.06	0.05	0.04	0.01	−0.03	**0.58**	0.15	0.09	0.02	−0.02	0.03	0.04	0.05	−0.01	0.03	0.01	**0.59**	0.12	0.12	0.02	−0.05	2.62	2.41	0.53	0.53	0.19
Anx3	−0.07	0.05	−0.04	0.01	0.02	−0.05	**0.48**	0.15	0.08	0.06	0.04	−0.07	0.01	0.03	0.03	0.02	−0.09	**0.54**	0.18	0.05	0.03	0.08	2.41	2.32	0.65	0.56	0.25
Anx4	−0.02	0.02	0.03	0.05	0.05	0.02	**0.63**	0.10	0.09	0.02	0.02	0.02	0.03	0.01	0.04	−0.02	0.03	**0.66**	0.05	0.10	0.04	0.00	1.96	1.96	0.51	0.49	0.22
Cnt (F8)																											
Cnt1	−0.01	0.01	0.02	−0.08	0.01	−0.08	0.04	**0.53**	0.05	0.07	−0.02	−0.01	0.03	0.00	−0.06	0.03	−0.08	0.04	**0.59**	0.06	0.05	0.01	2.23	2.04	0.62	0.56	0.01
Cnt2	−0.10	0.01	0.00	0.00	−0.04	0.04	0.10	**0.46**	0.11	0.12	0.11	−0.04	−0.04	−0.00	−0.02	−0.01	−0.03	0.11	**0.58**	0.15	0.06	0.07	1.96	1.87	0.59	0.47	0.23
Cnt3	−0.07	−0.01	0.03	0.01	−0.01	−0.01	0.04	**0.70**	0.03	0.04	0.05	−0.06	−0.01	0.00	−0.01	−0.04	−0.03	0.07	**0.70**	0.05	0.06	0.05	1.92	1.84	0.42	0.37	−0.07
Cnt4	−0.05	0.01	−0.01	0.03	−0.02	−0.06	0.05	**0.63**	0.08	0.04	0.06	−0.10	−0.03	−0.03	−0.01	−0.01	−0.05	0.06	**0.64**	0.05	0.10	0.06	1.87	1.82	0.46	0.41	0.08

	Factor loadings																						Inter		Uniq		Corr
	Time 1											Time 2															
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	Time 1	Time 2	Time 1	Time 2	Uniq
Avd (F9)																											
Avd1	−0.01	−0.03	0.01	−0.03	0.00	−0.01	0.00	0.07	**0.74**	0.03	−0.00	−0.03	−0.03	−0.01	0.03	0.03	−0.02	0.02	0.06	**0.77**	0.02	0.02	1.59	1.54	0.40	0.35	0.12
Avd2	−0.02	−0.00	−0.02	0.01	−0.01	−0.00	0.02	0.11	**0.76**	0.03	0.04	−0.05	−0.02	−0.01	−0.02	0.01	0.00	0.04	0.07	**0.79**	0.04	0.01	1.71	1.64	0.33	0.29	−0.06
Avd3	0.07	−0.07	0.02	−0.02	0.04	−0.07	0.19	0.00	**0.42**	0.09	0.17	0.11	−0.08	0.04	−0.07	0.05	−0.06	0.13	0.02	**0.41**	0.06	0.13	2.17	2.09	0.65	0.71	0.37
Avd4	−0.07	0.01	0.04	0.04	0.02	−0.01	0.13	0.01	**0.53**	0.12	0.05	0.02	−0.03	0.02	0.03	−0.01	0.00	0.08	0.01	**0.54**	0.16	0.06	1.79	1.69	0.60	0.61	0.20
Sh (F10)																											
Sh1	0.00	−0.04	−0.02	−0.08	0.05	−0.06	−0.05	0.11	0.07	**0.44**	0.09	0.08	−0.05	−0.03	−0.07	0.04	−0.09	−0.04	0.13	0.01	**0.49**	0.11	1.66	1.61	0.68	0.62	0.18
Sh2	−0.00	−0.08	0.03	−0.00	−0.06	−0.05	−0.04	0.10	0.04	**0.61**	0.07	−0.01	−0.02	−0.05	0.04	−0.06	−0.07	−0.02	0.08	0.05	**0.69**	0.12	1.68	1.61	0.49	0.35	0.04
Sh3	−0.06	−0.01	−0.04	0.01	−0.02	−0.04	0.01	0.04	0.05	**0.71**	0.07	−0.11	−0.02	−0.01	−0.01	−0.03	−0.05	0.01	0.04	0.07	**0.68**	0.05	1.72	1.64	0.38	0.39	0.05
Sh4	−0.03	0.04	−0.04	−0.10	−0.04	−0.06	0.04	0.06	0.03	**0.64**	0.04	−0.03	0.03	−0.02	−0.14	−0.05	−0.02	0.06	0.02	0.06	**0.65**	0.07	1.74	1.70	0.46	0.43	0.16
Deg (F11)																											
Deg1	0.06	−0.07	−0.05	−0.09	−0.05	−0.15	0.06	0.04	0.02	0.15	**0.44**	0.04	−0.07	−0.07	−0.07	−0.08	−0.11	0.06	0.10	0.03	0.05	**0.47**	1.60	1.54	0.58	0.56	0.19
Deg2	−0.05	−0.16	−0.05	−0.06	−0.05	−0.05	−0.05	0.03	−0.01	0.02	**0.65**	−0.04	−0.17	−0.05	−0.03	−0.02	−0.08	−0.06	0.01	0.01	0.01	**0.67**	1.45	1.42	0.35	0.32	0.16
Deg3	−0.08	−0.03	−0.09	−0.01	−0.02	−0.06	0.04	0.12	0.06	0.01	**0.45**	−0.11	−0.01	−0.09	−0.05	−0.01	−0.02	0.06	0.02	0.03	0.12	**0.47**	1.60	1.53	0.64	0.59	0.23
Deg4	−0.12	−0.07	−0.06	−0.04	−0.01	−0.05	−0.02	0.03	0.03	0.05	**0.65**	−0.09	−0.09	−0.04	−0.08	−0.04	−0.05	0.00	0.00	0.00	0.08	**0.66**	1.53	1.49	0.37	0.32	0.06

(continued)

Table 5. (continued)

	Factor correlations																					
	Time 1											Time 2										
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
Time 1																						
Eff (F1)	1.00																					
Val (F2)	0.38	1.00																				
Mas (F3)	0.39	0.31	1.00																			
Pln (F4)	0.22	0.24	0.21	1.00																		
Mng (F5)	0.26	0.19	0.19	0.37	1.00																	
Per (F6)	0.29	0.26	0.21	0.34	0.25	1.00																
Anx (F7)	−0.00	0.07	0.10	0.04	0.00	0.01	1.00															
Cnt (F8)	−0.20	−0.02	−0.00	−0.11	−0.08	−0.15	0.27	1.00														
Avd (F9)	−0.06	−0.05	0.01	−0.00	0.01	−0.06	0.23	0.24	1.00													
Sh (F10)	−0.11	−0.08	−0.07	−0.15	−0.13	−0.19	0.07	0.26	0.20	1.00												
Deg (F11)	−0.26	−0.32	−0.22	−0.20	−0.18	−0.25	0.02	0.19	0.15	0.25	1.00											
Time 2																						
Eff (F1)	**0.54**	0.16	0.17	0.11	0.13	0.17	−0.04	−0.11	−0.07	−0.10	−0.10	1.00										
Val (F2)	0.15	**0.49**	0.20	0.21	0.09	0.13	−0.02	−0.01	−0.04	−0.06	−0.19	0.37	1.00									
Mas (F3)	0.30	0.22	**0.50**	0.19	0.14	0.17	0.06	−0.02	0.00	−0.06	−0.23	0.37	0.37	1.00								
Pln (F4)	0.15	0.10	0.18	**0.57**	0.25	0.21	0.02	−0.09	0.01	−0.17	−0.10	0.20	0.23	0.21	1.00							
Mng (F5)	0.21	0.07	0.14	0.32	**0.60**	0.21	−0.00	−0.08	−0.02	−0.17	−0.15	0.22	0.19	0.21	0.34	1.00						
Per (F6)	0.24	0.16	0.17	0.24	0.13	**0.52**	−0.02	−0.15	−0.07	−0.22	−0.19	0.30	0.27	0.27	0.34	0.26	1.00					
Anx (F7)	0.04	0.07	0.09	0.05	0.02	0.03	**0.68**	0.22	0.16	0.04	−0.02	0.03	0.06	0.08	0.02	0.02	−0.01	1.00				
Cnt (F8)	−0.19	−0.01	0.02	−0.06	−0.09	−0.11	0.22	**0.59**	0.16	0.23	0.07	−0.15	−0.03	−0.03	−0.09	−0.05	−0.16	0.28	1.00			
Avd (F9)	−0.09	0.01	0.02	−0.03	0.04	−0.06	0.20	0.15	**0.49**	0.14	0.11	−0.05	−0.08	−0.02	−0.01	−0.02	−0.06	0.24	0.23	1.00		
Sh (F10)	−0.06	−0.04	0.00	−0.09	−0.07	−0.17	0.07	0.16	0.09	**0.47**	0.13	−0.13	−0.08	−0.10	−0.14	−0.13	−0.19	0.06	0.23	0.19	1.00	
Deg (F11)	−0.25	−0.24	−0.18	−0.20	−0.17	−0.23	0.02	0.06	0.07	0.20	**0.57**	−0.28	−0.35	−0.28	−0.23	−0.20	−0.29	0.02	0.16	0.13	0.28	1.00

Note: F1, F2, F3 … F11 indicate factor numbers, such as Factor 1, Factor 2, and Factor 3, and so on. Eff = self-efficacy, Val = valuing of school, Mas = mastery orientation, Pln = planning, Mng = study management, Per = persistence, Anx = anxiety, Cnt = uncertain control, Avd = failure avoidance, Sh = self-handicapping, Deg = disengagement; Uniq = item uniquenesses; Inter = item intercepts corr uniq =correlated uniquenesses.

Table 6. Patterns of Mean Differences Between Matching Motivation and Engagement Factors Between Girls and Boys and at Time 1 and Time 2 and Test–Retest Correlations Between Time-1 and Time-2 Factors

	Patterns of mean differences across gender										
Model	Eff (F1)	Val (F2)	Mas (F3)	Pln (F4)	Mng (F5)	Per (F6)	Anx (F7)	Ctl (F8)	Avd (F9)	Sh (F10)	Deg (F11)
Model 5 IN=FL, Inter	0.088	–0.091	–0.221	–0.171	–0.037	–0.087	–0.419	–0.168	0.078	–0.115	0.001
Model 7 IN=FL, Uniq, Inter	0.097	–0.092	–0.224	–0.168	–0.034	–0.085	–0.409	–0.163	0.081	–0.114	–0.002
Model 8 IN=FL, FVCV, Inter	0.078	–0.090	–0.222	–0.173	–0.040	–0.085	–0.414	–0.171	0.086	–0.104	0.008
Model 9 IN=FL, FVCV, Uniq, Inter	0.081	–0.092	–0.222	–0.173	–0.040	–0.085	–0.412	–0.171	0.086	–0.104	0.007
	Patterns of mean differences over time										
Model	Eff (F1)	Val (F2)	Mas (F3)	Pln (F4)	Mng (F5)	Per (F6)	Anx (F7)	Ctl (F8)	Avd (F9)	Sh (F10)	Deg (F11)
Model 5 IN=FL, Inter	0.041	–0.193	–0.063	–0.092	–0.038	0.012	0.008	–0.179	–0.021	–0.102	0.097
Model 7 IN=FL, Uniq, Inter	0.042	–0.192	–0.062	–0.092	–0.038	0.013	0.007	–0.183	–0.022	–0.104	0.095
Model 8 IN=FL, FVCV, Inter	0.039	–0.206	–0.061	–0.095	–0.042	0.011	0.005	–0.188	–0.023	–0.105	0.103
Model 9 IN=FL, FVCV, Uniq, Inter	0.038	–0.205	–0.060	–0.096	–0.042	0.011	0.005	–0.189	–0.023	–0.105	0.103
	Test–retest correlations between Tune 1 and Time 2 factors										
Model	Eff (F1)	Val (F2)	Mas (F3)	Pln (F4)	Mng (F5)	Per (F6)	Anx (F7)	Ctl (F8)	Avd (F9)	Sh (F10)	Deg (F11)
Model 0 IN=none, no CUs	0.58	0.55	0.54	0.60	0.62	0.55	0.73	0.60	0.51	0.49	0.61
Model 1 IN=none, CUs	0.54	0.49	0.50	0.57	0.60	0.52	0.68	0.59	0.49	0.47	0.57
Model 2 IN=FL	0.57	0.52	0.54	0.56	0.60	0.53	0.68	0.58	0.50	0.47	0.57
Model 3 IN=FL, Uniq	0.57	0.52	0.54	0.56	0.60	0.53	0.68	0.58	0.50	0.47	0.57
Model 4 IN=FL, FVCV	0.56	0.50	0.54	0.57	0.61	0.53	0.68	0.58	0.49	0.47	0.57

Note: F1, F2, F3 . . . F11 indicate factor numbers, such as Factor 1, Factor 2, and Factor 3, and so on. Eff = self-efficacy; Val = valuing of school; Mas = mastery orientation; Pln = planning; Mng = study management; Per = persistence; Anx = anxiety; Cnt = uncertain control; Avd = failure avoidance; Sh = self-handicapping; Deg = disengagement

Invariance Over Time: Latent Mean Structure Approach

Marsh et al. (2010, 2009) suggested that, with some adaptations, it is possible to apply the same set of 13-model taxonomy (Table 1) to test the invariance of ESEM factor structures with test–retest data. In this longitudinal ESEM application, we constrain an a priori set of 44 cross-wave-correlated uniquenesses (CUs) to account for the residual associations between matching items at 44 MES items at Time 1 and Time 2. Indeed, when the same item is used on multiple occasions, a correlation between the unique components of each item on the two occasions that cannot be explained by the correlations between the factors is likely to exist. The failure to include these CU is likely to systematically bias parameter estimates such that test–retest correlations among matching latent factors are systematically inflated (Marsh et al., 2004). Thus, we first tested the configural invariance of the responses to the MES with and without these 44 a priori CUs (Models LGI0 and LGI1). Fit indices for Model LGI0 and LGI1 clearly support the inclusion of 44 CUs (0.934 vs. 0.960 for TLI, 0.951 vs. 0.970 for CFI, 0.024 vs. 0.019 for RMSEA). Consistent with a priori expectations, the test–retest correlations for the 11 MES factors in LGI0 were inflated (.49 to .73, $M = 0.58$) compared to those in LGI1 (.47 to .68, $M = 0.55$). Based on these initial analyses, these a priori CUs are included in all subsequent models.

Tests of *weak factorial/measurement invariance* (LGI2 vs. LGI1) demonstrate the invariance of factor loadings over time. Fit indices controlling for model parsimony are slightly better for the more parsimonious LGI2 than for the less parsimonious LGI1 (TLI = 0.963 vs. 0.960, RMSEA = 0.018 vs. 0.019), whereas the CFIs are the same (0.970). Tests of *strong measurement invariance* (LGI2 vs. LGI5) requires that item intercepts—as well as factor loadings—to be invariant over time. Differences in fit indices for LGI5 and LGI2 (0.962 vs. 0.963 for TLI, 0.018 vs. 0.18 for RMSEA, 0.969 vs. 0.970 for CFI) are small in relation to traditional guidelines. Results based on LGI5 support strong measurement invariance of MES responses (and a lack of differential item functioning) and justify the comparison of latent means over time. Tests of s*trict measurement invariance* (LGI5 vs. LGI7) require the invariance of item uniquenesses as well as item intercepts and factors loadings. Fit indices for LGI7 are comparable to those of LGI5 (0.961 vs. 0.962 for TLI, 0.967 vs. 0.969 for CFI, 0.019 vs. 0.018 for RMSEA), supporting the more parsimonious model in relation to traditional cutoff values. Results based on LGI7 support the strict longitudinal invariance of the MES responses.

The invariance of the latent factor variance-covariance matrix (LGI4 vs. LGI2) provides good support for the invariance of these parameters (ΔCFI = 0.001, ΔTLI = 0.000, ΔRMSEA = 0.000). Models LGI10 to LGI13 each test the invariance of latent means in combination with the invariance of other sets of parameters. In each case, the fits of these models positing no latent mean differences is equally good and acceptable to the corresponding models in which latent mean differences are freely estimated: ΔCFIs (0.002), ΔTLIs (0.002), and ΔRMSEAs (0.000 to 0.001) based on comparisons between Models LGI10 and LGI5, between LGI11 and LGI7, between LGI12 and LGI8, and between LGI13 and LGI9. These findings suggest that the 11-MES-factor means do not differ systematically over time.

In summary, the ESEM approach applied to the motivation and engagement factors based on response to the MES provides reasonable support for the invariance of factor loadings, item intercepts, item uniquenesses, factor variances-covariances, and latent means over time.

Discussion

The main purpose of this study, which is based on a large database, is to compare factorial solutions of the ESEM and ICM-CFA approaches applied to a multidimensional measure of motivation and engagement—the MES (Martin, 2007, 2009). The pattern and the sizes of factor loadings are similar for the two approaches. Clearly the ESEM solution fits the data better than the traditional CFA solution. In addition to the difference in goodness of fit, the two solutions differed systematically in terms of the sizes of correlations among the factors, with the ESEM correlations (M absolute $r = .17$) being substantially smaller than those based on CFA factors (M absolute $r = .40$; see also Martin, 2007, 2009). The extent of this difference in correlations depends substantially on the appropriateness of the ICM-CFA assumption that nontarget loadings are all zero.

The smaller correlations among the 11 MES factors based on the ESEM approach relative to those of the CFA approach are important in applications when differential validity is important, for example, when the 11 MES factors are used to predict outcome variables so that multicollinearity is likely to become a problem. This difference is also consistent with the logic of the ESEM approach. In particular, when a large number of relatively small cross-loadings are constrained to be zero as in the ICM-CFA solution (Table 3), the only way that these cross-loadings can be represented is by inflating the factor correlations. In relation to the present investigation—and multivariate research generally—this is a particularly serious problem because it can undermine support for the multidimensionality of a construct and for the discriminant validity (or distinctiveness) of the multiple factors. Particularly when the multiple factors are used to predict outcome variables, as in regression or SEM analyses, the inflated factor correlations are likely to result in beta parameters distorted by multicollinearity, thus obscuring the predictive ability of individual motivation and engagement factors. Hence, the increased distinctiveness of ESEM factors is a major advantage over traditional CFA approach.

ESEM should generally be preferred to the ICM-CFA model when the factors are appropriately identified by ESEM, the goodness of fit for ESEM is meaningful better than for ICM-CFA, and factor correlations are meaningfully smaller for the ESEM model than for the ICM-CFA model. If the goodness of fit and factor correlations are nearly the same, then the CFA-ICM model is generally preferable. It is, of course, relevant to point out that the ESEM factors might not be consistent with the a priori design of the instrument while the CFA factors are forced to be consistent with it. Although clearly not the case here, this might call into question the appropriateness of the a priori model—particularly if the fit of the ICM-CFA model was not particularly good.

In terms of the methodological-measurement fruitfulness of ESEM, we demonstrated the applicability of the ESEM approach to testing full-measurement invariance: factor loading, variances-covariances, item uniquenesses, item intercepts, differential item functioning, and latent factor means. Evidence of strict measurement invariance is crucial before manifest scale scores can be validly compared using traditional approaches like *t* test or MANOVA across multiple groups or occasions. However, many psychoeducational assessment studies completely ignore all assumptions of measurement invariance, substantially undermining the validity of interpretations. It is important to note that, though the approach demonstrated here could be conducted with CFA models (but not traditional EFA models), the new ESEM approach offers potentially important advantages that integrate many of the advantages of EFA and CFA into a single analytic framework. In summary, the present investigation not only provides further support for the construct validity of the MES but also demonstrates the utility of the ESEM approach, especially when used to test the taxonomy of 13 factorial/measurement invariance models.

As ESEM is still a new statistical tool, there are a variety of issues in need of further research and evaluation in relation to applied practice. A major advantage of ESEM over CFA is that it typically provides a better fit to the data. However, like EFA, the researcher has less control over the a priori factor structure. Hence, more work is needed in the application of ESEM when the factor structure is not well defined. As demonstrated here, correlations among multiple factors are likely to be substantially higher in CFA solutions than in ESEM solutions whenever there are substantial cross-loadings. For tests of invariance we applied the Marsh et al.'s (2010, 2009) 13-model taxonomy of invariance tests. Although the taxonomy incorporates a richer selection of models than those typically considered, researchers might choose to emphasize specific models or a set of models that incorporate additional features.

The 13-model taxonomy emphasized here is equally appropriate for ESEM and CFA. However, to the extent that the ESEM solution provides an acceptable fit to the data, and the CFA solution does not, then the appropriateness of the taxonomy for CFA models is dubious. In this respect we present the ESEM model as a viable alternative to the ICM-CFA model, but we do not argue that the ESEM approach should replace the corresponding CFA approach. Indeed, justification of the ESEM approach should routinely begin with a rigorous comparison with ICM-CFA approaches, as in the present investigation. However, on the basis of Marsh et al.'s (2005, 2010, 2009) suggestion that few multidimensional psychological instruments widely used in practice provide an acceptable fit in relation to an a priori ICM-CFA structure, we suspect that, relative to ICM-CFA-based approaches, ESEM-based approaches are likely to generate better factorial solutions. In this situation, we suggest that advanced statistical strategies such as multigroup and MIMIC tests of measurement invariance and latent growth models in many applications are more appropriately conducted in an ESEM approach than in a traditional ICM-CFA approach.

Appendix

The MES Factors: Example Item and Coefficient Alpha Estimate of Reliability

1. Self-efficacy measures students' belief and confidence in their ability to understand or to do well in their school work and to perform to the best of their ability ($\alpha = .77$; for example, "If I try hard, I believe I can do my schoolwork well").
2. Valuing of school measures the extent to which students believe what they do and learn at school is useful, important, and relevant to them ($\alpha = .79$; for example, "Learning at school is important to me").
3. Mastery orientation measures students' focus on understanding, learning, and solving problems when learning ($\alpha = .82$; for example, "I feel very pleased with myself when I really understand what I am taught at school").
4. Planning measures the degree to which students plan their work and how much they keep track of their profess as they are doing it ($\alpha = .78$; for example, "Before I start an assignment I plan out how I am going to do it").
5. Study management measures the extent to which students manage their time, organize their timetables, and where they prepare for doing their school work ($\alpha = .83$; for example, "When I study, I usually study in places where I can concentrate").
6. Persistence measures students' capacity to persist in situations that are challenging and at times when they find it difficult to do what is required ($\alpha = .80$; for example, "If I can't understand my schoolwork at first, I keep going over t until I understand it").
7. Anxiety measures the extent to which students feel nervous when they think about their school work and are worried for not doing well in their school work ($\alpha = .77$; for example, "When exams and assignments are coming up, I worry a lot").
8. Uncertain control measures students' uncertainty about how to do well or how to avoid doing poorly ($\alpha = .78$; for example, "I'm often unsure how I can avoid doing poorly at school").
9. Failure avoidance measures the degree to which students believe that the main reason they try at school is to avoid doing poorly or to avoid being seen to do poorly ($\alpha = .80$; for example, "Often the main reason I work at school is because I don't want to disappoint my parents").

(continued)

Appendix (continued)

10. Self-handicapping measures the extent to which students reduce their chances of success at school so that they have a reason for not doing well at school (α = .81; for example, "I sometimes don't study very hard before exams so I have an excuse if I don't do as well as I hoped").
11. Disengagement measures the chances that students give up or are at risk of giving up at school and, in particular, school activities (α = .80; for example, "I often feel like giving up at school").

Author's Note

Requests for further information about this investigation should be sent to Professor Herbert W. Marsh, Department of Education, University of Oxford; E-mail: herb.marsh@education.ox.ac.uk. Requests for further information about the Motivation and Engagement Scale should be sent to Professor Andrew J. Martin, University of Sydney, Australia; E-mail: andrew.martin@sydney.edu.au

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by a grant to the first author from the UK Economic and Social Research Council and by a visiting fellowship from the University of Sydney where the research was conducted.

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*. Retrieved from http://www.statmodel.com/download/EFACFA810.pdf

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal, 13*(2), 186-203.

Chen, F. F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling, 14*, 464-504.

Cheung, G. W., & Rensvold, R. B. (2001). The effects of model parsimony and sampling error on the fit of structural equation models. *Organizational Research Methods, 4*, 236-264.

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling, 9*, 327-346.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47*, 309-326.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286-299.

Halpern, D. F. (2006). Assessing gender gaps in learning and academic achievement. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 635-653). Mahwah, NJ: Lawrence Erlbaum.

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.

Marsh, H. W., Balla, J. R., & Hau, K. T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical processes. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling techniques* (pp. 315-353). Mahwah, NJ: Lawrence Erlbaum.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McCardle (Eds.), *Psychometrics. A festschrift to Roderick P. McDonald*. Mahwah, NJ: Lawrence Erlbaum.

Marsh, H. W., Hau, K.-T., & Wen, Z., (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cut-off values for fit indices and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big-five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*, 471-491.

Marsh, H. W., Lüdtke, O., Robitzsch, A., Nagengast, B., Morin, A. J. S., & Trautwein, U. (in press). *Two wrongs do not make a right: Camouflaging misfit with item-parcels in CFA models*.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Applications to students' evaluations of university teaching. *Structural Equation Modeling, 16*, 439-476.

Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology, 77*, 413-440.

Martin, A. J. (2009). Motivation and engagement across the academic lifespan: A developmental construct validity study of elementary school, high school, and university/college students. *Educational and Psychological Measurement, 69*, 794-824.

Martin, A. J. (2010). *Building classroom success: Eliminating academic fear and failure*. London: Continuum.

Meredith, W. (1964). Rotation to achieve factorial invariance. *Psychometrika, 29*, 187-206.

Meredith, W., & Teresi, J. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*(Suppl. 3), 69-77.

Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical & Statistical Psychology, 38*, 171-189.

Muthén, L. K., & Muthén, B. (2008). *Mplus user's guide.* Los Angeles: Author.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*, 667-686.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*, 287-297.

Rhemtulla, M., Brosseau-Liard, P., & Savalei, V. (2010). How many categories is enough to treat data as continuous? A comparison of robust continuous and categorical SEM estimation methods under a range of non-ideal situations. Retrieved from http://www2.psych.ubc.ca/~mijke/files/HowManyCategories.pdf

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research, 45*, 73-103.

Bios

Herb Marsh is Professor of Education at Oxford University, having spent much of his career in Sydney after completing his PhD at UCLA. He is widely published and a "ISI highly cited researcher". He founded SELF Research Centre that has 450 members and satellite centres at leading Universities around the world. His research interests vary widely including both substantive and methodological pursuits, and their integration.

Alexandre J.S. Morin, Ph.D., is professor of Psychology at the University of Sherbrooke and member of the Educational Excellence and Equity (E^3) Research Program of the Center for Educational Research (University of Western Sydney). His research interests are centered on substantive methodological synergies aimed at illustrating the usefulness of powerful new statistical methods (among which exploratory structural equation models, mixture models for cross-sectional and longitudinal data and complex latent curve models) in the comprehension of substantively important research questions related to internalized disorders (including burnout, anxiety, and depression) and multidimensional conceptions of self-concept (including commitments).

Benjamin Nagengast is a post-doctoral research fellow at the Department of Education University of Oxford. He holds a PhD in Quantitative Methods in Psychology from the Friedrich-Schiller- UniversitÃ¤t Jena (Germany). His research interests focus on the application of statistical methods for causal inference, multilevel modelling and structural equation modelling to educational research.

Andrew Martin PhD is Professorial Research Fellow in the Faculty of Education and Social Work at the University of Sydney. He specializes in motivation, engagement, achievement and quantitative research methods applied to substantive questions and issues.

Gregory Arief D. Liem holds a doctorate from the National University of Singapore. He is presently on a research fellowship at the University of Sydney. Prior to this, he was an Assistant Professor at Nanyang Technological University, Singapore. His research interest is in the area of social psychology in education, with the main substantive foci on values, motivation, learning, social comparison in educational settings, and psychological aspects of civic behaviors.

Erratum

Journal of Psychoeducational Assessment
29(5) 497

DOI: 10.1177/0734282911423093
http://jpa.sagepub.com

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S., Nagengast, B. (2011). Methodological Measurement Fruitfulness of Exploratory Structural Equation Modeling (ESEM): New Approaches to Key Substantive Issues in Motivation and Engagement. *Journal of Psychoeducational Assessment.* 29, 322-346.
Original DOI: 10.1177/0734282911406657

In the article above, on page 322, the affiliations of the authors were published incorrectly at the bottom.

The correct academic affiliations are:

Herbert W. Marsh, Oxford University, UK; University of Western Sydney, Australia; and King Saud University, Saudi Arabia

Gregory Arief D. Liem, University of Sydney, Australia

Andrew J. Martin, University of Sydney, Australia

Benjamin Nagengast, Oxford University, UK

Alexander J. S. Morin, University of Sherbrooke, Canada; Educational Excellence and Equity (E3) Research Program, Center for Educational Research (University of Western Sydney)

The publisher SAGE regrets the error.