Towards an Integrative Perspective on the Structure of Teacher Work Engagement: Dimensionality, Measurement and Longitudinal Invariance, and Relations with Work Satisfaction

Harsha N. Perera

Lucie Vosicka

Helena Granziera

Peter McIlveen

Correspondence should be directed to Dr Harsha N. Perera, Department of Educational Psychology and Higher Education, College of Education, University of Nevada, Las Vegas; 4505 Maryland Pkwy, Box 453003; Las Vegas, NV 89154-3003; P: (702) 895-1110; E: Harsha.Perera@unlv.edu

Abstract

Towards an Integrative Perspective on the Structure of Teacher Engagement: Dimensionality, Measurement and Longitudinal Invariance, and Relations with Work Satisfaction

Recent years have witnessed an increasing interest in teachers’ work engagement in both scholarly and policy contexts. This increased interest has been attributed to accumulating evidence demonstrating the relations of teachers’ beliefs, feelings, and behaviors, such as efficacy beliefs (Goddard, Hoy, & Hoy, 2000; Klassen & Tze, 2014), emotional exhaustion (Arens & Morin, 2016), and work engagement (Klusmann et al., 2008; Roth, Assor, Kanat), with students’ academic and non-academic outcomes. At least another reason for increased attention to teachers’ work engagement is the increasing turnover of teachers in several modern education systems, including Australia, the United Kingdom, and the United States (Organization for Economic Cooperation and Development, 2002; Willett, Segal, & Walford, 2014). As well as placing considerable pressure on the teaching pipeline to meet expected growth in student numbers, teacher attrition has significant educational and social costs, not least the loss of expertise in, and destabilization of, schools that may undermine student outcomes. It is in this climate that researchers and policy-makers have turned their attention to teacher engagement (Bakker, Hakanen, Demerouti, & Xanthopoulou, 2007; Høigaard, Giske, & Sundsli, 2012; Klassen et al., 2012; McLennan, McIlveen, & Perera, 2017; McIlveen & Perera, 2015; Runhaar, Konermann, & Sanders, 2013; Skaalvik & Skaalvik, 2014), not as an immediate panacea for the turnover problem, but rather as an aspect of teacher effectiveness that may inform understandings of why and when teachers leave the profession. Notwithstanding accruing evidence supporting the role of teachers’ work engagement in both teacher and student outcomes, extant work is limited by a dearth of clarity about the theoretical and empirical structure of the teacher engagement construct, which complicates investigations of the construct and its relation to substantively meaningful antecedents and outcomes.

 Advances in teacher engagement research, and the utility of the construct for applied purposes, hinge on a theoretically meaningful and empirically robust structure of engagement. Accordingly, the present study seeks to shed light on the theoretical and empirical structure of teacher engagement by examining scores from an increasingly used measure of teacher work engagement (viz., the Engaged Teacher Scale [ETS]; Klassen, Yerdelen, & Dursken, 2013). First, the latent structure underlying ETS data is examined in a heterogeneous sample of teachers. Alternative structural representations are examined, including single-factor, correlated-factors, higher-order, and bifactor models, based on theoretical guidelines and prior empirical work. We examine these structures in both confirmatory factor analysis (CFA) and exploratory structural equation modeling (ESEM) frameworks. In addition, we examine the complete measurement and structural invariance of scores implied by the retained latent structure across primary and secondary teachers. We also examine gender differences in the latent engagement dimensions as well as the gender by teaching level interaction. Furthermore, the longitudinal invariance of the measurement structure and covariance stability are investigated. Finally, we examine the validity of scores based on the final engagement model for predicting teachers’ work satisfaction.

**Teacher Engagement: Theoretical Overview**

 In recent research on teacher engagement, generally two, related, models of work engagement have been used to conceptualize engagement (Hakanen, Bakker, & Scaufeli, 2006; Klassen et al., 2012; Skaalvik & Skaalvik, 2016). The first model centers on work engagement that is not specific to a particular vocation (e.g., teachers’ or nurses’ work engagement) (Bakker et al., 2011; Bakker & Demerouti, 2008) but instead assumes that there are dimensions of engagement that generalize across different vocational environments. From this perspective, work engagement is a motivational construct denoting the volitional allocation of physical, attentional, and affective resources to work-related tasks. Specifically, the model posits three dimensions of work engagement, namely vigor, dedication, and absorption. Workers with high vigor invest considerable energy and effort in their work and persist in work-related tasks even when confronting adversity. Dedicated workers perceive their work as significant and meaningful, and experience a sense of pride while performing work-related activities. Finally, individuals who are absorbed in their work devote considerable cognitive resources to, and maintain total concentration on, work-related tasks. These dimensions of work engagement are posited to be related but reflect distinct aspects of being engaged at work.

More recently, Klassen et al. (2013) proposed a teaching-specific model of work engagement, extending existing conceptualizations of engagement to reflect specific aspects of teachers’ work (e.g., interacting with students and colleagues). Specifically, Klassen et al. integrated dominant work engagement perspectives with the teacher-student relatedness literature to propose social engagement as part of the conceptual content domain of engagement in the context of teachers’ work given the centrality of interactions with students and colleagues to this work (Jennings & Greenberg, 2009; Rimm-Kaufman, Baroody, Larsen, & Curby, 2014). From this integrative perspective, teacher engagement comprises cognitive-physical, emotional, and social dimensions. Cognitive-physical engagement is the extent to which teachers attend to and invest effort in work tasks encountered and reflects both the vigor and absorption dimensions proposed in Bakker and Demerouti’s (2008) model. Emotional engagement refers to teachers’ positive affective responses, such as happiness, enjoyment, and excitement, to their work. Finally, social engagement, comprising both students and colleague domains, refers to teachers’ perceptions of their connection to, and concern for, students and colleagues, respectively. As with Bakker and Demeouti’s model, the dimensions proposed in Klassen et al’s conceptualization of teachers’ work engagement are posited to be distinct but related (Klassen et al., 2013).

**The Dimensionality of Teacher Engagement**

Although both the Bakker and Demerouti (2008) and Klassen et al. (2012) models of engagement are predicated on a multidimensionality perspective, evidence is unclear about the structure of data obtained from psychometric measures (e.g., Utrecht Work Engagement Scale [UWES], ETS) designed to operationalize the engagement constructs posited in these models. For instance, initial validation work on the UWES, intended to measure the three dimensions posited in Bakker and Demerouti’s model, found that a three-factor model provided a superior fit to data obtained from college students (sample 1) and private and public sector employees (sample 2) relative to unidimensional and two-factor models. However, correlations among the three dimensions were substantial (sample 1: *r* = .77-.98, *M* = .83; Sample 2: *r* = .84-93, *M* = .89) (Schaufeli, Salanova, Gonzalez-Roma, & Bakker, 2002). Likewise, Schaufeli, Bakker, and Salanova (2006) and Seppälä et al. (2009) found support for a correlated three-factor model based on data from a short-form version of the UWES (i.e., UWES-9); yet, akin to Schaufeli et al. (2002), correlations among the engagement dimensions approximated unity. Findings of appreciable correlations among these dimensions of work engagement have also been found in samples of teachers (Bakker et al., 2007; Hakanen, Bakker, & Schaufeli, 2006).

The magnitude of the relations among these engagement dimensions is indicative of the possibility of a general engagement factor underlying work engagement data. Indeed, Klassen et al. (2012) preferred a unidimensional model over a correlated three-factor structure based on UWES-9 data collected from teachers across five countries. The unidimensional solution has also been found to be tenable in other studies (Fong & Ho, 2015; Schimazu et al., 2008; Seppälä et al., 2009). The findings of substantial correlations, and accumulating evidence supporting a unidimensional structure of work engagement, undermine the discriminant validity of scores on these engagement dimensions and the multidimensional perspective on which the model of work engagement is predicated. Notably, as Seppälä et al. (2009) remark, based on the available evidence “the ultimate question of the one- vs. three-dimensionality of work engagement” remains unresolved (p. 476).

 The limited work examining the dimensionality of engagement data obtained from the ETS, which is designed to index work engagement in teachers specifically, also suggests a dearth of clarity on the structure of engagement. Seemingly in line with the multidimensionality perspective on engagement, Klassen et al. (2013) supported a four-factor model, with well-defined cognitive-physical and emotional engagement dimensions as well as social dimensions pertaining to colleagues and students. However, as with data obtained from the UWES, correlations among the factors were appreciable, suggesting the possibility of a general factor underlying the data. Notably, Klassen et al. (2013) found that a higher-order model, with a global engagement factor at the apex of the hierarchy and the four specific engagement dimensions at the first-order level, did not fit the ETS data significantly worse than the more complex correlated four-factor structure. This suggests that the engagement dimensions may share sufficient common variance to assume the presence of an underlying common cause (i.e., teachers’ global levels of engagement). On the basis of this evidence, Klassen et al. (2013) concluded that using the “four-factor or single-factor models” are viable for representing teacher engagement (p. 42). Although this may seem pragmatic for researchers, it points to unresolved concerns about the dimensionality of engagement data in teachers as with engagement data more generally (Seppälä et al., 2009).

The apparent tenability of a higher-order representation of teacher engagement obtained in Klassen et al. (2013) suggests the presence of construct-relevant multidimensionality in teacher engagement data due to hierarchically-ordered constructs. This higher-order model assumes that teacher engagement data simultaneously reflect specific engagement constructs as well as a global construct. In the higher order model, this is operationalized through the specification of the loading of each item onto one first-order factor, with each first-order factor, in turn, specified to load onto the global factor. Although this hierarchical model is a common specification in the motivation and engagement literature (references), it hinges on restrictive assumptions that may limit its utility as a plausible structural representation of teacher engagement data (Yung, McLeaod, & Thissen, 1999).

Higher-order models inherently impose stringent proportionality constraints that are unlikely to hold with motivational data. These constraints are reflected in the Schmid and Leiman (1957) decomposition procedure (SLP) for yielding a bifactor approximation from a hierarchical solution. Under the SLP procedure, two sets of indirect effects from a higher-order solution are usually computed to approximate bifactor general and specific factor model parameters. First, the indirect relation of the global factor with each item mediated via the first-order factor, which is computed as the product of the item loading on the first order factor and the loading of the first-order factor on the second-order factor, is intended to approximate general factor loadings. Second, the indirect association of the unique first-order term with each item mediated by the first-order factor, computed as the product of the first-order item loading and the disturbance variance of the first-order factor, is intended to approximate specific factor loadings. Notably, the second quantity in these product terms is a constant for both the global and specific variance components for all items associated with a single first-order factor. This constraint artificially restricts the decomposition of global and specific variance in all items within a specific first-order factor to be exactly proportional. Accordingly, in the higher-order model, the ratio of global to specific variance is assumed to be equal for all items within a particular first-order factor (Gignac, 2016), which is unlikely to hold with complex, multidimensional data (Reise, 2012). The implausibility of perfect proportionality raises concerns about the tenability of the higher-order conceptualization of teacher engagement; yet, evidence points to the possibility of a general engagement construct underlying teacher engagement data (Klassen et al., 2013).

The Bifactor model provides an alternative analytic approach to higher-order models for examining generality and specificity underlying data and may be particularly relevant for understanding the structure of teacher engagement data. For any set of items, a bifactor model posits a *f*-factor solution with one general factor (*G*-factor) and *f*-1 specific or group factors (*S*-factors). All items are specified to load onto the *G*-factor and only one S-factor, with correlations among the general and specific factors typically fixed to zero. As applied to teacher engagement data obtained from the ETS, the bifactor model would partition response variance into (a) a general engagement factor reflecting common variation shared by all items and (b) four *S*-factors (e.g., emotional engagement, social engagement with student) that account for additional common variance in item subsets above and beyond the *G*-factor. This variance decomposition is achieved through the specification of *direct* effects from the general and specific factors to the items. In this regard, the bifactor model does not rely on unrealistic proportionality restrictions that, although render the model less parsimonious than a higher-order structure, make it much more defensible from a theoretical standpoint. Indeed, the extent to which the bifactor model will provide a better representation to the data than a higher-order model depends on the magnitude of the violation of the proportionality constraint assumption underlying the higher-order solution (Gignac, 2016). Gignac (2016) showed that even at moderate levels of proportionality constraint violation, which might be expected to be routine in typical research settings, parsimony-corrected fit indices preferred the bifactor model over the higher-order model.

Quite apart from the statistical properties that render the bifactor model preferable over the higher-order structure, there is some evidence that a bifactor model may be a plausible structural representation of teacher engagement data. Although Klassen et al. (2012) supported a unidimensional engagement solution based on teachers’ responses to the UWES-SF, this model only reached acceptable levels of fit after the specification of several correlated uniquenesses, perhaps reflecting construct-relevant multidimensionality in the engagement data due to the presence of content specificities (e.g., cognitive engagement, emotional engagement) over and above general engagement. Comparable evidence was obtained in Fong and Ho (2015) who found support for a unidimensional Bayesian representation of engagement, with 11 statistically significant correlated residuals, based on UWES-SF data obtained from health-sector workers. As with Klassen et al., the presence of residual covariances suggests the possibility of multidimensionality in the data that, if construct-relevant, should almost always be modeled with substantive factors (Perera & Athanasou, 2016). However, across both Klassen et al. and Fong and Ho, it also possible that the correlated uniquenesses reflect construct-irrelevant multidimensionality due to item idiosyncrasies (e.g., parallel wording effects). Notably, in Fong and Ho, when data were estimated with maximum-likelihood, the bifactor model reached acceptable levels of fit in accordance with current standards and, importantly, provided an appreciably better fit to the data than unidimensional and correlated-factors models. A bifactor structure for work engagement data was also supported in de Bruin & Henn (2013).

The bifactor model may reconcile the multidimensionality perspective on engagement espoused by dominant conceptual models (references) with accumulating evidence for a general factor underlying teacher engagement data (Klassen et al., 2012; Klassen et al., 2013). From a conceptual standpoint, the presence of generality and specificity in teacher engagement data suggest that teachers’ inferences about their engagement may reflect the general or global extent of their engagement at work in addition to more specific or differentiated inferences about their engagement socially, emotionally, cognitively, and physically. For instance, it would seem to be entirely reasonable that a teacher feels generally engaged with work but perhaps not altogether connected with colleagues during work-related tasks. This differentiation of the generality of engagement from specificity is perhaps less true for cognitive-physical engagement, reflecting the behavioral intensity and attention in work, which would seem to be central to general inferences about engagement with work as well as engagement in the other domains. Altogether, the evidence and theory reviewed suggests that a comparative examination of competing one-factor, correlated-factors, higher-order, and bifactor representations of teacher engagement data, variously suggesting the possibility of the unidimensionality, multidimensionality, and generality and specificity of teachers’ inferences about their engagement at work, respectively, may be important to clarifying the structure of teacher engagement.

**Construct-Relevant Multidimensionality Beyond the Presence of General and Specific Constructs**

In addition to the purported construct-relevant multidimensionality in teacher engagement data attributable to the presence of both general and specific engagement constructs, another source of substantive multidimensionality in data obtained from the ETS may be psychometric complexity due to the fallibility of items as strictly unidimensional indicators of single engagement constructs (Morin, Marsh, & Arens, 2016; Perera, 2015a, 2015b). As items in the ETS, as with most measures of engagement and motivation, are designed to appraise closely related constructs (e.g., engagement with colleagues, engagement with students), systematic associations of these items with not only target constructs but also non-target constructs may be expected (Perera & Ganguly, 2016). In the factor analytic framework, these item-non-target-factor associations are manifested as small-to-moderate cross-loadings of items on non-target factors. Accordingly, the independent clusters model (ICM) of confirmatory factor analysis (CFA) typically used to examine the dimensionality of engagement data may not be an appropriate analytic model. The constraint of non-target loadings to zero in the ICM-CFA, even those that are small, may result in upwardly biased factor correlations (Asparouhov, Muthén, & Morin, 2016), potentially leading to erroneous inferences about the structure of engagement data. Indeed, sizeable correlations among the engagement dimensions have been obtained from previous applications of the ICM-CFA to ETS data (*r* = .49-.73, *M* = .61) (Klassen et al., 2013). Notably, these correlations were appreciably smaller in exploratory factor analyses of the ETS data (*r* = .33-.62, *M* = .45), which permits cross-loadings to be freely estimated, suggesting the possibility of construct-relevant multidimensionality in the ETS data due to item fallibility. In the present study, we test for this psychometric multidimensionality while simultaneously testing the generality and specificity of the engagement data in an integrative analytic framework.

**Differences in Engagement as a Function of Teacher Characteristics**

Models of teacher motivation and engagement posit that teacher characteristics, such as teaching level and gender, play a role in job-related beliefs, feelings, and behaviors (Klassen & Chiu, 2010; Duffy & Lent, 2010). Little research has directly examined differences in teacher engagement as a function of teaching level (e.g., primary vs. secondary teachers). Rey, Extremera, and Pena (2012) found that Spanish primary teachers reported (a) greater effort and energy in developing teaching tasks and interacting with students, (b) a greater sense of meaningfulness and significance of their work, and (c) higher cognitive absorption in their work relative to secondary teachers. A cognate body of research on teachers’ burnout as a function of teaching level has also found that that elementary teachers report lower depersonalization (i.e., feelings of callousness toward and detachment from students and colleagues) and higher personal accomplishment (i.e., feelings of competence and productivity) than their secondary counterparts (Byrne, 1994; Rey et al., 2012). These findings have been attributed to several mechanisms, including lower classes sizes, greater homogeneity, and more time spent with students in primary classes relative to secondary classes (Biddle & Berliner, 2002; Byrne, 1991; Rey et al., 2012).

Evidence also suggests that gender may be implicated in levels of teacher engagement. Extant teacher burnout research shows that female teachers are more likely to report greater feelings of emotional overextension (Antoniou, Polychroni, & Vlachkis, 2006; Grayson & Alvarez, 2008; Lau, Yuen, & Chan, 2005) and a reduced sense of productivity and effectiveness at work (Lau et al., 2005). Furthermore, female teachers have been shown to experience appreciably higher levels of workload and occupational stress than male teachers (Antoniou, Polychroni, & Vlachkis, 2006; Klassen & Chiu, 2010), which may lead them to disengage from their work. Although these findings from studies of burnout and stress point to potential gender differences in teacher engagement, recent research directly investigating the role of gender in teacher engagement yielded no effect. Specifically, Klassen et al. (2012) found no effects of gender on engagement in four of five country samples. One reason for the absence of appreciable effects may be the modeling of engagement as a unidimensional construct in Klassen et al. (2012), which may obscure meaningful gender differences on specific dimensions of teacher engagement. Indeed, from a theoretical standpoint, social role expectations for domestic caring may impose considerable demands on female teachers at home in addition to work demands (Byrne, 1991). We examine gender effects on the teacher engagement dimensions in a multiple indicator multiple causes framework (MIMIC). In addition, we test the gender by teaching level interaction effect in a hybrid MIMIC-multiple-group framework.

**Longitudinal Invariance, Stability, and Relations**

 Of considerable importance for the measurement of teacher engagement is the examination of the invariance of the supported measurement structure not just across known groups, such as distinct teaching level, but also over time. Support for longitudinal invariance of the measurement would provide strong evidence for the structural validity of the teacher engagement data derived from the ETS that extends structural validity evidence obtained from cross-sectional single-sample and multi-sample invariance tests. Beyond supporting the structural validity of data, longitudinal invariance tests are an important pre-requisite to examining stability and changes in constructs over time. Absent of longitudinal measurement invariance, changes in constructs may be confounded with the differential operation of a measurement instrument over time, leading to erroneous inferences about developmental stability and change. In the present study, we examine the longitudinal invariance of the retained ETS measurement model across two time points separated by five-months. We also extend this investigation of longitudinal variance to an examination of the covariance stability and cross-lagged relations of the engagement dimensions.

**Nomological Network: Relations with Job Satisfaction**

 Theories of motivation, career development, and workplace well-being postulate relations between work engagement and work satisfaction (Klassen, Aldhafri et al., 2012; Lent & Brown, 2008; Rothmann, 2008). For instance, the social cognitive career theory of work satisfaction (Lent & Brown, 2008) holds that individuals’ participation in goal-directed activity, encompassing their engagement with work-related activities, fosters work satisfaction. From this perspective, to the extent that individuals regulate their cognition, affect, and actions towards the pursuit of work-related goals, they establish favorable conditions for experiencing heightened levels of satisfaction with work (Lent & Brown, 2008). Consistent with this view, previous work has consistently shown that engaged workers are more likely to be satisfied (Shimazu et al., 2008; Klassen, Aldhafri et al., 2012). In teachers, specifically, Klassen, Aldhafri et al. found strong positive associations (*r* = .56-.77; *M* = .69) between teacher engagement and work satisfaction across five distinct national samples. In the present study, we examine the effects of teacher engagement work satisfaction, based on the empirically supported structure of engagement data.

**Method**

**Participants and Procedure**

Participants were 594 practicing teachers working in secondary (*n* = 239; 40.2%) and primary (*n* = 355; 59.8%) schools across all states and territories in Australia. The mean age of participants was 40.90 (*SD* = 13.58), and 89% (*n* = 512%) of the sample was female. Although this gender distribution is suggestive of the over-representation of women, the differences in gender proportions across the secondary (female: *n* = 196; 82.0%) and primary (female: *n* = 332; 93.5%) levels are largely consistent with Australian school population characteristics (ABS, 2013). The teaching experience of the sample was diverse, ranging from one month to 42 years (*M* = 11.82, *SD* = 10.67).

 Participants were recruited via visits to research sites (i.e., schools) and social media. First, the research team disseminated information about the study and participation requirements across several schools. Furthermore, invitations to participate were posted on several social media platforms, including pages dedicated to teacher professional development (e.g., “TeachMeet Sydney” and “TeachMeet Western Australia”. Participation involved the completion of an online battery of questionnaires during the first quarter of the academic year. The battery of questionnaires included demographic items and measures of work engagement and job satisfaction. In addition, participants completed a follow-up measure of engagement approximately five months after the initial administration. The protocols and procedures for the research were approved by the Institutional Review Board.

**Measures**

 **Teacher engagement**. Teachers’ work engagement was measured using the ETS (Klassen et al., 2013). The ETS is designed to appraise four dimensions of teachers’ engagement with work, namely cognitive-physical engagement (CE), emotional engagement (EE), social engagement with colleagues (SEC), and social engagement with students (SES). The measure is also intended to yield a composite or global score on global teacher engagement. The measure comprises 16 items, rated on a seven-point Likert-type scale (ranging from 1 = *never* to 7 = *always*), with each subscale comprising four items. Scores from the measure have been shown to be internally consistent, and evidence of convergent and divergent validity of scores has been found (Klassen et al., 2013). In the current sample, the coefficient alpha reliabilities for the total score (αT1 = .92, αT2 = .89) and CE (αT1 = .85, αT2 = .86), EE (αT1 = .91, αT2 = .92), SEC (αT1 = .84, αT2 = .82), and SES (αT1 = .87, αT2 = .79) scores were acceptable.

 **Work satisfaction**. Teachers’ work satisfaction was measured using the Brief Job Satisfaction Measure II (BJSM-II) (Judge, Locke, Durham, & Kluger, 1998). The BJSM-II consists of five items, rated on a seven-point Likert-type scale (ranging from 1 = strongly disagree to 7 = strongly agree), which as designed to measure individuals’ affective-cognitive evaluation of satisfaction with work. Prior work has shown that scores from the measure are internally consistent and possess construct validity (reference). In the present sample, the coefficient alpha reliabilities for the total scores were acceptable (α = .88).

**Statistical Analyses**

Analyses were conducted in five phases. First, CFA and ESEM analyses of the ETS T1 data were performed. The unidimensional model was specified to have all item loadings on a single engagement factor. For the correlated four-factor CFA, each ETS item was specified to load onto the factor it was designed to measure, per the a priori scoring key, with correlations among the four factors freely estimated. For the higher-order CFA model, items were specified to load onto one of the four specific engagement domains as per the correlated-factors model, and these first-order factors, in turn, were specified to index a higher-order global engagement factors. Disturbance covariances were fixed to zero. Finally, for the bifactor CFA, all items were specified to load onto a general QOL factor as well as one of the four domain-specific engagement factor per the scoring key. Null relations among the general and specific factor were specified. For the correlated-factors, higher-order, and bifactor ESEM models, the same pattern of target item factor loadings and factor correlations was specified as per their CFA analogues. However, the correlated-factors ESEM and bifactor ESEM solutions were rotated using target (oblique) and bifactor target orthogonal rotations, respectively, with all cross-loadings “targeted” to be approximately zero but not restricted to zero (Asparouhov & Muthén, 2009). For the higher-order ESEM, as current implementations of ESEM in software programs do not allow for the specification of second-order models, the ESEM-within-CFA approach was used (Morin, Marsh, & Nagengast, 2013; Perera, Izadikhah, O’Connor, & McIlveen, 2016).

 The second phase of the analysis involved subjecting the retained T1 measurement structure to multiple-group tests of full measurement and structural invariance across teaching level with secondary teachers as the baseline group. In the first instance, it should be noted that inspection of observed contingency tables with teaching level as the column variable revealed several cells with zero frequencies, primarily concerning the lowest category of response for the ETS. This spare data problem is common with ordered-categorical data at the extreme end of the response scale. However, a crucial assumption of models based on-ordered categorical data is that the same number of response categories is used across groups. Furthermore, empty cells can lead to model convergence problems and erroneous parameter estimates due to difficulties computing polychoric correlations under limited information estimation methods (Flora & Curran, 2004; Olsson, 1979). Thus, we collapsed adjacent spare categories across the implicated items (see Appendix A).

The multiple-group invariance tests were conducted in accordance with Millsap and Yun-Tein’s (2004) taxonomy of invariance tests for models estimated from ordered categorical data. This taxonomy involves additive and increasingly restrictive tests of configural invariance and the invariance of item factor loadings, thresholds, uniquenesses, factor variances and covariances, and factor means across the finite groups (Perera, McIlveen, Burton, & Corser, 2015). An important consideration in these tests of invariance across groups is the test of factor variance-covariance invariance in bifactor models. The invariance of the factor covariances cannot be examined in the bifactor CFA model given the orthogonality constraints on covariances typically imposed (Arens & Morin, 2016; Reise, 2012). However, for the bifactor ESEM with orthogonal bifactor target rotation, orthogonality is a function of the rotational criterion. Notably, invariance constraints may be imposed on the unrotated solution, thereby allowing the invariance of the factor covariances in the bifactor ESEM to be tested. However, as Arens and Morin (2016) note, the value of this test is questionable in light of the orthogonality of the final (rotated) solution. Nevertheless, the current operationalization of ESEM models requires invariance constraints to be simulatenaouly imposed on factor variances and covariances.

 In the third phase of the analyses, we examined latent mean differences in engagement across gender and its interaction with teaching level. For these models, the final teaching level invariance model was combined with MIMIC models to examine gender effects. MIMIC models may be preferable to complete multiple-group invariance models where a group sample size is small as in the present case with the male subsample (Morin et al., 2013). The combined multiple group-MIMIC modeling involved the examination of three models. First, a saturated model was tested with paths specified from gender (coded as a binary covariate with 0 = male and 1 = female) to all item indicators but not the latent variables in each teaching level group. Second, a threshold-invariant model was estimated in each teaching level group with paths from gender to the latent variables but not indicators. No appreciable degradation in fit of the threshold-invariance model relative to the saturated model is suggestive of support for the equivalence of item thresholds. A final model was estimated with paths from gender to the latent factors constrained to equality across teaching level to test for any gender interaction effects.

 Phase four of the analysis involved tests of the relations of the engagement dimensions with job satisfaction based on the retained T1 measurement structure of engagement. A general latent variable model was specified, including the retained ETS structure and a unidimensional work satisfaction factor indicated by the five items of the BJSM-II. Job satisfaction was regressed on the engagement dimensions. In addition, gender and teaching level were included as observed covariates.

 The final phase of the analysis involved tests of the longitudinal invariance, covariance stability and relations of the engagement dimensions based on the two waves of data using fully-latent autoregressive cross-lagged (ACL) panel analyses. This phase of the analysis was conducted in two distinct steps. First, we examined the longitudinal invariance of the retained bifactor ESEM structure as a necessary pre-condition to examining the covariance stability and lagged structural relations among the engagement constructs. These invariance tests were conducted in line with Liu et al’s (2016) novel taxonomy of longitudinal invariance tests for ordered-categorical data adapted for ESEM. This involved additive tests of a baseline model, in which the same factor model with a comparable pattern of free and fixed parameters was specified across time, and subsequent tests of the longitudinal invariance of factor loadings, thresholds, and uniquenesses. Across these models, each unique factor was free to correlate with itself overtime to account for residual associations between corresponding items administered at T1 and T2. These a priori correlated uniquenesses account for covariances between the same items administered across multiple waves beyond that which is explained by the cross-time factor correlations. The failure to model these cross-time correlated uniquenesses may lead to poor model fit and upwardly biased estimates of longitudinal stability coefficients (Marsh & Hau, 1996). Subject to support for longitudinal invariance, the retained invariance model was reparameterized as an ACL panel model, with all autoregressive and cross-lagged paths among the constructs freely estimated, to examine associations between engagement and satisfaction. Across both the longitudinal invariance measurement and ACL structural models, gender and teaching level were included as observed covariates. As contingency tables revealed a different number of categories observed for some items across time, we collapsed adjacent spare categories, such that all indicators had the same number of response categories (see Appendix B).

 Analyses were performed using Mplus 7.4 (Muthén & Muthén, 1998-2015). Solutions were estimated using robust diagonal weighted least squares, operationalized as the Weighted Least Squares Mean and Variance adjusted (WLSMV) estimator in Mplus. Model fit assessment involved an evaluation of fit indices, parameters estimates, and alternative structures. As the χ2 can be oversensitive to minor model misspecifications given even moderate-sized samples and contains a restrictive hypothesis test (i.e., exact fit), three approximate fit indices were considered: Root Mean Square Error of Approximation (RMSEA), < .050 and .080 for close and reasonable fit; Comparative fit index (CFI); and Tucker-Lewis Index (TLI), > .900 and .950 for acceptable and excellent fit, respectively (Marsh, Hau, & Wen, 2004). For nested model comparisons, because the adjusted χ2 difference (MD Δχ2) test appropriate for the WLSMV estimator also tends to be sensitive to even trivial differences, changes in the CFI (ΔCFI) and RMSEA (ΔRMSEA) were primarily used. A decrease in the CFI and increase in the RMSEA of less than .010 and .015, respectively, are indicative of support for a more restrictive model (Chen, 2007; Cheung & Rensvold, 2002). We hasten to add that these are rough guidelines for inferring model fit and model improvement (or degradation) and should not be interpreted as “golden rules” (Marsh et al., 2004), but instead used alongside an evaluation of parameter estimates.

 There were missing data across the observed indicators as is common in field studies with a longitudinal component. Missing data were largely attributable to attrition rather than within-wave non-response. The amount of within-wave missingness was trivial, ranging from 0.00% to 0.30% (*M* = 0.13%) for Time 1 [T1] and 0.00% to 1.00% (*M* = 0.33%) for Time 2 [T2]. Conversely, approximately 48% of the sample completed the follow-up measure of engagement. For the single-sample and multiple group tests of the competing latent structures without covariates based on the Time 1 [T1] data, pairwise present methods—the default under WLSMV estimation in Mplus when no covariates are included (Asparouhov & Muthén, 2010)—were used to manage the missing data. In the MIMIC analyses of gender effects and tests of the relations of the engagement dimensions with job satisfaction, including teaching level and gender as covariates, based on T1 data, the trivial missingness was permitted to be a function of the respective observed covariates. Finally, for the longitudinal tests of invariance, covariance stability and relations of the engagement dimensions, based on both waves of data, missingness was allowed to be a function of the observed covariates (i.e., gender and teaching level).

**Results**

**Latent Structure**

 Results of the tests of the measurement models are shown in Table 1. The test of the unidimensional model common to both CFA and ESEM specifications resulted in an unacceptable fit to the data. Of the CFA structures, the bifactor model provided the best fit; however, the fit of this model was comparable, particularly in terms of the parsimony-corrected indices, to the higher-order CFA. The bifactor CFA provided a higher degree of fit to the data than the correlated four-factor CFA, with changes in the CFI and RMSEA approaching appreciable levels. These results suggest the possibility of a general factor underlying the ETS data in addition to specificities. All three ESEM models provided a better fit to the data than their CFA analogues in terms of the corrected chi-square different test. However, inferences based on changes in fit indices were less clear. Whereas changes in the CFI were suggestive of appreciable or near appreciable improvement in fit of the bifactor (ΔCFI = +0.009), higher-order (ΔCFI = +0.008), and correlated-four-factor ESEM models (ΔCFI = +0.010) relative to their CFA analogues, changes in the RMSEA indicated a near-appreciable improvement in fit for only the bifactor ESEM model (ΔRMSEA = -0.011) over its CFA counterpart. Although the weight of evidence suggests that the ESEM structures provide a greater level of fit to the data than the corresponding CFA models, it is instructive to evaluate parameter estimates across the analytic approaches in the model retention process.

INSERT TABLE 1 ABOUT HERE

 In the general framework for testing construct-relevant multidimensionality in data (Morin et al., 2016; Perera et al., 2015), an important initial comparison is between estimates obtained from the correlated-four factor CFA and ESEM models. In general, the correlated-four-factor CFA loading estimates (|λ| = .643-.927, *M* = .829) were stronger than the corresponding ESEM target loadings (|λ| = .488-.919, *M* = .777). For the ESEM model, non-target loadings (|λ| = .001-.312, *M* = .076) were systematically smaller than target loadings, but there were sufficient non-trivial cross-loadings to suggest the presence of construct-relevant multidimensionality due to indicator fallibility. Although the construct-relevant multidimensionality reflected in these cross-loadings is not as substantial as observed in other motivation and engagement data (see e.g., Guay et al., 2015; Howard et al., 2016; Perera & Ganguly, 2016), the cross-loadings are sufficiently non-trivial to generate biases if fixed to zero as in the CFA model (Asparouhov, Muthen, & Morin, 2015). This inference is supported by systematically higher factor correlations in the correlated four-factor CFA (|*r*| = .474-.776, *M* = .614) relative to its ESEM analogue (|*r*| = .446-.657, *M* = .534). These findings, taken with the model fit results, support the superiority of the first-order ESEM representation of the ETS data relative to the CFA structure, suggesting the presence of construct-relevant multidimensionality due to indicator fallibility that necessitates adequate statistical control.

 The ESEM models were compared to determine the most appropriate solution. The test of the higher-order ESEM did not result in a decrement in fit relative to the more complex correlated-four-factor ESEM. In fact, practical fit indices controlling for parsimony indicated better fit of the higher-order ESEM, suggesting that the gain in parsimony with imposing greater constraints in the higher-order ESEM outweighs any loss in fit. Based on parsimony, the higher-order model should be preferred over the correlated four-factor ESEM. However, substantively, the higher-order ESEM is less desirable given the theoretically dubious proportionality constraints imposed on the model. Thus, parsimony alone should not be the only criterion for model selection. For the correlated-four-factor ESEM, although factor correlations were lower than corresponding estimates in the CFA analogue and more consistent with the multidimensionality perspective undergirding current views on teacher engagement, there were five instances of reasonably large (|λ| ≥ .150) cross-loadings across three of the four factors. A pattern of even slightly inflated cross-loadings may emerge in correlated first-order ESEM models where some global or general factor is unmodeled (Perera et al., 2016). In the first-order ESEM, generality underlying all items can be reasonably efficiently absorbed by inflated cross-loadings (Morin et al., 2016). The bifactor ESEM model can accommodate this construct-relevant multidimensionality due to the purported presence of a general factor underlying items in addition to systematic specificities captured by group or specific factors. Notably, unlike the higher-order structure, which also posits the presence of a global factor, the bifactor structure does not impose restrictive proportionality constraints, rendering the model more plausible from the perspective of contemporary theories of engagement.

 The BF-ESEM provided a good fit to the data, and an appreciable improvement in fit relative to the correlated four-factor ESEM (e.g., ΔRMSEA = -.019) and near-appreciable improvement relative to the higher-order ESEM (ΔRMSEA = -.012). The factor loading estimates from this solution are shown in Table 2. The general-factor was well-defined with uniformly moderate to strong and statistically significant loadings (|λ| = .334-.812, *M* = .643). The items designed to specifically index cognitive-physical engagement showed particularly strong loadings on the general factor (|λ| = .666-.812, *M* = .768). Indicators of emotional engagement (|λ| = .583-.645, *M* = .612) and engagement with students (|λ| = .572-.781, *M* = .678) and colleagues (|λ| = .334-.615, *M* = .515) also showed sizeable general-factor loadings. Beyond the general-factor, target loadings on the specific-factors (|λ| = .158-.702, *M* = .506) were systematically larger than cross-loadings (|λ| = .000-.236, *M* = .065). This suggests that the specific engagement factor possesses sufficient specificity above and beyond the general-factor. Indeed, the emotional engagement (|λ| = .553-.668, *M* = .634) and engagement with students (|λ| = .383-.645, *M* = .505) and colleagues (|λ| = .514-.702, *M* = .603) specific-factors were well-defined with uniformly moderate to strong target factor loadings. The cognitive-physical engagement specific-factor was less-well-defined (|λ| = .158-.469, *M* = .280); however, three of the four specific target factor loadings exceeded .200, suggesting adequate specificity beyond the general-factor. Importantly, specification of the general-factor in the bifactor-ESEM resulted in a slight decrease in the strength of cross-loadings relative to the correlated four-factor ESEM (|λ| = .001-.312, *M* = .076). This pattern of slightly reduced cross-loadings suggests that construct-relevant multidimensionality due to the co-existence of general and specific constructs underlying the engagement data, which was expressed as slightly inflated non-target loadings in the correlated-factor ESEM, is appropriately re-expressed via the general-factor in the bifactor-ESEM representation (Arens & Morin, 2016). The bifactor-ESEM was retained for further analysis.

INSERT TABLE 2 ABOUT HERE

**Teaching-Level Invariance**

 Table 3 shows the fit statistics for the tests of invariance across teaching level. The configural invariance model (MGM1) provided an excellent fit to the data. This baseline model was compared to the more restrictive weak-invariance model (MGM-2) in which the factor loadings were constrained to equality across the secondary and primary teacher groups. The test of this model resulted in a good fit to the data and, importantly, no appreciable decrement in fit relative to the configural model. The TLI and RMSEA increased and decreased, respectively, indicating that the gain in parsimony with the additive equality constraints imposed on the factor loadings offsets any loss in fit. In addition, support was found for the invariance of item thresholds (MGM-3), uniquenesses (MGM-4), and the factor variance-covariance matrix (MGM-5).[[1]](#footnote-1) However, the test of the model of factor mean invariance (MGM-6) resulted in an appreciable decrement in fit relative to MGM-5 in terms of changes in RMSEA (ΔRMSEA +.016) and near appreciable decrement with respect to changes in the CFI (ΔCFI = -.006), suggesting that latent means may be appreciably different across the groups. Accordingly, we retained MGM-5 for further examination of latent mean differences. Inspection of MGM-5 revealed that latent means for primary teachers, expressed as SD units from the secondary-group means (i.e., fixed to 0 for model identification), were significantly higher for general engagement (*M* = .410, *p* < .001) and engagement with students (*M* = .442, *p* < .001), but not significantly different for emotional engagement (*M* = -.207, *p* >.05), engagement with colleagues (*M* = -.191, *p* > .05), cognitive engagement (*M* = -.141, *p* *>* .05).

INSERT TABLE 3 ABOUT HERE

**Latent Mean Differences over Gender**

 Next, latent mean differences in the teacher engagement dimensions across gender and its interaction with teaching level were examined in the hybrid MG-MIMIC model. The final model of teaching level invariance was used as the baseline structure for these tests. The test of the saturated model resulted in an excellent fit, χ2 (246) = 301.950, *p* < .01, CFI = .996, TLI = .996, RMSEA = .028, 95% CI [.015, .038]. The threshold-invariant MIMIC model also provided a good fit to the data, χ2 (268) = 327.895, *p* < .01, CFI = .996, TLI = .996, RMSEA = .027, 95% CI [.015, .037], and, notably, no decrement in fit relative to the saturated model, MD χ2 (22) = 28.975, *p* > .05, ΔCFI = .000, ΔRMSEA = -.001. Finally, we tested an equality constrained model in which the effects of gender were constrained to be equal across the teaching level groups. This constrained solution provided an excellent fit to the data, χ2 (273) = 337.156, *p* < .01, CFI = .995, TLI = .995, RMSEA = .028, 95% CI [.016, .038], and, importantly, no degradation in fit relative to the threshold-invariant model, MD χ2 (5) = 9.640, *p* > .05, ΔCFI = -.001, ΔRMSEA = +.001. These findings suggest that there are no differences in the effects of the covariates on the engagement dimensions across teaching level. A significant effect of gender was found for the specific emotional engagement factor (γ = -0.359, *p* < .05), such that female teachers reported lower positive emotional responses to and experiences with teaching. No significant effects of gender on the general engagement (γ = 0.110, *p* > .05), specific cognitive-physical engagement (γ = 0.291, *p* > .05), and specific engagement with students (γ = -0.077, *p* > .05) and colleagues (γ = -0.028, *p* > .05) dimensions were found.

**Relations with Work Satisfaction**

A general LVM was specified to examine the relations of the engagement dimensions, per the retained BF-ESEM structure, with job satisfaction. Structural paths from each of the engagement dimensions to job satisfaction were freely estimated. This general LVM model provided am excellent fit to the data, χ2 (160) = 296.727, *p* < .01, CFI = .989, TLI = .983, RMSEA = .038, 95% CI [.016, .038]. Expressed as standardized estimates, the G-factor (γ = .411, *p* < .001) and emotional engagement (γ = .580, *p* < .001), and engagement with colleagues (γ = .168, *p* < .001) S-factors were positive predictors of job satisfaction. The engagement with students S-factor was significantly and negatively associated with job satisfaction (γ = -.096, *p* < .05). Specific cognitive-physical engagement was not significantly associated with job satisfaction (γ = -.071, *p* > .05). In totality, the engagement dimensions explained 55.0% of the variance in job satisfaction.

**Longitudinal Invariance, Stability, and Relations**

 Results of the tests of the longitudinal invariance of the bifactor ESEM engagement model are shown in Table 4. The baseline model (LI-M1) provided an excellent fit to the data. The baseline model was compared to the more restrictive loading invariance (LI-M2) in which the factor loadings were constrained to equality across time. This model provided a good fit to the data and, importantly, no appreciable decrement in fit relative to the baseline model. Next, we tested the more restrictive threshold invariance model (LI-M3) against the loading invariance structure. The test of the threshold-invariance model resulted in an excellent fit to the data and no degradation in fit as compared with the LI-M2. Finally, support was also found for longitudinal unique factor invariance as the model constraining the item uniquenesses to equality over time did not fit the data appreciable worse than the threshold-invariance model (see Table 4). Taken together, these results support the full longitudinal measurement invariance of the bifactor ESEM structure.

INSERT TABLE 4 ABOUT HERE

 The uniqueness longitudinal invariance model was reparameterized as

a ACL panel model, with all autoregressive and cross-lagged paths among the engagement dimensions. In this regard, the fully-latent ACL bifactor ESEM is equivalent to the corresponding measurement model in terms of degrees of freedom (i.e., the model is structurally saturated). Table 5 shows the results of the ACL model. Not surprisingly, the autoregressive effects were the strongest structural coefficients obtained, suggesting reasonable stability in teachers’ engagement over the five-month period. There are, however, several cross-lagged paths that are statistically significant and theoretically informative. T1 cognitive-physical engagement had a significant positive effect on general engagement at T2 and significant negative effect on specific engagement with students at T2. In addition, T1 emotional engagement had a significant negative effect on engagement with students at T2. No other significant results were obtained.

INSERT TABLE 5 ABOUT HERE

**Discussion**

 The present study represents a systematic attempt to examine the dimensionality of teacher engagement data and, in doing so, addresses the dearth of clarity about the theoretical and empirical structure of the teacher engagement construct. We examined the latent structure of ETS responses in a sample of practicing Australian teachers. We also tested the complete factorial invariance of the retained measurement structure of engagement across teaching level as well as latent mean differences in teacher engagement over gender. In addition, we examined the longitudinal invariance of the retained measurement structure across two-time points spanning a five-month period, and extended this examination to the investigation of the covariance stability and cross-lagged relations among the engagement dimensions. Finally, we examined the associations of the engagement dimensions, based on the retained measurement structure, with teachers’ work satisfaction. The results emerging from these analyses are discussed below with respect to engagement theory and extant evidence.

**Dimensionality of Teacher Engagement**

The results of the study support the BF-ESEM structure of engagement data. This model accounts for construct-relevant multidimensionality in engagement data due to (a) the fallibility of ETS items as indicators of the strictly unidimensional constructs they are purported to measure and (b) the presence of general and specific engagement constructs underlying responses. The first source of psychometric multidimensionality emerges in measures of constructs with closely-related conceptual dimensions, such as work engagement, and can be appropriately apportioned using first-order ESEM models (Morin et al., 2016). In the present study, the correlated-factors ESEM model yielded estimates of engagement factor correlations that were lower than corresponding ICM-CFA estimates, suggesting the presence of construct-relevant multidimensionality due to item fallibility. Indeed, in the ESEM structure, there were 24 statistically significant cross-loadings, of which 15 were non-trivial (|λ| ≥ .095). These cross-loadings are unsurprising given the conceptual relatedness between the engagement dimensions (Klassen et al., 2013). The restriction of the cross-loadings to zero in ICM-CFA models is a source of misspecification that may lead to not only lower model fit but also upwardly biased estimates of factor correlations (Asparouhov et al., 2016; Perera, 2015b) as observed in the present study as well as in previous research on ETS data (Klassen et al., 2013). The reduction in factor correlation estimates in the first-order ESEM model relative to its CFA counterpart provides strong support for the retention of ESEM structures (Morin et al., 2016). Notwithstanding the reasonable fit of, and theoretically-consistent factor correlations yielded by, the correlated-factor ESEM, there were several cross-loadings of an ample magnitude to suggest the possibility of a general engagement construct underlying the data. The correlated-factors ESEM cannot sufficiently account for the presence of generality underlying the data and, instead, redistributes this variance through slightly inflated cross-loadings (Perera et al., 2016).

Both higher-order and bifactor models can accommodate psychometric multidimensionality due to the presence of generality and specificity in engagement data. Although the higher-order ESEM model accounts for this second source of multidimensionality via the estimation of a global factor from the first-order engagement dimensions, and was shown to fit the data well in the present study as in previous research on the ETS (Klassen et al., 2013), the implicit proportionality restrictions imposed by this model render it a theoretically dubious structure for multidimensional engagement data. Indeed, from a theoretical standpoint, the perfectly proportional manner in which an indicator of engagement in a higher order model is assumed to contribute variance to global engagement and a specific factor disturbance across all indicators within a particular first-order factor cannot be justified. The theoretical dubiousness of the proportionality restrictions would seem to outweigh any advantages the higher-order model has in scientific parsimony. Yet, the higher-order conceptualization is a dominant representation of teacher engagement data (Klassen et al., 2013) and work-engagement data more broadly (Christian, Garza, & Slaughter, 2011). What is clear is that those who espouse a higher-order conceptualization of engagement in future work will need to theoretically justify the proportionality restrictions implied by higher-order models (Gignac, 2016).

 The bifactor-ESEM accommodates generality and specificity in data without imposing restrictive proportionality constraints, and, in the present study, provided an excellent fit to the engagement data. In the bifactor ESEM solution, the general factor was very well defined with all 16 standardized factor loadings exceeding .334, and 15 of the 16 at or exceeding .556. The strength of this general factor is remarkable given that the ETS items were designed to measure distinct dimensions of teachers’ work engagement in line with the multidimensionality perspective espoused by theories of work engagement (Bakker & Demerouti, 2008; Klassen et al., 2013). Items intended to specifically tap cognitive-physical engagement had particularly strong loadings on the general factor. Items measuring the remaining engagement dimensions also had moderate to strong loadings on the general factor, but these were generally weaker than the loadings from the cognitive-physical items. The pattern of loadings suggests that

teachers who are generally engaged with their work are, first and foremost, energetic and devote considerable cognitive resources to their work. This is consistent with the theoretical position that energy, involvement, and attention are the core components of work engagement (Bakker et al., 2011). Still, teachers’ perceptions of their engagement with students and colleagues, and their emotional responses to their work, constitute part of their perceptions of being generally engaged with work.

 Beyond the general engagement factor, there was non-trivial content specificity reflected in the specific engagement factors. For the emotional engagement, engagement with colleagues, and engagement with students S-factors, target loadings uniformly exceeded .350, suggesting considerable content specificity over and above the general factor. The cognitive-physical S-factor was less well-defined with loadings ranging from .158-.469 (*M* = .280). These generally weaker loadings are somewhat unsurprising given that the items indexing cognitive-physical engagement strongly loaded on the general factor, supporting the centrality of energy, involvement, and attention to teachers’ perceptions of general engagement with work (Bakker et al., 2011). However, the loadings are sufficiently non-trivial to suggest a minimal level of content specificity related to teachers’ sustained efforts to perform well while teaching.

The resultant bifactor-ESEM representation allows for the simultaneous consideration of teachers’ general level of engagement as well as their more specific perceptions of cognitive-physical engagement, emotional engagement, and engagement with students and colleagues. Notably, the orthogonality of the general and specific engagement dimensions implied in the bifactor ESEM suggests that individual domains of engagement may be experienced separately from one another and from perceptions of general engagement. For example, in practical terms, a teacher may perceive himself or herself as generally engaged with work but relatively unengaged with their colleagues or students. Similarly, a teacher may be engaged with their colleagues but perceive themselves as relatively unengaged with their students. The findings of a well-fitting model specifying generality and specify in teacher engagement data converge with prior work showing that a bifactor structure provides a reasonable account of engagement data in workers from the healthcare sector (Fong & Ho, 2015). Indeed, the retained bifactor-ESEM structure reconciles the extant multidimensionality perspective on work engagement (Klassen et al., 2013) with empirical findings supporting the generality of engagement data (references). These dimensionality results have important implications for theories of work engagement in suggesting that individuals’ general and specific perceptions of engagement may be experienced differentially and independent of one another.

Accounting for the general-factor in the bifactor-ESEM model resulted in weaker estimates of non-target loadings relative to the correlated-factors-ESEM. The reduction in the magnitude of these cross-loadings should be expected as any true general factor underlying an item set will be expressed via (inflated) cross-loadings in measurement models that do not explicitly control for generality (Perera et al., 2016). Yet, the small cross-loadings are theoretically meaningful, reflecting construct-relevant multidimensionality due to indicator fallibility, and serve to inform the estimation of the S-factors. For instance, Item 3 (“In class, I show warmth to my students”)—an indicator of engagement with students—cross-loaded non-trivially on the cognitive-physical engagement S-factor, reflecting the possibility that teachers’ perceptions of involvement, attention, and effort towards performing well while teaching implicate their perceptions of the extent to which they respond appropriately to students (O’Connor, 2008). Similarly, Item 1 (“At school, I connect well with my colleagues”)—an indicator of engagement with colleagues—cross-loaded non trivially on emotional engagement, which may reflect the possibility that teachers’ emotional responses to their work are informed by their relationships with co-workers. These cross-loadings serve to enhance construct estimation by using all available indicator information.

**Invariance and Latent Mean Differences**

 Support was found for the complete measurement invariance of the ETS ratings implied by the bifactor ESEM model across teaching level. Support was obtained for the equivalence of item loadings, thresholds, and factor uniquenesses suggesting, respectively, that (a) the extent to which the ETS items reflect the engagement dimensions is equivalent across primary and secondary teachers, (b) holding levels of the engagement dimensions constant, items have the same expected responses across teaching levels, and (c) the amount of ETS item variance unaccounted for by the engagement dimensions is equal across the groups. These results support the validity of scores obtained from the ETS by showing the generalizability of the retained latent structure across substantively meaningful subgroups. Importantly, the results suggest that the ETS functions equivalently across primary and secondary teachers.

Evidence was obtained for the invariance of latent variances and covariances across the teaching levels; however, meaningful differences in the latent means were obtained. Primary teachers reported significantly higher general engagement and specific engagement with students than their secondary counterparts. These findings replicate previous research showing that engagement is higher among teachers in lower grade levels (e.g., primary, elementary school levels) (Byrne, 1994; Rey et al., 2012). These effects of teaching level may be attributed to lower class sizes and greater constancy in primary school classes, which may allow primary teachers to invest greater physical, attentional, and affective resources in facilitating student learning and developing connections with colleagues and students ((Biddle & Berliner, 2002; Byrne, 1991; Blatchford, Bassett, & Brown, 2011), leading to inferences of greater general engagement with their work. Furthermore, it may be that characteristics of secondary students as compared with primary students, such as lower academic motivation and interest (Dotterer, McHale, & Crouter, 2009; Fredricks & Eccles, 2002; Wigfield & Eccles, 1994), and greater problem behavior (Vandenberghe & Huberman, 1999) place secondary teachers at a greater risk of general disengagement than primary teachers as physical, attentional, and affective resources are diverted away from instructional provisions and forging connections with colleagues and students to classroom management issues (Rey et al., 2012). Notably, the present findings extend prior work by demonstrating effects of teaching level on specific engagement with students after appropriately disentangling variance in general engagement from the specific engagement constructs. Greater time spent with students in primary classes and the greater emphasis on basic caregiving in primary schools may require primary teachers to be more aware of students’ feelings and show greater empathic understanding. Furthermore, primary teachers tend to report higher self-efficacy for student engagement (Klassen & Chiu, 2010; Wolters & Daugherty, 2007), which may lead them to perceive greater engagement with their students.

 Evidence was also obtained for a plausible latent mean difference in engagement as a function of gender. Female teachers were lower in specific emotional engagement, which was found to generalize across the teaching levels. This finding is consistent with previous research showing that female teachers report greater emotional exhaustion and workload stress than male teachers (Antoniou et al., 2006; Klassen & Chiu, 2010; Lau et al., 2005). Lower positive emotional engagement in females may be attributed to gender differences in domestic caring responsibilities, which may impose demands on female teachers in addition to work demands (Byrne, 1991; Greenglass &Burke, 2003). These dual demands—or as Byrne (1991) notes “double dose of caring” (p. 205)—may lead to female teachers’ emotional disengagement from work (Byrne, 1991; Grayson & Alvarez, 2008).

 The current research supported the longitudinal measurement invariance of the bifactor ESEM. Although the examination of invariance of measurement models over discrete groups is widely applied in SEM construct validity studies of motivation and engagement constructs, a less widely investigated issue is longitudinal measurement invariance. Yet, longitudinal invariance is crucial to ensuring that any changes in constructs over time reflect changes in the levels of the latent dimensions rather than the measurement instrument. In the present study, support was found for the invariance of factor loadings, thresholds, and uniquenesses across two time points separated by a five-month interval. Findings supporting the unique factor invariance model suggest that changes over time in expected means, variances, and within-wave covariances of the continuous latent response variates underlying the ETS indicators are attributable to the latent common factors over time. In addition, changes over time in the expected means and bivariate probabilities of the ETS indicators are attributable to the latent common factors over time (Liu et al., 2016).

 **Engagement and Work Satisfaction**

 A notable advantage of the bifactor-ESEM representation retained in the present study is the flexibility to model simultaneous relations of the general and specific engagement dimensions with substantively important outcomes.

Replicate and extend previous results reporting relations between teacher engagement and job satisfaction.

**Limitations and Future Directions**

Examined associations based on cross-sectional data. Says something about the directional ordering of engagement and job satisfaction and need for further work on this relation.

**Conclusion**

In summary, the results of this study show that the best representation of teacher engagement data is a BF-ESEM structure that accounts for two sources of construct-relevant multidimensionality common in multidimensional measures of motivation and engagement. In addition, evidence was obtained for the full measurement invariance of the bifactor ESEM structure across teaching level; however, some theoretically-consistent differences in mean levels of the constructs were obtained. Evidence was also obtained for a plausible gender difference in emotional engagement. Notably, evidence was acquired for the longitudinal invariance of the retained measurement structure as well as plausible auto-regressive and cross-lagged effects among the engagement dimensions. Finally, data were obtained supporting the validity of the engagement data, based on the retained measurement model, for predicting teachers’ work satisfaction.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | *χ*² | *df* | CFI | TLI | RMSEA | RMSEA 90% CI | MD *χ*² (df) |
| Unidimensional | 2592.871 | 104 | .807 | .777 | .201 | [.194, .207] |  |
| Four-factor CFA | 412.767 | 98 | .976 | .970 | .074 | [.066, .081] |  |
| Higher-Order CFA | 361.799 | 100 | .980 | .976 | .066 | [.059, .074] |  |
| Bifactor CFA | 299.004 | 88 | .984 | .978 | .064 | [.056, .072] |  |
| Four-factor ESEM | 250.894 | 62 | .985 | .972 | .072 | [.062, .081] | 220.641b (40)\*\*\* |
| Higher-order ESEMa | 223.477 | 64 | .988 | .977 | .065 | [.056, .074] | 167.975 (36)\*\*\* |
| Bifactor ESEM | 134.255 | 50 | .993 | .984 | .053 | [.042, .064] | 166.565 (38)\*\*\* |

Table 1

*Model Fit Statistics for the ICM-CFA and ESEM Measurement Structures of Teacher Engagement*

Note. *N* = 594. ICM-CFA = Independent clusters model of confirmatory factor analysis; ESEM = exploratory structural equation modeling; *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; 90% CI = 90% confidence interval for the RMSEA. a The higher-order ESEM specification was conducted in an EwC framework. b The nested model comparisons are between the ESEM structures and their CFA analogues. For these comparisons, the CFA models are nested in their ESEM counterparts.

Table 3

*Fit Statistics for the Multiple-Group Models of Teaching-Level Invariance*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | χ² | *df* | CFI | TLI | RMSEA | 90% CI | MD χ2 | ΔCFI | ΔRMSEA |
| MGM1 (Configural IN) | 166.561\*\*\* | 100 | .995 | .988 | .047 | [.034, .060] |  |  |  |
| MGM2 (IN FL) | 251.714\*\*\* | 155 | .993 | .989 | .046 | [.035, .056] | 100.142 (55)\*\*\* | -.002 | -.001 |
| MGM3 (IN FL + Th) | 299.513\*\*\* | 215 | .994 | .993 | .036 | [.026, .046] | 75.699 (60) | +.001 | -.010 |
| MGM4 (IN FL + Th + Uniq) | 313.300\*\*\* | 231 | .994 | .994 | .035 | [.024, .044] | 21.308 (24) | .000 | -.001 |
| MGM5 (IN FL + Th + Uniq + FVCV)  | 299.036\* | 246 | .996 | .996 | .027 | [.014, .037] | 22.903 (15) | +.002 | -.008 |
| MGM6 (IN FL + Th + Uniq + FVCV + FM) | 390.570\*\*\* | 251 | .990 | .990 | .043 | [.035, .051] | 56.636 (20)\*\*\* | -.006 | .016 |

*Note*. *df* = degrees of freedom; MD χ2 = change in χ2 relative to a more complex model computed using the Mplus DIFFTEST function; ΔCFI = change in comparative fit index; ΔRMSEA = change in root mean square error of approximation; MGM = multiple-group model; IN = invariance; FL = factor loadings; Th = Thresholds; Uniq = uniquenesses; Corr Uni = Correlated uniquenesses; FVCV = Factor variances and covariances; FM = Factor means. \* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001.

Table 4

*Fit Statistics for the Longitudinal Invariance Models*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | χ² | *df* | CFI | TLI | RMSEA | 90% CI | MD χ2 | ΔCFI | ΔRMSEA |
| LI-M1 (Baseline) | 540.608\*\*\* | 359 | .991 | .986 | .029 | [.024, .034] |  |  |  |
| LI-M2 (IN FL) | 578.138\*\*\* | 414 | .992 | .989 | .026 | [.021, .031] | 74.187 (55)\* | +.001 | -.003 |
| LI-M3 (IN FL + Th) | 628.177\*\*\* | 473 | .992 | .991 | .023 | [.023, .028] | 69.866 (59) | .000 | -.003 |
| LI-M4 (IN FL + Th + Uniq) | 677.943\*\*\* | 489 | .991 | .989 | .025 | [.021, .030] | 49.493 (16)\*\*\* | -.001 | +.002 |

*Note*. *df* = degrees of freedom; MD χ2 = change in χ2 relative to a more complex model computed using the Mplus DIFFTEST function; ΔCFI = change in comparative fit index; ΔRMSEA = change in root mean square error of approximation; MGM = multiple-group model; IN = invariance; FL = factor loadings; Th = Thresholds; Uniq = uniquenesses. \* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001.

1. Increases in CFI with increasingly restrictive models should be viewed as haphazard because, under WLSMV, which generates an adjusted χ2 test statistic to obtain appropriate *p*-values, the CFI (and χ2) can be non-monotonic with model complexity (Morin, Arens, Tran, & Caci, 2015; Perera et al., 2016). [↑](#footnote-ref-1)