



**HAL**  
open science

# Self-Belief Does Make A Difference: A Reciprocal Effects Model of the Causal Ordering of Physical Self-concept and Gymnastics Performance

Herbert W. Marsh, Julien Chanal, Philippe Sarrazin

► **To cite this version:**

Herbert W. Marsh, Julien Chanal, Philippe Sarrazin. Self-Belief Does Make A Difference: A Reciprocal Effects Model of the Causal Ordering of Physical Self-concept and Gymnastics Performance. *Journal of Sports Sciences*, 2006, 24 (1), pp.101-111. 10.1080/02640410500130920 . hal-00389009

**HAL Id: hal-00389009**

**<https://hal.science/hal-00389009>**

Submitted on 27 May 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Running Head: **Reciprocal Effects Model**

**Self-Belief Does Make A Difference: A Reciprocal Effects Model of the Causal Ordering of  
Physical Self-concept and Gymnastics Performance**

Herbert W. Marsh

SELF Research Centre, University of Western Sydney, Australia

Julien P. Chanal and Philippe G. Sarrazin

University J. Fourier of Grenoble, France

*Journal of Sports Sciences* (2006), 24(1), 101-111.

Author Note

Herbert Marsh is Director of the SELF Research Centre at the University of Western Sydney, Australia. Julien Chanal and Philippe Sarrazin are with the Laboratory “Sport et Environnement Social”, Université J. Fourier, Grenoble, France. We would like to thank Aïna Chalabaev and Damien Tessier who helped in the data collection of the study. Requests for further information about this investigation should be sent to Professor Herbert W. Marsh, Director, SELF Research Centre, University of Western Sydney, Bankstown Campus, Penrith South DC NSW 1797, Australia; E-Mail: [h.marsh@uws.edu.au](mailto:h.marsh@uws.edu.au) or to Dr Philippe Sarrazin, “Laboratoire Sport et Environnement Social” E.A. 540, UFRAPS - Université J. Fourier, Grenoble I. BP 53 - 38041 Grenoble Cedex 9 - France. E-mail: [philippe.sarrazin@ujf-grenoble.fr](mailto:philippe.sarrazin@ujf-grenoble.fr)

### Abstract

A large body of research in support of the reciprocal effects model of causal ordering demonstrates that prior academic self-concept predicts subsequent academic achievement beyond what can be explained in terms of prior achievement. Here we evaluate the generalizability of this support for the reciprocal effects model to a physical activity context in which achievement is reflected in gymnastics skills on a standardized gymnastics performance test evaluated by expert judges. Based on responses by 376 adolescents collected at the start (T1) and end (T2) of a gymnastics training program, findings support a reciprocal effects model in which there are significant paths leading from both T1 gymnastics self-concept to T2 gymnastics skills *and* from T1 gymnastics skills to T2 self-concept. Although there were gender and age effects (girls and older participants had better gymnastics skills, boys had higher self-concepts), multiple group structural equation models indicated that support for the reciprocal effects model generalized over responses by boys and girls. In summary, self-concept and performance are both determinants and consequences of each other.

As emphasized by Seligman and Csikszentmihalyi (2000), Vallerand et al. (2003), and many others, there is a revolution sweeping psychology, one that emphasizes a positive psychology and focuses on how normal individuals can get the most from life. Consistent with this emphasis, a positive self-concept is valued as a desirable outcome in many disciplines of psychology such as social, personality, education, health, sport/exercise, development and mental health. As is the case in many disciplines, in sport/exercise settings self-concept is frequently posited as a mediating variable that facilitates the attainment of other desired outcomes such as physical skills, health-related physical fitness, physical activity, and exercise adherence (Marsh, 1997, 2002). The rationale behind this research is that individuals who feel positively about themselves in a particular domain— in this study the physical domain—are more likely to pursue and achieve desirable outcomes in that domain than individuals who do not feel positively about themselves. We begin by briefly reviewing the extensive body of literature in support of this contention for academic self-concept and school performance measures, and then test its generalizability to the physical domain.

#### **Reciprocal Effects Model: Causal Ordering of Academic Self-concept and Achievement**

Empirical, theoretical, and methodological bases for the present investigation come in part from an extensive body of self-concept research, where there has been a long-standing debate about whether academic self-concept is a determinant or a consequence of academic achievement. This critical question has important theoretical and practical implications. Byrne (1984) for example, noted that much of the interest in the relations of academic self-concept to achievement stems from the belief that academic self-concept has motivational properties such that changes in academic self-concept will lead to changes in subsequent academic achievement. Calsyn and Kenny (1977) contrasted self-enhancement and skill development models. The *self-enhancement model* posits self-concept as a primary determinant of academic achievement (i.e. self-concept → achievement). In contrast, the *skill development model* implies that academic self-concept emerges principally as a consequence of academic achievement (i.e. achievement → self-concept). Due largely to limitations in statistical techniques used in the 1980s to test these models, researchers argued for “either-or” conclusions. In a review and critique of this research, Marsh (1990a, 1990b, 1993a; also see Marsh, Byrne & Yeung, 1999) argued that much of this research was methodologically unsound and inconsistent with academic self-concept theory. He

emphasized that it was widely accepted that prior academic achievement was one determinant of academic self-concept, so that the critical question was whether there also existed a link from prior academic self-concept to subsequent achievement. He further argued that a more realistic compromise between the self-enhancement and skill-development models was a *reciprocal effects model* in which prior self-concept affects subsequent achievement *and* prior achievement affects subsequent self-concept.

Because self-concept and academic achievement are not readily amenable to experimental manipulations, most research relies on longitudinal panel data in which both self-concept and achievement are measured on at least two occasions (i.e. a 2-wave, 2-variable design). With hindsight and 15 years' experience, Marsh et al. (1999) offered commentary on potential problems and how they can be avoided in tests of the reciprocal effects model. They demonstrated new, more defensible models of these data; emphasized more generally the role of researcher as substantive data detective; and updated Byrne's (1984) standards of an "ideal" study and directions for future research. Ideally, they argued, studies should: (a) measure academic self-concept and academic achievement (school performance, standardized test scores, or both) at least twice (i.e. a 2-wave study); (b) infer all latent constructs on the basis of multiple indicators; (c) consider a sufficiently large and diverse sample to justify the use of CFA and the generality of the findings, and; (d) fit the data to a variety of CFA models that incorporate measurement error and test for likely residual covariation among measured variables.

A growing body of research reviewed by Marsh, et al. (1999; Marsh & Craven, in press) has now established support for the reciprocal effects model of relations between academic self-concept and academic achievement, with major implications for the importance placed on self-concept as a means of facilitating other desirable outcomes, as well as being in itself an important outcome variable. In a recent meta-analysis of relevant research, Valentine (2002; Valentine, Dubois & Cooper, 2004) concluded that there was clear support for predictions based on the reciprocal effects model over those derived from self-enhancement and skill development models. Whereas the effect of prior self-concept on subsequent achievement after controlling for the effects of prior achievement was modest (mean standardized regression coefficient = .1), the effect was highly significant overall and positive in 90% of the studies in this meta-analysis. In support of the multidimensional

perspective that is so important to the reciprocal effects model, the effects of prior self-beliefs were significantly stronger when the measure of self-belief was based on a domain-specific measure rather than on global measures such as self-esteem, and when self-concept and achievement measures were matched in terms of subject area (e.g. mathematics achievement and math self-concept). Hence, this meta-analysis provides strong support for predictions based on the reciprocal effects model and for the construct validity of the multidimensional perspective of self-concept with its focus on domain specific measures. Recent extensions of this research demonstrate support for the reciprocal effects model generalizes to responses by young children (Guay, Marsh & Boivin, 2003) and to non-Western (Chinese) school settings responses (Marsh, Hau & Kong, 2002). In summary, when self-concept is based on a domain-specific measure that is logically related to the achievement measure, there is strong support for the effect of prior self-concept on subsequent achievement that is the critical prediction of the reciprocal effects model.

#### **The Present Investigation: Generalizability to a Physical Education Setting**

Reflecting historical trends in self-concept research more generally, self-concept instruments used in early sport/exercise research focused on global self-esteem (Marsh, 1997). More recently, however, there has been a stronger emphasis on a multidimensional perspective of self-concept with physical self-concept measures designed specifically for physical education, sport and exercise settings (Fox & Corbin, 1989; Marsh, 1997, 2002). They show that physical self-concept is distinct from other areas of self-concept (e.g. academic). This growing body of research also provides clear support for convergent and discriminant validity of physical self-concept responses in relation to a wide variety of sport/exercise outcome measures.

Whereas there now exists a large body of research in support of the reciprocal effects model for academic self-concept and achievement measures in traditional academic school subjects (e.g. mathematics and English) and strong support for a multidimensional perspective in sport/exercise psychology, there are apparently no tests of the reciprocal effects model to a sport/exercise setting with appropriate measures of physical self-concept and physical performance skills. This concern has important substantive and applied implications for sport/exercise research, but also has important theoretical implications for generalizability of the reciprocal effects model and for self-concept theory. For these reasons, the overarching purpose of the present investigation is to pursue tests of the

reciprocal effects model in relation to physical self-concept and physical performance skills. Specifically, we evaluate predictions from the reciprocal effects model of the causal ordering of gymnastics self-concept and gymnastics performance skills that are adapted from an extensive body of academic self-concept research. This model (Figure 1A) makes predictions about the effects of time 1 (T1) gymnastics self-concept and T1 gymnastics performance skills collected at the start of a gymnastics training program on time 2 (T2) gymnastics self-concept and T2 gymnastics performance skills collected at the end of the 10-week program. Not surprisingly, the reciprocal effects, self-enhancement, and skill development models all predict the effect of each T1 variable on the matching T2 variable to be substantial and positive (dashed horizontal lines in Figure 1A). The critical predictions distinguishing three theoretical models are the cross-paths relating T1 gymnastics self-concept to T2 gymnastics performance skills and T1 gymnastics performance skills to T2 gymnastics self-concept: the skill development model predicts that only path from T1 gymnastics performance skills to T2 gymnastics self-concept will be significant; the self-enhancement model posits that only the path from T1 gymnastics self-concept to T2 gymnastics performance skills will be significant; and the reciprocal effects model predicts that both paths will be significant (see Figure 1A).

We also extend the reciprocal effects model by adding gender, age, and their interaction to the structural equation model (Figure 1B). Whereas gender and age effects are of substantive interest, our main focus is on how their inclusion influences support for the reciprocal effects model and on evaluating the possibility of spurious, suppression, moderated, and mediated effects (e.g. Aiken & West, 1991; Baron & Kenny, 1986; Frazier, Tix, & Barron, 2004; Marsh, 1993b). *Spurious effects* are the effects of one variable that should be attributed to another variable that comes before it in the causal ordering. For example, part of the positive relation between physical self-concept and gymnastics performance may reflect gender and age differences in these variables. Typically, controlling for a pre-existing difference will result in weaker relations between subsequent variables, suggesting that part of the relation was spurious. It is possible, however, that controlling for a pre-existing difference will actually increase relations between subsequent variables—a *suppression* effect (see Marsh, 1993b; Smith, Ager & Williams, 1992). Previous research has demonstrated such an effect in the relation between physical fitness and physical self-concept when controlling for gender and age differences (Marsh, 1993b). Being able to establish a clear causal ordering of the

variables is a critical issue in demonstrating that an effect is spurious, or in the appropriate interpretation of a suppression effect. Thus, in our example, it is clear that gender and age are causally prior to T1 and T2 measures of physical self-concept and gymnastics performance, whereas T1 outcomes can be argued to be causally prior to T2 outcomes on the basis of their temporal ordering.

*Mediated effects* occur when at least some of the effect of a predictor variable on an outcome variable can be explained in terms of an intervening variable that comes between the two variables in the causal ordering. In the present investigation, for example, it is likely that the effects of gender and age on T2 variables are mediated substantially by corresponding T1 measures. Support for the mediation hypothesis, for example, requires that the effects of gender be reduced by the inclusion of the mediating variables. This can be total or partial mediation depending on whether there are gender effects at T2 beyond those mediated by T1 variables. The critical difference between spurious effects and mediated effects is the position of the variables in the causal order.

*Moderated effects* refer to interaction effects such that the effect of one variable depends on the level of another variable. To illustrate this issue in the present investigation, we evaluate whether the effects posited in the reciprocal effects model interact with (vary as a function of) gender and age (e.g. whether the positive effects of prior physical self-concept on subsequent exercise behavior are systematically larger or smaller for boys than for girls). In order to evaluate this possibility, we conducted multiple group structural equation and confirmatory factor analysis models in which we evaluated support for the reciprocal effects models in separate analyses of responses by boys and girls and by younger and older participants.

## **Method**

### ***Participants and Procedures***

Participants were 376 French adolescents (54% boys) from 17 seventh, eighth or ninth grade classes in six French High Schools situated in predominantly middle class areas in the south of France. The mean age of participants was 13.6 years ( $SD = 1.6$ ). In France, physical education is a compulsory subject for all high school students. Generally, physical education teachers teach each of several physical and sporting activities in 10-week cycles. The study was conducted during gymnastics cycles in scheduled physical education lessons. Data were collected twice, the first (T1) and last (T2) class period of the gymnastics cycle. Self-concept, gender, age, and gymnastics skills were assessed by one



of the second authors and two assistants using a questionnaire and a standardized test (see below) at T1 and again at the end of the program (T2). The anonymity of the responses was guaranteed and it was emphasized to participants that there were no right or wrong responses and that each participant should answer honestly. Participants' initial gymnastic skills were assessed by videotaping each participant on a gymnastics skills test that they completed three times. In order to ensure anonymity of responses, participants were only identified in terms of class, gender and birthdate. However, because T1 and T2 responses by some participants could not be matched on these variables (due to missing or inaccurate demographic information), the final sample was reduced from 406 to 376 participants.

### *Measures*

Participants' T1 and T2 gymnastics self-concepts were assessed with a 3-item questionnaire adapted from previous research by Nicholls and Duda (Duda & Nicholls, 1992; Nicholls, Cheung, Lauer, & Patashnick, 1989; e.g. "When you are in gymnastics lessons, you consider yourself..."). Participants responded on 7-point scales that varied from 1 (very bad) to 7 (very good). This questionnaire had shown good construct validity, internal consistency, and predictive validity (e.g. Sarrazin, Roberts, Cury, Biddle, & Famose, 2002) in previous research. In the present investigation, this scale had good internal consistency ( $\alpha = .85$  at T1,  $.87$  at T2). Reliability estimates were reasonably similar for younger participants ( $\alpha = .84$  at T1,  $.85$  at T2) and older participants ( $\alpha = .86$  at T1,  $.87$  at T2), and for girls ( $\alpha = .81$  at T1,  $.88$  at T2) and for boys ( $\alpha = .89$  at T1,  $.86$  at T2). Support for the construct validity of the self-concept measure is presented as part of the present investigation.

An evaluation of participants' gymnastics skills was based on a gymnastics performance test consisting of basic performance exercises. Three judges expert in gymnastics rated the videotaped performance of each participant on a 1 (low level) to 7 (high level) response scale. The test consisted of five gymnastic exercises (e.g. stretched handstand back drop, backward roll, cartwheel). Each judge made a single global rating of the participant after viewing the participant's performance on all five exercises. As participants completed the set of exercises three times, each judge appraised each participant three times and the highest rating was used in subsequent analyses. Different sets of judges were used at T1 and T2. Based on responses by the three different raters, inter-judge reliability was very good ( $\alpha = .96$  at T1,  $.97$  at T2), and similar for younger participants ( $\alpha = .95$  at T1,  $.96$  at T2)

and older participants ( $\alpha = .95$  at T1,  $.97$  at T2), and for girls ( $\alpha = .96$  at T1,  $.97$  at T2) and for boys ( $\alpha = .94$  at T1,  $.97$  at T2).

### *Statistical Analyses*

CFAs and SEMs were conducted with LISREL (Version 8.54) using maximum likelihood estimation (for further discussion see Bollen, 1989; Byrne, 1998; Joreskog & Sorbom, 1996). Following Marsh and colleagues (Marsh, Balla & Hau, 1996; Marsh, Balla & McDonald, 1988; Marsh, Hau & Wen, 2004) we consider the Tucker-Lewis index (TLI), the relative noncentrality index (RNI), and the root mean square error of approximation (RMSEA) to evaluate goodness of fit as well as presenting the normal theory  $\chi^2$  test statistic (the default in LISREL) and an evaluation of parameter estimates. The TLI and RNI vary along a 0-to-1 continuum in which values greater than  $.90$  and  $.95$  are typically taken to reflect acceptable and excellent fits to the data. RMSEA values of less than  $.05$  and  $.08$  are taken to reflect a close fit and a reasonable fit, respectively. The RNI contains no penalty for a lack of parsimony so that improved fit due to the introduction of additional parameters may reflect capitalization on chance, whereas the TLI and RMSEA contain penalties for a lack of parsimony.

As recommended by Marsh et al. (1999) in their guidelines for the evaluation of reciprocal effects models, all analyses presented in this study included correlated uniquenesses posited a priori to account for method effects associated with the same self-concept items administered on different occasions in longitudinal research. Marsh and Hau (1996; Joreskog, 1979) emphasize that if the same measurements are used on multiple occasions, then the corresponding residual error variables will tend to be correlated and, in order to get accurate estimates of relations among the constructs, correlations among errors must be included in the model. In preliminary analyses, the inclusion of these correlated uniquenesses was supported by modestly better fits to the data and, in particular, their exclusion would have positively biased the corresponding test-retest correlation estimates. Their inclusion, however, had no substantively important effect on the pattern of parameter estimates, suggesting that the inclusion of correlated uniquenesses in this study was not a critical issue. In order to facilitate the substantive import of the results, only the models with correlated uniquenesses were presented. It is important to emphasize that all correlated uniquenesses were for the same self-concept items administered at T1 and T2, were hypothesized a priori, and did not include correlated uniquenesses between different (non-matching) items. Because different judges made ratings of

gymnastics performances at T1 and T2, no correlated uniquenesses were posited for their ratings (as would have been the case if the same judges had been used at T1 and T2).

In confirmatory factor analysis studies with multiple groups it is possible to test the invariance of any one, any set, or all parameter estimates across the multiple groups. In the present investigation, two such sets of multiple group comparisons were conducted: one comparing parameter estimates based on separate analyses of responses by the 172 girls and 204 boys, and one based on responses by the 257 Year 7 participants and the 119 participants in Years 8 and 9. Tests of factorial invariance (see Bollen, 1989; Byrne, 1998; Joreskog & Sorbom, 1996; Marsh, 1994) traditionally posit a series of nested models in which the end-points are the least restrictive model with no invariance constraints and the most restrictive (total invariance) model with all parameters are constrained to be the same across all groups. In applying this approach, there is a well developed methodology in which the goodness of fit of alternative models are compared, including the least restrictive model that does not require any of the parameter estimates to be the same in different groups and the most restrictive model that requires all parameter estimates to be the same in the different groups. Typically, the minimal condition for *factorial invariance* is the equivalence of all factor loadings in the multiple groups and this is typically the first test of invariance (following the test with no invariance constraints) in the sequence of invariance models. There is no clear consensus in recommendations about the ordering of subsequent invariance constraints (e.g. Bentler, 1988; Bollen, 1989; Byrne, 1998; Joreskog & Sorbom, 1996), although Bentler (1988; Byrne, 2003) noted that the equality of parameters associated with measurement errors is typically the least important hypothesis to test and is unlikely to be met in most applications. The relative importance of invariance tests for different parameters also depends in part on the intent of the study. In the present investigation, for example, we are particularly interested in the invariance over gender and age of path coefficients that are the basis of predictions used to test the reciprocal effects model. Hence, support for the invariance of these parameters provides a strong test of the generalizability of the results over boys and girls, and over participants differing in age.

Whereas differences between nested models, under appropriate conditions, can be tested for statistical significance (Bentler, 1990), others (see Marsh, Hau & Grayson, in press) have argued that it is typically more appropriate to compare them in terms of indices of fit like those emphasized here.

Testing for factor invariance essentially involves comparing a number of models in which aspects of the factor structure are systematically held invariant across groups and assessing fit indices when elements of these structures are constrained. If the introduction of increasingly stringent invariance constraints results in little or no change in goodness of fit, then there is evidence in support of the invariance of the factor structure. If, on the other hand, the introduction of a set of invariance constraints results in a substantial decrement in goodness of fit, then there is evidence against the appropriateness of those invariance constraints. Thus, for example, Cheung and Rensvold (2002) suggested that for incremental fit indices (e.g. TLI and RNI) decreases in fit greater than .01 might be important. For present purposes, we focus on goodness of fit indices that control for parsimony (e.g. TLI and RMSEA). We also note that in tests of factorial invariance, tests of the relative fit of models testing more or fewer invariance constraints are of greater importance than the absolute level of fit for any one model. Whereas tests of statistical significance and indices of fit aid in the evaluation of the fit of a model, there is ultimately a degree of subjectivity and professional judgment required in the selection of a “best” model (Marsh, Balla & McDonald, 1988).

In the final set of analyses we evaluated the effects of gender, age, and their interaction. We initially constructed three single-degree-of-freedom contrast variables to represent the effects of age, gender (1 = female, 2 = male) and age x gender interaction. In order to minimize the potential confounding effects of multicollinearity, all variables were standardized ( $M = 0$ ,  $SD = 1$ ) before constructing the product terms. In the actual SEM, these three contrast variables were considered as three single-item factors measured without error and regressed on the gymnastics self-concept and performance factors. This is analogous to a typical multiple regression approach to ANOVA (Aiken & West, 1991), but the CFA/SEM approach is stronger (see Marsh, Ellis & Craven, 2002) in that it is based on latent constructs that are purged of measurement error (as well as providing a test of the underlying model rather than just assuming that it is correct in the formation of scale scores). Of particular relevance in the present investigation is the effect of the inclusion of these participant background variables on support for the reciprocal effects models and the interpretations of effects as spurious, suppressed, moderated, and mediated effects.

## **Results**

### ***Tests of the Reciprocal Effects Model***

Results from two models designed to test the reciprocal effects model for gymnastics self-concept and gymnastics performance are summarized in Tables 1 and 2. In Model 1A, a measurement model, the four latent factors (self-concept at T1 and T2; gymnastics performance at T1 and T2) are merely allowed to be correlated without positing a particular causal ordering. The goodness of fit index demonstrated that the ability of the model to fit the data is extremely good (e.g. TLI = .995). Each of the four factors is well defined in that all factor loadings are statistically significant and substantial (.65 to .98;  $Md = .92$ ). In support of the construct validity of the self-concept and gymnastic performance factors, the correlation between the two constructs is substantial ( $r = .51$  at both T1 and T2). Although the test-retest stabilities are only moderate (.77 for self-concept and .67 for gymnastics performance), this is reasonable given that participants were actively participating in a gymnastics training program. In summary, the results of Model 1A provide good support for the psychometric properties of the measures.

*Insert Tables 1 and 2 about here*

In Model 1B, a structural equation model is evaluated (see Figure 1A) in order to test the reciprocal effects model. Because Models 1A and 1B are equivalent models (i.e. the number of parameters estimates and goodness of fit statistics are necessarily the same), the critical parameter estimates are the four path coefficients leading from the two T1 factors to the two T2 factors. Not surprisingly, the largest effect on T2 self-concept is T1 physical self-concept (.70) and the largest effect on T2 gymnastics performance is T1 gymnastics performance (.59). These are represented by the horizontal paths (dashed lines) in Figure 1A. Of greater interest, however, are the cross paths (represented by solid black lines in Figure 1A). Consistent with the predictions from the reciprocal effects model, T1 self-concept has a significant effect on T2 gymnastics performance (.15) and the T1 gymnastics performance has a significant effect on T2 self-concept (.15). In summary, in terms of the overarching purpose of the present investigation, the results of this model provide clear support for generalizability of predictions based on the reciprocal effects model to this physical setting.

#### *Tests of Invariance Over Gender and Year in School*

Self-concept researchers have lamented that inadequate attention is given to tests of the invariance of the factor structure in different subgroups. Unless there is at least reasonable support for invariance over groups, comparisons of mean level responses between groups may be dubious. Also, differences

in the structure of self-concept for different groups may be substantively important in their own right. Particularly in the present investigation, tests of invariance over groups provide strong tests of the generalizability of support for the reciprocal effects model across responses by boys and girls, and across responses by younger and older participants.

In this study we pursued a parallel set of five models to evaluate invariance across age groups (Models 2A-2E in Table 1) and gender groups (Models 3A-3E). In each set of models we began with the least restrictive model (Models 2A and 3A) that imposed no invariance constraints (i.e. no parameter estimates were constrained to be the same in different groups) and progressed to the most restrictive model (Models 2E and 3E) that imposed complete factorial invariance (i.e. all parameter estimates are required to be the same in each of the groups). Intermediate models required only the factor loadings to be the same across groups (Models 2B, 3B and 4B) or factor loadings in combination with path coefficients, factor residual variances and covariances, and uniquenesses (error variances).

The results of this complex set of 10 multiple group SEM invariance models are easy to summarize. For each set of models representing year in school (models 2A-2E) and gender (models 3A-3E), there is good support for invariance of all parameter estimates. This support is particularly strong for tests of invariance over year in school, in that fit indexes that include a penalty for lack of parsimony (TLI and RMSEA) are as good or better for Model 2E, which requires all parameters to be invariant across groups, than Model 2A, which imposes no invariance constraints. Because the model of complete invariance is so much more parsimonious and fits the data as well as or better than less parsimonious models, the “best” model is Model 2E. For tests of the invariance of over gender, there is good support for the invariance of factor loadings and path coefficients (e.g. Models 3A, 3B and 3C all have TLIs = .996). Although there is a small decrement in fit when additional invariance constraints are added for factor residual variances and covariances (Model 3D) and uniquenesses (Model 3E), even the total invariance model provides an excellent fit to the data (TLI = .992 for Model 3E) and is only modestly poorer than Model 3A, which imposes no invariance constraints. The reason for this slight decrement in fit is that boys’ gymnastic performance scores were consistently more variable than were those of girls according to all three expert judges, particularly at T1, whereas T1 self-concept responses were slightly more reliable for girls than boys at T1. In the SEM models

this was reflected in slightly larger factor variances and measured variable uniqueness scores for boys than girls. However, of particular importance to the present investigation, there was support for the invariance of the path coefficients used to evaluate the reciprocal effects model (Model 3C). In summary, particularly for the path coefficients used to evaluate reciprocal effects model, there was good support for the generalizability of the results across responses by girls and boys, and across responses by younger and older participants.

*Tests of the Extended Reciprocal Effects Model: The Addition of Gender and Age*

In Models 4A and 4B (Table 3), we extend Models 1A and 1B by the inclusion of gender, age, and their interaction. In Model 4A, the measurement model, our focus is an evaluation of the pattern of relations between gender, age, gymnastics self-concept, and gymnastics performance. However, the main focus of this section is Model 4B, the SEM model used to evaluate the reciprocal effects model.

In Model 4A (Table 3), age is positively correlated with gymnastics performance at both T1 (.23) and T2 (.27), but not significantly correlated with self-concept at either T1 or T2. This is consistent with the suggestion that participants compare their gymnastic skills with those of other participants their own age (i.e. a frame of reference effect; Marsh, 1993b). Interestingly, particularly at T1 but also at T2, boys have significantly higher gymnastics self-concepts than girls, but boys have significantly lower gymnastics performance scores. This suggests that, relative to girls, boys' self-concepts scores are positively biased in relation to their actual performance levels. However, although still evident at T2, the size of this apparent bias is smaller, suggesting that boys' gymnastic self-concepts became more realistic in relation to actual performance through participation in the gymnastics training program. Although there were main effects of gender and age, there were no statistically significant age x gender interactions for any of the self-concept or the gymnastic performance outcomes.

In Model 4B, the direct effects of gender and age—the path coefficients in Table 3—are systematically smaller than the corresponding correlations in Model 4A. Hence, not surprisingly, much of the effect of the background variables on T2 outcomes is mediated through the corresponding T1 variables.

The primary interest in Model 4B is the path coefficients relating T1 self-concept and gymnastics performance to T2 self-concept and gymnastic performances. The general pattern and

even the sizes of these path coefficients are very similar to those reported in Model 1B (Table 1), which did not control for gender and age effects. Although the differences are small, it is interesting to explore the direction of these differences. The effects of T1 gymnastics performance on T2 outcomes in Model 4B are somewhat smaller than the corresponding effects in Model 1B—particularly for the stability of the gymnastics performance over time. A small amount of the apparent stability in performance can be explained in terms of higher scores for girls and higher scores for older participants. Because gender and age come before gymnastics performance in the causal ordering, they are interpreted as spurious effects associated with a potential bias in not controlling for these variables. However, controlling for the effects of gender and age actually increases the effects of T1 self-concept, although the sizes of these differences are small. While suppression effects are apparently unusual in the social sciences, other examples have been previously demonstrated in a similar situation (Marsh, 1993b). Because age was nearly unrelated to self-concept responses, controlling for age had almost no effect on the stability of self-concept over time. Gender, however, was negatively related to self-concept but positively related to achievement. Hence, a small portion of the positive effect of T1 self-concept on T1 gymnastics performance was suppressed by gender so that the effect was somewhat larger when gender was controlled. In summary, controlling for gender and age had little effect on support for the reciprocal effects model and actually led to a small increase in the size of the path leading from T1 self-concept to T2 gymnastics performance. These results support the reciprocal effects model.

### **Discussion**

There now exists an extensive body of research in support of the reciprocal effects model of relations between academic self-concept and academic achievement in traditional academic settings (Marsh & Craven, in press). In the present investigation we began by evaluating support for the generalizability of these results to relations between gymnastics self-concept and gymnastics performance skills. Consistent with academic self-concept research, we found clear support for the reciprocal effects model. Of particular importance to self-concept theory, more positive levels of prior self-concept led to higher subsequent levels of gymnastics performance than could be explained by prior levels of gymnastics performance. However, higher levels of prior gymnastics performance also led to higher levels of subsequent gymnastics self-concept than could be explained in terms of prior



levels of gymnastics self-concept. This finding has important implications for physical education teachers and physical fitness leaders, but also for self-concept theory. In particular, the results support the claims by self-concept researchers that a positive self-concept is an important goal in its own right as well as being important in facilitating other desirable outcomes—gymnastics performance skills, in the present investigation. In this sense, physical self-concept and exercise behavior are logically distinct, and each is considered to be an important outcome in physical education settings.

Results of the present investigation have significant implications for the importance placed on physical self-concept as a means of facilitating performance skills in a physical setting, as well as being an important outcome variable. If the direction of causality had been from self-concept to performance (self-enhancement model), then class leaders might be justified in placing most of their effort into enhancing participants' self-concepts rather than fostering gymnastics skills per se. On the other hand, if the direction of causality had been from gymnastics performance to self-concept (skill development model), then class leaders should focus primarily on improving performance skills of the participants, in that this is also the best way to improve physical self-concept. In contrast to both these apparently over-simplistic (either-or) models, the reciprocal effects model implies that self-concept and gymnastics performance are reciprocally related and mutually reinforcing. Improved physical self-concepts will lead to improved performance skills *and* improved performance skills will lead to better physical self-concepts. Hence, if class leaders enhance participants' physical self-concepts without improving corresponding performance skills, then the gains in self-concept are likely to be short-lived. However, if class leaders improve participants' performance skills without also fostering participants' self-beliefs in their physical capabilities, then the performance gains are also unlikely to be long lasting. If class leaders focus on either one of these constructs to the exclusion of the other, then both are likely to suffer. Hence, according to the reciprocal effects model, class leaders should strive to improve simultaneously both physical self-concept and performance skills. Importantly, the generalizability of support for these predictions in a physical setting from the well-established pattern of results-based on academic self-concept and traditional measures of academic achievement provides a much stronger basis of support for the reciprocal effects model.

## References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51 (6), 1173-1182.
- Bentler, P. M. (1988). *Theory and implementation of EQS: A structural equations program*. Los Angeles: BMDP Statistical Software.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238-246.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley and Sons.
- Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. *Review of Educational Research*, 54, 427-456.
- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications and programming*. Mahwah, NJ: Erlbaum.
- Byrne, B. M. (2003). Testing for equivalent self-concept measurement across culture: Issues, caveats, and application. In H. W. Marsh, R. G. Craven, & D. McInerney (Eds.), *International Advances in Self Research* (Vol. 1, pp. 291-314). Greenwich, CT: Information Age.
- Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? *Journal of Educational Psychology*, 69, 136-145.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
- Duda, J.L., & Nicholls, J.G. (1992). Dimensions of achievement motivation in schoolwork and sport. *Journal of Educational Psychology*, 84, 290-299.
- Fox, K. R., & Corbin, C. B. (1989). The physical self-perception profile: development and preliminary validation. *Journal of Sport & Exercise Psychology*, 11(4), 408-430.
- Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing Moderator and Mediator Effects in Counseling Psychology Research. *Journal of Counseling Psychology*, 51, 115-134.

- Guay, F., Marsh, H. W. & Boivin, M. (2003). Academic Self-concept and Academic Achievement: Development Perspectives on Their Causal Ordering. *Journal of Educational Psychology, 95*, 124-136.
- Joreskog, K. G. (1979). Statistical estimation of structural models in longitudinal investigations. In J. R. Nesselroade & B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 303-351). New York: Academic Press.
- Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.
- Marsh, H. W. (1990a). The causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal path analysis. *Journal of Educational Psychology, 82*, 646-656.
- Marsh, H. W. (1990b). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2*, 77-172.
- Marsh, H. W. (1993a). Academic self-concept: Theory measurement and research. In J. Suls, (Ed.) *Psychological Perspectives on the self* (Vol. 4, pp. 59-98). Hillsdale, NJ: Erlbaum.
- Marsh, H. W. (1993b). Physical fitness self-concept: Relations to field and technical indicators of physical fitness for boys and girls aged 9-15. *Journal of Sport and Exercise Psychology, 15*, 184-206.
- Marsh, H. W. (1994). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling, 1*, 5-34.
- Marsh, H. W. (1997). The measurement of physical self-concept: A construct validation approach. In K. Fox (Ed.), *The physical self: From motivation to well-being* (pp. 27-58). Champaign, IL: Human Kinetics.
- Marsh, H. W. (2002). A multidimensional physical self-concept: A construct validity approach to theory, measurement, and research. *Psychology: The Journal of the Hellenic Psychological Society, 9*, 459-493.
- Marsh, H. W., Balla, J. R., & Hau, K. T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical processes. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *Advanced structural equation modeling techniques* (pp. 315-353). Hillsdale, NJ: Erlbaum.
- Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor

- analysis: The effect of sample size. *Psychological Bulletin*, 102, 391-410.
- Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal Ordering of Academic Self-concept and Achievement: Reanalysis of a Pioneering Study and Revised Recommendations. *Educational Psychologist*, 34, 155-167.
- Marsh, H.W. & Craven, R. G. (in press). What comes first?: A reciprocal effects model of the mutually reinforcing effects of academic self-concept and achievement. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.). *International Advances in Self Research* (Volume 2, pp. xx-xx). Greenwich, CT: Information Age.
- Marsh, H. W., & Hau, K-T. (1996). Assessing goodness of fit: Is parsimony always desirable? *Journal of Experimental Education*, 64, 364-390.
- Marsh, H. W., Hau, K. T., & Kong, K. W. (2002). Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English vs. Chinese) for Hong Kong Students. *American Educational Research Journal*, 39, 727-763.
- Marsh, H. W., Ellis, L., & Craven, R. G. (2002). How do preschool children feel about themselves? Unravelling measurement and multidimensional self-concept structure. *Developmental Psychology*, 38, 376-393.
- Marsh, H. W., Hau, K-T & Grayson, D. (in press). Goodness of Fit Evaluation in Structural Equation Modeling. In A. Maydeu-Olivares & J. McCardle (Eds.), *Psychometrics. A Festschrift to Roderick P. McDonald*. NJ: Erlbaum.
- Marsh, H. W., Hau, K.T. & Wen, Z., (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modelling*, 11, 320-341.
- Nicholls, J. G., Cheung, P., Lauer, J., & Patashnick, M. (1989). Individual differences in academic motivation: Perceived ability, goals, beliefs, and values. *Learning and Individual Differences*, 1, 63-84.
- Sarrazin, P., Roberts, G., Cury, F., Biddle, S., & Famose, J.-P. (2002). Exerted effort and performance in climbing among boys: The influence of achievement goals, perceived ability, and task difficulty. *Research Quarterly for Exercise and Sport*, 73(4), 425-436
- Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American*

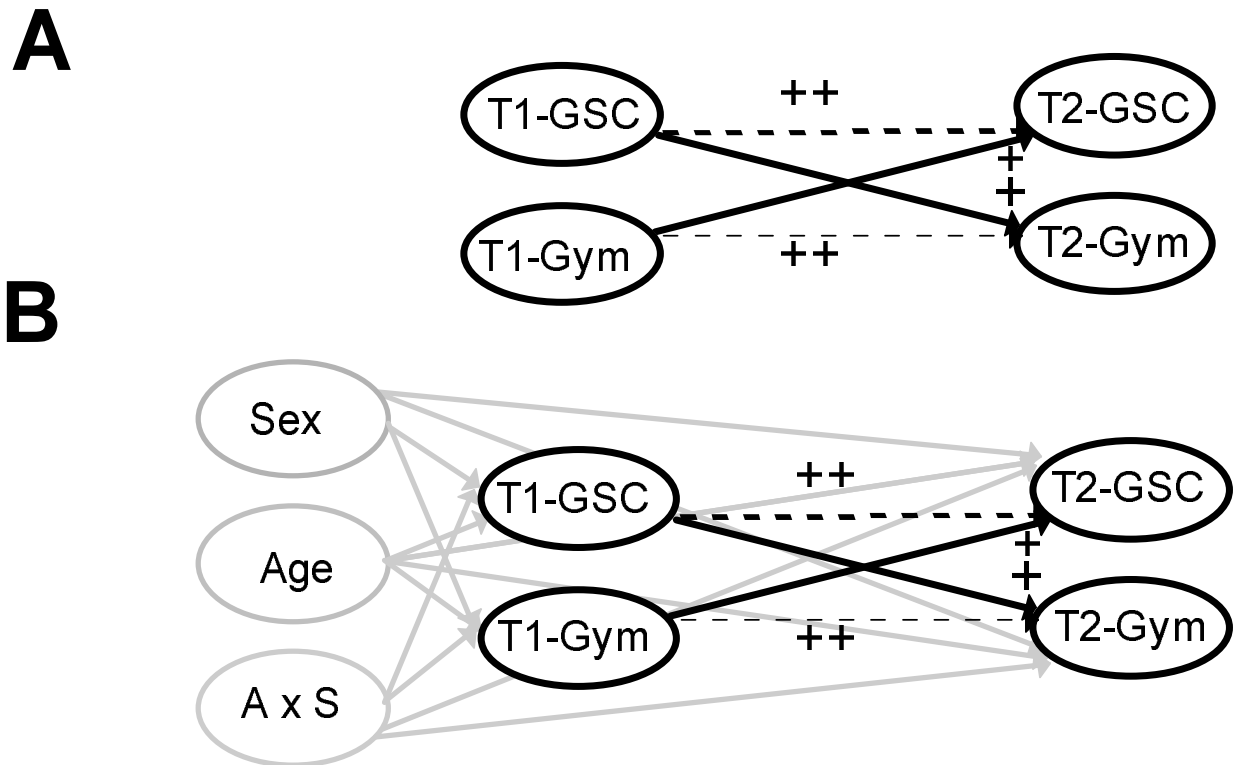
*Psychologist*, 55, 5–14.

Smith, R. L., Ager, J. W., & Williams, D. L. (1992). Suppressor variables in multiple regression/correlation. *Educational & Psychological Measurement*, 52, 17-29.

Valentine, J. C. (2002). The relation between self-concept and achievement: A meta-analytic review. *Dissertation Abstracts International: Section B: The Sciences & Engineering*, 62, 4278.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relations between self-beliefs and academic achievement: A systematic review. *Educational Psychologist*, 39, 111-133.

Vallerand, R. J., Blanchard, C., Mageau, G. A., Koestner, R., Ratelle, C., Leonard, M., Gagne, M., & Marsolais, J. (2003). Les passions de l'ame: On obsessive and harmonious passion. *Journal of Personality & Social Psychology*, 85, 756-767.



*Figure 1.* Reciprocal Effects Model. (A) Path model of relations between gymnastics self-concept (GSC) and gymnastics performance skills (Gym) at Times 1 and 2 (T1 & T2). The skill development model predicts that the T1 Gym  $\rightarrow$  T2 GSC path will be significantly positive (+). The self-enhancement models posits that the T1 GSC  $\rightarrow$  T2 Gym path will be significantly positive (+). The reciprocal effects model predicts that both T1 Gym  $\rightarrow$  T2 GSC and T1 GSC  $\rightarrow$  T2 Gym paths will be significantly positive. All three theoretical models predict that the horizontal paths (represented by dashed lines) will be highly significant (++). (B) Reciprocal effect model extended to include gender (sex), age, and their interaction (A x S). Because the main focus of this study is on the reciprocal effects model, paths associated with new variables in the extended model are represented as gray lines.

**Table 1**

*Basic Reciprocal Effects Model: Corresponding Factor Solutions for the Corresponding Confirmatory Factor Analysis (CFA; Model 1A) and Structural Equation (SEM; Model 1B) Models*

**Factors (CFA Solution, Model 1A, Table 2)**

Variable	T1PSC	T1ACH	T2PSC	T2ACH	UNIQ	CU
<b><u>Factor Loadings (Model 1A &amp; 1B)</u></b>						
T1PSC1	.92*				.15*	
T1PSC2	.88*				.22*	
T1PSC3	.65*				.58*	
T1ACH1		.92*			.16*	
T1ACH2		.98*			.05*	
T1ACH3		.91*			.18*	
T2PSC1			.91*		.16*	.05*
T2PSC2			.93*		.14*	.02
T2PSC3			.66*		.57*	.27*
T2ACH1				.97*	.07*	
T2ACH2				.95*	.10*	
T2ACH3				.94*	.11*	
<b><u>Factor Correlations (Model 1A)</u></b>						
T1PSC	1.00					
T1ACH	.51*	1.00				
T2PSC	.77*	.51*	1.00			
T2ACH	.45*	.67*	.51*	1.00		
<b><u>Factor Residuals (Model 1B)</u></b>						
T1PSC	1.00					
T1ACH	.51*	1.00				
T2PSC			.39*			
T2ACH			.09*	.54*		
<b><u>Factor Path Coefficients (Model 1B)</u></b>						
T1PSC						
T1ACH						
T2PSC*	.70*	.15*				
T2ACH*	.15*	.59*				

*Note.* PSC = physical self-concept, ACH = gymnastics performance, T1 = Time 1, T2 = Time 2, uniq = measured variable uniqueness, CU = correlated uniqueness. Model 1A was a confirmatory factor analysis (CFA) model in which the four latent constructs were allowed to be correlated. Model 1B was a structural equation (SEM) model, the reciprocal effects model, in which T1 constructs were posited to influence T2 constructs. Because these are equivalent models, the factor loadings, uniquenesses, and correlated uniquenesses are the same for both models, whereas the factor residuals and paths coefficients in Model 1B are merely a mathematical transformation of the factor correlations in Model 1A.

\*  $p < .05$

Table 2

*Summary of Goodness of Fit for All Models*

<u>MODEL</u>	<u><math>\chi^2</math></u>	<u>DF</u>	<u>TLI</u>	<u>RNI</u>	<u>RMSEA (90% CI)</u>	<u>Description</u>
<b>Model 1: CFA/SEM Total group (N = 376)</b>						
1a	67.7	45	.995	.997	.037 (.016-.054)	Confirmatory Factor Analysis Model(CFA)
1b	67.7	45	.995	.997	.037 (.016-.054)	Structural Equation Model (SEM)
<b>Model 2: SEM Year Group Invariance: Younger (N = 257) Older (N = 119)</b>						
2a	116.6	90	.994	.996	.040 (.013-.059)	CFA: No Invariance
2b	126.0	98	.994	.996	.039 (.013-.057)	CFA: Factor Loading (FL) invariant
2c	132.0	102	.994	.995	.040 (.015-.058)	SEM: FL, Path Coeff (PC) invariant
2d	133.3	108	.995	.996	.035 (.000-.054)	CFA: FL, PC, Factor Resid (FR) invariant
2e	150.3	123	.994	.995	.034 (.004-.052)	CFA: Total (FL, PC,FR, Uniq) invariant
<b>Model 3: SEM Gender Invariance: Male (N = 204) Female (N = 172)</b>						
3a	105.2	90	.996	.998	.030 (.000-.052)	CFA: No Invariance
3b	116.0	98	.996	.997	.031 (.000-.052)	CFA: Factor Loading (FL) invariant
3c	123.0	102	.996	.997	.033 (.000-.053)	SEM: FL, Path Coeff (PC) invariant
3d	140.7	108	.994	.995	.040 (.017-.058)	CFA: FL, PC, FR invariant
3e	178.8	123	.992	.992	.049 (.032-.064)	CFA: Total (FL, PC,FR, Uniq) invariant
<b>Model 4: CFA/SEM Extended Model with Gender and Year (N = 376)</b>						
4a	102.1	69	.993	.995	.036 (.020-.050)	CFA: Total Group
4b	102.1	69	.993	.995	.036 (.020-.050)	SEM: Total Group

*Note.* RNI = relative noncentrality index, TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation; 90% CI = 90% confidence interval for the RMSEA, DF = degrees of freedom. All tests of invariance (Models 2, 3 and 4) compared the effects of constraining some or all parameter estimates to be the same in two groups based on age (Model 2), gender (Model 3) or class placement (Model 4). In Model 5 a structural equation model was constructed by adding to Model 1 four single-item factors representing the effects of Age, Gender, Age x Gender interaction, and Placement (see Table 4).



**Table 3**

*Basic Reciprocal Effects Model: Corresponding Factor Solutions for the Corresponding Confirmatory Factor Analysis (CFA; Model 4A) and Structural Equation (SEM; Model 4B) Models*

Factor	Age	Gen	AxG	T1PSC	T1ACH	T2PSC	T2ACH
<b>Factor Correlations (Model 4A)</b>							
Age (A)	1.00						
Gender (G)	.19*	1.00					
A x G	.05	-.03	1.00				
T1PSC	.03	.22*	.08	1.00			
T1ACH	.23*	-.16*	.02	.51*	1.00		
T2PSC	.08	.12*	.07	.77*	.51*	1.00	
T2ACH	.27*	-.13*	.06	.45*	.67*	.51*	1.00
<b>Factor Residuals (Model 4B)</b>							
Age (A)	1.00						
Gender (G)	.19	1.00					
A x G	.05	-.03	1.00				
T1PSC				.94*			
T1ACH				.55*	.90*		
T2PSC						.39*	
T2ACH						.09*	.52*
<b>Factor Path Coefficients (Model 4B)</b>							
Age (A)							
Gender (G)							
A x G							
T1PSC	-.02	.22*	.09				
T1ACH	.27*	-.21*	.01				
T2PSC	.03	-.02	.01	.71*	.14*		
T2ACH	.11*	-.11*	.03	.20*	.52*		

*Note.* Age = age in years; Gender 1 = female, 2 = male; Gender x Age = Gender by age interaction. PSC = physical self-concept, ACH = gymnastics performance, T1 = Time 1, T2 = Time 2. Models 3A and 3B were constructed by adding to Models 1A and 1B, respectively, three single-item factors representing the effects of Age, Gender, and the Age x Gender interaction. Because the factor structure for the physical self-concept and gymnastics achievement was essentially the same as those from Models 1A and 1B (see Table 1), factor loadings are not repeated. All parameter estimates are presented in completely standardized form.

\*  $p < .05$