Curve of Factors Model: A Latent Growth Modeling Approach for Educational Research

Educational and Psychological
Measurement
2018, Vol. 78(2) 203–231
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0013164416677143
journals.sagepub.com/home/epm

Marilu Isiordia¹ and Emilio Ferrer¹

Abstract

A first-order latent growth model assesses change in an unobserved construct from a single score and is commonly used across different domains of educational research. However, examining change using a set of multiple response scores (e.g., scale items) affords researchers several methodological benefits not possible when using a single score. A curve of factors (CUFFS) model assesses change in a construct from multiple response scores but its use in the social sciences has been limited. In this article, we advocate the CUFFS for analyzing a construct's latent trajectory over time, with an emphasis on applying this model to educational research. First, we present a review of longitudinal factorial invariance, a condition necessary for ensuring that the measured construct is the same across time points. Next, we introduce the CUFFS model, followed by an illustration of testing factorial invariance and specifying a univariate and a bivariate CUFFS model to longitudinal data. To facilitate implementation, we include syntax for specifying these statistical methods using the free statistical software R.

Keywords

curve of factors model, latent growth models, longitudinal data analysis

Assessing change in an unobservable construct (e.g., motivation, academic interest, peer relations) over time is essential for understanding developmental processes in education. Under a structural equation modeling (SEM) framework, latent growth modeling (LGM) has become a prominent method among educational researchers for

¹University of California, Davis, CA, USA

Corresponding Author:

Marilu Isiordia, Department of Psychology, University of California, Davis, CA 95616, USA.
Email: misiordia@ucdavis.edu

evaluating trajectories of such constructs (Marsh & Hau, 2007). However, the majority of work in educational research involving these longitudinal methods has focused on first-order latent growth models (1LGMs). In 1LGMs, change is assessed from a single composite score (e.g., total sum or averaged scale score) that is used to represent the construct of interest at a given measurement occasion (Biemer, Christ, & Wiesen, 2009). Unfortunately, analyzing change with composite scores and 1LGMs limits assessment of important data characteristics such as measurement invariance, partitioning of time-specific and item-residual variance, and detecting varying item-residual covariance patterns (Bishop, Geiser, & Cole, 2015; Sayer & Cumsille, 2001).

When multiple scale items are available at each measurement occasion, the aforementioned analytical limitations of a 1LGM can be addressed with a second-order growth model, specifically the curve of factors model (CUFFS; McArdle, 1988). An extension of the 1LGM, the CUFFS model characterizes the relation between the multiple items and their underlying construct at each time point, as well as the construct's growth trajectory. Although the CUFFS model was introduced almost 30 years ago and despite the analytical advantages it offers over the 1LGM, its application in educational research and the social sciences has been limited (Geiser, Keller, & Lockhart, 2013). Consequently, the benefits of the CUFFS model remain unfamiliar to many educational researchers working with longitudinal data.

The main objective of this article is to advocate the use of the CUFFS model for modeling trajectories of latent constructs in educational research. We first provide a review of measurement invariance in longitudinal designs within the SEM framework, termed *factorial invariance*. Factorial invariance is a necessary condition for evaluating change in the construct's mean across repeated measurement occasions (Horn & McArdle, 1992). Next, we describe the CUFFS model, emphasizing its methodological advantages. We conclude with an empirical example of these methods, illustrating the use of univariate and bivariate CUFFS models.

Longitudinal Measurement Invariance

Constructs such as students' motivation in school, self-perceived competence, affect, and interest in reading, are not directly observable. Assessment of these unobserved variables typically requires researchers to use scales with various items that represent the intended construct (Riley, 1963). In longitudinal studies, researchers commonly use the same scales to ensure that the same construct is being measured over time. However, using the same scale does not guarantee that the same construct is being represented across measurement occasions. For example, as participants get older (or, say, the nature of the assessment is different across occasions), the respondents' interpretations of the various items might change. Thus, the construct being measured is no longer the same as the one measured in previous occasions (Embretson, 2007). Consequently, any changes over time might actually be due to changes in the factor structure rather than to growth changes in the construct (Horn & McArdle,

1992). In other words, the observed changes are qualitative, rather than quantitative, in nature. Therefore, measurement invariance needs to be demonstrated in order to ensure that the same construct is being represented across repeated measurements (Meredith, 1964, 1993).

Distributional assumptions of the data determine which analytical framework is appropriate to evaluate measurement invariance. For instance, when dealing with categorical data, measurement invariance is typically tested using item response theory and is termed *differential item function*. For continuous data, measurement invariance is assessed within the SEM framework and is known as factorial invariance. In this article, we focus on the latter framework for testing measurement invariance over time. Information about approaches for differential item functioning is available elsewhere in the literature (e.g., Camilli & Shepard, 1994).

Evaluation of factorial invariance is possible when multiple items representing a construct are available at each time point. For example, suppose that a set of three items in a scale is used to measure students' perceived competence in verbal ability in first grade, fifth grade, and 10th grade:

$$Y_{jti} = \tau_{jt} + \lambda_{jt}f_{it} + e_{jti}, \quad (1)$$

where Y_{jti} is the response to survey item j at measurement occasion t that varies across individuals i and is a function of an intercept τ , a factor loading λ linking the item response score to the common construct or factor f_{it} , and residual variance e_{jti} .

To test whether this set of scale items is measuring the same construct (i.e., perceived competence in verbal ability) across school grades, a series of models with increasing invariance constraints is assessed using confirmatory factor analysis (CFA; Widaman & Reise, 1997). The first model assesses configural invariance, which requires that the factor structure is represented by the items in the same manner across repeated measurements. According to this specification, a latent construct comprises the same items at all time points, irrespective of any numerical value. The path diagram in Figure 1A represents this level of invariance. In this figure, we use standard SEM path diagram notation (McArdle & McDonald, 1984).

The next level of invariance is weak invariance, in which the factor loadings for each items are specified to be the same across measurement occasions (Figure 1B; $\lambda_{11} = \lambda_{12} = \lambda_{13}$, $\lambda_{21} = \lambda_{22} = \lambda_{23}$, and $\lambda_{31} = \lambda_{32} = \lambda_{33}$). Weak invariance ensures that the observed variables measure the common construct in the same quantitative manner over time. A strong invariance model additionally requires equality constraints on the corresponding item intercepts across all times of measurement (Figure 1C¹; $\tau_{11} = \tau_{12} = \tau_{13}$, $\tau_{21} = \tau_{22} = \tau_{23}$, and $\tau_{31} = \tau_{32} = \tau_{33}$). Establishing strong invariance indicates that mean changes are implied at the construct level and not in the item intercepts. In other words, any changes detected at the factor level reflect differences in the students' perceived verbal ability across school grades, and not in their interpretation of the measurement tool over time. Finally, a strict invariance model is fit, which additionally requires equality constraints on the item residual variances across time (Figure 1D; $e_{11} = e_{12} = e_{13}$, $e_{21} = e_{22} = e_{23}$, and $e_{31} = e_{32} = e_{33}$). With strict

Figure 1. A series of path diagrams of a confirmatory factor analysis (CFA) model for one construct across three occasions. (A) depicts a configural invariance model with minimum constraints; (B) depicts a weak invariance model with factor loading equality constraints; (C) depicts a strong invariance model with factor loadings and intercepts equality constraints; and (D) depicts a strict invariance model with factor loadings, intercepts, and residual variance equality.

invariance, the entire function relating the observed variables to the common construct is equivalent across time. However, strong invariance is sufficient for comparing changes in the construct mean over measurement occasions (Meredith, 1964, 1993). Meeting factorial invariance criteria ensures that the qualitative and quantitative nature of the construct remains the same across occasions (Horn & McArdle, 1992). In other words, demonstrating factorial invariance ensures that students' interpretation of the scale items continues to be about perceived competence in verbal ability in 1st, 5th, and 10th grades, and not about, say, reading motivation or interest in school learning.

Modeling Growth Trajectories

In educational research, most theoretical and applied work focused on assessing a construct's trajectory over time has used 1LGMs. This is the case, for instance, in research assessing change in student achievement (Bianconcini & Cagnone, 2012), self-efficacy (Phan, 2013), abilities and skill acquisition (Voelkle, Wittmann, & Ackerman, 2006), problem-based learning (Wimmers & Lee, 2015), bullying perpetration and peer victimization (Turner, Reynolds, Lee, Subasic, & Bromhead, 2014), and career preparation and adjustment from high school to early adulthood (Stringer, Kerpelman, & Skorikov, 2011). To get a sense of how often such research used 1LGMs versus multivariate models, we reviewed 100 articles in the Education Resources Information Center database (U.S. Department of Education, 2016) from 2005 to 2015 by relevance using the following search terms: *applied longitudinal data analysis*, *latent growth model*, *curve of factors model*, and *multiple-item latent growth model*. None of the studies in which multiple items were available tested longitudinal factorial invariance or used a CUFFS model to measure change over time. Instead, all 100 studies assessed scale reliability, made composite scores, and used 1LGMs. As described previously, 1LGM does not provide researchers the methodological benefits inherent in modeling the relations between the multiple items and the underlying construct (Sayer & Cumsille, 2001).

A more appropriate way to model a construct's trajectory when multiple items are available at each time of measurement is with a CUFFS model. The CUFFS model combines a measurement model and a growth model in one single specification. Once strong invariance is demonstrated, researchers can assess a construct's latent trajectory from the measurement model's common factors at each time point. For instance, the extracted first-order factors from the measurement model in Equation (1) can vary across individuals i and are a function of a latent growth process:

$$f_{ti} = f_{0i} + \beta_t f_{si} + \sigma_{f_{ti}}^2. \quad (2)$$

Each common factor f_{ti} is a function of the level or intercept f_{0i} , the slope f_{si} that is scaled by a basis parameter that determines the shape of the curve β_t , and a factor disturbance parameter $\sigma_{f_{ti}}^2$ that represents reliable time-specific variability at time

Figure 2. A path diagram of a curve of factors model (CUFFS) across three measurement occasions. Factor loadings, intercepts, and residual variances are represented as invariant across measurement occasions.

measurement occasion t for individual i . The intercept and slope factors can be expressed as

$$f_{0i} = \mu_{0i} + \sigma^2_{f0} \tag{3}$$

and

$$f_{Si} = \mu_{Si} + \sigma^2_{fS}. \tag{4}$$

Figure 2 depicts a path diagram representing a CUFFS model. The mean structure for the intercept and slope provides information about the participants' average initial level μ_{0i} and average change over time μ_{Si} . Estimating the variance of the intercept σ^2_{f0} and slope σ^2_{fS} will provide information on individual differences in these parameters. The association between a construct's intercept and slope over time can also be estimated σ_{f0s} .

The methodological advantages of the CUFFS model over the 1LGM have been previously documented (Ferrer, Balluerka, & Widaman, 2008; Geiser et al., 2013;

Leite, 2007; Murphy, Beretvas, & Pituch, 2011; Sayer & Cumsille, 2001; von Oerzen, Hertzog, Lindenberger, & Ghisletta, 2010; Widaman, Ferrer, & Conger, 2010). We summarize the most important strengths of the CUFFS model. One, once factorial invariance is shown to hold in the data, researchers can conveniently use the existing measurement models to assess growth from the common factors (Horn & McArdle, 1992). Two, the measurement component of the CUFFS model preserves the measurement characteristic between the items and the common construct (Sayer & Cumsille, 2001; Widaman et al., 2010). Three, the CUFFS model partitions observed score variance into reliable construct variance σ_{f0}^2 and σ_{fs}^2 , reliable time-specific variance σ_{fi}^2 , and residual variance σ_e^2 . In a 1LGM, reliable time-specific variance and residual variance are confounded (Ferrer et al., 2008). Given that in a CUFFS model, the residual variance is parsed out of the common construct over time, a more accurate interpretation of the latent trajectory is possible. That is, using a CUFFS model prevents an error-saturated observed score that may result from 1LGM, and instead yields a theoretically error-free construct for evaluating a growth trajectory (Hancock, Kuo, & Lawrence, 2001). Four, the CUFFS model has been shown to have greater statistical power for detecting individual differences across occasions (von Oerzen et al., 2010). Five, because the CUFFS model uses multiple items to evaluate change over time, different item-specific variance-covariance patterns, also known as error structures, can be specified (Grilli & Varriale, 2014). A particular error structure may be warranted, for example, when there is a shared method (e.g., specific items answered by only students, teachers, or parents) within and across time points. Furthermore, specification of an appropriate error structure increases model convergence and fit between the hypothesized model and observed data (Liu, Rovine, & Molenaar, 2012).

To advocate the practice of testing for longitudinal factorial invariance and using the CUFFS model, we provide an empirical example next.

Empirical Example

Data for this article are from the “Motivation in High School Project” (Ferrer & McArdle, 2003), a study aimed at assessing a range of self-perceptions among high school students. Participants in this study included 261 adolescent students from three high schools (mean age of 14.4 years, $SD = 0.57$; 149 males and 112 females) who identified themselves as White (68.8%), Black (18.8%), Asian (1.3%), Hispanic (2.7%), Native American (0.4%), or Other (0.7%). Data collection involved four measurement occasions, with the first assessment taking place in the first week of school and subsequent observations measured at succeeding 6-week intervals. Of the 261 students, 144 completed all scale items at all four time points. In the original study, Ferrer and McArdle (2003) found that students with complete data had marginally higher motivation scores than students with lower motivation scores at the first and last measurement occasion; however, no other differences were found.

Measures

Students completed the Self-Perception Profile for Adolescents (Harter, 1988) scale to report their perceived physical competence and global self-worth. To measure perceived physical competence, four items were used on a 4-point Likert-type scale (1 = *Not true for me* to 4 = *Really true for me*). A representative item from this scale is "Some teenagers feel that they are better than others their age at physical activities BUT other don't feel they can do as well." For each time point, coefficient alpha reliability estimates for this scale were .83, .83, .86, and .86. For global self-worth, four scale items on a 4-point Likert-type scale (1 = *Not true for me* to 4 = *Really true for me*) were used. A representative item from this scale is "Some teenagers are often disappointed with themselves BUT other teenagers are pretty pleased with themselves." For the Global Self-Worth scale, coefficient alpha estimates for each time point were .83, .77, .84, and .83.

We selected these study variables for the analyses for two reasons. First, factorial invariance was demonstrated for each of these scales, which is not only a requirement of the CUFFS model (McArdle, 1988) but also a necessary condition for comparing changes in the construct mean in any developmental process (Meredith, 1964). Second, extant research supports a developmental interrelation between students' global self-esteem and perception of competence across multiple academic domains (Harter, 1985). Specifically, higher self-ratings of global self-worth have been associated with higher self-competence ratings in physical athletic activities across time (e.g., Chan, 2002; Noordstar, van der Net, Jak, Helders, & Jongmans, 2016).

Table 1 shows the means and standard deviations as well as skewness and kurtosis values for students' academic competence and global self-worth scores across the four measurement occasions. For perceived competence across time points, these statistics indicate a slight increase in the mean of each scale item. For global self-worth, these statistics indicate that some of these items marginally increase and decrease across occasions. Tables 2 and 3 present scale items' correlations between items for each scale, respectively. Items for each scale that are within each time point strongly correlate with one another. These correlations support the assumption that their association is due to an underlying construct; however, this hypothesis needs to be formally tested. Furthermore, items from each scale correlate across time points. Examination of data distribution for the Perceived Physical Competence scale did not show evidence of excessive nonnormality (skewness: $Mdn = -0.43$, range = -0.58 to -0.14 ; kurtosis: $Mdn = 1.92$, range = 1.54 to 2.11). This was also the case for the Global Self-Worth scale (skewness: $Mdn = -0.78$, range = -1.11 to -0.51 ; kurtosis: $Mdn = 2.16$, range = 1.73 to 2.50). Thus, we treated items from both scales as continuous and evaluated measurement invariance under the SEM framework. The subsequent analyses are aimed at investigating longitudinal factorial invariance and the change in each construct, as well as the interrelations among these two processes over time.

Table 1. Means, Standard Deviations, Skewness, and Kurtosis Estimates for Each Scale Item Across Measurement Occasions.

Perceived Competence																
<i>j</i>	Time 1				Time 2				Time 3				Time 4			
	<i>l</i>	2	3	4	<i>l</i>	2	3	4	<i>l</i>	2	3	4	<i>l</i>	2	3	4
<i>M</i>	2.87	2.77	2.66	2.79	2.89	2.86	2.85	2.85	2.99	2.90	2.91	2.95	3.05	2.94	2.84	2.97
<i>SD</i>	1.00	0.93	0.88	0.98	0.77	0.80	0.77	0.96	0.78	0.85	0.76	0.95	0.79	0.86	0.82	0.91
<i>S</i>	-0.51	-0.15	-.14	-0.22	-0.46	-0.45	-0.41	-0.36	-0.40	-0.58	-0.35	-0.56	-0.50	-0.56	-0.31	-0.47
<i>K</i>	1.97	2.06	2.03	1.94	2.01	1.85	1.96	2.11	1.69	2.06	1.82	2.04	1.97	1.75	1.54	2.07
Global Self-Worth																
<i>j</i>	Time 1				Time 2				Time 3				Time 4			
	<i>l</i>	2	3	4	<i>l</i>	2	3	4	<i>l</i>	2	3	4	<i>l</i>	2	3	4
<i>M</i>	3.10	3.25	3.32	3.19	2.89	3.15	3.33	3.03	3.02	3.23	3.41	3.16	3.00	3.17	3.24	3.18
<i>SD</i>	0.81	0.77	0.83	0.82	0.82	0.76	0.71	0.81	0.82	0.74	0.71	0.82	0.84	0.83	0.76	0.78
<i>S</i>	-0.59	-0.98	-1.02	-0.81	-0.52	-0.74	-0.83	-0.51	-0.54	-1.04	-1.11	-0.67	-0.57	-0.92	-0.82	-0.75
<i>K</i>	1.77	2.45	2.25	2.07	1.92	2.47	2.40	1.93	1.78	2.50	2.35	1.73	1.74	2.43	2.40	1.92

Note. *j* = item; *S* = skewness; *K* = kurtosis.

Table 2. Zero-Order Correlations for Perceived Competence Across Time.

	Time 1				Time 2				Time 3				Time 4			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Time 1	1	1.00														
	2	.66	1.00													
	3	.54	.57	1.00												
	4	.45	.53	.54	1.00											
Time 2	1	.49	.55	.49	.51	1.00										
	2	.52	.55	.44	.42	.58	1.00									
	3	.46	.51	.55	.54	.52	.49	1.00								
	4	.50	.55	.49	.60	.51	.54	.66	1.00							
Time 3	1	.54	.47	.52	.50	.52	.53	.54	.57	1.00						
	2	.53	.54	.52	.43	.50	.61	.43	.47	.68	1.00					
	3	.56	.50	.51	.51	.45	.43	.64	.66	.64	.57	1.00				
	4	.57	.61	.62	.60	.52	.46	.54	.63	.62	.56	.64	1.00			
Time 4	1	.64	.61	.58	.61	.58	.54	.51	.51	.74	.64	.63	.63	1.00		
	2	.61	.61	.46	.51	.47	.52	.48	.43	.68	.70	.65	.66	.67	1.00	
	3	.52	.47	.55	.49	.46	.41	.48	.35	.58	.47	.54	.55	.62	.57	1.00
	4	.55	.54	.54	.62	.52	.49	.52	.57	.66	.51	.57	.66	.66	.58	.50

Table 3. Zero-Order Correlations for General Self-Worth Across Time.

	Time 1				Time 2				Time 3				Time 4			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Time 1	1	1.00														
	2	.59	1.00													
	3	.43	.52	1.00												
	4	.55	.63	.56	1.00											
Time 2	1	.40	.45	.42	.49	1.00										
	2	.45	.53	.42	.54	.48	1.00									
	3	.34	.45	.47	.46	.38	.45	1.00								
	4	.44	.53	.37	.64	.40	.56	.43	1.00							
Time 3	1	.48	.53	.44	.48	.47	.49	.35	.43	1.00						
	2	.51	.44	.42	.40	.46	.50	.46	.39	.52	1.00					
	3	.41	.42	.51	.49	.36	.43	.51	.33	.44	.69	1.00				
	4	.47	.49	.49	.54	.42	.52	.51	.53	.60	.57	.56	1.00			
Time 4	1	.41	.40	.35	.42	.45	.42	.40	.38	.53	.45	.45	.52	1.00		
	2	.40	.43	.40	.52	.41	.51	.31	.39	.49	.56	.49	.45	.56	1.00	
	3	.25	.27	.30	.30	.40	.43	.35	.28	.40	.45	.40	.42	.48	.54	1.00
	4	.29	.33	.35	.43	.33	.43	.42	.41	.49	.48	.48	.54	.52	.63	.57

Evaluating Longitudinal Factorial Invariance

To test factorial invariance, we specified a succession of increasingly constrained longitudinal measurement models for each scale. We used the marker variable method to identify all CFA models and scale latent variables in our analyses (see Bollen & Curran, 2006, for scaling methods in SEM). Specifically, at each time point we fixed the factor loading of the first item to 1.0 and its intercept to 0. First, we evaluated a configural invariance model for each scale with minimal constraints. In this model, all scale items at each measurement occasion were regressed onto one common factor. With the exception of the reference item, all remaining factor loadings and intercepts were freely estimated. Given that we estimated the item intercepts, estimating the latent variable means would result in redundant information (i.e., a latent variable's mean is the average of its item's intercepts). Hence, all latent variable means were fixed to 0. However, all latent variable variances and covariances were freely estimated. The second level of invariance for each scale was weak invariance. In this model, we constrained the corresponding items' factor loadings across time to be invariant. No additional model parameter modifications were done going from the configural to weak invariant model. The third level of invariance we evaluated for each scale was strong invariance. For the strong invariant model, we added equality constraints to the item intercepts. Also, we now estimated the factor means, however, the factor mean structure requires additional scaling. In order for the factor mean parameter estimates to be interpretable, one of the factor means needs to serve as the reference time point and fixed to 0. In our illustration, we fixed the first factor mean (i.e., Time 1) to 0. This resulted in the second and third factor mean being scaled in reference to the first occasion. We did not modify any other model parameters from the former model. The fourth level of invariance examined for each scale was strict invariance. In this strict invariant model, we added equality constraints to the corresponding item residuals over time. No additional model parameter modifications were specified going from the strong to strict invariant model.

As the aim of this empirical example was to demonstrate how to evaluate longitudinal factorial invariance and apply and interpret CUFFS models, we did not construct a theoretical *a priori* error structure. However, for all analytical models in this study with the exception of the strict invariant measurement models, we freely estimated all error variances (i.e., specified a strong invariant measurement model). In addition, we allowed corresponding item residuals at adjacent time points to freely covary.

Estimation and Fit Criteria

Analyses were conducted using the Lavaan package (Appendices A-E) in the free statistical software R 3.0.2 (R Development Core Team, 2013). Full information maximum likelihood (FIML) estimation was used to fit all statistical models to the data because initial evaluation of the data did not show major violations of normality (Table 1). It has been shown that FIML within the CFA framework for ordered

categorical items, such as Likert-type scale items, is likely to perform as well as limited information estimators when items are not highly skewed (Browne & Arminger, 1995; Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009).² Furthermore, FIML adequately handles missing data when data are either missing at random or completely missing at random. FIML assumes that measurement occasions for an individual are associated across time, as a result, this estimator uses all observed data in the measurement model to estimate parameters and standard errors (Ferron & Hess, 2007). Since we assumed that missing data were a result of a random process and FIML's capability to adjust estimates based on all available data, FIML was deemed an appropriate estimation method for the analyses done in this study.³ Finally, given that the main objective of this article was to illustrate the utility of the CUFFS model, no additional analysis was run to examine missing data. For a detailed explanation of modern missing data theory and estimation procedures, see Enders (2010).⁴

When using the SEM framework, it is common practice to assess global model fit before interpreting any estimated parameters in order to ensure that the specified model as a whole fits the data. Therefore, we used several common fit statistics to evaluate model fit: the model chi-square test (χ^2 ; Bentler & Bonnet, 1980), the comparative fit index (CFI; Bentler, 1990), the Tucker–Lewis Index (TLI; Tucker & Lewis, 1973), the absolute model fit index root mean square error of approximation (RMSEA; Browne & Cudeck, 1993), and the standardized root mean square residual (SRMR; Hu & Bentler, 1999).

To support the tenability of a particular level of factorial invariance, fit indices are used to show that the successive stricter equality constrained model does not substantially worsen model fit to the data (Widaman et al., 2010). The likelihood ratio chi-square difference test ($\Delta\chi^2$) is often used to test factorial invariance, as the series of hierarchically constraint measurement models used for invariance testing are nested within the former model (Bentler & Bonnet, 1980). A particular level of invariance holds in the data when the $\Delta\chi^2$ between two nested models is not significant (i.e., $\Delta\chi^2$ in $p < .05$), meaning statistical model equivalence is supported. However, minor nonconformities of normality in the data can result in a statistically significant $\Delta\chi^2$ test and thus rejection of the model (West, Finch, & Curran, 1995). In addition, the $\Delta\chi^2$ is highly powerful with large sample sizes, and any negligible model differences might result in model rejection. Consequently, differences in fit between nested models with increasing invariance constraints should also be assessed using differences in practical fit indices (Widaman et al., 2010). A simulation study conducted to examine model fit measures for use in tests of invariance found that CFI performed well for determining the tenability of invariance (Cheung & Rensvold, 2002). Therefore, we also recommend Cheung and Rensvold's guideline for supporting factorial invariance by using the change in comparative fit index (Δ CFI). Specifically, a Δ CFI greater than .01 indicates a meaningful change in model fit in factorial invariance.⁵ In summary, we used both the $\Delta\chi^2$ and Δ CFI criteria to support a particular level of factorial invariance in these data.

Results

Table 4 presents fit indices for each level of factorial invariance tested in each scale. For perceived physical competence, configural invariance was met, as all four items loaded positively on only one factor at each measurement occasion. Given that model fit indices indicated strong model fit for these data, findings also supported the assumption that these scale items represent one underlying construct at each time point. Next, results showed weak invariance, as there was not a significant increase in misfit $\Delta\chi^2$ ($p > .05$) and only a very small change in CFI ($\Delta\text{CFI} < .01$) going from the configural invariance model. Next, results supported strong invariance in these data, as a nonsignificant $\Delta\chi^2$ ($p > .05$) between weak and strong invariance was found, as well as a slight change in CFI ($\Delta\text{CFI} < .01$). Finally, we examined the tenability of strict invariance in these data. A test for strict invariance exponentially increased the χ^2 fit test and not surprisingly, results showed a significant $\Delta\chi^2$ ($p < .05$). Moreover, a noticeable change in CFI ($\Delta\text{CFI} > .01$) was found. Thus, these findings indicate that longitudinal strong invariance is the highest level of factorial invariance tenable for the Perceived Physical Competence scale.

For global self-worth, four items loaded positively on one factor at each time point, supporting configural invariance. Model fit indices indicated strong model fit for these data, thus supporting the assumption that one underlying construct is represented by the scale items at each time point. Next, weak invariance was tenable in these data, as there was a nonsignificant $\Delta\chi^2$ ($p > .05$) and a slight change in CFI ($\Delta\text{CFI} < .01$) going from the configural to the weak invariance model. Results also supported strong invariance in these data, as findings showed a nonsignificant $\Delta\chi^2$ ($p > .05$) between weak and strong invariance and a slight change in CFI ($\Delta\text{CFI} < .01$). Finally, we evaluated strict invariance. The model with strict invariance constraints resulted in a significant $\Delta\chi^2$ ($p < .05$) but a minor change in CFI ($\Delta\text{CFI} > .01$). These results strongly support that at least strong invariance is established in the Global Self-Worth Scale. Thus, as results showed strong invariance for each scale, we proceeded to fit CUFFS models.

Univariate CUFFS Models

To measure students' perceived physical competence and global self-worth, univariate CUFFS models for each scale were specified. A CUFFS model combines a measurement model and a growth model in a single underlying structure. In the first-order structure (i.e., measurement model), we imposed longitudinal strong invariance constraints since we demonstrated that the same construct is represented at each time point. Again, we used the marker variable method (i.e., the first item at each time point was our reference variable) to identify the models and scale the latent variables at each time point. It is important to note that when at least strong invariance holds in the data, the same growth trajectory of a construct is estimated irrespective of which item in the measurement model is used as the marker variable (Ferrer et al., 2008). However, when strong invariance is not tenable, the arbitrary selection of a marker

Table 4. Evaluation of Longitudinal Factorial Invariance.

	Configural invariance	Weak invariance	Strong invariance	Strict invariance
Perceived Physical Competence				
χ^2/df fit	167/86	179/95	191/104	247/116
$\Delta\chi^2/\Delta df$ fit	—	12/9*	12/9*	56/12
CFI	.962	.96	.959	.938
ΔCFI	—	.002*	.001*	.021
TLI	.947	.95	.952	.936
RMSEA (CI)	.06 (.046, .074)	.058 (.045, .071)	.05 (.044, .069)	.06 (.054, .076)
SRMR	.04	.057	.058	.066
Global Self-Worth				
χ^2/df fit	142/86	148/95	164/104	188/116
$\Delta\chi^2/\Delta df$ fit	—	6/9*	16/9*	24/12
CFI	.966	.968	.963	.956
ΔCFI	—	.002*	.005*	.007*
TLI	.952	.96	.957	.955
RMSEA (CI)	.05 (.035, .065)	.046 (.031, .06)	.048 (.033, .061)	.049 (.036, .061)
SRMR	.043	.051	.054	.059

Note. CFI = comparative fit index; df = degrees of freedom; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; TLI = Tucker–Lewis index; CI = confidence interval. An asterisk (*) next to the $\Delta\chi^2$ indicates that the model was not statistically significant from prior specified model, change less than .05. An asterisk next to the ΔCFI indicates a change in CFI less than .01.

variable may result in different estimated values of that construct's associations with other variables, as well as its growth trajectory over time.

Now that we specify a second-order model structure (i.e., growth model) and estimate the intercept and slope means, we fixed the first-order factors' means to 0. Additionally, equality constraints are specified on the variance of the first-order factors. Also, now that the interrelation of the first-order factor covariances are being accounted by the growth factors (intercept and slope), their interrelation is fixed to 0. Given that the first item intercept was fixed to 0, the latent intercept is scaled to this item. Specifically, we ran latent basis CUFFS models in which some of the basis coefficients are estimated rather than specified, as in the case, for instance, when one specifies in a growth model a linear or quadratic trajectory for a construct (see Bollen & Curran, 2006; McArdle, 1988, for alternative ways of specifying latent basis growth models). In this example, the first and the last latent basis coefficients were fixed to 0 and 1, respectively. Thus, the intercept represents the initial mean, the slope denotes the change from the first to the last occasion, and the second and third basis coefficients, which are estimated, represent percentage of the overall change. In addition to imposing all the aforementioned specifications, the intercepts and slopes for each of the two scales were set to covary.

Table 5 shows fit indices and parameter estimates from the univariate CUFFS model for each of the constructs. According to the fit indices, results adequately

Table 5. Model Fit Indices and Parameter Estimates (Standard Errors) for Students' Perceived Physical Competence (PPC) and Global Self-Worth (GSW) Scores Across School Grades.

	Univariate CUFFS		Bivariate CUFFS	
	PPC	GSW	PPC	GSW
Basis coefficients				
β_1	.00 (—)	.00 (—)	.00 (—)	.00 (—)
β_2	.44 (.18)**	.33 (.15)**	.34 (.18)*	.22 (.17)***
β_3	.80 (.12)***	.55 (.11)***	.84 (.14)***	.54 (.13)***
β_4	1.00 (—)	1.00 (—)	1.00	1.00
Means				
μ_{f0}	2.83 (.05)***	3.02 (.05)***	2.83 (.05)***	3.00 (.05)***
μ_{fs}	.19 (.03)***	-.04 (.06)	.18 (.04)***	-.01 (.06)
Variances				
σ^2_{f0}	.43 (.02)***	.31 (.04)***	.41 (.05)**	.30 (.04)***
σ^2_{fs}	.04 (.03)***	.17 (.04)***	.03 (.03)*	.15 (.05)***
Covariances				
$\sigma_{f0,fs} [\rho]$	-.04 (.03) [-.33 (.16)]	-.07 (.03)* [-.31 (.11)]*	-.02 (.03) [-.19 (.16)]	-.05 (.04) [-.23 (.13)]
$\sigma_{f01,f02} [\rho]$	—	—	.22 (.03) [.66 (.06)]***	
$\sigma_{f01,fs2} [\rho]$	—	—	-.07 (.03) [-.30 (.12)]**	
$\sigma_{fs1,f02} [\rho]$	—	—	-.04 (.02) [-.40 (.19)]	
$\sigma_{fs1,fs2} [\rho]$	—	—	.05 (.04) [.85 (.22)]**	
Model fit				
χ^2/df fit	205/110	197/110	873/456	
RMSEA	.058	.055	.059	
(95% CI)	(.045, .07)	(.042, .067)	(.053, .065)	
CFI	.955	.947	.899	
TLI	.951	.95	.901	
SRMR	.071	.063	.084	

Note. CFI = comparative fit index; df = degrees of freedom; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; TLI = Tucker-Lewis index; CUFFS = curve of factors model; CI = confidence interval.

supported a univariate CUFFS model for perceived physical competence across time (e.g., RMSEA = .058, CFI = .955, TLI = .951, SRMR = .071). The estimated latent basis coefficients were statistically significant, indicating unequal change in the trajectory from measurement occasion to measurement occasion. Specifically, the difference in the slope for students' perceived physical competence scores from the first to the second occasion was .44 ($SE = .18$), about 44% of the overall change. From the second to the third occasion, the change was .80 ($SE = .12$), or 36% of the change. Furthermore, findings showed a statistically significant mean in students' perceived physical competence at the initial measurement occasion ($\mu_{f0} = 2.83$, $SE = .05$), which also varied across students ($\sigma_{f0}^2 = .43$, $SE = .02$). Results also showed a statistically significant overall change in physical competence across all time points ($\mu_{fs} = .19$, $SE = .03$), which varied across students over time ($\sigma_{fs}^2 = .04$, $SE = .03$).

Together, these values indicate that in the first measurement occasion, students' perceived physical competence scores are an average of 2.83 and then increase an average of .19 from the first to fourth time point. Findings showed no significant covariation between the intercept and slope ($\sigma_{f0,fs} = -.04$, $SE = .03$). However, a significant correlation could be interpreted, as students who had higher perceived physical competence scores at the first occasion had lower rates of change in these scores across time points ($\rho_{f0,fs} = -.33$, $SE = .16$).

For global self-worth, fit indices indicate that this CUFFS model is an acceptable representation of the data (e.g., RMSEA = .055, CFI = .947, TLI = .951, SRMR = .063). The latent basis coefficients indicate that the trajectory of general self-worth also significantly changes nonlinearly across occasions. Specifically, the difference in the slope for students' global self-worth scores from the first to second time points was .33 ($SE = .15$), about 33% of the overall change. From the second to the third occasion, the change was .55 ($SE = .11$) or 22% of the change. Overall, the estimated latent basis coefficients indicate that the change in form of students' global self-worth scores changed across all repeated measurements. Furthermore, there was a significant mean in students' global self-worth scores in the initial measurement occasion ($\mu_{f0} = 3.02$, $SE = .05$), which also varied across students ($\sigma_{f0}^2 = .31$, $SE = .04$).

Furthermore, the overall change in global self-worth across time was not a statistically significant latent trajectory ($\mu_{fs} = -.04$, $SE = .06$), but there was variation across students ($\sigma_{fs}^2 = .17$, $SE = .04$). In other words, whereas the average trajectory was flat, this was not the case for everyone. Finally, there was a significant covariation between the intercept and slope ($\sigma_{f0,fs} = -.07$, $SE = .03$), indicating that students who had higher global self-worth scores at time one had lower rates of change in these scores across measured time points ($\rho_{f0,fs} = -.31$, $SE = .11$).

Bivariate CUFFS Model

With a bivariate LGM, one can model the interrelation of two processes by simultaneously correlating the respective intercepts and slopes. These bivariate models are

also known as parallel process models because they involve two variables over time. For instance, these models have been used to evaluate the longitudinal relation between adolescents' substance use and school conduct problems (Wu, Witkiewitz, McMhon, & Dodge, 2010), physical education cognition and physical activity participation (Yli-Piipari, Barkoukis, Jaakkola, & Liukkonen, 2013), and morphological awareness and vocabulary knowledge (Kieffer & Lesaux, 2012). However, using a second-order bivariate LGM, specifically a bivariate CUFFS model, researchers capitalize on the advantages inherent in modeling a construct's change over time with multiple items. With regard to our empirical example, this model could inform us about the interrelation of students' perceived physical competence and global self-worth from across the four measurement occasions. Thus, to conduct a bivariate CUFFS model, we allow the intercept and slope of each construct to covary.

Table 5 shows fit indices and parameter estimates from the bivariate CUFFS of students' perceived physical competence and global self-worth. Together, the fit indices indicate that the specified bivariate CUFFS model does not fit these data very well (e.g., RMSEA = .059, CFI = .889, TLI = .901, SRMR = .084). It is worth noting that fit indices are global measures of how well the hypothesized model reproduces the observed data. Thus, in cases in which fit indices do not support the hypothesized model, researchers should not only return to the statistical model for evidence of misfit but also to theory to help guide any model modifications (see Little, 1997, and McCallum & Austin, 2000, for more on dealing with model misfit). However, for pedagogical purposes we will proceed to interpret parameters from this model. Individual latent mean and variance parameter estimates for these constructs would be interpreted in the same way as the univariate CUFFS model. Therefore, interpretation will be focused on the associations, specifically correlations, of the intercepts and slopes of these two processes over time.

According to this model, results showed a significant and strong positive correlation between the two intercepts ($\rho_{f01,f02} = 0.66, SE = .06$). That is, students who had higher scores in perceived physical competence also had higher scores in global self-worth at the first measurement occasion. Findings also indicated a significant and moderate negative correlation between the intercept of perceived physical competence and slope of global self-worth scores ($\rho_{f01,fs2} = -0.30, SE = .12$). That is, students who scored higher in perceived physical competence scores in the first time point had a slower rate of growth in global self-worth scores across repeated occasions. However, this bivariate CUFFS model does not support an association between the slope of students' perceived physical competence and the intercept of global self-worth scores ($\rho_{fs1,f02} = -0.40, SE = .19$). Finally, and most important, results showed a strong statistically significant positive correlation between the slopes of both constructs ($\rho_{fs1,fs2} = .85, SE = .22$). This can be interpreted as students who changed more rapidly in perceived physical competence also had a faster rate of change in global self-worth. Overall, this model supports the link between students' perceived physical competence and global self-worth in high school.

Discussion

In this article, we describe the CUFFS model as a tool for modeling a construct's trajectory over time in educational research. Through a series of measurement models with increasing invariance constraints, we showed that strong invariance, a required condition for comparing changes in the construct mean, was tenable in these data. Next, we specified two univariate CUFFS models to show how change in a single construct (e.g., perceived physical competence) can be evaluated with these techniques. Finally, we ran a bivariate CUFFS model to analyze the association between the two constructs (i.e., students' perceived physical competence and global self-worth) over time.

Several prerequisites are required prior to conducting analyses using CUFFS models. One, longitudinal data, such as repeated measures data are necessary to detect within-person changes and between-person differences (Sayer & Cumsille, 2001). Two, multiple items representing a construct are necessary to test factorial invariance and provide assurance that the same intended construct is being measured at each time point. Careful development of a test battery could ensure that two or more scales for each construct are available (Widaman et al., 2010). Three, CUFFS models rely on model assumptions from SEM with latent variables. Therefore, measurement property issues such as large sample size and power are relevant to identify change and individual differences in growth trajectories (Hertzog, von Oertzen, Ghisletta, & Lindenberger, 2008).

Extensions to CUFFS models can be specified to increase the capacity of researchers to answer complex questions of change over time in educational contexts. For instance, we specified a latent basis growth model; however, just as with 1LGs, researchers can also test an array of hypothesized forms of change, such as linear, quadratic, or stepwise, to name a few (see McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002; Ram & Grimm, 2007). Although we illustrated the application of univariate and bivariate CUFFS models only, educational researchers can use a multivariate CUFFS model to measure more multifaceted representations of growth between multiple constructs (e.g., self-perceptions, motivation, affect), or groups (e.g., students, parents, teachers) over time. In addition, the CUFFS model can accommodate analyses of time-invariant covariates (e.g., gender, race, ethnicity) or time-varying covariates (e.g., school GPA, employment status, literacy skills; Hancock & Buehl, 2008). These variables can be used in CUFFS models as predictors or outcomes of change, thus permitting researchers to address a range of questions related to antecedents and consequences of development in education.

Using multiple items and a CUFFS model to analyze change in longitudinal data has a number of benefits not possible when using composite scores and 1LGs. For instance, using a CUFFS model, researchers are able to test the assumption of measurement invariance, further partition the observed score variance (i.e., time-specific and item-residual variance) that is confounded in 1LGs, and assess competing error structures (Bishop et al., 2015; Sayer & Cumsille, 2001). As mentioned previously, specification of an appropriate error structure increases fit between the implied model and observed data (Liu et al., 2012). Common error structures for second-

order LGMs include: unstructured, random coefficients, covariation between specific variance of same items at adjacent occasions, and first-order autoregression (Grilli & Varriale, 2014). To help select an appropriate error structure, researchers can (a) use theory and knowledge of the data collection process to justify item-specific associations, (b) use model fit indices to compare different longitudinal error structures to determine which structure better represents the observed data, and (c) compare the explicability and substantive conclusions of contending error structures.

CUFFS models offer educational researchers a more accurate tool of modeling change than has been used in previous research. These models also provide researchers flexibility to address a number of different research questions about a construct's latent trajectory over time. For instance, once measurement invariance has been established, CUFFS models can be used to test hypotheses about change in students' motivation, academic interest, and peer relations across school grades. We hope that by providing model descriptions, parameterization, and syntax of free statistical software, researchers in education will be encouraged to assess longitudinal factorial invariance and apply CUFFS models in their work.

Appendix A

R Syntax for Evaluating Longitudinal Configural Invariance in Lavaan

For an introduction on how to specify structural equation models in Lavaan, see Rosseel (2012). In this configural model, all scale items for perceived physical competence (e.g., pc1_1, pc2_1) at each measurement occasion were regressed onto one common factor (i.e., f1, f2, f3, and f4). To identify and scale the factors at each time point, we used the marker variable method (the default in Lavaan). Thus, no additional specifications need to be done to the item's factor loadings; however, the marker item's intercept does need to manually be fixed to 0.

```
configural <- ‘
```

```
#Creating first-order factors (no constraints)
```

```
f1 =~ pc1_1 + pc2_1 + pc3_1 + pc4_1; f2 =~ pc1_2 + pc2_2 + pc3_2 + pc4_2;
f3 =~ pc1_3 + pc2_3 + pc3_3 + pc4_3; f4 =~ pc1_4 + pc2_4 + pc3_4 + pc4_4;
```

```
#Item intercepts (no constraints)
```

```
pc1_1~0*1; pc1_2~0*1; pc1_3~0*1; pc1_4~0*1;
pc2_1~1; pc2_2~1; pc2_3~1; pc2_4~1;
pc3_1~1; pc3_2~1; pc3_3~1; pc3_4~1;
pc4_1~1; pc4_2~1; pc4_3~1; pc4_4~1;
```

```
#Item residual variances (no constraints)
```

```
pc1_1~~pc1_1; pc1_2~~pc1_2; pc1_3~~pc1_3; pc1_4~~pc1_4;
pc2_1~~pc2_1; pc2_2~~pc2_2; pc2_3~~pc2_3; pc2_4~~pc2_4;
pc3_1~~pc3_1; pc3_2~~pc3_2; pc3_3~~pc3_3; pc3_4~~pc3_4;
pc4_1~~pc4_1; pc4_2~~pc4_2; pc4_3~~pc4_3; pc4_4~~pc4_4;
```

```

#Item residual covariances (no constraints)
pc1_1~~pc1_2; pc1_2~~pc1_3; pc1_3~~pc1_4;
pc2_1~~pc2_2; pc2_2~~pc2_3; pc2_3~~pc2_4;
pc3_1~~pc3_2; pc3_2~~pc3_3; pc3_3~~pc3_4;
pc4_1~~pc4_2; pc4_2~~pc4_3; pc4_3~~pc4_4;

#First-order factor means (all fixed to zero)
f1 ~ 0*1; f2 ~ 0*1; f3 ~ 0*1; f4 ~ 0*1;

#First-order factor variances (no constraints)
f1 ~~ f1; f2 ~~ f2; f3 ~~ f3; f4 ~~ f4;

#First-order factor covariance (no constraints)
f1~~f2; f1~~f3; f1~~f4;
f2~~f3; f2~~f4
f3~~f4;
'

fit_configural <- sem(configural, data=my.data, meanstructure=TRUE, missing='fiml')
summary(fit_configural, fit.measures=TRUE, standardized=TRUE)

```

Appendix B

R Syntax for Evaluating Longitudinal Weak Invariance in Lavaan

In this weak invariant model, equality constraints to the factor loadings are incorporated to the former model. This is accomplished by premultiplying the same label (e.g., l1, l2, l3) to each parallel item over time. No other changes to the former model syntax are necessary.

```

weak <- '
#Creating first-order factors (factor loading equality constraints)
f1 =~ l1*pc1_1 + l2*pc2_1 + l3*pc3_1 + l4*pc4_1;
f2 =~ l1*pc1_2 + l2*pc2_2 + l3*pc3_2 + l4*pc4_2;
f3 =~ l1*pc1_3 + l2*pc2_3 + l3*pc3_3 + l4*pc4_3;
f4 =~ l1*pc1_4 + l2*pc2_4 + l3*pc3_4 + l4*pc4_4;

#Item intercepts (no constraints)
pc1_1~0*1; pc1_2~0*1; pc1_3~0*1; pc1_4~0*1;
pc2_1~1; pc2_2~1; pc2_3~1; pc2_4~1;
pc3_1~1; pc3_2~1; pc3_3~1; pc3_4~1;
pc4_1~1; pc4_2~1; pc4_3~1; pc4_4~1;

#Item residual variances (no constraints)
pc1_1~~pc1_1; pc1_2~~pc1_2; pc1_3~~pc1_3; pc1_4~~pc1_4;
pc2_1~~pc2_1; pc2_2~~pc2_2; pc2_3~~pc2_3; pc2_4~~pc2_4;
pc3_1~~pc3_1; pc3_2~~pc3_2; pc3_3~~pc3_3; pc3_4~~pc3_4;
pc4_1~~pc4_1; pc4_2~~pc4_2; pc4_3~~pc4_3; pc4_4~~pc4_4;

```

```
#Item residual covariances (no constraints)
pc1_1~~pc1_2; pc1_2~~pc1_3; pc1_3~~pc1_4;
pc2_1~~pc2_2; pc2_2~~pc2_3; pc2_3~~pc2_4;
pc3_1~~pc3_2; pc3_2~~pc3_3; pc3_3~~pc3_4;
pc4_1~~pc4_2; pc4_2~~pc4_3; pc4_3~~pc4_4;

#First-order factor means (all fixed to zero)
f1 ~ 0*1; f2 ~ 0*1; f3 ~ 0*1; f4 ~ 0*1;

#First-order factor variances (no restrictions)
f1 ~~ f1; f2 ~~ f2; f3 ~~ f3; f4 ~~ f4;

#Time factor covariance (no restrictions)
f1~~f2; f1~~f3; f1~~f4;
f2~~f3; f2~~f4;
f3~~f4;
'

fit_weak <- sem(weak, data=my.data, meanstructure=TRUE, missing='fiml')
summary(fit_weak, fit.measures=TRUE, standardized=TRUE)
```

Appendix C

R Syntax for Evaluating Longitudinal Strong Invariance in Lavaan

In this strong invariant model, additional equality constraints to the item intercepts are specified. This is done by premultiplying the same label (e.g., m1, m2, m3) to each parallel item over time. To scale the factor means to the first time point, we fixed the first factor mean to 0 by premultiplying it by 0. No other changes to the former model syntax are necessary.

```
strong <- '
#Creating first-order factors (equality constraints)
f1 =~ l1*pc1_1 + l2*pc2_1 + l3*pc3_1 + l4*pc4_1
f2 =~ l1*pc1_2 + l2*pc2_2 + l3*pc3_2 + l4*pc4_2
f3 =~ l1*pc1_3 + l2*pc2_3 + l3*pc3_3 + l4*pc4_3
f4 =~ l1*pc1_4 + l2*pc2_4 + l3*pc3_4 + l4*pc4_4

#Item intercepts (intercept equality constraints)
pc1_1~0*1; pc1_2~0*1; pc1_3~0*1; pc1_4~0*1;
pc2_1~m2*1; pc2_2~m2*1; pc2_3~m2*1; pc2_4~m2*1;
pc3_1~m3*1; pc3_2~m3*1; pc3_3~m3*1; pc3_4~m3*1;
pc4_1~m4*1; pc4_2~m4*1; pc4_3~m4*1; pc4_4~m4*1;

#Item residual variances (no constraints)
pc1_1~~pc1_1; pc1_2~~pc1_2; pc1_3~~pc1_3; pc1_4~~pc1_4;
pc2_1~~pc2_1; pc2_2~~pc2_2; pc2_3~~pc2_3; pc2_4~~pc2_4;
pc3_1~~pc3_1; pc3_2~~pc3_2; pc3_3~~pc3_3; pc3_4~~pc3_4;
pc4_1~~pc4_1; pc4_2~~pc4_2; pc4_3~~pc4_3; pc4_4~~pc4_4;
```

```

#Item residual covariances (no restrictions)
pc1_1~~pc1_2; pc1_2~~pc1_3; pc1_3~~pc1_4;
pc2_1~~pc2_2; pc2_2~~pc2_3; pc2_3~~pc2_4;
pc3_1~~pc3_2; pc3_2~~pc3_3; pc3_3~~pc3_4;
pc4_1~~pc4_2; pc4_2~~pc4_3; pc4_3~~pc4_4;

#First-order factor means
f1 ~ 0*1; f2 ~ 1; f3 ~ 1; f4 ~ 1;

#First-order factor variances
f1 ~~ f1; f2 ~~ f2; f3 ~~ f3; f4 ~~ f4;

#First-order factor covariance
f1~~f2; f1~~f3; f1~~f4; f2~~f3; f2~~f4; f3~~f4;
,

fit_strong <- sem(strong, data=my.data, meanstructure=TRUE, missing='fiml')
summary(fit_strong, fit.measures=TRUE, standardized=TRUE)

```

Appendix D

R Syntax for Evaluating Longitudinal Strict Invariance in Lavaan

In this strict invariant model, equality constraints to the residuals are incorporated to the former model. This is accomplished by premultiplying the same label (e.g., r1, r2, r3) to each parallel residual variance over time. No other changes to the former model syntax are necessary.

```

strict <- ‘
#Creating first-order factors (equality constraints)
f1 =~ I1*pc1_1 + I2*pc2_1 + I3*pc3_1 + I4*pc4_1
f2 =~ I1*pc1_2 + I2*pc2_2 + I3*pc3_2 + I4*pc4_2
f3 =~ I1*pc1_3 + I2*pc2_3 + I3*pc3_3 + I4*pc4_3
f4 =~ I1*pc1_4 + I2*pc2_4 + I3*pc3_4 + I4*pc4_4

#Item intercepts (equality constraints)
pc1_1~0*1; pc1_2~0*1; pc1_3~0*1; pc1_4~0*1;
pc2_1~m2*1; pc2_2~m2*1; pc2_3~m2*1; pc2_4~m2*1;
pc3_1~m3*1; pc3_2~m3*1; pc3_3~m3*1; pc3_4~m3*1;
pc4_1~m4*1; pc4_2~m4*1; pc4_3~m4*1; pc4_4~m4*1;

#Item residual variances (equality constraints)
pc1_1~~r1*pc1_1; pc1_2~~r1*pc1_2; pc1_3~~r1*pc1_3; pc1_4~~r1*pc1_4;
pc2_1~~r2*pc2_1; pc2_2~~r2*pc2_2; pc2_3~~r2*pc2_3; pc2_4~~r2*pc2_4;
pc3_1~~r3*pc3_1; pc3_2~~r3*pc3_2; pc3_3~~r3*pc3_3; pc3_4~~r3*pc3_4;
pc4_1~~r4*pc4_1; pc4_2~~r4*pc4_2; pc4_3~~r4*pc4_3; pc4_4~~r4*pc4_4;

```

```

#Item residual covariances (no restrictions)
pc1_1~~pc1_2; pc1_2~~pc1_3; pc1_3~~pc1_4;
pc2_1~~pc2_2; pc2_2~~pc2_3; pc2_3~~pc2_4;
pc3_1~~pc3_2; pc3_2~~pc3_3; pc3_3~~pc3_4;
pc4_1~~pc4_2; pc4_2~~pc4_3; pc4_3~~pc4_4;

#First-order factor means
f1 ~ 0*1; f2 ~ 1; f3 ~ 1; f4 ~ 1;

#First-order factor variances
f1 ~~ f1; f2 ~~ f2; f3 ~~ f3; f4 ~~ f4;

#First-order factor covariance
f1~~f2; f1~~f3; f1~~f4;
f2~~f3; f2~~f4;
f3~~f4;
'

fit_strict <- sem(strict, data=my.data, meanstructure=TRUE, missing='fiml')
summary(fit_strict, fit.measures=TRUE, standardized=TRUE)

```

Appendix E

R Syntax for Specifying a Univariate Curve of Factors Model in Lavaan

In this univariate curve of factors model, strong invariance constraints are imposed on the measurement model. The first-order factors' means are fixed to 0. Equality constraints are specified on the variance of the first-order factors (e.g., $f1 \sim v1*f1$ and $f2 \sim v1*f2$). Also, now that the interrelations of the first-order factor covariances are accounted for by the growth factors (intercept and slope), their associations are fixed to 0. The latent basis growth process is specified by fixing the first and last latent basis parameters to 0 and 1, respectively, and freely estimating the second and third basis parameters ($Slope = 0*f1 + f2 + f3 + 1*f4$).

```

pc_cuffs <- '
#Creating first-order factors (first loading fixed to 1.0, equality constraints for rest of loadings)
f1 =~ 1*pc1_1 + l2*pc2_1 + l3*pc3_1 + l4*pc4_1;
f2 =~ 1*pc1_2 + l2*pc2_2 + l3*pc3_2 + l4*pc4_2;
f3 =~ 1*pc1_3 + l2*pc2_3 + l3*pc3_3 + l4*pc4_3;
f4 =~ 1*pc1_4 + l2*pc2_4 + l3*pc3_4 + l4*pc4_4;

#Item intercepts (first item intercept fixed to 0, remaining item intercepts constraint to equality)
pc1_1~0*1; pc1_2~0*1; pc1_3~0*1; pc1_4~0*1;
pc2_1~m2*1; pc2_2~m2*1; pc2_3~m2*1; pc2_4~m2*1;
pc3_1~m3*1; pc3_2~m3*1; pc3_3~m3*1; pc3_4~m3*1;
pc4_1~m4*1; pc4_2~m4*1; pc4_3~m4*1; pc4_4~m4*1;

```



```

#Item residual variances (no constraints)
pc1_1~~pc1_1; pc1_2~~pc1_2; pc1_3~~pc1_3; pc1_4~~pc1_4;
pc2_1~~pc2_1; pc2_2~~pc2_2; pc2_3~~pc2_3; pc2_4~~pc2_4;
pc3_1~~pc3_1; pc3_2~~pc3_2; pc3_3~~pc3_3; pc3_4~~pc3_4;
pc4_1~~pc4_1; pc4_2~~pc4_2; pc4_3~~pc4_3; pc4_4~~pc4_4;

#Item residual covariances (no restrictions)
pc1_1~~pc1_2; pc1_2~~pc1_3; pc1_3~~pc1_4;
pc2_1~~pc2_2; pc2_2~~pc2_3; pc2_3~~pc2_4;
pc3_1~~pc3_2; pc3_2~~pc3_3; pc3_3~~pc3_4;
pc4_1~~pc4_2; pc4_2~~pc4_3; pc4_3~~pc4_4;

#First-order factor means (first factor means set to 0)
f1 ~ 0*1; f2 ~ 0*1; f3 ~ 0*1; f4 ~ 0*1;

#First-order factor variances (equality constraints)
f1 ~~ v1*f1; f2 ~~ v1*f2; f3 ~~ v1*f3; f4 ~~ v1*f4;

#Second-order growth factors
Level=~1*f1 + 1*f2 + 1*f3 + 1*f4;
Slope=~0*f1 + f2 + f3 + 1*f4;

#Second-order factor means and variances (no constraints)
Level ~ 1; Slope ~ 1; Level~~Level; Slope~~Slope;

#First- and second- order factors covariation (all fixed to 0)
f1~~0*f2; f1~~0*f3; f1~~0*f4; f2~~0*f3; f2~~0*f4; f3~~0*f4;
Level~~0*f1 + 0*f2 + 0*f3 + 0*f4;
Slope~~0*f1 + 0*f2 + 0*f3 + 0*f4;

```

```

fit_pc_cuffs <- sem(pc_cuffs, data=my.data, meanstructure=TRUE, missing='fiml')
summary(fit_pc_cuffs, fit.measures=TRUE, standardized=TRUE)

```

Note. To conduct a bivariate CUFFS model in Lavaan, the same univariate CUFFS model syntax specifications can be used for each construct. The only additional specifications required are the covariations of the intercepts and slopes.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by grants from the National Science Foundation (BCS-05-27766 and BCS-08-27021) and NIH-NINDS (R01 NS057146-01) to Emilio Ferrer.

Notes

1. The two-headed arrow with value of 1 in the constant is required to generate the appropriate mean expectations based on the path diagram. When a structural equation model involves means, a constant with such a fixed value of 1 is needed to represent the sums of squares and cross-products (or covariances and mean squares; McArdle & McDonald, 1984). This feature allows the algebraic derivation of expected means and variances.
2. A simulation study comparing continuous and categorical estimation methods in SEM models found that maximum likelihood produced unbiased estimates of the structural model parameters with items containing two to four response options (Rhemtulla, Brosseau-Liard, & Savalei, 2012). This finding is critical to this article because we focus on the application and interpretation of the growth estimates of the CUFFS model (i.e., structural model parameters).
3. Liu, Millsap, West, Tein, Tanaka, and Grimm (2017) review estimation issues related to testing measurement invariance in longitudinal data with ordered-categorical measures.
4. To examine possible differences when using the items as categorical, we carried out all statistical models in our study using weighted least squares estimation. In accordance with previous literature (e.g., Forero et al., 2009), the differences between parameter estimates using this method and those reported in the article were negligible.
5. It is important to note that the recommendation to use the $\Delta\text{CFI} > .01$ is based on work involving multiple group CFA models (Cheung & Rensvold, 2002); thus, its adequacy for establishing invariance in longitudinal CFA models is still uncertain. However, in our experience in evaluating numerous scales for longitudinal invariance, and that of other prominent researchers in the field (e.g., Little, 2013, pp. 154-156), the ΔCFI criteria has been consistent with the $\Delta\chi^2$ test in establishing a certain level of invariance over time.

References

- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Bentler, P. M., & Bonnet, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bianconcini, S., & Cagnone, S. (2012). A general multivariate latent growth model with applications to student achievement. *Journal of Educational and Behavioral Statistics*, 37, 339-364.
- Biemer, P. P., Christ, S. L., & Wiesen, C. A. (2009). A general approach for estimating scale score reliability for panel survey data. *Psychological Methods*, 14, 400-412.
- Bishop, J., Geiser, C., & Cole, D. A. (2015). Modeling latent growth with multiple indicators: A comparison of three approaches. *Psychological Methods*, 20, 43-62.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.
- Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean-and covariance-structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185-249). New York, NY: Springer.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sociological Methods and Research*, 21, 230-258.

- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Chan, D. W. (2002). Perceived domain-specific competence and global self-worth of primary students in Hong Kong. *School Psychology International*, 23, 355-368.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
- Embrettsen, S. E. (2007). Impact of measurement scale in modeling developmental processes and ecological factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 63-87). Mahwah, NJ: Lawrence Erlbaum.
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.
- Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 4, 22-36.
- Ferrer, E., & McArdle, J. J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling*, 10, 493-524.
- Ferron, J. M., & Hess, M. R. (2007). Estimation in SEM: A concrete example. *Journal of Educational and Behavioral Statistics*, 32, 110-120.
- Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16, 625-641.
- Geiser, C., Keller, B., & Lockhart, G. (2013). First versus second order latent growth curve models: Some insights from latent state-trait theory. *Structural Equation Modeling*, 20, 479-503. doi:10.1080/10705511.2013.797832
- Geiser, C., & Lockhart, G. (2012). A comparison of four approaches to account for method effects in latent state-trait analyses. *Psychological Methods*, 17, 255-283.
- Grilli, L., & Varriale, R. (2014). Specifying measurement error correlations in latent growth curve models with multiple indicators. *Methodology*, 10, 117-125.
- Hancock, G. R., & Buehl, M. M. (2008). Second-order latent growth models with shifting indicators. *Journal of Modern Applied Statistical Methods*, 7, 39-55.
- Hancock, G. R., Kuo, W., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling*, 8, 470-489.
- Harter, S. (1985). Competence as a dimension of self-evaluation: Toward a comprehensive model of self-worth. In R. E. Leahy (Ed.), *The development of the self* (pp. 55-121). Orlando, FL: Academic Press.
- Harter, S. (1988). *Manual for the self-perception profile for adolescents*. Denver, CO: University of Denver.
- Hertzog, C., von Oertzen, T., Ghisletta, P., & Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual difference in change. *Structural Equation Modeling*, 15, 541-563.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117-144.
- Hu, L., & Bentler, P. M. (1999). Fit indices in covariance structure modeling: Sensitivity to under-parameterized model misspecification. *Psychological Methods*, 3, 424-453.
- Kieffer, M. J., & Lesaux, N. K. (2012). Development of morphological awareness and vocabulary knowledge in Spanish-speaking language minority learners: A parallel process latent growth curve model. *Applied Psycholinguistics*, 33, 23-54.

- Leite, W. L. (2007). A comparison of latent growth models for constructs measured by multiple items. *Structural Equation Modeling*, 14, 581-610.
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32, 53-76.
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.
- Liu, S., Rovine, M. J., & Molenaar, P. C. M. (2012). Selecting a linear mixed model for longitudinal data: Repeated measures analysis of variance, covariance pattern model, and growth curve approaches. *Psychological Methods*, 17, 15-30.
- Liu, Y., Millsap, R. E., West, S. G., Tein, J.-Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, 22, 486-506. doi:10.1037/met0000075
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in macCallum research. *Annual Review of Psychology*, 51, 201-226.
- Marsh, H. W., & Hau, K.-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, 32, 151-170.
- McArdle, J. J. (1988). Dynamic but structural modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *The handbook of multivariate psychology* (2nd ed.; pp. 561-614). New York, NY: Springer.
- McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life-span. *Developmental Psychology*, 38, 115-142.
- McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology*, 37, 234-251.
- Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29, 177-186.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525-543.
- Murphy, D. L., Beretvas, N., & Pituch, K. A. (2011). The effects of autocorrelation on the curve-of-factors growth model. *Structural Equation Modeling*, 18, 430-448.
- Noordstar, J. J., van der Net, J., Jak, S., Helders, P. J. M., & Jongmans, J. M. (2016). Global self-esteem, perceived athletic competence, and physical activity in children: A longitudinal cohort study. *Psychology of Sport and Exercise*, 22, 83-90.
- Phan, H. P. (2013). Examination of self-efficacy and hope: A developmental approach using latent growth modeling. *Journal of Educational Research*, 106, 93-104.
- R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ram, N., & Grimm, K. J. (2007). Using simple and complex growth models to articulate developmental change: Matching method to theory. *International Journal of Behavioral Development*, 31, 303-316.
- Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17, 354-373.
- Riley, M. W. (1963). *Sociology research: A case approach*. New York, NY: Harcourt Brace Jovanovich.

- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2). Retrieved from <https://www.jstatsoft.org/article/view/v048i02>
- Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 179-200). Washington, DC: American Psychological Association.
- Stringer, K., Kerpelman, J., & Skorikov, V. (2011). Career preparation: A longitudinal, process-oriented examination. *Journal of Vocational Behavior*, 79, 158-169.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
- Turner, I., Reynolds, K. J., Lee, E., Subasic, E., & Bromhead, D. (2014). Well-being, school climate, and the social identity process: A latent growth model study of bullying perpetration and peer victimization. *School Psychology Quarterly*, 29, 320-335.
- U.S. Department of Education. (2016). *Education Resources Information Center (ERIC)*. Retrieved from <http://search.proquest.com/eric?accountid=14505>
- Voelkle, M. C., Wittmann, W. W., & Ackerman, P. L. (2006). Abilities and skill acquisition: A latent growth curve approach. *Learning and Individual Differences*, 16, 303-319.
- von Oerzen T., Hertzog, C., Lindenberger, U., & Ghisletta, P. (2010). The effect of multiple indicators on the power to detect inter-individual differences in change. *British Journal of Mathematical and Statistical Psychology*, 63, 627-646.
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.
- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, 4, 10-18.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association.
- Wimmers, P. F., & Lee, M. (2015). Identifying longitudinal growth trajectories of learning domains in problem-based learning: A latent growth curve modeling approach using SEM. *Advances in Health Sciences Education*, 20, 467-478.
- Wu, J., Witkiewitz, K., McMahon, R. J., & Dodge, K. A. (2010). A parallel process growth mixture model of conduct problems and substance use with risky sexual behavior. *Drug and Alcohol Dependence*, 111, 207-214.
- Yli-Piipari, S., Barkoukis, V., Jaakkola, T., & Liukkonen, J. (2013). The effect of physical education goal orientations and enjoyment in adolescent physical activity: A parallel process latent growth analysis. *Sport, Exercise, and Performance Psychology*, 2, 15-31.