Preventive Medicine **31,** 584–594 (2000)
doi:10.1006/pmed.2000.0735, available online at http://www.idealibrary.com on IDEAL®

Factorial Validity and Invariance of Questionnaires Measuring Social-Cognitive Determinants of Physical Activity among Adolescent Girls

R. W. Motl,* R. K. Dishman,*[,1] S. G. Trost,† R. P. Saunders,† M. Dowda,† G. Felton,† D. S. Ward,‡
and R. R. Pate†

**University of Georgia, Athens, Georgia 30602-6554; †University of South Carolina, Columbia, South Carolina 29208; and*
‡University of North Carolina, Chapel Hill, North Carolina 27599

Background. **There are few theoretically derived questionnaires of physical activity determinants among youth, and the existing questionnaires have not been subjected to tests of factorial validity and invariance. The present study employed confirmatory factor analysis (CFA) to test the factorial validity and invariance of questionnaires designed to be unidimensional measures of attitudes, subjective norms, perceived behavioral control, and self-efficacy about physical activity.**

Methods. **Adolescent girls in eighth grade from two cohorts ($N = 955$ and 1,797) completed the questionnaires at baseline; participants from cohort 1 ($N = 845$) also completed the questionnaires in ninth grade (i.e., 1-year follow-up). Factorial validity and invariance were tested using CFA with full-information maximum likelihood estimation in AMOS 4.0. Initially, baseline data from cohort 1 were employed to test the fit and, when necessary, to modify the unidimensional models. The models were cross-validated using a multigroup analysis of factorial invariance on baseline data from cohorts 1 and 2. The models then were subjected to a longitudinal analysis of factorial invariance using baseline and follow-up data from cohort 1.**

Results. **The CFAs supported the fit of unidimensional models to the four questionnaires, and the models were cross-validated, as indicated by evidence of multigroup factorial invariance. The models also possessed evidence of longitudinal factorial invariance.**

Conclusions. **Evidence was provided for the factorial validity and the invariance of the questionnaires designed to be unidimensional measures of attitudes, subjective norms, perceived behavioral control, and self-efficacy about physical activity among adolescent girls.**

Key Words: **measurement; confirmatory factor analysis; attitude; social norms; perceived control; self-efficacy; race.**

INTRODUCTION

There is broad consensus that physical inactivity is a public health burden in the United States [*1,2*]. Although youth are the most active segment of the U.S. population, the 1997 Youth Risk Behavior Survey (YRBS) [*3*] indicated that 35% of high school students did not meet established guidelines [*4*] for participation in vigorous physical activity (VPA). Not only is the low level of participation in VPA troublesome for the contemporary health of children, but physical activity during childhood might influence adult physical activity [*5,6*]. Hence, it is important to understand the determinants of physical activity among youth to guide interventions designed to increase physical activity [*7*]. Knowledge in this area is limited.

Age, gender, and race clearly are associated with physical activity among youth. Based on 1997 YRBS data [*3*], the percentage of high school students that participates in VPA 3 or more days per week declines steadily with increasing age, and it declines more rapidly among girls than among boys. Only 44% of 12th-grade girls report regular participation in VPA in contrast to 68% of 12th-grade boys. Among adolescent girls, physical activity levels are markedly lower among African-Americans, with only 41% of African-American girls reporting regular participation in VPA compared with 58% of white girls.

The research was supported by Grant NIH HL 57775 from the National Heart, Lung, and Blood Institute.

[1] To whom reprint requests should be addressed at the Department of Exercise Science, The University of Georgia, Ramsey Student Center, 300 River Road, Athens, Georgia 30602-6554. Fax: (706) 542-3148. E-mail: rdishman@coe.uga.edu.

0091-7435/00 $35.00

Although other correlates of physical activity among youth have not been clearly established, a growing literature has suggested that social-cognitive factors such as attitudes, subjective norms, perceived behavioral control, and self-efficacy influence the decision to become physically active among youth. However, theory-based research examining which of those social-cognitive factors are the most important influences of physical activity among youth has been limited [*8–10*]. Three prominent theoretical models employed to study physical activity determinants among youth are the Theories of Reasoned Action (TRA) [*11*] and Planned Behavior (TPB) [*12*] and Social-Cognitive Theory (SCT) [*13*].

Unfortunately, research examining components of TRA, TPB, and SCT as determinants of physical activity has been limited by measurement problems [*14*]. The TRA, TPB, and SCT typically have been tested using single items as observed indicators of latent constructs. Further, factorial validity and factorial invariance have not been tested for the few theoretically derived questionnaires of physical activity determinants among youth, particularly among adolescent black and white girls. Evidence of factorial validity is necessary to demonstrate whether a set of items measures a specific latent variable, and it is directly tested using confirmatory factor analysis (CFA) [*15*]. Factorial validity also is necessary in order to test theories using structural equation modeling (SEM). Utilizing questionnaires without factorial validity (i.e., poor-fitting measurement models) in SEM analyses compromises the fit of the model and impacts the size, significance, and meaning of relationships among variables. Establishing evidence of multigroup or longitudinal factorial invariance (e.g., equivalence of the factor structure and factor loadings across groups or time) using CFA is necessary to determine whether differences in mean scores indicate true group differences or effects of an intervention on the underlying construct, rather than a change in the factor structure and loadings of the questionnaires. It is possible, for example, that differences in mean scores across groups could be attributable to differences in the factor structure of a questionnaire (i.e., group differences in either the number of factors underlying responses to the questionnaire or the magnitude of factor loadings) rather than a difference in the latent construct of interest.

The present study tested the factorial validity and invariance of questionnaires that were designed to be unidimensional measures of attitudes, subjective norms, perceived behavioral control, and self-efficacy about physical activity, and the questionnaires were designed to be consistent with the unidimensional conceptualization of constructs within TRA, TPB, and SCT. The tests of factorial validity and invariance were performed using data from two cohorts of black and white adolescent girls in eighth and ninth grades. The factorial validity of the questionnaires was established by specifying and testing tentative measurement models and then respecifying the measurement models and testing the modifications using baseline data from cohort 1. The final measurement models were cross-validated using a multigroup analysis of factorial invariance on the baseline data from cohorts 1 and 2. The longitudinal factorial invariance of the measurement models was tested using the baseline and follow-up data from the first cohort.

METHODS

Participants

Participants were black and white girls in the eighth and ninth grades from 24 schools in South Carolina. The girls were from two cohorts. The girls in cohort 1 ($N = 955$) had a mean age of 13.7 years ($SD = 0.7$) with racial proportions of 46.7% black, 48.8% white, and 4.5% other; 845 of the girls from cohort 1 completed assessments 1 year later in the ninth grade. Girls in cohort 2 ($N = 1{,}797$) had a mean age of 13.6 years ($SD = 0.6$) with racial proportions of 49.9% black, 45.8% white, and 3.6% other; 0.7% of the girls in cohort 2 did not report race. There was a statistically significant, though trivial, difference between cohorts in age, $t(2{,}733) = 5.71$, $P < 0.001$, but not in the distribution of race, χ^2 (2, $N = 2{,}740$) = 3.52, $P = 0.17$. The procedures were approved by the University of South Carolina Institutional Review Board. All participants and their parent or guardian provided written informed consent.

Questionnaires

The questionnaires were designed to be unidimensional measures of attitudes, subjective norms, perceived behavioral control, and self-efficacy about physical activity. The unidimensional models were consistent with the conceptualization of constructs within TRA, TPB, and SCT. The questionnaires consisted of items that were either modified from previously published instruments [*16*] or specifically developed for the present study. Items on the attitude and self-efficacy measures were phrased similarly to identify a subset of items that best tapped the constructs under investigation and that possessed factorial validity; we were confident that the items on the subjective norm and perceived behavioral control measures adequately tapped the constructs of interest. The items were reviewed by experts and then subjected to a series of pilot studies with eighth-grade girls of similar demographic characteristics to further modify and improve the initial item pool.

The attitude questionnaire included 22 items that consisted of beliefs about the consequences of being physically active and a corresponding positive or negative evaluation of the consequences. The belief statements were rated on a 5-point Likert-type scale anchored by 1 (Disagree a lot) and 5 (Agree a lot); value statements were rated on a 5-point scale ranging from 1 (Very bad) to 5 (Very good). The attitude items were formed as a product of the belief and corresponding value item scores. The subjective norm questionnaire included 8 items that consisted of normative beliefs about the expectations of others toward being physically active and the corresponding motivation to comply with the expectations. The items were rated on a 5-point scale anchored by 1 (Disagree a lot) and 5 (Agree a lot). The subjective norm item scores were formed as a product of the normative belief and motive to comply item scores. The formation of the item scores for the attitude and subjective norm measures based on a multiplicative approach is in accordance with theory and previous research [*17*]. The perceived behavioral control questionnaire included 4 items that pertained to perceptions of the ease/difficulty of being physically active. The items were rated on a 5-point scale. The anchors were 1 (Very easy/Agree a lot) and 5 (Very difficult/Disagree a lot). The self-efficacy questionnaire contained 15 items that pertained to confidence in one's ability to be physically active. The items were rated on a 5-point scale ranging from 1 (Very easy or Disagree a lot) to 5 (Very difficult or Agree a lot).

Procedure

Baseline testing was conducted with cohorts 1 and 2 in the Spring 1998 and 1999 semesters when students were in the eighth grade. Follow-up testing was conducted with cohort 1 in the Spring 1999 semester when students were in the ninth grade. Questionnaires were administered to participants in small group settings (6 to 10 girls) by trained data collectors.

Data Analysis

The data were analyzed in three steps. Initially, the measurement models were tested and, when necessary, modified using baseline data from cohort 1. The final measurement models then were cross-validated using a multigroup analysis of factorial invariance on the baseline data from cohorts 1 and 2. Then the longitudinal factorial invariance of the final measurement models were tested using the baseline and follow-up data from cohort 1.

Confirmatory factor analysis. The measurement models were tested using CFA with full-information maximum likelihood (FIML) estimation in AMOS 4.0 (SmallWaters Corp., Chicago, IL) [*18*]. FIML was selected because there were missing responses to items on the questionnaires, which is a common problem in school-based, longitudinal research using multiple large samples, and it can be attributed to item nonresponse, differential attrition, and absenteeism on the scheduled day of testing.[2] FIML is an optimal method for the treatment of missing data in CFA. It is a theory-based approach [*19*], and it has resulted in more accurate absolute and relative fit indices with simulated missing data than other approaches to missing data such as pairwise deletion, list-wise deletion, and mean imputation [*20*]. Although simulation studies have not been conducted to evaluate fit indices from FIML with ordered categorical data, ML has resulted in accurate absolute and relative fit indices with ordered categorical data of varying degrees of skewness and/or kurtosis [*21*]. ML also is a commonly accepted estimation technique for use with ordered categorical data [*22*]. Standard procedures were employed to establish the fixed, freed, and constrained parameters in the factor loading, factor variance–covariance, and uniqueness matrices. The sample size was adequate based on two criteria: (1) sample size larger than 500 and (2) ratio of sample size to number of freely estimated parameters greater than 10:1 [*23,24*].

Model fit. Model fit was assessed according to multiple indices. The χ^2 statistic assessed absolute fit of the model to the data, but it is sensitive to sample size and assumes the correct model [*24,25*]. Accordingly, other "ad hoc" indices also were employed to judge model fit. The root mean square error of approximation (RMSEA) represents closeness of fit, and values approximating 0.08, 0.05, and 0 demonstrate reasonable, close, and exact fit, respectively [*26*]. The 90% confidence interval (CI) around the RMSEA point estimate also should contain 0.05 and/or 0 to indicate the possibilities of close and/or exact fit [*26*]. The Relative Noncentrality Index (RNI) and Non-Normed Fit Index (NNFI) are incremental fit indices and test the proportionate improvement in fit by comparing the target model with a more restricted, baseline or null model with no structure or correlations among observed variables [*27,28*]. The RNI is non-centrality based and monotonic with model complexity, while the NNFI compensates for the effect of

---

[2] The extent of missing data for the attitude, perceived behavioral control, and self-efficacy questionnaires ranged between 9 and 15% of cases per variable in cohort 1 (M = 10%, Mdn = 10%), 5 and 7% of cases per variable in cohort 2 (M = 5%, Mdn = 5%), and 19 and 21% of cases per variable across time in cohort 1 (M = 20%, Mdn = 20%). The extent of missing data for the subjective norm questionnaire ranged between 10 and 37% of cases per variable in cohort 1 (M = 18%, Mdn = 13%), 5 and 29% of cases per variable in cohort 2 (M = 13%, Mdn = 10%), and 20 and 39% of cases per variable across time in cohort 1 (M = 26%, Mdn = 22%). The largest percentage of missing data on the subjective norm questionnaire was attributable to two items pertaining to sister's (29–39%) or brother's (28–36%) beliefs. The missing data for the two items were reasonable because some subjects did not have female or male siblings.

model complexity [*27–30*]. Both RNI and NNFI values are nonnormed and can exceed 1. Minimally acceptable fit was based on threshold RNI and NNFI values of 0.90 [*24,29,30*]; values approximating 0.95 were indicative of good fit [*27*]. The estimates of item loadings, uniquenesses, standard errors, *t* values, and squared multiple correlations (SMCs) also were inspected for appropriate sign and/or magnitude [*31,32*].

Model modifications. Modifications to the measurement models were performed based on empirical and substantive information to improve the factorial validity of the questionnaires by identifying a subset of items that best tapped the latent variables. Model modifications were conducted by computing a variance–covariance matrix using FIML in AMOS 4.0 and then inputting the matrix into LISREL 8.20 (Scientific Software International, Inc., Chicago, IL) [*33*]. We employed LISREL to compute standardized residuals and modification indices (MIs) in the uniqueness matrix because FIML in AMOS 4.0 does not provide the information needed to perform model modifications. The models were modified based on large standardized residuals (i.e., greater than ±2.58) [*15*] and MIs in the uniqueness matrix in combination with similarity of item content. Large standardized residuals and MIs in the theta–delta matrix identified pairs of items that were not accurately predicted by the model [*24,33*]. One of the two items was removed based on redundant content or the correlation between uniquenesses was estimated when the contents were similar, but not entirely redundant (e.g., items referring to either a girl's mother or her father). The CFA then was rerun in AMOS 4.0 to determine whether the modification resulted in an improved fit. The process of identifying model modifications via LISREL and testing the modification in AMOS was continued until a reasonable model was generated as indicated by the fit indices, but modifications were made only when substantively appropriate [*33*].

The Aikake Information Criterion (AIC) [*34*] and the Expected Cross-Validation Index (ECVI) [*26*] were employed to test modifications, because χ^2 difference tests cannot be legitimately performed on nonnested models (i.e., models that were modified based on deletion of items). The AIC value was computed based on the χ^2 value for the model plus two times the number of estimated parameters. The ECVI is a single-sample estimate that indicates how well the current solution would fit in an independently drawn sample [*26*]. The AIC and ECVI are not normed on a zero-to-one scale; reductions in AIC and ECVI values in comparison with other competing models indicated an improved and more parsimonious fit of a model [*26,34*]. χ^2 difference tests and the AIC and ECVI were employed to test modifications among nested models (i.e., models that were modified by freeing parameters to be estimated).

Cross-validation. The final measurement models applied to the social-cognitive questionnaires were cross-validated because post hoc modifications performed in a single sample might capitalize on chance, particularly in samples smaller than 800 [*16,17,33*]. The final model, therefore, needs to be tested in an independent, cross-validation sample [*24,25,35*]. Typically, cross-validation has involved only an independent test of the measurement model after post hoc modifications were performed using a calibration sample from a single cohort. We employed a more stringent method of cross-validating the measurement models by performing a multigroup analysis of factorial invariance using the baseline data from cohorts 1 and 2. The factorial invariance provides information about the model's external validity, particularly the robustness of parameter estimates across independent samples.

The invariance of the measurement models across groups were tested using a multistep procedure [*15,24,33,36*]. The invariance routine involved initial CFAs to test the model in cohorts 1 and 2 separately. The next analysis assessed whether the variance–covariance matrices (Equal Sigmas) underlying the item responses were invariant across cohorts. The test of Equal Sigmas may produce inconsistent results as an initial test of invariance [*15*], and it may not necessarily be an indication that the measurement parameters were invariant across the cohorts. Accordingly, it was necessary to further test the equality of parameters across the cohorts [*15*].

The final portion of the invariance routine involved four nested CFAs in which successive analyses contained the previous restriction(s) plus one additional restriction. The first CFA tested the equality of the factor structure across cohorts (i.e., same dimensions or location of fixed, freed, and constrained parameters; Model 1). The subsequent two CFAs tested the invariance of the factor loadings (i.e., equality of coefficients linking the observed and latent variables; Model 2) and factor variances (i.e., Model 3) across cohorts. The final, most restrictive CFA tested the invariance of item uniquenesses (i.e., equality of measurement and specific error variance associated with each item; Model 4). Model 2 is considered to be the minimal evidence of factorial invariance, with the other models demonstrating increased evidence of invariance [*24,33,36*]. Invariance was evaluated by a χ^2 difference test, RMSEA with a 90% CI, RNI, and NNFI. The ad hoc fit indices were employed based on problems of biased χ^2 values with large samples [*36*], particularly the increased power for detecting the effect of small and potentially meaningless differences in model parameters constrained to be invariant across cohorts.

Longitudinal factorial invariance. We also tested the longitudinal invariance of the measurement models using the baseline and follow-up data from cohort 1. The longitudinal invariance routine involved a single-group, two-factor correlated measurement model with autocorrelations specified between identical indicators on opposing factors. The two factors represented the baseline and follow-up versions of the unidimensional measurement models. The longitudinal invariance routine involved four nested CFAs. The first CFA served as the baseline model and tested the equality of the factor structure (i.e., same dimensions or location of fixed, freed, and constrained parameters; Model 1) without equality constraints across the two factors. The subsequent two CFAs tested the invariance of the factor loadings (i.e., equality of coefficients linking the observed and latent variables; Model 2) and factor variances (Model 3) across time. The final CFA tested the invariance of item uniquenesses (i.e., equality of measurement and specific error variance associated with each item; Model 4). Model 2 is considered to be the minimal evidence of factorial invariance, with the other models demonstrating increased evidence of invariance. Invariance was evaluated by a χ^2 difference test, RMSEA with a 90% CI, RNI, and NNFI. The interfactor correlation from the model demonstrating acceptable evidence of invariance was employed as a measure of temporal stability of the factors across a 1-year period, which is analogous to a test–retest reliability coefficient.

RESULTS

Descriptive Statistics

Table 1 contains the mean of the univariate skewness and kurtosis estimates for the items forming the four questionnaires. The univariate skewness and kurtosis values, which were obtained using list-wise deletion of missing data in PRELIS 2.20 (Scientific Software International, Inc.) [*37*], did not identify any serious violations of normality with the subjective norm, perceived behavioral control, and self-efficacy questionnaires. There were some items on the attitude questionnaire that were very leptokurtic, indicated by the large mean for univariate kurtosis across samples. Multivariate normality was evaluated using Mardia's coefficient for skewness and kurtosis [*38*], computed using list-wise deletion of cases in PRELIS 2.20. The normalized estimates of skewness and kurtosis are provided in Table 1; the values are interpreted similar to Z scores. The normalized estimates indicated that the data, particularly the items on the attitude questionnaire, violated the assumption of multivariate normality. However, Mardia's coefficient is largely influenced by sample size [*24*,*38*]. The values are reported for descriptive purposes as recommended by others [*15*,*32*].

TABLE 1

Statistics Evaluating the Univariate and Multivariate Skewness and Kurtosis for Items on the Four Social-Cognitive Questionnaires

	Univariate		Multivariate	
Measure	Skewness	Kurtosis	Skewness	Kurtosis
Attitude (22 items)				
Cohort 1—baseline	0.50	2.72	66.52	39.45
Cohort 1—follow-up	0.48	2.64	72.73	45.39
Cohort 2	0.58	2.67	104.31	65.88
Subjective norm (8 items)				
Cohort 1—baseline	0.36	−0.56	15.30	8.75
Cohort 1—follow-up	0.39	−0.55	16.10	14.28
Cohort 2	0.46	−0.56	22.64	18.37
Perceived behavioral control (4 items)				
Cohort 1—baseline	−1.33	1.64	28.43	16.96
Cohort 1—follow-up	−1.11	0.77	23.45	15.46
Cohort 2	−1.11	0.83	32.54	20.00
Self-efficacy (15 items)				
Cohort 1—baseline	−0.99	0.59	43.87	33.23
Cohort 1—follow-up	−0.63	0.38	41.16	31.58
Cohort 2	−0.83	0.21	56.87	46.26

Confirmatory Factor Analyses

Attitude. The one-factor model did not represent an acceptable fit to the 22-item measure of attitudes [χ^2 = 1405.68, *df* = 209, RMSEA = 0.075 (90% CI 0.071–0.079), RNI = 0.50, NNFI = 0.45]. We modified the one-factor model because three items possessed nonsignificant factor loadings and other items possessed multiple large residuals and MIs in the matrix of uniquenesses. Through an iterative process of removing a single item with either a nonsignificant item loading or multiple large residuals and MIs in the theta–delta matrix and reestimating the model, the final solution to the attitude questionnaire contained 8 items forming a single factor. The final model possessed an acceptable fit [χ^2 = 47.36, *df* = 20, RMSEA = 0.037 (90% CI 0.027–0.050), RNI = 0.97, NNFI = 0.95]. The AIC and ECVI were decreased from 1537.68 and 1.51 to 95.36 and 0.09 for the 22-item and 8-item, one-factor models, respectively. The reduction in AIC and ECVI demonstrated that the 8-item model represented an improved and parsimonious fit. The estimates of item loadings, uniquenesses, standard errors, *t* values, and SMCs were of the appropriate sign and/or magnitude.

The one-factor model was then cross-validated. As indicated in Table 2, the model fit acceptably in the separate analyses on the baseline data from cohorts 1 and 2. The test of Equal Sigmas was not rejected and indicated that the variance–covariance matrix underlying the attitude items was invariant across cohorts. Models 1 and 2 were not significantly different based on the χ^2 difference test and the ad hoc fit indices, which provided evidence that the factor structure and factor loadings were invariant across cohorts. Models 2 and 3 and Models 3 and 4 were different based on

TABLE 2

Results of the CFAs Testing the Factorial Invariance of the 8-Item Unidimensional Model Applied to the Attitude Questionnaire across Cohorts

Model	*df*	χ^2	RMSEA (90% CI)	RNI	NNFI
Cohort 1	20	47.36	0.037 (0.023–0.050)	0.97	0.95
Cohort 2	20	118.50	0.052 (0.043–0.062)	0.96	0.94
Equal Sigmas	36	67.13	0.018 (0.011–0.024)	0.99	0.98
Model 1	40	165.86	0.033 (0.028–0.039)	0.96	0.94
Model 2	47	176.52	0.031 (0.026–0.036)	0.96	0.95
Model 3	48	186.04	0.032 (0.027–0.037)	0.96	0.95
Model 4	56	201.90	0.030 (0.026–0.035)	0.95	0.95

Model comparisons	*df*	χ^2_{diff}	*P* value
Models 1 and 2	7	10.66	ns
Models 2 and 3	1	9.52	$P < 0.05$
Models 3 and 4	8	15.86	$P < 0.05$

the χ^2 difference tests, but the RMSEA, RNI, and NNFI values were similar and acceptable across models—the factor variances and the uniquenesses were invariant across cohorts. The factor loadings ($M = 0.51$, range = 0.31–0.66) and SMCs ($M = 0.24$, range = 0.06–0.44) are from Model 4 because it was acceptable, and it is considered the most restrictive test of factorial invariance [*15,24,33,36*].

The longitudinal factorial invariance of the one-factor model then was tested using baseline and follow-up data from cohort 1 (see Table 3). Models 1 and 2 were not significantly different based on all fit indices and indicated that the factor structure and factor loadings were invariant across time. Models 2 and 3 and Models 3 and 4 were significantly different based on the χ^2 difference tests, but the RMSEA, RNI, and NNFI values were similar and acceptable across models. Therefore, the factor variances and the uniquenesses were invariant across time. The factor loadings ($M = 0.50$, range = 0.26–0.70) and SMCs ($M = 0.27$, range = 0.07–0.49) are from Model 4 because it represented an acceptable fit. The interfactor correlation from Model 4 was 0.55, which demonstrated acceptable temporal stability of the factors across a 1-year period.

TABLE 3

Results of the CFAs Testing the Factorial Invariance of the 8-Item Unidimensional Model Applied to the Attitude Questionnaire across Time

Model	*df*	χ^2	RMSEA(90% CI)	RNI	NNFI
Model 1	95	208.45	0.034 (0.028–0.041)	0.96	0.95
Model 2	102	213.39	0.033 (0.027–0.039)	0.96	0.95
Model 3	103	217.76	0.033 (0.027–0.039)	0.95	0.95
Model 4	111	244.54	0.034 (0.029–0.040)	0.95	0.94

Model comparisons	*df*	χ^2_{diff}	*P* value
Models 1 and 2	7	4.94	ns
Models 2 and 3	1	4.37	$P < 0.05$
Models 3 and 4	8	26.78	$P < 0.05$

Subjective norm. The one-factor model resulted in a poor fit to the 8-item subjective norm scale [$\chi^2 = 456.69$, $df = 20$, RMSEA = 0.146 (90% CI 0.135–0.158), RNI = 0.81, NNFI = 0.73]. The MIs in the theta–delta matrix and the standardized residuals identified four pairs of items that resulted in the poor fit of the model. Inspection of the item content indicated that the pairs of items tapped similar, but not identical content. For example, items 5 and 6 both pertained to parental beliefs and values, but were specifically oriented toward the normative beliefs of either a girl's mother or her father. We opted to estimate the correlations between uniquenesses, to account for the systematic residual variance between items [*31*], rather than remove an item and compromise content representativeness. The final model contained correlated uniquenesses between items 1 and 2 (peers/friends), items 3 and 4 (teachers), items 5 and 6 (parents), and items 7 and 8 (siblings), and it represented an acceptable fit [$\chi^2 = 83.91$, $df = 16$, RMSEA = 0.064 (90% CI 0.051–0.078), RNI = 0.97, NNFI = 0.95]. The χ^2 difference test indicated that the one-factor model with correlated uniquenesses represented a significantly better fit than the one-factor model without correlated uniquenesses ($\chi^2_{\text{diff}} = 372.78$, $df = 4$, $P < 0.05$). The AIC and ECVI were reduced from 504.69 and 0.49 for the one-factor model without correlated uniquenesses to 139.91 and 0.14 for the model with correlated uniquenesses. The 8-item model with correlated uniquenesses demonstrated improved and parsimonious fit. The estimates of item loadings, uniquenesses, standard errors, *t* values, and SMCs were of the appropriate sign and/or magnitude.

The final model then was cross-validated using a multigroup analysis of factorial invariance. Results are presented in Table 4. The one-factor model with correlated uniquenesses fit acceptably in the separate analyses on the baseline data from cohorts 1 and 2. The test of Equal Sigmas was not rejected and indicated that the variance–covariance matrix underlying the items was invariant across cohorts. Models 1 and 2, Models 2 and 3, and Models 3 and 4 were not significantly different based on the χ^2 difference test and the ad hoc fit indices, which demonstrated that the factor structure, factor loadings, factor variances, and uniquenesses were invariant across cohorts. The factor loadings ($M = 0.63$, range = 0.54–0.67) and SMCs ($M = 0.40$, range = 0.29–0.44) are from Model 4 because it demonstrated a reasonable fit.

The longitudinal factorial invariance of the one-factor model with correlated uniquenesses then was tested using the baseline and follow-up data from cohort 1. See results in Table 5. Models 1 and 2, Models 2 and

TABLE 4

Results of the CFAs Testing the Factorial Invariance of the 8-Item Unidimensional Model with Correlated Uniquenesses Applied to the Subjective Norm Questionnaire across Cohorts

Model	*df*	χ^2	RMSEA (90% CI)	RNI	NNFI
Cohort 1	16	83.91	0.064 (0.051–0.078)	0.97	0.95
Cohort 2	16	108.11	0.057 (0.047–0.067)	0.98	0.97
Equal Sigmas	36	67.85	0.018 (0.011–0.024)	1.00	0.99
Model 1	32	192.02	0.042 (0.036–0.048)	0.98	0.96
Model 2	39	200.61	0.038 (0.033–0.044)	0.98	0.97
Model 3	40	201.40	0.038 (0.033–0.043)	0.98	0.97
Model 4	48	207.06	0.034 (0.030–0.039)	0.98	0.97

Model comparisons	*df*	χ^2_{diff}	*P* value
Models 1 and 2	7	8.59	ns
Models 2 and 3	1	0.79	ns
Models 3 and 4	8	5.66	ns

TABLE 6

Results of the CFAs Testing the Factorial Invariance of the 4-Item Unidimensional Model Applied to the Perceived Behavioral Control Questionnaire across Cohorts

Model	*df*	χ^2	RMSEA (90% CI)	RNI	NNFI
Cohort 1	2	8.05	0.054 (0.019–0.096)	0.98	0.95
Cohort 2	2	5.69	0.032 (0.000–0.064)	1.00	0.99
Equal Sigmas	10	33.22	0.029 (0.018–0.040)	0.98	0.98
Model 1	4	13.74	0.029 (0.013–0.047)	0.99	0.98
Model 2	7	20.89	0.027 (0.014–0.040)	0.99	0.98
Model 3	8	25.29	0.028 (0.016–0.040)	0.98	0.98
Model 4	12	45.73	0.032 (0.022–0.042)	0.97	0.97

Model comparisons	*df*	χ^2_{diff}	*P* value
Models 1 and 2	3	7.15	ns
Models 2 and 3	1	4.40	$P < 0.05$
Models 3 and 4	4	20.44	$P < 0.05$

3, and Models 3 and 4 were not significantly different based on all the fit indices—the factor structure, factor loadings, factor variances, and uniquenesses were invariant across time. The factor loadings (M = 0.63, range = 0.57–0.68) and SMCs (M = 0.40, range = 0.33–0.47) are from Model 4 because it represented a good fit. The interfactor correlation from Model 4 was 0.51, and it demonstrated acceptable temporal stability of the factors across a 1-year period.

Perceived behavioral control. One factor adequately fit the 4-item measure of perceived behavioral control [χ^2 = 8.05, *df* = 2, RMSEA = 0.054 (90% CI 0.019–0.096), RNI = 0.98, NNFI = 0.95]. The estimates of item loadings, uniquenesses, standard errors, *t* values, and SMCs were of the appropriate sign and/or magnitude.

The measurement model was cross-validated using a multigroup analysis of factorial invariance. As indicated in Table 6, the one-factor model fit acceptably in the CFAs on the baseline data from cohorts 1 and 2. The test of Equal Sigmas was not rejected—the variance–covariance matrix underlying the items was invariant across cohorts. Models 1 and 2 were not significantly different based on all fit indices, which provided evidence that the factor structure and factor loadings were invariant across cohorts. Models 2 and 3 and Models 3 and 4 were different based on the χ^2 difference tests, but the RMSEA, RNI, and NNFI values were similar and acceptable across models—the factor variances and the uniquenesses were invariant across cohorts. The factor loadings (M = 0.57, range = 0.51–0.64) and SMCs (M = 0.32, range = 0.26–0.41) are from Model 4 because it possessed a reasonable fit.

The longitudinal factorial invariance of the one-factor model then was tested using the baseline and follow-up data from cohort 1. See results in Table 7. Models 1 and 2 were not significantly different based on all fit indices, which provided evidence that the factor structure and factor loadings were invariant across time. Models 2 and 3 and Models 3 and 4 were different based on the χ^2 difference tests and RMSEA, RNI, and NNFI

TABLE 5

Results of the CFAs Testing the Factorial Invariance of the 8-Item Unidimensional Model with Correlated Uniquenesses Applied to the Subjective Norm Questionnaire across Time

Model	*df*	χ^2	RMSEA(90% CI)	RNI	NNFI
Model 1	87	257.27	0.044 (0.038–0.050)	0.97	0.95
Model 2	94	265.94	0.042 (0.036–0.048)	0.96	0.96
Model 3	95	265.98	0.042 (0.036–0.048)	0.97	0.96
Model 4	103	275.92	0.041 (0.035–0.046)	0.96	0.96

Model comparisons	*df*	χ^2_{diff}	*P* value
Models 1 and 2	7	8.67	ns
Models 2 and 3	1	0.04	ns
Models 3 and 4	8	9.95	ns

TABLE 7

Results of the CFAs Testing the Factorial Invariance of the 4-Item Unidimensional Model Applied to the Perceived Behavioral Control Questionnaire across Time

Model	*df*	χ^2	RMSEA(90% CI)	RNI	NNFI
Model 1	15	49.57	0.048 (0.033–0.063)	0.97	0.95
Model 2	18	55.11	0.045 (0.032–0.059)	0.97	0.96
Model 3	19	82.96	0.057 (0.045–0.070)	0.95	0.93
Model 4	23	136.12	0.069 (0.058–0.081)	0.92	0.90

Model comparisons	*df*	χ^2_{diff}	*P* value
Models 1 and 2	3	5.44	ns
Models 2 and 3	1	27.85	$P < 0.05$
Models 3 and 4	4	53.16	$P < 0.05$

TABLE 8

Results of the CFAs Testing the Factorial Invariance of the 8-Item Unidimensional Model Applied to the Self-Efficacy Questionnaire across Cohorts

Model	*df*	χ^2	RMSEA (90% CI)	RNI	NNFI
Cohort 1	20	39.93	0.031 (0.017–0.045)	0.98	0.98
Cohort 2	20	123.38	0.054 (0.045–0.063)	0.96	0.95
Equal Sigmas	36	51.50	0.012 (0.002–0.020)	1.00	0.99
Model 1	40	163.31	0.033 (0.028–0.038)	0.97	0.96
Model 2	47	171.10	0.031 (0.026–0.036)	0.97	0.96
Model 3	48	173.34	0.030 (0.026–0.035)	0.97	0.96
Model 4	56	195.75	0.030 (0.025–0.034)	0.97	0.97

Model comparisons	*df*	χ^2_{diff}	*P* value
Models 1 and 2	7	7.79	ns
Models 2 and 3	1	2.24	ns
Models 3 and 4	8	22.41	$P < 0.05$

values. Although the RNI and NNFI were acceptable across models, the factor variances and uniquenesses may not be invariant across time. The factor loadings (baseline $M = 0.55$, range = 0.49–0.54; follow-up $M = 0.66$, range = 0.58–0.74) and SMCs (baseline $M = 0.36$, range = 0.29–0.50; follow-up $M = 0.38$, range = 0.27–0.51) are from Model 2 because it was acceptable. The interfactor correlation from Model 2 was 0.55, indicating acceptable temporal stability of the factors across a 1-year period.

Self-efficacy. The fit of the one-factor model to the 15-item measure of self-efficacy was not acceptable [$\chi^2 = 801.33$, $df = 87$, RMSEA = 0.088 (90% CI 0.082–0.094), RNI = 0.80, NNFI = 0.76], and it could be improved. Using an iterative approach, one item was removed based on the standardized residuals, MIs in the theta–delta matrix, and similar content. The model then was reestimated. The final model contained 8 items forming a single factor and it demonstrated acceptable fit [$\chi^2 = 39.93$, $df = 20$, RMSEA = 0.031 (90% CI 0.017–0.045), RNI = 0.98, NNFI = 0.98]. The AIC and ECVI were reduced from 891.33 and 0.87 to 87.93 and 0.09 for the 15-item and 8-item, one-factor models, respectively. The 8-item model demonstrated improved and parsimonious fit. The estimates of item loadings, uniquenesses, standard errors, *t* values, and SMCs were of the appropriate sign and/or magnitude.

We then cross-validated the final measurement model. According to the results in Table 8, the one-factor model fit acceptably in the separate analyses on the baseline data from both cohorts. The test of Equal Sigmas was not rejected and indicated that the variance–covariance matrix underlying the items was invariant across cohorts. Models 1 and 2 and Models 2 and 3 were not significantly different based on all fit indices, which provided evidence that the factor structure, factor loadings, and factor variances were invariant across cohorts. Models 3 and 4 were different based on the χ^2 difference test, but the RMSEA, RNI, and NNFI values were similar and acceptable across models—the uniquenesses appeared to be invariant across cohorts. The factor loadings ($M = 0.56$, range = 0.38–0.63) and SMCs ($M = 0.32$, range = 0.14–0.39) are from Model 4 because it was acceptable.

The longitudinal factorial invariance of the one-factor model then was tested using the baseline and follow-up data from cohort 1 (see Table 9). Models 1 and 2 and Models 2 and 3 were not significantly different based on all fit indices—the factor structure, factor loadings, and factor variances were invariant across time. Models 3 and 4 were different based on the χ^2 difference test, but the RMSEA, RNI, and NNFI values were similar and acceptable across models. Therefore, there was evidence that the uniquenesses were invariant across time. The factor loadings ($M = 0.57$, range = 0.39–0.61) and SMCs ($M = 0.32$, range = 0.15–0.40) are from Model 4 because it represented a reasonable and acceptable fit. The interfactor correlation from Model 4 was 0.61, and it indicated acceptable temporal stability of the factors across a 1-year period.

DISCUSSION

Overview

Using two samples of eighth-grade girls and one sample of ninth-grade girls, the present study established the factorial validity and factorial invariance of unidimensional measurement models applied to four social-cognitive questionnaires, which were developed to be consistent with the unidimensional conceptualization of constructs within the Theory of Reasoned Action, Theory of Planned Behavior, and Social Cognitive Theory. The factorial validity of the unidimensional measurement models initially was established using baseline data from cohort 1 and then was further

TABLE 9

Results of the CFAs Testing the Factorial Invariance of the 8-Item Unidimensional Model Applied to the Self-Efficacy Questionnaire across Time

Model	*df*	χ^2	RMSEA (90% CI)	RNI	NNFI
Model 1	95	225.61	0.037 (0.031–0.043)	0.96	0.95
Model 2	102	230.23	0.035 (0.029–0.041)	0.96	0.95
Model 3	103	230.59	0.035 (0.029–0.041)	0.96	0.95
Model 4	111	254.08	0.036 (0.030–0.041)	0.95	0.95

Model comparisons	*df*	χ^2_{diff}	*P* value
Models 1 and 2	7	4.62	ns
Models 2 and 3	1	0.36	ns
Models 3 and 4	8	23.49	$P < 0.05$

demonstrated using a multigroup analysis of factor invariance on baseline data from cohorts 1 and 2. The longitudinal factorial invariance of the measurement models was established using the baseline and follow-up data from cohort 1. Final versions of the questionnaires are in the Appendix.

Attitude

The original attitude questionnaire consisted of 22 items. It was reduced to 8 items forming a single factor using the baseline data from cohort 1. The final measurement model was cross-validated using a multigroup analysis of factorial invariance on the baseline data from cohorts 1 and 2, which provided evidence that the variance–covariance matrices, factor structure, factor loadings, factor variances, and uniquenesses were invariant across cohorts. The 8-item unidimensional measure of attitude also demonstrated evidence of longitudinal factorial invariance; the factor structure, factor loadings, factor variances, and uniquenesses were invariant across time in cohort 1. Therefore, the unidimensional model applied to the 8-item measure of attitude possessed factorial validity and factorial invariance across groups and time in the samples of adolescent girls. Inspection of the items in the Appendix supports the content validity of the attitude questionnaire.

Subjective Norm

The 8-item measure of subjective norms consisted of a single factor with correlated uniquenesses among items tapping similar, but not redundant content [e.g., parental (mother and father) or sibling (brother and sister) norms and values]. The unidimensional model with correlated uniquenesses was supported in the first cohort and then was cross-validated using a multigroup analysis of factorial invariance across two cohorts. The invariance analysis provided evidence that the variance–covariance matrices, factor structure, factor loadings, factor variances, and uniquenesses were invariant across cohorts. The one-factor model with correlated uniquenesses also demonstrated evidence of longitudinal factorial invariance as the factor structure, factor loadings, factor variances, and uniquenesses were invariant across time in cohort 1. The unidimensional model applied to the 8-item measure of subjective norm possessed factorial validity and factorial invariance across groups and time in the samples of adolescent girls. The subjective norm items presented in the Appendix seem to possess content validity.

Perceived Behavioral Control

Initially, the one-factor model applied to the 4-item measure of perceived behavioral control was supported using the baseline data from cohort 1. The 4-item unidimensional measure of perceived behavioral control demonstrated evidence of multigroup factorial invariance across cohorts 1 and 2 (i.e., invariant variance–covariance matrices, factor structure, factor loadings, factor variances, and uniquenesses), which provided evidence of cross-validation. The factor structure and factor loadings also were invariant across time in cohort 1. The results, therefore, support the factorial validity and invariance of the unidimensional model applied to the 4-item measure of perceived behavioral control in adolescent girls. The perceived behavioral control items are presented in the Appendix and appear to possess content validity.

Self-Efficacy

Using the baseline data from cohort 1, the final model applied to the self-efficacy questionnaire contained 8 items, which conformed to a single factor. The model then was cross-validated using a multigroup analysis of factorial invariance on the baseline data from cohorts 1 and 2, which provided evidence that the variance–covariance matrices, factor structure, factor loadings, factor variances, and uniquenesses were invariant across cohorts. It also possessed evidence of invariant factor structure, factor loadings, factor variances, and uniquenesses across time in cohort 1. Accordingly, the unidimensional model applied to the 8-item measure of self-efficacy possessed factorial validity and invariance in the samples of adolescent girls. The items on the self-efficacy measure are presented in the Appendix and seem to possess content validity.

Implications

The unidimensional measurement models applied to the four questionnaires, which were consistent with the unidimensional conceptualization of constructs within the Reasoned Action, Planned Behavior, and Social-Cognitive theories, demonstrated acceptable evidence of factorial validity and multigroup and longitudinal factorial invariance. Hence, a comparison of mean scores across groups and time can be interpreted according to a change in the underlying construct (e.g., attitude toward physical activity or physical activity self-efficacy). The invariance analysis further indicated that the constructs of attitude, subjective norms, perceived behavioral control, and self-efficacy about physical activity are measured similarly across groups and time among adolescent girls by the questionnaires generated in the present study.

The questionnaires can be employed in SEM analyses to test the explanatory power of TRA, TPB, and SCT in understanding participation in physical activity by adolescent girls using both cross-sectional and longitudinal designs. The use of questionnaires having established measurement models facilitates model estimation using SEM, evaluation of model fit, and correct and meaningful estimation and interpretation of paths coefficients within SEM. Estimating models using SEM is an advance over traditional regression approaches because it enables simultaneous estimation of paths among multiple exogenous and endogenous variables, and it performs the structural analysis on latent variables rather than observed variables.

The present study extended previous research and, we hope, provides a heuristic for future researchers interested in establishing and testing the measurement properties of newly designed instruments using CFA. The present study: (1) developed theoretically based questionnaires of social-cognitive determinants of physical activity in black and white adolescent girls, (2) integrated empirical and substantive information to test and modify the measurement models, (3) employed CFA rather than EFA to directly test the fit of measurement models to the questionnaires, (4) employed a multigroup analysis of factorial invariance to cross-validate the measurement models, and (5) utilized procedures for testing the longitudinal factorial invariance or stability of the measurement models across time.

CONCLUSIONS AND FUTURE RESEARCH

Based on contemporary theory, the factorial validity and invariance of unidimensional measurement models applied to questionnaires measuring attitude, subjective norm, perceived control, and self-efficacy about physical activity were demonstrated in two cohorts of adolescent black and white girls. There are several possible directions for future research. For example, the factorial validity and invariance of the questionnaires should be tested between races and in other populations. The present study established only the structural and content validity of the four questionnaires, and research is necessary to further examine the validity of inferences from scores on the questionnaires. The questionnaires were developed based on the unidimensional conceptualization of constructs within TRA, TPB, and SCT. Future researchers that use these theories may be interested in developing questionnaires based on multidimensional conceptualizations of their constructs, which might help to better predict and explain physical activity participation.

APPENDIX

Attitude Questionnaire

If I were to be physically active during my free time on most days . . .

1. It would help me cope with stress.
2. It would be fun.
3. It would help me make new friends.
4. It would get or keep me in shape.
5. It would make me more attractive.
6. It would give me more energy.
7. It would make me hot and sweaty.
8. It would make me better in sports, dance, or other activities.

Subjective Norm Questionnaire

1. My fellow students think I should be physically active during my free time on most days.
2. My best friend thinks I should be physically active during my free time on most days.
3. My physical education teacher thinks I should be physically active during my free time on most days.
4. My other teachers think I should be physically active during my free time on most days.
5. My mother or female guardian thinks I should be physically active during my free time on most days.
6. My father or male guardian thinks I should be physically active during my free time on most days.
7. My sister/sisters think I should be physically active during my free time on most days.
8. My brother/brothers think I should be physically active during my free time on most days.

Perceived Behavioral Control Questionnaire

1. For me to be physically active during my free time on most days would be . . .
2. I have control over my being physically active during my free time on most days.
3. I believe I have all the things I need to be physically active during my free time on most days.
4. If I want to be I can be physically active during my free time on most days.

Self-Efficacy Questionnaire

1. I can be physically active during my free time on most days.
2. I can ask my parent or other adult to do physically active things with me.
3. I can be physically active during my free time on most days even if I could watch TV or play video games instead.
4. I can be physically active during my free time on most days even if it is very hot or cold outside.
5. I can ask my best friend to be physically active with me during my free time on most days.
6. I can be physically active during my free time on most days even if I have to stay at home.
7. I have the coordination I need to be physically active during my free time on most days.
8. I can be physically active during my free time on most days no matter how busy my day is.

ACKNOWLEDGMENT

We acknowledge the helpful comments and suggestions of an anonymous reviewer.

REFERENCES

1. McGinnis JM. The public health burden of a sedentary lifestyle. Med Sci Sports Exerc 1992;24(Suppl):S196–200.
2. Powell KE, Blair SN. The public health burdens of sedentary living habits: theoretical but realistic estimates. Med Sci Sports Exerc 1994;26:851–6.
3. Centers for Disease Control and Prevention. Youth risk behavior surveillance—United States, 1997. MMWR 1998;47(SS-3):1–89.
4. U.S. Department of Health and Human Services. Healthy People 2010: understanding and improving health. Washington: U.S. Govt. Printing Office, 2000:1–70. [No. 017-001-00543-6]
5. Malina RM. Tracking of physical activity and physical fitness across the lifespan. Res Q Exerc Sport 1996;57:48–57.
6. Stone EJ, McKenzie TL, Welk GJ, Booth ML. Effects of physical activity interventions in youth: review and synthesis. Am J Prev Med 1998;15:298–315.
7. Baranowski T, Anderson C, Carmack C. Mediating variable framework in physical activity interventions. How are we doing? How might we do better? Am J Prev Med 1998;15:266–97.
8. Dishman RK, Sallis JF. Determinants and interventions for physical activity and exercise. In: Bouchard C, Shephard RJ, Stephens T, editors. Physical activity, fitness, and health: international proceedings and consensus statement. Champaign (IL): Human Kinetics, 1994:214–38.
9. Sallis JF, Zakarian JM, Hovell MF, Hofstetter CR. Ethnic, socioeconomic, and sex differences in physical activity among adolescents. J Clin Epidemiol 1996;49:125–34.
10. Sallis JF, Prochaska JJ, Taylor WC. A review of correlates of physical activity of children and adolescents. Med Sci Sports Exerc 2000;32:963–75.
11. Ajzen I, Fishbein M. Attitude–behavior relations: a theoretical analysis and review of empirical research. Psychol Bull 1977; 84:888–918.
12. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckman J, editors. Action–control: from cognition to behavior. Heidelberg: Springer-Verlag, 1985:11–39.
13. Bandura A. Social foundations of thought and action. Englewood Cliffs (NJ): Prentice Hall, 1986.
14. Dishman RK. The measurement conundrum in exercise adherence research. Med Sci Sports Exerc 1994;26:1382–90.
15. Byrne BM. Structural equation modeling with LISREL, PRELIS, and SIMPLIS: basic concepts, applications, and programming. Mahwah (NJ): Erlbaum, 1998.
16. Saunders RP, Pate PR, Felton G, Dowda M, Weinrich MC, Ward DS, et al. Development of questionnaires to measure psychosocial influences on children's physical activity. Prev Med 1997;26: 241–7.
17. Ajzen I. The theory of planned behavior. Organ Behav Hum Decision Processes 1991;50:179–211.
18. Arbuckle JL, Wothke W. Amos 4.0 user's guide. Chicago: SmallWaters, 1999.
19. Wothke W. Longitudinal and multigroup modeling with missing data. In: Little TD, Schnabel KU, Baumert J, editors. Modeling longitudinal and multilevel data: practical issues, applied approaches, and specific examples. Mahwah (NJ): Erlbaum, 2000:219–40.
20. Arbuckle JL. Full information estimation in the presence of incomplete data. In: Marcoulides GA, Schumacker RE, editors. Advanced structural equation modeling: issues and techniques. Mahwah (NJ): Erlbaum, 1996:243–77.
21. Hutchinson SR, Olmos A. Behavior of descriptive fit indices in confirmatory factor analysis using ordered categorical data. Struct Equat Model 1998;5:344–64.
22. Hoyle RH, Panter AT. Writing about structural equation models. In: Hoyle RH, editor. Structural equation modeling: concepts, issues, and applications. Thousand Oaks (CA): Sage, 1995: 158–76.
23. Bentler PM, Chou C. Practical issues in structural modeling. Soc Methods Res 1987;16:78–117.
24. Bollen KA. Structural equations with latent variables. New York: Wiley, 1989.
25. Jöreskog KG. Testing structural equation models. In: KA Bollen, JS Long, editors. Testing structural equation models. Newbury Park (CA): Sage, 1993:294–316.
26. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: KA Bollen, JS Long, editors. Testing structural equation models. Newbury Park (CA): Sage, 1993:136–62.
27. Hu L, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. Struct Equat Model 1999;6:1–55.
28. McDonald RP, Marsh HW. Choosing a multivariate model: noncentrality and goodness-of-fit. Psychol Bull 1990;107:247–55.
29. Marsh HW, Balla JR, Hau K. An evaluation of incremental fit indices: a clarification of mathematical and empirical properties. In: Marcoulides GA, Schumacker RE, editors. Advanced structural equation modeling: issues and techniques. Mahwah (NJ): Erlbaum, 1996:315–51.
30. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychol Bull 1980;88: 588–606.
31. Kline RB. Principles and practice of structural equation modeling. New York: Guilford, 1998.
32. Ullman JB. Structural equation modeling. In: Tabachnick BG, Fidell LS, editors. Using multivariate statistics. 3rd ed. New York: Harper Collins, 1996:709–811.
33. Jöreskog KG, Sörbom D. LISREL 8: user's reference guide. Chicago: Scientific Software International, 1996.
34. Aikake H. Factor analysis and AIC. Psychometrika 1987;52: 317–32.
35. MacCallum RC, Roznowski M, Necowitz LB. Model modifications in covariance structure analysis: the problem of capitalization on chance. Psychol Bull 1992;111:490–504.
36. Marsh HW. Confirmatory factor analysis models of factorial invariance: a multifaceted approach. Struct Equat Model 1994;1:5–34.
37. Jöreskog KG, Sörbom D. PRELIS 2: user's reference guide. Chicago: Scientific Software International, 1996.
38. Mardia KV. Measures of multivariate skewness and kurtosis with application. Biometrika 1970;57:519–30.