Published in final edited form as:

J Appl Soc Psychol. 2013 April 1; 43(4): 721–729. doi:10.1111/jasp.12000.

Longitudinal measurement invariance of psychosocial measures in physical activity research: An application to adolescent data

Scott C. Roesch,

Department of Psychology, San Diego State University

Greg J. Norman,

Department of Family and Preventive Medicine, University of California, San Diego

Erin L. Merz,

San Diego State University/University of California, San Diego Joint Doctoral Program in Clinical Psychology

James F. Sallis, and

Department of Psychology, San Diego State University

Kevin Patrick

Department of Family and Preventive Medicine, University of California, San Diego

Abstract

The current study served as a practical and substantive guide to establishing longitudinal measurement invariance of psychosocial measures commonly used in adolescent physical activity (PA) research. Psychosocial data on an initial sample of 878 adolescents (ages 11 – 15) recruited through primary care providers were provided at baseline, 6, 12, and 24 months. The target measures included family support, peer support, decisional balance (pros, cons), self-efficacy, and behavioral strategies. Five of the six psychosocial measures exhibited strict longitudinal measurement invariance, with the 6th measure (self-efficacy) exhibiting strong longitudinal measurement invariance. These findings support the equivalence of these measures across time, and provide the foundation to substantively interpret group differences and associations involving these measures and PA.

Keywords

Structural equation modeling; measurement; physical activity

Because of the multiple physical and mental health benefits of physical activity in youth, at least 60 minutes per day are recommended (Strong et al., 2005). However, youth physical activity (PA) generally declines over the course of school age years (Wharf-Higgins, Gaul, Gibbons, & Van Gyn, 2003), and particularly during adolescence (Duncan, Duncan, Strycker, & Chaumeton, 2007). This decline in PA, and accompanying sedentary living, has raised serious public health concerns (Berry & Wharf-Higgins, 2005) largely because PA-related chronic disease risk factors begin in childhood and adolescence (Grunbaum et al., 2001) and continue into adulthood. Because of the beneficial effects of PA, intervention studies are critical for the identification of effective methods of health behavior change (Baranowski, Anderson, & Carmack, 1998). Specifically, Baranowski et al. and a recent review by Lubans, Foster, and Biddle (2008) have emphasized the importance of evaluating the effect of PA interventions on targeted mediator psychosocial constructs to better understand the mechanisms of PA change.

Understanding mechanisms of PA change and establishing *causal* relationships requires three methodological/statistical elements: (a) a statistical association between a target variable and PA; (b) elimination of alternative explanations for the association (primarily reduced by the use of random assignment and/or inclusion of covariates in predictive models); and (c) temporal precedence, the cause (e.g., psychosocial antecedent/mediator variable) must precede the effect (PA) in time. Before causal statements about the relationships between psychosocial variables and PA can be definitively made, the measurement validity and invariance of these psychosocial variables across time needs to be evaluated. Only after it has been shown that the constructs of interest are being measured similarly at each time-point (referred to as stationarity) can one attribute differences/associations to substantive factors rather than changes in the measurement instrument over time.

To establish the invariance of a measure over time, confirmatory factor analysis is traditionally implemented (Little, Preacher, Selig, & Card, 2007). Consider the measurement model presented in Figure 1. Assume that we are interested in establishing the longitudinal measurement invariance of a 4-item self-efficacy measure across two time-points. Because measures are assessed repeatedly over time, a mean augmented covariance matrix of the observed variables (e.g., items from a particular psychosocial measure) is typically used for the CFA (Vandenberg & Lance, 2000). Continuing with our self-efficacy example, this matrix contains 8 observed means (4 observed variables at 2 time-points) and the variances/covariances among these 8 observed variables. Importantly, CFA models based on this matrix can take into account the lagged associations (or autocorrelated errors) between observed variables (e.g., allowing an observed variable at time 1 to be correlated with the same variable at time 2).

Four general steps (models) are required to establish the longitudinal measurement invariance of a target measure (Little, 1997): (a) configural invariance, (b) metric invariance, (c) scalar invariance, and (d) item uniqueness invariance. Configural invariance establishes a best-fitting baseline model for the target measure(s) of interest. So, for example, a self-efficacy factor (latent variable) would be specified at each time-point, with direct paths from each latent variable to the four observed variables (items) that were measured (see Figure 1). Moreover, the item means (intercepts) and item uniqueness terms would be freely estimated at each time-point. In short, configural invariance is established if the same set of observed variables load on the target latent variable at each time-point. Once a configural invariance model has been established, a metric invariance model can be specified and tested. The metric invariance model constrains factor loadings at each time-point to equivalence. These equality constraints reflect the hypothesis that the strength of the relationship between each item and the latent variable is of the same (statistical) magnitude. In other words, the factor loadings remain consistent across time. According to Ployhart and Oswald (2004), equal factor loadings are a prerequisite for making meaningful comparisons of latent constructs across time. This is referred to as weak factorial invariance (Meredith, 1993).

The necessity of models c and d in establishing longitudinal measurement equivalence is arguable (Byrne & Stewart, 2006). Building on the metric invariance model, a scalar invariance model also constrains the mean of each observed variable (i.e., item intercepts) to equivalence across time. Intercept differences across time can suggest the presence of a systematic response bias (Vandenberg & Lance, 2000). If item intercepts are invariant across time-points, strong measurement invariance is achieved (Meredith, 1993). Although intercept differences may reflect a systematic response bias, there is the possibility that they reflect “real” or expected group differences (Ployhart & Oswald, 2004; Byrne & Stewart, 2006), particularly in the context of an intervention study (Pentz & Chou, 1994), or in studies where psychosocial constructs are expected to change for developmental reasons.

For example, it is perfectly reasonable to expect that mean levels of self-efficacy will be higher across time (relative to baseline levels) in an intervention study, given that the intervention targets changes self-efficacy beliefs as a proximal predictor of PA, rather than indicate that a particular response bias is present.

Building on the scalar invariance model, the item uniqueness (residual variation or error) invariance model constrains the error variance term of each observed variable to equivalence across time. The item uniqueness terms reflect the sum of the item-specific and random sources of measurement error for each observed variable. In other words, this variability is *not* common variability shared among the items as reflected in the factor loadings, but rather unique variability for each item. If this model fits well, strict longitudinal measurement invariance will have been achieved (Meredith, 1993). Some suggest that invariance of these model parameters should not be tested because it is based on the unrealistic assumption that random measurement error will behave in a systematic fashion (Little, 1997); others (Deshon, 2004) argue that these item uniqueness terms also reflect systematic effects of unmodeled variables (e.g., method effects).

At each step of this process, the overall fit of each target model must be assessed. Traditionally, the likelihood ratio chi-square test has been reported but sparingly used to determine whether or not a model fits well. This test statistic has been identified as unsatisfactory for numerous reasons, including the heavy reliance of this statistic on sample size (Hoyle, 2000). Because of this, descriptive fit indices have been developed and used to determine model fit. While the use descriptive fit indices and the values used to determine overall model fit is certainly contentious (Marsh, Hau, & Wen, 2004; *Personality and Individual Differences*, 2007), three descriptive fit indexes have been generally recommended (Bentler, 2007): (a) the Comparative Fit Index (CFI; Bentler, 1990), a relative index of model fit compared to the null model with values greater than .95 indicating good model fit and values greater than .93 indicating acceptable model fit; (b) the root mean square error of approximation (RMSEA; Steiger, 1990), an absolute index of overall model fit with values less than .08 indicative acceptable model fit and values less than .05 indicative of good model fit; and (c) the standardized root mean square residual (SRMR), also an absolute descriptive index of overall model fit with values less than .08 indicative of acceptable model fit and values less than .05 indicative of good model fit.

The four step invariance testing outlined above reflected a hierarchical series of models, with models becoming more restrictive at each step of the procedure. Importantly, this hierarchy reflects nested models that can be statistically (and descriptively) compared to determine if each more restrictive step of measurement invariance is plausible. For example, the metric invariance model (step 2) is nested within the configural invariance model (step 1). These models differ only in that the factor loadings are freely estimated at each time-point for the former but constrained to be equivalent across time-points for the latter. Traditionally, statistical model comparisons have been conducted using the chi-square difference test ($\Delta\chi^2$), with a non-statistically significant $\Delta\chi^2$ value ($p > .05$) reflecting that the nested model (metric invariance model) fits as well as the configural invariance model. Thus, a determination would be made that the factor loadings are invariant across time based on the metric invariance model being more parsimonious. However, like the overall likelihood ratio χ^2 test, the chi-square difference test is also dependent on sample size (Kelloway, 1995). Longitudinal measurement invariance testing, then, is biased for invariance with low sample sizes (i.e., lower statistical power) and biased against invariance with large sample sizes (i.e., higher statistical power; MacCallum, Browne, & Cai, 2005). For this reason, researchers have begun to use differences in the descriptive fit indices to determine comparative model fit, although the chi-square difference test is traditionally still reported. For example, Δ CFI differences that are less than .01 and RMSEA values less

than .015 indicate no difference between nested models (Chen, 2007; Cheung & Rensvold, 2002).

It is possible that nested model comparisons will not support longitudinal measurement invariance at a particular step. This is the case when a less restrictive model (e.g., configural invariance model) fits statistically (and descriptively) better than a more restrictive model (e.g., metric invariance model). In model comparisons such as these, individual equality constraints can be statistically evaluated in the more restrictive model to determine the plausibility of equivalence across time for a subset of the constrained parameters. For example, the factor loading for item 2 of the self-efficacy measure represented in Figure 1 might be significantly higher at time 2 relative to time 1, thus rendering the equality constraint in the metric invariance model for this factor loading inaccurate. This equality constraint can be removed and the metric invariance model re-estimated—this model is now referred to as a partial metric invariance model. Some degree of invariance is tolerated across the metric, scalar, and item uniqueness variance models. Importantly, at least partial measurement invariance must be established before subsequent more restrictive measurement invariance models (steps) can be undertaken (Vandenberg & Lance, 2000).

Of interest in the current study, was the longitudinal measurement invariance of target psychosocial constructs hypothesized to be related to PA. Measures of these constructs included family support, peer support, decisional balance (pros, cons), self-efficacy, and behavioral strategies assessed at four time-points (pre-intervention, midpoint of intervention, immediate post-intervention, *long-term* follow-up). The current study used data and psychosocial variables from a health behavior change intervention for adolescents (the Patient-centered Assessment and Counseling for Exercise + Nutrition [PACE+], 26). Participants were randomized to one of two intervention conditions. The PACE+ intervention targeted PA, sedentary behavior, total fat intake, and servings per day of fruits and vegetables (Patrick et al., 2006). Participants in the comparison condition received a sun protection intervention consisting of an initial primary care office-based computer assessment of sun protection behaviors.

Methods

Participants

Adolescents between the ages of 11 and 15 were recruited through their primary care providers from May 2001 to June 2002. Details about the recruitment process and interventions have been previously published (Patrick et al., 2006). The study was approved by the participating healthcare organizations and university institutional review boards. The interventions were conceptualized as two 1-year doses comprised of monthly modules and counselor calls delivered in year 1 and then repeated in year 2. Participants completed assessments at baseline (pre-intervention, $n=877$), 6-months (midpoint of year 1 intervention, $n=742$), 12-months (end of year 1-intervention, $n=703$), and 24-months (end of year 2 intervention, $n=620$). The average age of adolescents in the sample was 12.7 (SD = 1.3) years. The race/ethnicity distribution of the sample was 58% white non-Hispanic, 13% Hispanic, 6% African-American, 3% Asian or Pacific Islander, and 20% multi-ethnic or other race/ethnicities. Highest adult household education level was distributed as 34% some college or less, 29% Bachelor's degree, and 37% graduate or professional degree.

Measures

Psychosocial measures—A survey battery was created that included items assessing psychosocial constructs that the intervention was designed to impact, including change strategies, pros and cons of change, self-efficacy, family support, and peer support. The

Flesch-Kincaid formula estimated the surveys to be at a 7th grade reading level. Some measures were based on previous measures for adults but were adapted for adolescents. All of the survey instruments are available at www.paceproject.org/measures.

Family support—Four items from the Amherst Health and Activity Study assessed family influences on PA (Sallis, Taylor, Dowda, Freedson, & Pate, 2002). The items asked participants to rate the frequency a household member encouraged, participated, provided transportation, or watched PA. Items were asked in reference to a typical week and participants responded using a 5-point scale from 1 ‘Never’ to 5 ‘Every Day.’

Peer support—Items similar to the family support items assessed peer influences related to PA. Four items assessed the frequency that friends provided encouragement and support for participating in PA. Items were asked in reference to a typical week and participants responded using a 5-point scale from 1 ‘Never’ to 5 ‘Every Day.’

Decisional balance—Decisional balance consists of two constructs labeled the ‘Pros’ and ‘Cons’ of change that address the consequences of behavior change that influence decision-making. Five pros and five cons of PA were developed by modifying items previously used with adults (Marcus, Rakowski, & Rossi, 1992). Participants rated the importance of each item using a 5-point Likert scale with anchors of 1 ‘not at all important’ to 5 ‘extremely important.’

Self-efficacy—Situational self-efficacy represents the person’s confidence he or she can meet a behavioral criterion in situations that may present barriers to the behavior. A 6-item PA self-efficacy scale was created based on previous scales (Marcus & Owen, 1992; Sallis, Pinski, Grossman, Patterson, & Nader, 1988). Participants responded to each item on a 5-point Likert scale ranging from 1 ‘I’m sure I can’t’ to 5 ‘I’m sure I can’.

Behavior change strategies—Fifteen items reflected thoughts and activities related to behavior change and were based on items developed by Saelens et al. (2000). The specific strategies were selected to match the content of the PACE+ intervention. The response format assessed how often each strategy was used, with a 5-point Likert scale ranging from 1 ‘Never’ to 5 ‘Many Times’.

Results

Ranges for the means, standard deviations, and Cronbach alpha values are presented in Table 1. Descriptively, the mean and standard deviations were similar for each measure across the four time-points. Moreover, each measure was internally consistent and similar in value across the four time-points. All skewness and kurtosis values for each item at each time-point were less than two.

Model Testing

Single-factor CFA models were simultaneously tested for each measure at each time-point. The factor loading for the first item of every measure at each time-point was specified as 1 to set the metric for the latent variable. Preliminary analyses (not reported here) showed that the factor loading for the first item of all 6 measures were invariant across time, supporting this practice. Moreover, to test for invariance, the baseline values for target model parameters (factor loadings, item intercepts, and item uniqueness) were used as reference values (for comparison) to the values at 6-, 12-, and 24-months, respectively. For all models the maximum likelihood missing data procedure employed by EQS 6.1 (Bentler, in press) was used. This maximum likelihood procedure uses all available data from each participant

and assumes that data are missing at random. All parameter estimates reported are unstandardized values. Because of space limitations, we present the ranges of model parameters across items. Specific parameter estimates at each time-point for each item can be obtained from the first author.

Family support—The configural invariance model for family support fit extremely well, $\chi^2(75) = 239.50$, $p < .05$, CFI = .952, RMSEA = .050, SRMR = .043. All factor loadings at each time-point were statistically significant (.92–1.15; item intercept values ranged from 2.30–3.36 and item uniqueness values ranged from 0.58–1.37). The metric invariance model also fit well overall, $\chi^2(84) = 247.82$, $p < .05$, CFI = .951, RMSEA = .047, SRMR = .043, and did not differ in overall model fit from the configural invariance model statistically, $\Delta\chi^2(9) = 8.32$, $p > .05$, and descriptively (all Δ values $< .01$). The scalar invariance also fit well overall, $\chi^2(96) = 302.56$, $p < .05$, CFI = .952, RMSEA = .050, SRMR = .044. While the metric invariance model fit *statistically* better than the scalar invariance model, $\Delta\chi^2(12) = 54.744$, $p < .05$, these models did not differ when the descriptive indexes were considered (all Δ values $< .01$). The invariant item uniqueness model also fit well, $\chi^2(108) = 328.17$, $p < .05$, CFI = .948, RMSEA = .048, SRMR = .046. While the scalar invariance model fit *statistically* better than the invariant item uniqueness model, $\Delta\chi^2(12) = 25.61$, $p < .05$, these models did not differ when the descriptive indexes were considered (all Δ values $< .01$). Thus, strict longitudinal measurement invariance was supported for this measure.

Peer support—The configural invariance model for peer support fit extremely well, $\chi^2(75) = 144.02$, $p < .05$, CFI = .973, RMSEA = .032, SRMR = .037. All factor loadings at each time-point were statistically significant (.44–1.33; item intercept values ranged from 2.18–3.74 and item uniqueness values ranged from 0.63–1.48). Next, factor loadings were constrained to equivalence across time. This metric invariance model also fit well overall, $\chi^2(84) = 145.99$, $p < .05$, CFI = .974, RMSEA = .029, SRMR = .038, and did not differ in overall model fit from the configural invariance model statistically, $\Delta\chi^2(9) = 1.97$, $p > .05$, and descriptively (all Δ values $< .01$). To the metric invariance model, item intercepts were constrained to equivalence across time. This scalar invariance also fit well overall, $\chi^2(96) = 203.98$, $p < .05$, CFI = .974, RMSEA = .036, SRMR = .038. While the metric invariance model fit *statistically* better than the scalar invariance model, $\Delta\chi^2(12) = 57.99$, $p < .05$, these models did not differ when the descriptive indexes are considered (all Δ values $< .01$). Thus, scalar invariance is supported. Finally, to the scalar invariance model, the variances of the item uniqueness terms were constrained to equivalence across time. This model also fit well overall, $\chi^2(108) = 223.77$, $p < .05$, CFI = .971, RMSEA = .035, SRMR = .040. While the scalar invariance model fit *statistically* better than the invariant item uniqueness model, $\Delta\chi^2(12) = 19.79$, $p < .05$, these models did not differ when the descriptive indexes were considered (all Δ values $< .01$). Thus, strict longitudinal measurement invariance was supported for this measure.

Decisional balance-pros—The configural invariance model for Decisional Balance-Pros fit extremely well, $\chi^2(134) = 224.99$, $p < .05$, CFI = .971, RMSEA = .028, SRMR = .030. All factor loadings at each time-point were statistically significant (.71–1.56; item intercept values ranged from 3.13–4.39 and item uniqueness values ranged from 0.34–1.43). The metric invariance model also fit well overall, $\chi^2(146) = 247.65$, $p < .05$, CFI = .969, RMSEA = .028, SRMR = .035. This model differed from the configural invariance model statistically, $\Delta\chi^2(12) = 22.62$, $p < .05$, but not descriptively (all Δ values $< .01$). The scalar invariance also fit well overall, $\chi^2(161) = 454.27$, $p < .05$, CFI = .967, RMSEA = .046, SRMR = .036. While the metric invariance model fit *statistically* better than the scalar invariance model, $\Delta\chi^2(15) = 209.42$, $p < .05$, these models did not differ when the CFI and SRMR were considered (all Δ values $< .01$). There was a marginal difference between the

two models when the RMSEA was considered ($\Delta\text{RMSEA} = .018$). The invariant item uniqueness model also fit well, $\chi^2(176) = 564.44$, $p < .05$, CFI = .947, RMSEA = .050, SRMR = .045. While the scalar invariance model fit *statistically* better than the invariant item uniqueness model, $\Delta\chi^2(15) = 110.17$, $p < .05$, these models did not differ when the RMSEA and SRMR were considered (all Δ values $< .01$). There was a marginal difference between the two models when the CFI was considered ($\Delta\text{CFI} = .020$). While there was some disagreement among the descriptive fit indexes in the nested model comparisons, in general, strict longitudinal measurement was supported for this measure.

Decisional balance-cons—The configural invariance model for Decisional Balance-Cons fit extremely well, $\chi^2(134) = 220.93$, $p < .05$, CFI = .958, RMSEA = .027, SRMR = .034. All factor loadings at each time-point were statistically significant (.86–1.58; item intercept values ranged from 1.40–2.18 and item uniqueness values ranged from 0.39–1.13). The metric invariance model also fit well overall, $\chi^2(146) = 228.03$, $p < .05$, CFI = .959, RMSEA = .025, SRMR = .035, and did not differ from the configural invariance model statistically, $\Delta\chi^2(12) = 7.90$, $p > .05$, or descriptively (all Δ values $< .01$). The scalar invariance also fit well overall, $\chi^2(161) = 281.19$, $p < .05$, CFI = .959, RMSEA = .029, SRMR = .035. While the metric invariance model fit *statistically* better than the scalar invariance model, $\Delta\chi^2(15) = 53.16$, $p < .05$, these models did not differ according to the descriptive fit indices (all Δ values $< .01$). The invariant item uniqueness model fit reasonably well, $\chi^2(176) = 458.88$, $p < .05$, CFI = .911, RMSEA = .043, SRMR = .050). While the scalar invariance model fit *statistically* better than the invariant item uniqueness model, $\Delta\chi^2(15) = 177.69$, $p < .05$, these models did not differ when the RMSEA and the SRMR were considered (all Δ values $< .015$). There was a difference between the two models when the CFI was considered ($\Delta\text{CFI} = .048$). While there was some disagreement among the descriptive fit indexes in the nested model comparisons, in general, strict longitudinal measurement was supported for this measure.

Self-efficacy—The configural invariance model for self-efficacy fit well, $\chi^2(211) = 286.78$, $p < .05$, CFI = .977, RMSEA = .020, SRMR = .027. All factor loadings at each time-point were statistically significant (.67–1.49; item intercept values ranged from 3.04–4.15 and item uniqueness values ranged from 0.51–1.24). The metric invariance model also fit well overall, $\chi^2(226) = 321.83$, $p < .05$, CFI = .974, RMSEA = .022, SRMR = .037. The configural model did fit statistically better than the configural invariance model, $\Delta\chi^2(15) = 35.05$, $p < .05$, but the two models did not differ according to the descriptive fit indexes (all Δ values $< .01$). The scalar invariance also fit well overall, $\chi^2(244) = 435.00$, $p < .05$, CFI = .974, RMSEA = .030, SRMR = .038). While the metric invariance model fit statistically better than the scalar invariance model, $\Delta\chi^2(18) = 113.17$, $p < .05$, these models did not differ according to the descriptive fit indices (all Δ values $< .01$). The invariant item uniqueness model fit reasonably well, $\chi^2(262) = 599.28$, $p < .05$, CFI = .952, RMSEA = .038, SRMR = .059. The scalar invariance model fit statistically better than the invariant item uniqueness model, $\Delta\chi^2(18) = 164.28$, $p < .05$, and both the CFI and SRMR suggested that the scalar invariance model fit better (both Δ s $> .02$). Because of this the LaGrange Multiplier Test for Constraints were evaluated for the variance of the item uniqueness terms. Nine of these 18 parameters statistically differed: (a) item 1 time 1-time 2 (.82 vs. .67, $p = .009$), (b) item 1 time 1-time 3 (.82 vs. .67, $p < .001$), (c) item 2 time 1-time 4 (.63 vs. .51, $p < .001$), (d) item 3 time 1-time 2 (.75 vs. .57, $p = .005$), (e) item 3 time 1-time 4 (.75 vs. .52, $p < .001$), (f) item 4 time 1-time 4 (1.12 vs. .83, $p < .001$), (g) item 5 time 1-time 3 (1.09 vs. .86, $p = .007$), (h) item 5 time 1-time 4 (.79 vs. .69, $p = .001$), and (i) item 6 time 1-time 4 (.90 vs. .65, $p = .002$). These equality constraints were removed and partial measurement invariance model for the item uniqueness terms was evaluated. This model fit well overall, $\chi^2(253) = 501.80$, $p < .05$, CFI = .965, RMSEA = .033, SRMR = .043. Moreover, while the

scalar invariance model still fit statistically better than this revised model, $\Delta\chi^2(15) = 53.16$, $p < .05$, these two models did not differ according to the descriptive fit indices (all Δ values $< .01$). At a minimum, strong longitudinal measurement equivalence was established for this measure.

Behavior change strategies—The configural invariance model for strategies fit well, $\chi^2(1616) = 3441.22$, $p < .05$, CFI = .863, RMSEA = .036, SRMR = .045, although the CFI value did not reach specified levels for good model fit. All factor loadings at each time-point were statistically significant (.71–2.26.; item intercept values ranged from 1.84–4.05 and item uniqueness values ranged from 0.61–1.20). The metric invariance model also fit reasonably well overall, $\chi^2(1658) = 3525.37$, $p < .05$, CFI = .859, RMSEA = .036, SRMR = .051. While the configural invariance model fit statistically better, $\Delta\chi^2(42) = 84.15$, $p < .05$, the model fit equally well descriptively (all Δ values $< .01$). The scalar invariance also fit well overall, $\chi^2(1703) = 3792.16$, $p < .05$, CFI = .861, RMSEA = .037, SRMR = .051. While the metric invariance model fit statistically better than the scalar invariance model, $\Delta\chi^2(45) = 266.79$, $p < .05$, these models did not differ when the descriptive indexes were considered (all Δ values $< .01$). The invariant item uniqueness model also fit reasonably well, $\chi^2(1748) = 3931.16$, $p < .05$, CFI = .855, RMSEA = .038, SRMR = .052. While the scalar invariance model fit *statistically* better than the invariant item uniqueness model, $\Delta\chi^2(45) = 139.00$, $p < .05$, these models did not differ when the descriptive indexes were considered (all Δ values $< .01$). Thus, strict longitudinal measurement invariance was supported for this measure.

Discussion

The current study served as a practical and substantive guide to establishing longitudinal measurement invariance of measures commonly used in adolescent PA research. Five of the six target psychosocial measures exhibited strict longitudinal measurement invariance, with the 6th measure (self-efficacy) exhibiting strong longitudinal measurement invariance. These findings provide firm evidence that the relationship between each factor (i.e., psychosocial “scale”) and its’ respective items, the mean of each item, and the variance of each item uniqueness term were *statistically equivalent* across time. Thus, any research involving group comparisons (e.g., as a function of intervention/control group differences), growth trajectories, or concurrent/predictive associations of PA at specific time-point or across time with these psychosocial measures are valid.

The current study is certainly not the first to introduce the concept and procedures of longitudinal measurement invariance into adolescent PA research. Most recently, Dishman et al. (2010) examined the measurement equivalence/invariance of psychosocial measures of PA in a sample of adolescent girls at two time-points. Similar, but not identical, measures of social support, self-efficacy, and self-management strategies were used to those used in the current study. General longitudinal measurement invariance was established for all psychosocial measures of interest. Motl, Dishman, and colleagues found similar measurement invariance across two time-points (1-year apart) in a sample of adolescent girls for a measure of social provisions (Motl, Dishman, Saunders, Dowda, & Pate, 2004), with some statistical differences in the variances of the item uniqueness terms, and across a 3-year period for measures of self-efficacy, perceived behavioral control, attitudes, subjective norms, and intentions (Dishman et al., 2006). While there are subtle differences between the measures of the current study and those of Motl and Dishman, the evidence suggests that the psychometric properties of these psychosocial measures are robust across time.

Interestingly, longitudinal measurement invariance was established in the current intervention study when the data for both the target intervention group and the comparison

group were aggregated. However, two caveats exist. First, no true control group was used. The comparison group (i.e., Sunsmart group) received a sun protection intervention consisting of an initial primary care office-based computer assessment of sun protection behaviors followed by tailored computer-generated printed recommendations for improvement, such as wearing protective clothing and use of sunscreen (Norman et al., 2007). Brief counseling phone calls were also provided by a health counselor at 3 months and 6 months followed by a mailed feedback report and tip sheet to encourage continued sun protection behavior. One could argue that the PACE+ group and the Sunsmart group were similarly impacted by their respective interventions with respect to responding on the psychosocial PA measures. Second, compliance with the PACE+ intervention was not explicitly monitored. This could have also resulted in more similarities between the two groups.

The focus of the current paper was to establish the measurement equivalence of target psychosocial measures across time. That is, to determine if each psychosocial measure exhibited stationarity. A natural extension of this work is to examine the *structural* invariance of these psychosocial measures. While it is not necessary to establish invariance of structural parameters (factor variances/covariances, latent factor means) to exhibit measurement equivalence (Little, 1997), these parameters can be used to answer substantive questions of interest. Testing for the invariance of factor covariances (or autoregressive paths), for example, across 3 or more occasions would determine if the relative rank order of participants' scores remains the same across time—often referred to as differential stability. Building on this model, one could add lagged paths from a psychosocial mediator at an earlier time-point (e.g., self-efficacy at time 1), for example, to PA at a later time-point (time 2) while controlling for the autoregressive effects of PA at time 1. With 3 or more time-points one could statistically compare whether or not the lagged effects are invariant across time, and test for mediation at multiple time points. Differences found at the structural level are substantively interesting, as they identify group differences/predictive associations that differ as a function of time—assuming that longitudinal measurement invariance has already been established.

Acknowledgments

This project was supported by the National Cancer Institute (R01 CA113828).

References

- Baranowski T, Anderson C, Carmack C. Mediating variable framework in physical activity interventions: How are we doing? How might we do better? *American Journal of Preventive Medicine*. 1998; 15:266–297. [PubMed: 9838973]
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107:238–246. [PubMed: 2320703]
- Bentler PM. On tests and indices for evaluating structural models. *Personality and Individual Differences*. 2007; 45:825–829.
- Bentler, PM. EQS 6 Structural Equations Program Manual. Encino, CA: Multivariate Software, Inc; (in press)
- Berry T, Naylor PJ, Wharf-Higgins J. Stages of change in adolescents: An examination of self-efficacy, decisional balance, and reasons for relapse. *Journal Adolescent Health*. 2005; 37:452–459.
- Byrne BM, Stewart SM. The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*. 2006; 13:287–321.
- Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*. 2007; 14:464–504.

- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 2002; 9:233–255.
- Deshon RP. Measures are not invariant across groups with error variance homogeneity. *Psychological Science*. 2004; 6:137–49.
- Dishman RM, Hales DP, Sallis JF, Saunders R, Dunn AL, Bedimo-Rung AL, Ring KB. Validity of social-cognitive measures of physical activity in middle-school girls. *Journal of Pediatric Psychology*. 2010; 35:72–88. [PubMed: 19433571]
- Dishman RM, Saunders R, Felton F, Ward DS, Dowda M, Pate RR. Goals and intentions mediate efficacy beliefs and declining physical activity in high school girls. *American Journal of Preventive Medicine*. 2006; 31:475–483. [PubMed: 17110077]
- Duncan SC, Duncan TE, Strycker LA, Chaumeton NR. A cohort-sequential latent growth model of physical activity from ages 12 to 17 years. *Annals of Behavioral Medicine*. 2007; 33:80–89. [PubMed: 17291173]
- Grunbaum JA, Kann L, Kinchen SA, Williams B, Ross JG, Lowry R, Kolbe L. Youth Risk Behavior Surveillance--United States, 2001. *Morbidity and Mortality Weekly Report*. 2002; 51:1–64.
- Hoyle, RH. Confirmatory factor analysis. In: Tinsley, HEA.; Brown, SD., editors. *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. San Diego, CA: Academic Press; 2000. p. 465–497.
- Kelloway EK. Structural equation modeling in perspective. *Journal of Organizational Behavior*. 1995; 16:215–224.
- Little TD. Means and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*. 1997; 32:53–76.
- Little TD, Preacher KJ, Selig JP, Card NA. New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*. 2007; 31:357–365.
- Lubans DR, Foster C, Biddle SJH. A review of mediators of behavior in interventions to promote physical activity among children and adolescents. *Preventive Medicine*. 2008; 47:463–470. [PubMed: 18708086]
- Marcus BH, Owen N. Motivational readiness, self-efficacy and decision-making for exercise. *Journal of Applied Social Psychology*. 1992; 22:3–16.
- Marcus BH, Rakowski W, Rossi JS. Assessing motivational readiness and decision making for exercise. *Health Psychology*. 1992; 11:257–261. [PubMed: 1396494]
- Marsh HW, Hau K, Wen Z. In search of golden rules: Comment on hypothesis testing approaches to setting cut-off values for fit indexes and dangers in overgeneralization Hu and Bentler's (1999) findings. *Structural Equation Modeling*. 2004; 11:320–341.
- MacCallum RC, Browne MW, Cai L. Tested differences between nested covariance structure models: Power analyses and null hypotheses. *Psychological Methods*. 2005; 11:19–35.
- Meredith W. Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*. 1993; 58:525–543.
- Motl RW, Dishman RK, Saunders RP, Dowda M, Pate RR. Measuring social provisions for physical activity among adolescent black and white girls. *Educational and Psychological Measurement*. 2004; 64:682–706.
- Norman GJ, Adams MA, Calfas KJ, Covin J, Sallis JF, Rossi JS, Redding CA, Cella J, Patrick K. A Randomized trial of a multicomponent intervention for adolescent sun protection behaviors. *Archives of Pediatric and Adolescent Medicine*. 2007; 161:146–152.
- Patrick K, Calfas KJ, Norman GJ, Zabinski MF, Sallis JF, Rupp J, Covin J, Cella J. A randomized controlled trial of a primary care and home-based intervention for physical activity and nutrition behaviors: PACE+ for adolescents. *Archives of Pediatric and Adolescent Medicine*. 2006; 160:128–136.
- Pentz MA, Chou CP. Measurement invariance in longitudinal clinical research assuming change from development and intervention. *Journal of Consulting and Clinical Psychology*. 1994; 62:450–462. [PubMed: 8063972]
- Personality and Individual Differences. Special issue on structural equation modeling. *Personality and Individual Differences*. 2007; 42:811–898.

- Ployhart RE, Oswald FL. Applications of mean and covariance structure analysis: Integrating correlational and experimental approaches. *Organizational Research Methods*. 2004; 7:27–65.
- Saelens BE, Gehrman CA, Sallis JF, Calfas KJ, Sarkin JA, Caparosa S. Use of self-management strategies in a 2-year cognitive-behavioral intervention to promote physical activity. *Behavior Therapy*. 2000; 31:365–379.
- Sallis JF, Pinski RB, Grossman RM, Patterson TL, Nader PR. The development of self-efficacy scales for health-related diet and exercise behaviors. *Health Education Research*. 1988; 3:283–292.
- Sallis JF, Taylor WC, Dowda M, Freedson PS, Pate RR. Correlates of vigorous physical activity for children in grades 1 through 12: Comparing parent-reported and objectively measured physical activity. *Pediatric Exercise Science*. 2002; 14:30–44.
- Steiger JS. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*. 1990; 25:173–180.
- Strong WB, Malina RM, Blimkie CJ, Daniels SR, Dishman RK, Gutin B, Hergenroeder AC, Must A, Nixon PA, Pivarnik JM, Rowland T, Trost S, Trudeau F. Evidence based physical activity for school-age youth. *Journal of Pediatrics*. 2005; 146:732–737. [PubMed: 15973308]
- Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organization research. *Organizational Research Methods*. 2000; 3:4–70.
- Wharf-Higgins J, Gaul C, Gibbons S, Van Gyn G. Factors influencing physical activity levels among Canadian youth. *Canadian Journal of Public Health*. 2003; 94:45–51.

Figure 1.
Single-factor self-efficacy model across two time-points.

Table 1

Descriptive statistics and Cronbach alpha values for each measure across time-points

Measure	<i>M (SD)</i>	<i>α</i>
Family Support	2.79–2.93(0.97–1.03)	0.74–0.79
Peer Support	2.70–2.78(0.98–1.04)	0.70–0.77
Decisional Balance-Pros	3.91–3.97(0.75–0.80)	0.70–.077
Decisional Balance-Cons	1.64–1.76(0.61–0.68)	0.64–0.70
Self-Efficacy	3.45–3.73(0.86–0.92)	0.74–0.85
Behavior Change Strategies	3.08–3.23(0.66–0.79)	0.86–0.91

Note. Ranges for these summary statistics include the minimum and maximum value for a measure across the 4 time-points.