Published in final edited form as:

Ann Behav Med. 2010 October ; 40(2): 164–175. doi:10.1007/s12160-010-9208-2.

Validity of Processes of Change in Physical Activity Among College Students in the TIGER Study

Rod K. Dishman, Ph.D.,

Department of Kinesiology, The University of Georgia, Ramsey Student Center, 330 River Road, Athens, GA 30602-6554, USA

Andrew S. Jackson, H.S.D., and

University of Houston, Houston, TX, USA

Molly S. Bray, Ph.D.

University of Alabama-Birmingham, Birmingham, AL, USA

Rod K. Dishman: rdishman@uga.edu

Abstract

Objective—To test the factorial validity and measurement equivalence/invariance of scales used to measure processes of change derived from the Transtheoretical Model (TTM) applied to physical activity.

Methods—Confirmatory factor analysis of questionnaire responses obtained from a diverse sample ($N=1,429$) of students enrolled in the Training Interventions and Genetics of Exercise Response (TIGER) Study at the University of Houston during academic years 2004–2005 through 2007–2008. Cohorts of students ($N=1,163$) completed the scales at the beginning and end of each Fall semester, permitting longitudinal analysis.

Results—Theoretically and statistically sound models were developed that support the factorial validity of nine of the ten hypothesized 1st-order factors. A structure of nine correlated 1st order factors or a hierarchical structure of those factors subordinate to two correlated 2nd-order factors were each defensible. Multi-group invariance of each model was confirmed across race/ethnicity groups (African American, Hispanic, non-Hispanic White), gender, age, BMI levels, employment status, physical activity level, and study adherence. Longitudinal invariance across the semester was also confirmed.

Conclusion—The scores from the scales provide valid assessments that can be used in observational studies of naturally occurring change or in interventions designed to test the usefulness of TTM processes as mediators of change in physical activity among college students. Item content and factor structure require further evaluation in other samples in order to advance TTM theory applied to physical activity.

Keywords

African American; Confirmatory factor analysis; Hispanic/Latino; Measurement equivalence/invariance; Mediators; Transtheoretical model

Introduction

The processes of change derived from the transtheoretical model (TTM) [1, 2] are putative mediators of change in physical activity [3, 4] that have been understudied [5–7]. The TTM theorizes that people use experiential (i.e., cognitive–affective) and behavioral (i.e., overt tactics) processes to alter their experiences and environment in ways to prompt or support their attempts to move between progressive stages of change from building intention to subsequent adoption and maintenance of regular physical activity [8, 9]. The experiential and behavioral processes are conceptualized as two correlated second-order factors that each consist of five first-order constructs [2, 9].

A few prospective, observational studies of adults reported mixed evidence that the processes are related to stage progression [10–13]. Likewise, experimental evidence has been mixed as to whether the processes mediate the effects of stage-based interventions to increase physical activity [14–18].

Resolution of the equivocal evidence about the usefulness of the TTM for understanding and changing physical activity [4, 5, 7] has been hampered by the absence of confirming evidence to support the original conceptualization of TTM processes derived from psychotherapy systems [1] and smoking cessation [9] that were then applied to physical activity [2, 19, 20]. It remains to be determined whether the TTM processes can be validly assessed in population subgroups that differ according to characteristics that might modify the effectiveness of physical activity interventions (e.g., race/ethnicity, gender, age, body mass index (BMI), employment status, and physical activity history) [5, 20]. We are aware of only two studies that tested the factorial validity of existing measures of TTM processes for physical activity. Rhodes et al. [21] did not confirm the hypothesized structure of the original processes measure developed for physical activity [19], which was modified for use with adolescents. Paxton et al. [22] used a different measure [23] with a large multi-ethnic sample of adults and reported that several items loaded significantly on more than one construct and that the constructs of reinforcement management and self-liberation were not supported as separate factors. Hence, despite the large literature on applications of the TTM to physical activity [4–6], the validity of existing measures used to assess processes of change applied to physical activity is not established. It is also not known whether the scale's measurement properties hold across time.

The three purposes of the study were as follows: (1) to examine the factorial validity of the original measures of processes of physical activity change (i.e., a test of the hypothesized ten-factor structure subordinate to two correlated higher-order factors) [19]; (2) to subsequently test re-specified structural models based on observed fit to the data in this sample; and (3) to determine whether the structure and scaling metrics of the re-specified models were equivalent/invariant across time and between groups of college students who differed according to race/ethnicity, gender, age, BMI, employment status, physical activity level, and program adherence/dropout status. Our overall goals were twofold. First, we wanted to derive scales that yield scores with adequate reliability and factorial validity for use in the ongoing TIGER study, and possibly other diverse samples of college students. Second, we wanted to provide a starting point for other investigators to evaluate whether the poor fit [19, 21, 22], and uneven performance [4, 5], of the original processes model that has been reported previously is because of population specificity, poor item content, or improper conceptualization of the TTM constructs when they are applied to physical activity.

Furthermore, we evaluated the measurement equivalence/invariance of the scales. The establishment of multi-group and longitudinal factorial invariance is necessary before meaningful inferences between groups or across time can be made about the meaning of

scores on psychometric tests. Factorial validity is the degree to which the structure of a measure conforms to the theoretical definition of its construct [24]. Multi-group factorial invariance is the degree to which factor structure (i.e., configural), factor loadings (i.e., metric), factor variances/covariances, item intercepts or means (i.e., scalar), and item errors (i.e., uniquenesses) are similar between different types of people [25]. Longitudinal factorial invariance is the degree to which those measurement properties are similar across points in time and is necessary for the proper interpretation of change across time in tests of mediation or moderation of change [26]. Without evidence for factorial invariance, differences between groups or across time in scores on a measure might reflect differences in the measurement properties of the self-report instrument used rather than true differences in the latent variable.

Here, we report on the factorial validity, multi-group invariance (i.e., African American, Hispanic, and non-Hispanic White; male and female; age levels; low, average, and high BMI; employment status, physical activity level, and study adherence/dropout status), and longitudinal invariance (i.e., across each Fall semester (15 weeks) from 2004 to 2007) of TTM process scales in a large, diverse sample of college students enrolled in an online health and weight management course at the University of Houston. The group comparisons reflect major features of diversity among the students that are known to be associated with physical activity levels in other populations [27, 28].

Methods

Participants

Participants were enrolled in KIN1304, Public Health Issues in Physical Activity and Obesity, which is an interactive, online course in the university core requirement for all undergraduate degree programs. Students were excluded from participation in the study if they had a physical or physiological contra-indication to aerobic exercise as identified by their physician (e.g., cardiomyopathy), a known metabolic disorder that may alter body composition (e.g., lipodystrophy), were pregnant or lactating, or were already actively participating in a regular exercise program within the previous 30 days. The course was developed as the education arm of an NIH funded research project entitled TIGER: Training Interventions and Genetics of Exercise Response. The broad aim of the TIGER study is to investigate the effect of genetic variation on exercise responses in a multiracial/ethnic cohort of college-age men and women, both at baseline and after 15- or 30-week exercise programs. The course includes eight bi-weekly lectures on physical activity and health (i.e., cardiovascular health, body composition, nutrition, genetics, energy balance, exercise program design, stretching, and exercise tracking), content examinations, online forums, and a two-semester, exercise training program 3 days/week. Each exercise session lasts at least 40 min and includes a 5-min warm up, a 30-min workout, and a 5-min cool down. Participants perform their choice of aerobic exercises (i.e., stationary cycling, treadmill or track running/walking, elliptical stepping, rowing, stair stepping, and arm cranking), which are documented by direct observation. Intensity is regulated using computerized heart rate monitors (Polar E600) with audible feedback set between 65% and 85% of each person's age- and gender-specific predicted maximum heart rate reserve. Participants are required to complete at least 25 min within that intensity range for an exercise session to be considered complete. Heart rate monitors are downloaded via infrared readers directly to a database record of the intensity and duration of each exercise session. To track participation, each student is given a barcode ID tag that is scanned at each exercise session. In each of the first 5 years of the study, a cohort of participants (about 250±20 students per semester) has undergone the exercise protocol, as well as three physical examinations during the year-long course: prior to beginning the exercise training protocol, at the midpoint after completing the

first semester of exercise training (approximately 15 weeks) and once more after completing the 30-week training protocol.

The student body at the University of Houston is one of the most ethnically diverse among research universities in the USA. It is comprised of approximately 35,000 undergraduate and graduate students, more than 70% of whom come from the Houston metropolitan area. Approximately 12% of the student population and the TIGER participants live on campus. The course development and curriculum has been described more fully elsewhere [29]. The multiracial/ethnic sample for this report was 28.6% non-Hispanic White, 27.5% African American, 22.8% Hispanic, 7.3% Asian, 3.8% Asian Indian, 0.2% American Indian or Native Alaskan, 2.4% other, and 6.2% multiracial. Other participant characteristics are described in Table 1.

Study Design

Data Collection Procedures—All measurement protocols were reviewed and approved by the Institutional Review Boards at the University of Alabama at Birmingham, Baylor College of Medicine, the University of Houston, and the University of Texas Health Science Center at Houston. Participants signed a consent form prior to measurements. The psychometric forms were completed in a quiet room on paper teleforms that were individually identified by a barcode. Once completed, a member of the study staff quickly checked to make sure that participants had answered all the questions. After this, the form was scanned into the database using an automated system (the staff member then visually checked the information that the scanner was entering and made corrections as needed, entry 1). A different member of the study staff scanned the same form as described previously (entry 2). Entries 1 and 2 were then compared using Access database queries. If both entries matched, then the data were transferred to a master database. If the entries did not match, the forms were visually checked to determine where the error was located, and errors were corrected prior to transfer to the master database.

Measures—Students responded to questions about their gender, age, race/ethnicity, physical activity level, and their marital and employment status. They were asked whether they thought of themselves as non-Hispanic White, Hispanic White, African American or Black, Native American or Alaskan, Asian Indian, Asian, or other (e.g., multiracial).

BMI was computed as Quetelet's index (weight in kg/height in m²). Body weight was measured to the nearest 1/10 kg using a digital scale (Seca 770, Hanover, MD, USA), and height was assessed to the nearest millimeter with a portable stadiometer (Seca Road Rod). Participants were measured in their bare feet or wearing socks after removing all excess clothing and any heavy accessories. The stability of BMI between semesters (ICC-2) was 0.994.

Physical activity history was assessed using an ordinal eight-category self-report measure that has been validated using maximal aerobic capacity as the criterion measure [30, 31]: (0=none; 1=minimal; infrequent activity, 2=moderate; 10–60 min/week, 3=moderate; over 1 h/week, 4=vigorous; run less than 1 mile or 30 min of comparable physical activity per week, 5=vigorous; run 1–5 miles or 30–60 min of comparable physical activity per week, 6=vigorous; run 5–10 miles or 1–3 h of comparable physical activity per week, 7=vigorous; run over 10 miles or over 3 h of comparable physical activity per week). The stability of physical activity history from beginning to end of the semester (ICC-2) was 0.571.

Processes of change for physical activity were assessed using a 39-item self-report measure previously validated for use with adults [19]. The measure is conceptualized as consisting of two correlated second-order factors which each include five first-order factors. The

experiential processes included (1) consciousness raising, e.g., seeking information; (2) dramatic relief, e.g., emotional aspects of change; (3) environmental reevaluation, e.g., assessment of how inactivity affects society; (4) self-reevaluation, e.g., assessment of personal values; (5) social liberation, e.g., awareness, availability, and acceptance of active lifestyles in society. Behavioral processes consisted of (1) counter conditioning, e.g., substituting physical activity for other leisure choices; (2) helping relationships, e.g., using social support during change; (3) reinforcement management, e.g., self-reward for change; (4) self-liberation, e.g., commitment and efficacy beliefs about change; (5) stimulus control, e.g., managing situations that prompt inactivity or activity. Items from the processes of change questionnaire [19] were rated by participants using a five-point Likert-type response format. The order of item presentation was randomized to minimize response bias of items clustered within their hypothesized construct, but item numbers from the original publication [19] are used in this report. A more recent version of the scale has modified the wording of some items, references physical activity rather than exercise, and includes another dramatic relief item [32].

Data Analysis—Confirmatory factor analysis (CFA) models were tested with full-information maximum likelihood estimation using *Mplus* 5.1 [33]. The proportion of missing item responses for each scale ranged from 6.5% to 7.3%. Overall, missingness was 6.9% (3,821 of 55,731 responses). In contrast to other techniques such as pairwise and listwise deletion of cases, full-information maximum likelihood estimation yields accurate fit indices and parameter estimates with up to 25% simulated missing data [34]. Covariances were computed for >98.4% of the variables for cross-sectional analyses and >75% for longitudinal analyses. Listwise deletion per scale would have retained 85.9% participants. Item/scale descriptive statistics were obtained using SPSS 16.0. Internal consistency reliability of each scale was estimated by the Cronbach alpha coefficient and by composite reliability based on CFA. Alpha underestimates the reliability of a composite score, especially for a multidimensional scale, because it assumes uncorrelated errors among the indicators [35]. Hence, composite reliability was also estimated from each factor structure [$\sum \text{factor loadings}^2 / (\sum \text{factor loadings}^2 + \sum [1 - (\text{factor loading})^2])$]. Full-information maximum likelihood parameter estimates are robust to non-normality [36]. No items had skewness or kurtosis values greater than 1.4 (see Table 2). Factor models were adjusted for nesting effects of students within semester cohorts by correcting the standard errors of the adjusted parameter estimates for between-semester variance using the Huber–White sandwich estimator [33].

Analysis and Fit: The comparative fit index (CFI), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and the chi-square (χ^2) statistic were used to evaluate and compare model fit. The χ^2 statistic was used to assess absolute fit of the model to the data. This statistic is very sensitive to sample size and suggests rejection of the hypothesized model in most cases [35]. For this reason, it is reported but is not used alone to draw specific conclusions about model fit [37].

The CFI tests the proportionate improvement in fit by comparing the target model to a baseline model [38]. Values of the CFI around 0.90 are considered acceptable, while values ≥ 0.95 indicate good fit [37, 38]. The RMSEA is a standardized estimate that represents closeness of fit of population data to the model, is sensitive to mis-specified factor loadings, and is widely considered one of the most informative fit criteria [25]. Values of the RMSEA ≤ 0.06 and ≤ 0.08 reflect close and acceptable fit of the model [37, 39]. The 90% confidence interval (CI) for the RMSEA is also presented. The SRMR represents the average error between the observed and specified covariances and is sensitive to mis-specified latent structure. Concurrent values ≥ 0.96 for CFI and ≤ 0.08 for SRMR provide optimal protection against type I and type II error rates, especially in sample sizes ≤ 250 [37]. Although factors

such as the number of indicators and non-normal distributions affect statistical power, the available sample size was adequate for model tests in the overall sample and for subgroup analyses according to condition [40].

Nested models were compared based on χ^2 difference tests and changes in the values of the CFI, RMSEA, and SRMR [25]. Although χ^2 difference tests were conducted, differences in the RMSEA and CFI have been found to be superior to interpretations based strictly on χ^2 difference tests [37]. The main criterion used to judge significant model differences was a change in CFI of >0.01 between nested models [41]. Overlap in the RMSEA point estimates and 90% CIs between two nested models, as well as the fit of each successive model, were also used to judge meaningful change in fit between models.

Models: The factor validity of each scale was examined first by fitting the hypothesized model of ten correlated first-order factors subordinate to two correlated second-order factors and an alternative model of ten correlated first-order factors [19] to the baseline data from a random holdout sample of 500 students using CFA [42]. If the hypothesized model was not supported, modification indices, cross-loadings of items on other factors, covariances between items, standardized residuals, and squared multiple correlations were examined to determine whether misfit was a function of a problem item or the hypothesized factor structure [42]. The revised model was then tested in the full sample. After establishing a good fitting model, multi-group and longitudinal factor invariance was examined. The primary analyses involved testing the factor invariance across males and females and between African American, Hispanic, and non-Hispanic White students. Secondary analyses were conducted to determine whether the instruments were invariant across age (<20 , ≥ 20 –21, and ≥ 22 –35 years), BMI categories (<25 , ≥ 25 –29, and ≥ 30), employment status (yes or no) of full-time students, physical activity levels (≤ 2 , 3–4, and ≥ 5 on the physical activity category scale) [31], and between study adherents (those who maintained enrollment throughout the semester) and dropouts (those who dropped the course after week 1 but before the end of the second semester) during each academic year. The age groups correspond roughly to comparison of first year, second through third year, and fourth year, non-traditional or graduate students. The BMI groups correspond to CDC standards for normal weight, overweight, and obese classifications among adults. Longitudinal invariance was also tested for the whole sample and between the student subgroups to determine whether the measurement properties of the scales were equivalent across the first semester of enrollment (i.e., 15 weeks) between students who differed according to the aforementioned characteristics.

Factor invariance for each scale was examined by testing and comparing a series of nested models using standard procedures [25]. The first step was to fit the model for a given instrument to the data from each group separately (e.g., African American, Hispanic, and non-Hispanic White for the race analysis). This allowed the adequacy of the model to be assessed within each group prior to the multi-group invariance analysis. Sample sizes were too small (e.g., <200) to estimate stable parameters for other racial groups [40, 43]. The invariance analysis involved testing and comparing four successive models to determine whether model fit was affected by constraining sets of parameters to be equal across groups. Each successive model (models 1 to 4) included previous model restrictions (i.e., model 3 included restrictions from model 2) plus additional constraints, resulting in a series of nested models; model 1 tested the equivalence of the hypothesized pattern of paths, factor variances, item means, and item errors across groups. In this model, all hypothesized parameters were freely estimated in the groups. Model 2 had restricted paths from the factor(s) to the observed items (factor loadings). Model 3 constrained item intercepts (means) to be equal, while in model 4, the item uniquenesses (errors) were constrained across groups. The most constrained, equivalent model is reported in the tables. Item errors

reflect random variance or systematic variance otherwise not explained by the factor model. Testing the equivalence of item means and errors is very restrictive, and equivalence of factor structure (configural invariance) and loadings (metric invariance) is conventionally considered sufficient criteria for concluding factorial invariance across groups [25].

The general model used to test longitudinal invariance of each instrument was a two-wave (times 1 and 2) single-factor model, which includes auto-correlated errors [44]. The measurement error terms (item uniquenesses) are allowed to co-vary because some of the systematic variance unaccounted for by the latent factor should be the same over time. Comparisons of successive, nested models model 1 – model 4 tested the stationarity of the scales (i. e., are measurement properties of the scales equivalent across time?). The stability (Do participants remain in the same rank order over time?) was also assessed. The stability coefficient is estimated as the correlation between factor scores at two time points.

Results

General Descriptives

Scale means, standard deviations, reliabilities, and kurtosis values are shown in Table 2. The internal consistency reliabilities of the re-specified factors ranged from 0.54 (social liberation) to 0.87 (consciousness raising). Most were above 0.70 or 0.80. They are comparable or slightly lower than those reported in the original validation samples [19], despite a nearly one-third reduction in the number of items.

Factor Validity

The hypothesized model of ten first-order factors subordinate to two correlated second-order factors ($r=0.85$) had poor fit to the data at baseline ($X^2=4,590.8$ $df=691$, $p<0.001$, CFI=0.813, RMSEA=0.065 (95% CI, 0.063–0.067), SRMR=0.077). Likewise, an alternative model of ten correlated first-order factors had inadequate fit ($X^2=3,832.6$ $df=657$, $p<0.001$, CFI=0.848, RMSEA=0.060 (95% CI, 0.058–0.062), SRMR=0.067).

Modification indices revealed that several items should be removed because of cross-loadings on other factors or large covariances with other items, indicating redundant content. These included items 32, 34, and 35 of the self-liberation factor. Item 32 (“I tell myself I am able to keep exercising if I want to”) had negative cross-loadings on all factors except self-reevaluation, counter conditioning, and helping relationships and had large covariances with item 20 (“Instead of remaining inactive, I engage in some physical activity”) and item 33 (“I tell myself that if I try hard enough I can keep exercising”). Item 34 (“I make commitments to exercise”) cross-loaded on counter conditioning and had large covariances with items 20 and 33. Item 35 (“I remind myself that I am the only one who is responsible for my health and well-being, and that only I can decide whether or not I will exercise”) cross-loaded on environmental reevaluation, self-reevaluation, and social liberation and had large covariances with item 12 (“I am considering the idea that regular exercise would make me a healthier, happier person to be around”) and item 30 (“When I exercise, I tell myself that I am being good to myself by taking care of my body in this way”). Removal of items 32, 34, and 35 left item 33 as a specific factor, so it was also removed, resulting in deletion of the self-liberation factor.

Other items also cross-loaded on other scales or had large covariances. Item 1 of consciousness raising (“I recall information people have personally given me on the benefits of exercise”) cross-loaded on environmental reevaluation, self-reevaluation, social liberation, and helping relationships and had large covariances with item 12 of self-reevaluation (“I am considering the idea that regular exercise would make me a healthier, happier person to be around”) and item 24 of helping relationships (“I have someone on

whom I can depend when I am having problems with exercising”). Item 9 of environmental reevaluation (“I wonder how my inactivity affects those people who are close to me”) cross-loaded on dramatic relief and had a large covariance with item 17 of social liberation (“I am aware of more and more people encouraging me to exercise these days”), which cross-loaded on all factors except counter conditioning.

Item 15 of self-reevaluation (“I consider the fact that I would feel more confident in myself if I exercised regularly”) cross-loaded on environmental reevaluation, had negative cross-loadings on counter conditioning, reinforcement management, and stimulus control and large covariances with items 12 and 8 of environmental reevaluation (“I feel I would be a better role model for others if I exercised regularly”). Item 23 of counter conditioning (“When I’m feeling tense, I find that exercise is a great way to relieve my worries”) had a large covariance with item 21 (“Rather than viewing exercise as simply another task to get out of the way, I try to use it as my special time to relax and recover from the day’s worries”), which was retained because its content was more specific to the factor than that of item 23. Item 26 of helping relationships (“I have someone who points out my rationalizations for not exercising”) cross-loaded on dramatic relief and had a large covariance with item 17.

A re-specified model with the self-liberation factors and the other items removed still fit poorly. Item 30 of reinforcement management (“When I exercise, I tell myself that I am being good to myself by taking care of my body in this way”) had large cross-loadings on environmental reevaluation, self-reevaluation, and social liberation. With item 30 removed, item 29 of reinforcement management (“I try to set realistic goals for myself rather than setting myself up for failure by expecting too much”) cross-loaded on social liberation, counter conditioning, and stimulus control. Item 37 of stimulus control (“I keep things around my place of work that remind me to exercise”) had a large covariance with item 36 (“I put things around my home to remind me of exercising”), which was retained because of better model fit.

Re-specified Model

After removing these items, a re-specified model of nine first-order factors subordinate to two correlated second-order factors ($r=0.83$) provided an acceptable fit to the data at baseline ($X^2=1,232.8$ $df=289$, $p<0.001$, CFI=0.928, RMSEA=0.049 (95% CI, 0.047–0.052), SRMR=0.050). See Fig. 1. The second-order factor loadings indicate that the experiential structure accounted for 33% to 83% of the variance in its first-order factors, and the behavioral structure accounted for 45% to 88% of the variance in its first-order factors. An alternative model of nine correlated first-order factors (r ranged from 0.21 to 0.82) had better fit ($X^2=890.1$ $df=263$, $p<0.001$, CFI=0.952, RMSEA=0.042 (95% CI, 0.039–0.045), SRMR=0.039) (χ^2 difference= 342.7 df 26, $p<0.01$; CFI difference=0.024). Factor intercorrelations are presented in Table 3. However, the hierarchical structure was acceptable and was also retained for subsequent tests of factor invariance because of TTM theory [1, 2, 19].

Multi-group Invariance

The revised hierarchical and first-order models each had acceptable fit (CFI $\geq$ 0.90, RMSEA<0.06, SRMR<0.08), for males and females, for African American, Hispanic, and non-Hispanic White students, and for age, BMI, employment status, physical activity, and adherence/dropout groups (see Table 4). Invariance tests indicated that factor structure, factor loadings, item means, and item errors were equivalent across groups according to age, BMI, employment status, and adherence/dropout status (see Table 5). Factor structure, loadings, and item means were equivalent across race/ethnicity groups. Factor structure and

loadings were invariant between males and females and between low, moderate, and high physical activity groups.

Longitudinal Invariance

Factor structure, factor loadings, item means, and item errors were equivalent longitudinally across the 15-week semester for the revised hierarchical and first-order models (see Table 6). The stability coefficients were 0.72 for the experiential factor and 0.71 for the behavioral factor, $p < 0.001$. Stability coefficients for the first-order factors ranged from 0.58 to 0.74 (mean = 0.68), $p < 0.001$, and are presented in Table 3. For tests of multi-group longitudinal invariance, Table 6 contains the fit of the base model (all parameters free; model 1) and the most constrained model judged to be invariant for each analysis (e.g., model 2 is presented if factor loadings, but not item means, were invariant). Configural (i.e., factor structure) and metric (i.e., factor loadings) invariance was supported in all multi-group longitudinal analyses (i.e., change in CFI ≤ 0.01 and values of RMSEA and SRMR were very similar across models). However, overall model fit of the second-order model was less than optimal for race/ethnicity, age, BMI, and physical activity level because of the stringent restrictions of fitting longitudinal invariance models across three subgroups for each variable.

Discussion

The results, obtained from a diverse sample of students enrolled in an online health and weight management course at the University of Houston, mainly confirm the factorial validity of revised self-report scales used to measure TTM processes of change applied to physical activity [19]. Evidence supported the measurement equivalence/invariance of the factor structure and factor loadings in subgroups of the students characterized according to their race/ethnicity, gender, age, BMI, employment status, physical activity level, and their adherence/dropout status during the study. Item means were also invariant between all groups except gender and physical activity level. Item errors were invariant between groups based on age, BMI, employment status, and program adherence. The equivalence of factor structure and factor loadings indicates that scores from the scales can be interpreted as having similar meaning among African American, Hispanic, and non-Hispanic White college students regardless of their gender, age, BMI, employment status, physical activity level, and adherence status [25]. Longitudinal invariance of the scales across the semester was also confirmed, a necessary feature of measurement instruments when they are used to assess change in a variable over time [26]. Hence, scores on the scales obtained as long as 15 weeks apart can be interpreted as having the same meaning. The scales thus appear suitable for use and further evaluation in studies of change in physical activity during an academic semester. The stability of factor scores across the semester (i.e., the extent to which students' rank order of scores stayed the same across time) was moderately high but indicated a considerable amount of naturally occurring change, making the processes feasible targets for intervention.

There is no "control" group in the TIGER program with which to compare factor structures or measurement equivalence/invariance across time, so it is not possible to experimentally test the "effect" of the online course on physical activity or the TTM processes. However, the fact that the re-specified models held up at baseline in all groups and retained at least their structure and factor loadings in those groups across the semester (including a comparison of those who dropped the course and those who adhered) indicates that varying exposures to the TIGER program did not affect factor validity.

The factor solution for the original 39-item scale was not supported. After removing 13 items that had large cross-loadings on other factors or large covariances with other items, re-specified models of nine first-order factors subordinate to two correlated hierarchical factors

or nine first-order correlated factors (which each excluded the originally hypothesized self-liberation factor) had adequate fit to the data. All re-specified factors were identified with three indicator items, except reinforcement management which was under-identified by two items. Nonetheless, the composite reliability of the two-item factor was 0.76, which was better than the highest reliability of 0.69 for any alternative three-item solution for that scale. In addition to their cross-loadings with other factors or high covariances with other items, some items might have performed poorly in this sample because of restricted variance. Deleted items had, on average, lower variance (mean 1.32, SD 0.23, range 0.98–1.69) than retained items (mean 1.52, SD 0.26, range 0.96–2.18), although four deleted items had a variance higher than the average of the retained items. Some of the deleted items might have higher variance and contribute, as initially hypothesized, to the factor structure in other population samples.

The present findings extend an earlier report on the measurement properties of a different measure of processes of change derived from the TTM [22]. In that report on 700 adult residents of Hawaii, two different models were proposed to describe the processes of change. One model was comprised of five fully identified, correlated first-order factors. Another was a two-factor, second-order model that preserved seven of the originally conceptualized processes of change factors. However, those factors were under-identified (i.e., each had only two indicator items). The experiential construct of self-reevaluation was deleted in that cohort because of co-linearity or cross-loadings observed with two processes (i.e., reinforcement management and self-liberation, which were subsequently collapsed into a single factor) from the second-order behavioral factor. Although not reported, the authors also explored the measurement model for an eight-factor correlated model with no second-order factors. Similar to the results we report here, their eight-factor correlated model provided a better fit to the data when compared to two-factor second-order model, but the authors proceeded with the higher-order model because of TTM theory. Also, similar to our present findings, the correlation between the experiential and behavioral factors was 0.82, which supports that the experiential and behavioral factors appear to represent independent but highly related constructs.

Another study [21] of the factor structure of a modified version of the scales we used here [19] on a small sample of 15–17-year-old high school students confirmed the structure of all factors except social liberation. However, reinforcement management, self-liberation, and stimulus control were under-identified by just two items each. In that sample, a model of nine correlated first-order factors (CFI= 0.91, RMSEA=0.06) was a slightly better fit than the hypothesized model of two correlated second-order factors (CFI=0.89, RMSEA=0.06).

Similar to those two reports, a model of nine correlated first-order factors had better fit in our sample of college students, but the hierarchical structure was acceptable and was also retained for subsequent tests of factor invariance because of TTM theory [1, 2, 19] and the extant literature that has applied the processes of change to physical activity [14–18]. The original validation paper on the TTM processes for physical activity change found that the model of ten correlated factors exhibited better fit than the hierarchical model of two correlated second-order factors (χ^2 difference=1,686, *df* 44, *p*<0.001), but the hierarchical model has been perpetuated on the basis of TTM theory derived mainly from psychotherapy and smoking cessation research [1, 9, 19]. The authors of the scales acknowledged that further research would help determine whether the higher-order structure is appropriate when applied to exercise [19]. Unfortunately, that research has been slow to develop and has lacked the requisite reevaluation of the measurement structure of the process variables.

Construct validation is an evolving process that involves the evaluation and re-specification of external structure, tests of stationarity and growth over time, and relations with external

criteria or other variables within nomological networks stipulated by theory [45, 46]. Factorial validity corresponds to the external structure of a test and is one aspect of construct validity. More broadly though, validity refers to the meaning and appropriate use of inferences drawn from test scores [24]. It is the extent to which empirical evidence and theoretical rationales converge to support or refute the adequacy and appropriateness of interpretations and actions based on test scores [46]. The use of confirmatory factor analysis to test hypothesized structures that explain the covariance of responses to scale items is an important aspect of building evidence for the validity of scores derived from the items. However, validity theorists generally agree that judgments about whether a model of factor structure is adequate should be based on multiple criteria that consider theoretical, statistical, and practical considerations about social acceptance and impact [46]. Either the second-order, two-factor model or the nine correlated first-order factor model of the revised process scales is an advance in the development of valid scales because previous models of these process items reported worse fit than we report here [19, 21].

More research is needed to determine whether the original conceptualization of second-order experiential and behavioral factors is a better predictor of change in physical activity than are the individual first-order factors. For example, the second-order model, but not the nine-factor model, helps predict adherence to the TIGER program when the TTM processes are tested in a nomological network that includes self-efficacy and self-motivation, two established predictors of physical activity maintenance (unpublished observations). With one exception that we know of [10], tests of the TTM processes in longitudinal studies of physical activity change have evaluated only the first-order factors [11, 18] or the second-order factors [12–17] without establishing the factorial validity or measurement equivalence/invariance for the populations that were sampled.

Given the limited and inconsistent evidence for the structure of the existing scales used to measure the processes, we recommend renewed attempts to develop contemporary item content that may better indicate the conceptual basis of the hypothesized factors as they apply to physical activity behavior change generally or in specific populations or settings. Meaningful tests of the usefulness of the theorized TTM processes when applied to physical activity cannot occur without a valid technology to measure the hypothesized process constructs. Our results suggest that can be done among college students who are racially/ethnically diverse. Whether that can be done in other populations by a generalized form remains to be determined. Nonetheless, we conclude that, until the structure we report here is verified or new item content can be developed, the scores from these revised scales can provide valid assessments of TTM processes of change as putative mediators of change in physical activity. The scales can be used in observational studies of naturally occurring change or interventions designed to increase physical activity among college students regardless of race/ethnicity, gender, age, BMI, employment status, physical activity level, and adherence/dropout status, especially those who identify themselves as African American or Black, Hispanic/Latino, or non-Hispanic White.

Acknowledgments

This study was funded by NIDDK grant R01DK062148

References

1. Prochaska, J. Systems of psychotherapy: a transtheoretical analysis. Homewood: Dorsey; 1979.
2. Prochaska, JO.; Marcus, BH. The transtheoretical model: applications to exercise. In: Dishman, RK., editor. Advances in exercise adherence. Champaign: Human Kinetics; 1994. p. 161-180.

3. Lewis BA, Marcus BH, Pate RR, Dunn AL. Psychosocial mediators of physical activity behavior among adults and children. *American Journal of Preventive Medicine*. 2002; 23:26–35. [PubMed: 12133735]
4. Hutchison AJ, Breckon JD, Johnston LH. Physical activity behavior change interventions based on the transtheoretical model: a systematic review. *Health Education & Behavior*. 2009; 36:829–845. [PubMed: 18607007]
5. Marshall SJ, Biddle SJ. The transtheoretical model of behavior change: a meta-analysis of applications to physical activity and exercise. *Annals of Behavioral Medicine*. 2001; 23:229–246. [PubMed: 11761340]
6. Spencer L, Adams TB, Malone S, Roy L, Yost E. Applying the transtheoretical model to exercise: a systematic and comprehensive review of the literature. *Health Promotion Practice*. 2006; 7 (4):428–443. [PubMed: 16840769]
7. Brug J, Conner M, Harré N, Kremers S, McKellar S, Whitelaw S. The transtheoretical model and stages of change: a critique. *Health Education Research*. 2005; 20:244–258. [PubMed: 15618324]
8. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. *American Psychologist*. 1992; 47:1102–1114. [PubMed: 1329589]
9. Prochaska JO, Velicer WF, DiClemente CC, Fava J. Measuring processes of change: applications to the cessation of smoking. *Journal of Consulting and Clinical Psychology*. 1988; 56:520–528. [PubMed: 3198809]
10. Lowther M, Mutrie N, Scott EM. Identifying key processes of exercise behaviour change associated with movement through the stages of exercise behaviour change. *Journal of Health Psychology*. 2007; 12:261–272. [PubMed: 17284490]
11. Marcus BH, Simkin LR, Rossi JS, Pinto BM. Longitudinal shifts in employees' stages and processes of exercise behavior change. *American Journal of Health Promotion*. 1996; 10:195–200. [PubMed: 10163299]
12. Plotnikoff RC, Hotz SB, Birkett NJ, Courneya KS. Exercise and the transtheoretical model: a longitudinal test of a population sample. *Preventive Medicine*. 2001; 33:441–452. [PubMed: 11676586]
13. Rhodes RE, Plotnikoff RC, Courneya KS. Predicting the physical activity intention-behavior profiles of adopters and maintainers using three social cognition models. *Annals of Behavioral Medicine*. 2008; 36:244–252. [PubMed: 19057976]
14. Calfas KJ, Sallis JF, Oldenburg B, Ffrench M. Mediators of change in physical activity following an intervention in primary care: PACE. *Preventive Medicine*. 1997; 26:297–304. [PubMed: 9144753]
15. Greaney ML, Riebe D, Ewing Garber C, et al. Long-term effects of a stage-based intervention for changing exercise intentions and behavior in older adults. *Gerontologist*. 2008; 48:358–367. [PubMed: 18591361]
16. Napolitano MA, Papandonatos GD, Lewis BA, et al. Mediators of physical activity behavior change: a multivariate approach. *Health Psychology*. 2008; 27:409–418. [PubMed: 18642998]
17. Pinto BM, Lynn H, Marcus BH, DePue J, Goldstein MG. Physician-based activity counseling: intervention effects on mediators of motivational readiness for physical activity. *Annals of Behavioral Medicine*. 2001; 23:2–10. [PubMed: 11302351]
18. Woods C, Mutrie N, Scott M. Physical activity intervention: a transtheoretical model-based intervention designed to help sedentary young adults become active. *Health Education Research*. 2002; 17:451–460. [PubMed: 12197590]
19. Marcus BH, Rossi JS, Selby VC, Niaura RS, Abrams DB. The stages and processes of exercise adoption and maintenance in a worksite sample. *Health Psychology*. 1992; 11:386–395. [PubMed: 1286658]
20. Dishman RK. Increasing and maintaining exercise and physical activity. *Behavior Therapy*. 1991; 22:345–378.
21. Rhodes RE, Berry T, Naylor P, Higgins SJ. Three-step validation of exercise behavior processes of change in an adolescent sample. *Measurement in Physical Education and Exercise Science*. 2004; 8:1–20.

22. Paxton RJ, Nigg CR, Motl RW, et al. Are constructs of the transtheoretical model for physical activity measured equivalently between sexes, age groups, and ethnicities? *Annals of Behavioral Medicine*. 2008; 35:308–318. [PubMed: 18607667]
23. Nigg CR, Norman GJ, Rossi JS, Benisovich SV. Processes of exercise behavior change: Redeveloping the scale. *Annals of Behavioral Medicine*. 1998:S79.
24. Messick S. Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*. 1995; 50:741–749.
25. Vandenberg RJ, Lance CE. A review and synthesis of the management invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*. 2000; 3:4–69.
26. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. *Annual Review of Psychology*. 2007; 58:593–614.
27. Nelson TF, Gortmaker SL, Subramanian SV, Cheung L, Wechsler H. Disparities in overweight and obesity among US college students. *American Journal of Health Behavior*. 2007; 31:363–373. [PubMed: 17511571]
28. Crespo CJ, Smit E, Andersen RE, Carter-Pokras O, Ainsworth BE. Race/ethnicity, social class and their relation to physical inactivity during leisure time: results from the Third National Health and Nutrition Examination Survey, 1988–1994. *American Journal of Preventive Medicine*. 2000; 18:46–53. [PubMed: 10808982]
29. McFarlin BK, Jackson AS. Development of an online university-based physical activity and obesity education program. *Diabetes Educator*. 2008; 34:766–770. 772–4. [PubMed: 18832283]
30. Jackson AS, Blair SN, Mahar MT, Wier LT, Ross RM, Stuteville JE. Prediction of functional aerobic capacity without exercise testing. *Medicine and Science in Sports and Exercise*. 1990; 22:863–870. [PubMed: 2287267]
31. Jurca R, Jackson AS, LaMonte MJ, et al. Assessing cardiorespiratory fitness without performing exercise testing. *American Journal of Preventive Medicine*. 2005; 29:185–93. [PubMed: 16168867]
32. Marcus, BH.; Forsyth, LH. Motivating people to be physically active. 2. Champaign: Human Kinetics; 2009. p. 1-216.
33. Muthén, LK.; Muthén, BO. Mplus: statistical analysis with latent variables. 51. Los Angeles: Muthén and Muthén; p. 1998-2008.
34. Enders CK, Bandalos DL. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*. 2001; 8:430–457.
35. Bollen, KA. Structural equations with latent variables. New York: Wiley; 1989. p. 1-514.
36. Enders CK. The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*. 2001; 6:352–370. [PubMed: 11778677]
37. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
38. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance-structures. *Psychological Bulletin*. 1980; 88:588–606.
39. Browne MW, Cudeck R. Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*. 1989; 24:445–455.
40. Kaplan D, George R. A study of power associated with testing factor mean differences under violations of factorial invariance structural equation modeling. *Structural Equation Modeling*. 1993; 2:101–118.
41. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 2002; 9:233–255.
42. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin*. 1988; 103:411–423.
43. Comrey, AL.; Lee, HB. A first course in factor analysis. 2. Hillsdale: Erlbaum; 1992. p. 1-430.

44. Pitts SC, West SG, Tein J. Longitudinal measurement models in evaluation research: examining stability and change. *Evaluation and Program Planning*. 1996; 19:333–350.
45. Nunnally, JC. *Psychometric theory*. 2. New York: McGraw-Hill; 1978. p. 86-113.p. 431-436.
46. Messick, S. Validity. In: Linn, RL., editor. *Educational measurement*. New York: Macmillan; 1989. p. 13-103.

Fig. 1.

Proposed model for the processes of change. A hierarchical model of nine first-order factors subordinate to two correlated second-order factors of experiential and behavioral processes is depicted. In the re-specified model, reinforcement management was under-identified by two item indicators, and self-liberation could not be identified. Correlations among the nine first-order factors are reported in the text. Factor loadings and variances for enumerated items are presented

Table 1

Participant characteristics at baseline

Characteristic		Sample (N=1,429)	
		Mean (SD)	Frequency (%)
Age (years)	n=1,429	21.3 (3.0)	
BMI (kg/m ²)	n=1,406	26.4 (6.2)	
Physical activity (0–9)	n=1,372	3.2 (2.0)	
Gender			
Male			563 (39.4)
Female			866 (60.6)
Race/ethnicity			
Non-Hispanic White			408 (28.6)
Hispanic			326 (22.8)
African American or Black			393 (27.5)
Native American or Alaskan			3 (0.2)
Asian Indian			54 (3.8)
Asian			105 (7.3)
Other			34 (2.4)
Multiracial			89 (6.2)
Not reported			17 (1.2)
Marital Status			
Never married (single)			1,226 (85.8)
Married			104 (7.3)
Separated			7 (0.5)
Divorced			21 (1.5)
Domestic partner			30 (2.1)
Widowed			2 (0.1)
Not reported			39 (2.7)
Employment status			
Full-time student employed			815 (57.0)
Full-time student not employed			614 (43.0)

Table 2

Scale means, standard deviations (SD), reliabilities, and kurtosis values

Scale	Samples	Mean	SD	α	CR	Item Kurtosis
Change processes						
Experiential	1,279	3.11	0.74	0.873	0.896	
Consciousness raising	1,327	2.97	1.17	0.870	0.873	-1.15 to -1.00
Dramatic relief	1,318	2.60	1.06	0.846	0.851	-0.91 to -0.76
Environmental reevaluation	1,323	3.33	0.97	0.721	0.736	-0.95 to -0.61
Self-reevaluation	1,311	3.82	0.87	0.672	0.681	-0.90 to 1.15
Social liberation	1,323	2.87	0.97	0.542	0.542	-1.39 to -0.78
Behavioral	1,294	2.82	0.77	0.851	0.861	
Counter conditioning	1,326	3.08	0.97	0.773	0.779	-0.93 to -0.43
Helping relations	1,322	2.84	1.13	0.793	0.793	-1.19 to -1.09
Reinforcement management	1,317	2.89	1.08	0.757	0.760	-0.84 to -0.82
Stimulus control	1,325	2.52	0.89	0.662	0.680	-0.76 to -0.52

Internal consistency calculated as Cronbach α reliability CR composite reliability

Table 3

Standardized correlations among the nine first-order factors

	1	2	3	4	5	6	7	8	9
1. Consciousness raising	<i>0.707</i>								
2. Dramatic relief	0.425	<i>0.578</i>							
3. Environmental reevaluation	0.491	0.446	<i>0.637</i>						
4. Self-reevaluation	0.611	0.497	0.819	<i>0.727</i>					
5. Social liberation	0.644	0.556	0.702	0.662	<i>0.644</i>				
6. Counter conditioning	0.525	0.210	0.320	0.449	0.605	<i>0.727</i>			
7. Helping relationships	0.338	0.363	0.470	0.475	0.604	0.401	<i>0.673</i>		
8. Reinforcement management	0.441	0.367	0.483	0.574	0.671	0.555	0.625	<i>0.686</i>	
9. Stimulus control	0.671	0.421	0.544	0.617	0.799	0.773	0.592	0.671	<i>0.739</i>

Stability coefficients are presented in italics in the diagonal

Table 4

Model fit of change processes measured at baseline

	2nd-order factor structure				9 correlated 1st-order factors			
	χ^2 (289)	CFI	RMSEA (90% CI)	SRMR	χ^2 (263)	CFI	RMSEA (90% CI)	SRMR
Change processes, $N=1,337$	1,233	0.928	0.049 (0.047–0.052)	0.050	890	0.952	0.042 (0.039–0.045)	0.039
Sample groups								
Males, $n=522$	690	0.922	0.052 (0.047–0.057)	0.059	539	0.946	0.045 (0.039–0.050)	0.050
Females, $n=815$	777	0.930	0.045 (0.042–0.049)	0.050	594	0.952	0.039 (0.035–0.043)	0.038
African American, $n=370$	616	0.904	0.055 (0.049–0.061)	0.059	492	0.933	0.049 (0.042–0.055)	0.047
Hispanic, $n=308$	555	0.925	0.055 (0.048–0.061)	0.060	462	0.944	0.050 (0.042–0.057)	0.051
Non-Hispanic White, $n=390$	608	0.910	0.053 (0.047–0.059)	0.064	482	0.938	0.046 (0.040–0.053)	0.047
Age <20, $n=553$	756	0.919	0.054 (0.049–0.059)	0.055	610	0.940	0.049 (0.044–0.054)	0.045
Age 20–21, $n=409$	524	0.939	0.044 (0.038–0.051)	0.054	417	0.960	0.038 (0.031–0.045)	0.041
Age 22–35, $n=375$	566	0.917	0.051 (0.044–0.057)	0.061	453	0.943	0.044 (0.037–0.051)	0.048
BMI <25, $n=679$	867	0.910	0.054 (0.050–0.058)	0.051	723	0.928	0.051 (0.046–0.055)	0.042
BMI 25–29, $n=391$	568	0.917	0.051 (0.045–0.057)	0.063	456	0.945	0.043 (0.037–0.050)	0.050
BMI ≥30, $n=261$	407	0.949	0.040 (0.030–0.048)	0.061	324	0.973	0.030 (0.017–0.040)	0.048
Employed, $n=756$	834	0.925	0.050 (0.046–0.054)	0.052	622	0.951	0.042 (0.038–0.047)	0.040
Not employed, $n=581$	740	0.923	0.052 (0.047–0.056)	0.056	590	0.944	0.046 (0.041–0.051)	0.046
Low active, $n=543$	639	0.925	0.047 (0.042–0.052)	0.055	465	0.957	0.038 (0.032–0.043)	0.041
Moderately active, $n=366$	544	0.921	0.049 (0.043–0.055)	0.060	453	0.941	0.044 (0.037–0.051)	0.050
High active, $n=412$	596	0.913	0.051 (0.045–0.057)	0.055	496	0.934	0.046 (0.040–0.053)	0.046
Adherents, $n=281$	481	0.922	0.049 (0.041–0.056)	0.066	383	0.952	0.040 (0.031–0.049)	0.053
Dropouts, $n=997$	1,030	0.923	0.051 (0.047–0.054)	0.051	764	0.948	0.044 (0.040–0.047)	0.039

χ^2 chi-square test statistic, CI confidence interval, CFI comparative fit index, $RMSEA$ root mean square error of approximation, $SRMR$ standardized root mean square residual

Table 5

Model fit of change processes and measurement equivalence invariance between groups according to student characteristics

Model	2nd-order factor structure				9 correlated 1st-order factors			
	χ^2 (df)	CFI	RMSEA (90% CI)	SRMR	χ^2 (df)	CFI	RMSEA (90% CI)	SRMR
Gender (N=1,337)	1 1,469 (578)	0.927	0.048 (0.045–0.051)	0.054	1,134 (526)	0.950	0.042 (0.038–0.045)	0.043
	2 1,507 (602)	0.926	0.047 (0.044–0.050)	0.055	1,155 (543)	0.950	0.041 (0.038–0.044)	0.043
Race/ethnicity (N=1,068)	1 1,777 (867)	0.914	0.054 (0.051–0.058)	0.061	1,435 (789)	0.939	0.048 (0.044–0.052)	0.048
	3 1,983 (967)	0.904	0.054 (0.051–0.058)	0.070	1,624 (875)	0.929	0.049 (0.045–0.053)	0.055
Age (N=1,337)	1 1,833 (867)	0.924	0.050 (0.047–0.053)	0.056	1,470 (789)	0.946	0.044 (0.041–0.047)	0.045
	4 2,075 (1,019)	0.917	0.048 (0.045–0.051)	0.064	1,690 (927)	0.940	0.043 (0.040–0.046)	0.053
BMI (N=1,331)	1 1,827 (867)	0.921	0.050 (0.047–0.053)	0.057	1,485 (789)	0.943	0.045 (0.041–0.048)	0.046
	4 2,073 (1,019)	0.913	0.048 (0.045–0.051)	0.067	1,697 (927)	0.937	0.043 (0.040–0.047)	0.054
Employment (N=1,337)	1 1,575 (578)	0.925	0.051 (0.048–0.054)	0.054	1,212 (526)	0.948	0.044 (0.041–0.047)	0.042
	4 1,674 (654)	0.923	0.048 (0.045–0.051)	0.056	1,300 (595)	0.947	0.042 (0.039–0.045)	0.045
Physical activity (N=1,321)	1 1,776 (867)	0.920	0.049 (0.046–0.052)	0.056	1,411 (789)	0.946	0.042 (0.039–0.046)	0.045
	2 1,832 (915)	0.920	0.048 (0.045–0.051)	0.062	1,452 (823)	0.945	0.042 (0.038–0.045)	0.048
Adherence (N=1,278)	1 1,477 (578)	0.922	0.049 (0.046–0.052)	0.054	1,126 (526)	0.948	0.042 (0.039–0.046)	0.043
	4 1,519 (654)	0.925	0.045 (0.043–0.048)	0.057	1,160 (595)	0.951	0.039 (0.035–0.042)	0.045

Models 1–4=nested models 1 through 4 are described in the text. The most constrained, equivalent model is reported here

χ^2 chi-square test statistic, df degrees of freedom, CFI comparative fit index, RMSEA root mean square error of approximation, CI confidence interval, SRMR standardized root mean square residual

Table 6

Model fit of change processes and longitudinal measurement equivalence invariance between groups according to student characteristics

Model	2nd-order factor structure				9 correlated 1st-order factors			
	χ^2 (df)	CFI	RMSEA (90% CI)	SRMR	χ^2 (df)	CFI	RMSEA (90% CI)	SRMR
Change processes								
1	3,097 (1,215)	0.924	0.038 (0.036–0.039)	0.048	2,478 (1,095)	0.942	0.035 (0.033–0.036)	0.038
4	3,468 (1,291)	0.912	0.039 (0.038–0.041)	0.054	2,821 (1,164)	0.931	0.037 (0.035–0.038)	0.046
Sample groups								
Gender								
1	4,618 (2,430)	0.912	0.041 (0.039–0.042)	0.054	3,886 (2,190)	0.930	0.038 (0.036–0.040)	0.043
2	4,752 (2,502)	0.914	0.041 (0.039–0.042)	0.057	3,929 (2,241)	0.931	0.038 (0.036–0.040)	0.045
Race/ethnicity								
1	6,100 (3,645)	0.884	0.049 (0.046–0.051)	0.066	5,265 (3,285)	0.904	0.047 (0.044–0.049)	0.054
2	6,106 (3,765)	0.891	0.047 (0.045–0.049)	0.070	5,344 (3,370)	0.904	0.046 (0.044–0.048)	0.056
Age								
1	6,456 (3,645)	0.893	0.046 (0.044–0.048)	0.066	5,695 (3,285)	0.906	0.046 (0.044–0.048)	0.049
2	6,533 (3,765)	0.896	0.045 (0.043–0.047)	0.063	5,757 (3,370)	0.907	0.045 (0.043–0.047)	0.051
BMI								
1	6,482 (3,645)	0.890	0.047 (0.045–0.048)	0.061	5,641 (3,285)	0.906	0.045 (0.043–0.047)	0.050
2	6,542 (3,765)	0.896	0.047 (0.043–0.047)	0.064	5,702 (3,370)	0.907	0.045 (0.043–0.046)	0.051
Employment status								
1	4,454 (2,430)	0.915	0.040 (0.038–0.041)	0.054	3,780 (2,190)	0.933	0.037 (0.035–0.039)	0.044
4	4,995 (2,658)	0.901	0.041 (0.039–0.042)	0.063	4,267 (2,397)	0.921	0.038 (0.036–0.040)	0.053
Physical activity								
1	6,399 (3,645)	0.886	0.047 (0.045–0.049)	0.061	5,615 (3,285)	0.904	0.045 (0.043–0.047)	0.049
2	6,532 (3,765)	0.886	0.046 (0.044–0.048)	0.064	5,684 (3,370)	0.905	0.045 (0.043–0.047)	0.052
Adherence								
1	4,771 (2,430)	0.908	0.042 (0.040–0.044)	0.054	3,998 (2,190)	0.927	0.039 (0.038–0.041)	0.044
2	4,841 (2,502)	0.911	0.041 (0.040–0.043)	0.057	4,062 (2,241)	0.927	0.039 (0.037–0.041)	0.046

Models 1–4=nested models 1 through 4 are described in the text. The most constrained, equivalent model is reported here

χ^2 chi-square test statistic, df degrees of freedom, CFI comparative fit index, RMSEA root mean square error of approximation, CI confidence interval, SRMR standardized root mean square residual