Chapter 6

Invariance Testing Across Samples and Time: Cohort-Sequence Analysis of Perceived Body Composition

Order of Authorship and Contact Information

Herbert W. Marsh
Institute for Positive Psychology and Education,
Australian Catholic University, Strathfield Campus,
Locked Bag 2002, Strathfield, NSW 2135, Australia
E-mail: Herb.Marsh@acu.edu.au

Philip D. Parker
Institute for Positive Psychology and Education,
Australian Catholic University, Strathfield Campus,
Locked Bag 2002, Strathfield, NSW 2135, Australia
E-mail: Philip.Parker@acu.edu.au

Alexandre J. S. Morin,
Institute for Positive Psychology and Education,
Australian Catholic University, Strathfield Campus,
Locked Bag 2002, Strathfield, NSW 2135, Australia
E-mail: Alexandre.Morin@acu.edu.au

Acknowledgements

This chapter was made possible in part by grants from the Australian Research Council (DP130102713; DP140101559).

Marsh, A.J.S., Parker, P.D., & Morin, A.J.S. (2015, In Press). Invariance Testing Across Samples and Time: Cohort-Sequence Analysis of Perceived Body Composition. In N. Ntoumanis, & N. Myers (Eds.), An Introduction to Intermediate and Advanced Statistical Analyses for Sport and Exercise Scientists. Wiley.

Invariance Testing Across Samples and Time:

Cohort-Sequence Analysis of Perceived Body Composition

Abstract

Longitudinal invariance of perceived body composition for elite athletes and non-athletes was assessed on the basis of responses from 4 age cohorts (grades 7–10 in high school) who completed the same multi-item instrument 4 times over a 2-year period (multiple waves). Cohort-sequence designs such as this provide a stronger basis for assessing developmental stability/change than either cross-sectional (multi-cohort, single occasion) or longitudinal (single-cohort, multiple occasion) designs. Here we demonstrate different approaches to evaluating measurement invariance across the 5-year span (grades 7–11) and 4 cohorts, leading to tests of longitudinal stability/change based on latent means as a function of athlete/non-athlete and gender groups.

Keywords: Measurement invariance; cohort-sequence design; growth models; differential item functioning; goodness-of-fit

Some twenty years ago, Schutz and Gessaroli (1993) argued that while for many applications, confirmatory factor analysis (CFA) and structural equation modeling (SEM) should be the methodology of choice, they had seen almost no application in sport/exercise psychology. Even a casual perusal of the major sport/exercise journals shows that the popularity of CFA/SEM has increased substantially. However, in the 2007 *Handbook of Sport Psychology*, Marsh (2007) argued that despite this growing popularity, there appears to be an ever-widening gap between "state-of-the-art" methodological and statistical techniques that should be part of the repertoire of quantitative researchers, and the actual skill levels of many applied researchers in sport/exercise sciences. Among other issues described by Marsh were tests of factorial invariance of parameter estimates across independent groups (e.g., men and women, different age groups, or elite vs. non-elite athletes) or over time, that can be formulated as a set of fully (or partially) nested set of models in which the endpoints are models with no invariance constraints and models with all parameter estimates invariant (i.e., constrained to be equivalent) across all groups or occasions. In this chapter we expand upon the presentation of invariance across groups and time, introducing new and evolving design issues and strategies for evaluating these data.

We briefly introduce the importance of invariance testing generally, and its relevance in sports and exercise science. In particular we focus on the use of the cohort-sequence design (also called accelerated longitudinal design; Bell, 1953; Mehta & West, 2000; Nesselroade & Baltes, 1979) to evaluate longitudinal invariance over time, and multi-group invariance over samples. Then we briefly review substantive issues that underpin the demonstrations in this chapter. Finally, we demonstrate evolving statistical approaches to the evaluation of cohort-sequence designs.

General Introduction to the Importance of Measurement Invariance

Tests of measurement invariance evaluate the extent to which measurement properties generalize over multiple groups, situations, or occasions. Of particular substantive importance for sport/exercise research are the evaluations of differences across multiple groups (e.g., athlete versus nonathlete; male versus female, age groups, exercise treatment groups versus control groups) or over time (i.e., observing the same group of participants on multiple occasions, perhaps before and after an intervention). The need for rigorous tests of whether the underlying factor structure is the same for different groups or occasions has often been ignored in sport/exercise research. However, such comparisons assume the invariance of at least factor loadings and, in some cases, item intercepts. Indeed, unless the underlying factors are measuring the same construct in the same way, and the measurements themselves are operating in the same manner across groups or time, the comparison of parameter estimates is potentially invalid. For example, if gender or longitudinal differences vary substantially for different items used to infer a construct, in a manner unrelated to respondents' true levels on the latent construct, then the observed differences might be idiosyncratic to the particular items used. From this perspective, it is important to be able to evaluate the full measurement invariance of participants' responses.

Marsh et al. (2009, 2010) operationalized a taxonomy of 13 models (see Appendix 6.1) designed to test measurement invariance that integrates traditional CFA approaches to factor invariance (e.g., Jöreskog & Sörbom, 1988, 1993; Marsh 1994, 2007; Marsh & Grayson 1994) with item-response-theory approaches to measurement invariance (e.g., Meredith 1964, 1993; also see Millsap 2011, Vandenberg & Lance 2000). Key models test the goodness of fit of models with no invariance constraints (configural invariance, Model 1); invariance of factor loadings (metric or weak invariance, Model 2), factor loadings and item intercepts (scalar or strong invariance, Model 5), or factor loadings, item intercepts and item uniquenesses (strict invariance, Model 7). The final four models (Models 10–13) all constrain mean differences between groups to be zero—in combination with the invariance of other parameters.

Essentially the same logic and taxonomy of models can be used to test the invariance of parameters across multiple occasions for a single group. One distinctive feature of longitudinal analyses is that they should normally include correlated uniquenesses between responses to the same item on different occasions (see Jöreskog 1979, Marsh 2007, Marsh & Hau 1996). Although occasions are the most typical test of invariance over a within-person construct like time (i.e., multiple occasions), this is easily extended to include other within-subject variables (e.g., coach or teammates' ratings of the same athlete; e.g., Marsh, Liem, Martin, Morin, & Nagengast,2011; Marsh, Nagengast, Morin, Parada, Craven & Hamilton, 2011; Marsh, Morin, Parker & Kaur, 2014). Indeed, it is possible to extend these models to test the invariance over multiple grouping variables or combinations of multigroup (between-person) and within-person variables (e.g., Marsh, Abduljabbar, Parker et al. 2014; Marsh, Morin, Parker & Kaur, 2014; Marsh, Nagengast & Morin, 2013).

Although application of the full taxonomy of models is useful, in the present investigation we focus primarily on three models that are central for latent variable models and the evaluation of latent means:

- configural invariance (whether the a priori factor structure fits when no invariance constraints are imposed over time or groups; see Model Con-1a and 1b in Table 1, and Mplus syntax in Appendix 6.2:).
- metric or weak factorial invariance (tests of the invariance of factor loadings over time and/or groups; Model Met-2a, 2b and 2C in Table 1, and Mplus syntax in Appendix 6.3). and
- scalar or strong (tests of the invariance of factor loadings and item intercepts over time and/or groups; see Models Scl-3a to Scl-3d in Table 1, and Mplus syntax in Appendix 6.4 and 6.5).

Differential item functioning (DIF) is a critical issue in tests of invariance: whether item-level parameter estimates are the same over groups, time, or combinations of groups and time. The metric invariance model tests for DIF in relation to factor loadings, whereas the scalar invariance model tests for DIF in relation to item intercepts. If the purpose of a research study is merely to evaluate how relations among constructs vary over multiple groups or occasions, then factor loading (metric) invariance may be sufficient. However, if the purpose of the study is to compare latent means over groups or occasions, then scalar (factor loading and intercept) invariance is required. For example, if differences over groups or time are not consistent across the items associated with a particular latent factor, then the results are likely to depend on the mix of items considered. A subset of the items actually used, or a new sample of items designed to measure the same factor, could give different results.

We note that it is also possible to test the invariance of the uniqueness terms (including random measurement error) associated with individual items (see Models Unq-4a in Table 1, and Mplus syntax in Appendix 6.6.) However, if the focus is on the evaluation of latent relations or latent means based on latent variable models, then uniqueness invariance is not a necessary condition. This follows in that measurement error is controlled in latent-variable models in which each of the constructs is based on multiple indicators (typically items). Nevertheless, although the valid comparison of latent means does not depend on the invariance of item uniquenesses, there are limitations associated with the non-invariance of item uniquenesses (e.g., manifest means, item variances, and scale variances, as well as relations across manifest constructs, are not comparable across groups). Thus, if the comparison over groups or occasions is based on manifest variables (or scale scores), then the necessary assumptions are the invariance of item uniquenesses, together with the invariance of the factor loadings and item intercepts. Hence, the comparison of manifest variables is considerably more demanding than comparisons based on latent-variable models controlling for measurement error.

It is also possible to test the invariance over groups or occasions of latent factor variances and covariances, particularly when there are multiple factors. Although the invariance of factor variances and covariances is not a necessary assumption underlying the comparison of latent means, relations among factors (or the latent factors and other variables) are likely to be of fundamental interest in many applied research studies.

Cohort Sequential Designs: Longitudinal Invariance Across Samples and Time

As argued by Marsh (1998; Marsh, Craven & Debus, 1999; Parker, Marsh, Morin, Seaton & Van Zanden, in press), a multiwave-multicohort design often provides a stronger basis for evaluating developmental differences than cross-sectional comparisons based on many age cohorts, or longitudinal comparisons based on a single age cohort. Whilst educational and particularly developmental psychologists often extol the virtues of true longitudinal designs over cross-sectional designs, ultimately, support for the generality of developmental effects requires convergence of results across multiple approaches. Hence, juxtaposition of longitudinal and cross-sectional approaches to developmental change provides an important basis for cross-validating the results based on each approach. For example, multi-cohort sequential designs, as in the present investigation, have the advantage of providing tests for history and cohort effects (i.e., based on overlapping data collection waves collected from multiple cohorts of participants) that would not be possible with pure longitudinal designs based on a single cohort or in pure cross-sectional designs based on multiple cohorts.

Increasingly, developmental research with school-aged children relies on large-scale longitudinal data sets. However, many of these databases extend over no more than two or three years, making it difficult to fully explore the growth of key educational constructs over the course of major developmental periods such as high school/adolescence. Here we address this issue through the use of cohort sequential designs in which multiple waves of data are collected simultaneously from multiple age cohorts. This strategy provides sport and exercise researchers with a feasible, cost-effective alternative to exploring growth over the course of an entire developmental period (see Brodbeck, Bachmann, Croudace, & Brown, 2012; Enders, 2010; Graham, 2012; Marsh, 1998; Marsh, Craven & Debus, 1999). Furthermore, simulation studies have shown that cohort sequential designs have greater power than standard longitudinal designs when the same number of time waves is collected in each cohort (Graham, 2012). The increasing prevalence of such designs and the development of statistical procedures for analysing such data is a substantive–methodological synergy (Marsh & Hau, 2007) in which complex substantive issues stimulate the development of stronger methodological tools.

In our demonstration, four waves of data were collected, six months apart, for each of four age cohorts. The cohort-sequential design thus provided a total of 10 waves of data covering five years of high school, in which the multiple waves of data overlap for each successive cohort (see Figure 6.1a for structure and Figure 6.1b for measurement model for this design). One of the most critical aspects of such a design however, is how to deal with the inevitable large amount of data that is missing by design (white cells in Figure 6.1), even in the absence of sample attrition and data holes. The advantage of cohort-sequential designs, however, is that the missing time points in all cohorts are missing due to the design of the study, not as a function of participants' characteristics. Thus, these missing data fully correspond to missing-completely-at-random assumptions of modern missing data techniques (Enders, 2010). This suggests that modern missing data techniques can provide unbiased parameter estimates even in the presence of missing data (Enders, 2010). There are essentially two approaches to estimating growth models with cohort sequential data that aim to overcome this missing by design component. A common approach is to use full information maximum likelihood estimation on data that is stacked and merged across cohorts. In other words, this approach involves re-organizing the data set so that each participant (each line) is specified as having 10 measurement points, with 6 of those being missing (see Figure 6.1). In this approach however, some cells of the variance covariance matrix have zero coverage (i.e., all cases are missing for the entire cohort) and thus, full-information-maximum likelihood estimation of these covariances becomes problematic (see Enders, 2010).

A second approach, and that used here, is to make use of a multi-group approach to model estimation. Modern structural equation modeling packages are becoming increasingly powerful and flexible in estimating complex models. For models such as those proposed here, we take advantage of the ability of Mplus to fit differing models in each cohort. Essentially, models are fitted in each cohort (treated as a separate group in a multiple group design), in relation to their relative position in the developmental period of interest. In the current research, models were specified such that the first cohort reflected growth over Years 7 and 8, the second Years 8 and 9, the third Years 9 and 10, and the last Years 10 and 11. Through the inclusion of invariance constraints across these multiple groups for the overlapping time points (the cells in boxes in Figure 6.1a), the resulting full model covers five years of high school, even though any one participant only has data for two years. In Appendices 6.2 to 6.7, we provide annotated Mplus scripts for the estimation of a variety of models using a cohort sequential design (see also Brown, Croudace, & Heron, 2011).

Substantive Application: Physical Self-Concept

Here we briefly review research literature underpinning the substantive focus of this largely methodological chapter. Physical self-concept is a key construct in sport and exercise psychology. Here we focus on stability/change over time of perceived body composition (PBC) during the potentially turbulent adolescent period, which is critical for the development of boys and girls, and of athletes and non-athletes. An essential underlying assumption in such comparisons is measurement invariance.

Age and gender effects on self-concept have theoretical, practical, and methodological implications. Historically, self-concept researchers have been particularly interested in stability and change in self-concept during the potentially volatile adolescent period (Dusek & Flaherty, 1981; Wylie, 1979), although most of this research has focused on self-esteem, rather than on physical components of self-concept. Predictions about how self-concept develops with age have been proposed from a variety of theoretical perspectives. Marsh (1989; 1990; Marsh, Craven, & Debus, 1991) proposed that the self-concepts of very young children are consistently high but that with increasing life experience children learn their relative strengths and weaknesses, so that mean levels of self-concept decline, multiple dimensions of self-concept become more differentiated, and self-concepts become more highly correlated with external indicators of competence (e.g., skills, accomplishments, and self-concepts inferred by significant others). Eccles, Wigfield, Harold, and Blumfeld (1993) similarly proposed that the declines in mean levels of self-concept, particularly during the pre-adolescent/early-adolescent period, reflected an optimistic bias for young children and increased accuracy in responses as they grow older. Based on a large empirical study, Marsh (1989) reported that there was a reasonably consistent pattern of self-concepts declining from a young age through early adolescence, leveling out, and then increasing, at through at least to early adulthood.

Historically, gender differences in self-concept have focused on typically small differences in self-esteem favouring boys (e.g., Feingold, 1994). Marsh (1989), however, demonstrated that these small differences in total scores reflect larger, counterbalancing gender differences in specific components of self-concept. The gender differences in specific scales tended to be consistent with traditional gender stereotypes: (a) boys had higher self-concepts for Physical Ability, Appearance, Math, Emotional Stability, Problem Solving, and Esteem; (b) girls had higher self-concepts for Verbal/Reading, School, Honesty/Trustworthiness, and Religion/Spiritual Values; and (c) there were no gender differences for the Parents scale. For all three age ranges, the largest gender differences were the higher scores for boys in Physical Ability and Appearance self-concepts. Whereas the gender differences for preadolescents were larger for Physical Ability than Appearance, the Appearance differences were larger for adolescents and late-adolescents.

In this chapter our focus is on development/change in self-perception of body composition for adolescent boys and girls *(N* = 1268) attending a prestigious sports high school in Sydney Australia. Each year, elite athlete students from across the state compete for enrollment in most major sports. However, the school also admits students (62%) from the local catchment area who typically do not participate in the elite athletic program unless they choose to try out for the program. Although we refer to these as the elite-athlete and non-athlete groups respectively, some non-athletes may participate in athletic activities, and all are students in compulsory physical education classes, even though they are not part of the elite athlete program. Hence, we are interested in development or change in PBC over the adolescent period in relation to gender, athletic group, and their interaction.

Based on previous research we anticipated that boys would have better PBCs than girls, and that athletes would have better PBCs than non-athletes. However, we left as a research question whether gender differences would interact with athlete groups (i.e., whether gender differences in PBC are similar for athlete and non-athlete groups). Of particular relevance to the present investigation is whether there is development and change in these group differences over the adolescent period under consideration—whether gender differences for athletes and non-athletes become larger or smaller during adolescence.

These data also provide a unique opportunity to evaluate how the physical self-concepts of elite athletes develop/change over the five-year period, starting from the time when they are first selected to participate in this highly selective sporting program. Based on the theoretical self-concept theory underpinning the big-fish-little-pond effect (BFLPE; e.g., Marsh, Abduljabbar, et al., 2014; Marsh, Kuyper et al., 2014; Marsh, Seaton, et al., 2008), there is a juxtaposition of reflected glory effects (assimilation; “if I am good enough to be selected to be in this program, I must be pretty good”), leading to more positive self-concepts, and social comparison (contrast; “relative to all these other elite athletes, maybe I am not as good as I thought”) leading to more negative effects. Also, there is a temporal aspect to this juxtaposition, wherein positive assimilation effects might dominate when participants are first selected to be in an elite athlete program, but contrast effects become increasingly strong over time as their frame of reference shifts to other elite athletes in the same program. Although there is ample evidence of the negative effects on academic self-concept of attending academically selective high schools and programs, causing long-lasting negative effects on many academic outcomes (e.g., Marsh, 1991; Marsh, Seaton, et al., 2008), there is apparently little research in relation to sport and exercise (but see Marsh, Chanal, Sarrazin & Bois, 2006). Nevertheless, based on research in the academic domain, the athlete/non-athlete difference might be expected to be largest at the start of high school and to decline over this five-year period.

Methodology

The PSDQ Instrument. The short version of the PSDQ (Marsh, Martin & Jackson, 2010) measures nine specific components and two global components of physical self-concept, based on responses to 3 or 4 items for each scale. The theoretical basis of the PSDQ is derived in part from the Shavelson et al. (1976) multidimensional, hierarchical model of self-concept, but also from Marsh's (1993) confirmatory factor analysis of physical fitness indicators, based on Fleishman’s (1964) structure of physical fitness. PSDQ research (see reviews by Byrne, 1996; Marsh, 1997; Marsh, Martin & Jackson, 2010; Marsh & Cheng, 2012) has provided good support for the PSDQ factor structure and its generalizability over gender, and different groups of elite and non-elite athletes (Marsh, Perry, Horsely & Roche, 1995; Marsh, Richards, et al., 1994), but also for convergent and discriminant validity in relation to a set of external validity criteria, including body composition, physical activity levels, and physical fitness tests of cardiovascular endurance, strength, and flexibility (Marsh, 1996a, 1996b, 1997). For the purposes of the present investigation, we only considered responses to the PBC factor: self-perceptions of body composition measured by a set of three items (“My waist is too large”; “I have too much fat on my body”; “I am overweight”).

Statistical Analyses. Across all models, PBCs were specified as latent variables estimated from multiple items. This requires relatively complex identification constraints. In some models we used a non-arbitrary metric for factor loadings (which were constrained to average 1 for each factor) and item intercepts (which were constrained to sum to 0 for one factor), allowing results to be interpreted according to the original 6-point Likert scale rather than according to a standardized metric (when the model is identified by constraining factor variances to equal 1 and factor means to equal 0) or as a function of the scale of a referent indicator (when the model is identified by constraining the factor loading of a referent indicator to equal 1 and its intercept to equal 0) (see Little, Slegers, & Card, 2006). The PSDQ was administered on four occasions during a two-year period of time at approximately six-month intervals (see Figure 6.1a, b). As is typical in large longitudinal field studies, a substantial portion (23%) of the sample had missing data for at least one of the four occasions, due primarily to absence or the provision of an illegible (or fictitious) name on at least one of the four occasions. All models were fitted using the robust Maximum Likelihood estimator (MLR) available in Mplus 7.2, which has the advantage of being robust to non-normality of data and remains equivalent to ML when normality assumptions are met. Multiple imputation was used to handle missing data. The multiple imputation was based on a data file (see Chapter 6 data file in Appendix 6.9, the multiple imputation file used for analyses in the present chapter) in which each student had four waves of data, so that imputation was used to fill in missing values within the four waves actually completed by each cohort of students, but not in the data from waves not completed by the cohort (i.e., imputation for the Year 7 cohort had imputed data for the four waves of data in Years 7 and 8, but not for Years 9 and 10). The analyses were done on five imputed data sets and the results were combined automatically by Mplus using the Rubin (1987; Schafer, 1997) strategy to obtain unbiased parameter estimates, standard errors, and goodness of fit statistics.

Longitudinal models that evaluate development/change of latent means across multiple waves assume strong/scalar invariance in which factor loadings and item intercepts are assumed to be equal across time waves. In a cohort sequential design, these parameters are thus assumed to be invariant across both time waves and cohorts (thus providing a direct test of possible historical/cohort effects). Furthermore, to estimate the growth trajectories from all time waves and cohorts, cohort sequential designs also assume that overlapping latent means (e.g., Waves 3 and 4 for the Year 7 cohort, when these students are in Year 8, with waves 1 and for the Year 8 cohort, when they are also in Year 8; see boxes in Figure 6.1a) are invariant across cohorts. Thus, the model constrains loadings and item intercepts to be invariant over time (strong/scalar invariance), and latent means for overlapping time points to be invariant across time points—providing an additional test of cohort-specific historical effects (see syntax in for Scalar Invariance Model SCl-3d in Appendix 6.5). This is an inherent assumption of cohort sequential designs for models involving latent means, but is rarely tested in applied research.

Goodness of Fit. Historically, applied SEM researchers sought universal "golden rules" to justify objective interpretations of their data, rather than being forced to defend subjective interpretations (Marsh, Hau, & Wen, 2004). Many fit indices have been proposed (e.g., Marsh, Balla, & McDonald, 1988), but there is even less agreement today than in the past as to what constitutes an acceptable fit. Some still treat the indices and recommended cut-offs as golden rules; others argue that fit indices should be discarded altogether; a few argue that we should rely solely on chi-square goodness-of-fit indices; and many (like us) argue that fit indices should be treated as rough guidelines to be interpreted cautiously in combination with other features of the data. Generally, given the known sensitivity of the chi-square test to sample size, to minor deviations from multivariate normality, and to minor misspecifications, applied SEM research generally focuses on indices that are sample-size independent (Hu & Bentler, 1999; Marsh, Balla, & Hau, 1996; Marsh et al., 2004; Marsh et al., 2005), such as the Root Mean Square Error of Approximation (RMSEA), the Tucker-Lewis Index (TLI), and the Comparative Fit Index (CFI). Population values of TLI and CFI vary along a 0-to-1 continuum in which values greater than .90 and .95 typically reflect acceptable and excellent fits to the data, respectively. Values smaller than .08 and .06 for the RMSEA are typically interpreted as acceptable and good model fits respectively.

For the comparison of two nested models, the chi-square difference test can be used, but it suffers from even more problems than the chi-square test for single models, which led to the development of other fit indices (see Marsh, Hau, Balla, & Grayson, 1998). Cheung and Rensvold (2002) and Chen (2007) suggested that if the decrease in fit for the more parsimonious model is less than .01 for incremental fit indices like the CFI, then there is reasonable support for the more parsimonious model. Chen (2007) suggested that when the RMSEA increases by less than .015, there is support for the more constrained model. For indices that incorporate a penalty for lack of parsimony, such as the RMSEA and the TLI, it is also possible for a more restrictive model to result in a better fit than a less restrictive model. However, we emphasize that these cut-off values only constitute rough guidelines.

Results

Basic Cohort-Sequence Model: Four Cohort Groups and Four Waves

We begin with the basic cohort-sequence model depicted in Figure 6.1a and 6.1b.The critical features of this design are that there are four cohorts (year in school groups) and four waves of data for each cohort. In the first two models (Models Con-1A and Con-1B in Table 6.1; see syntax in Appendix 6.2), we establish that the a priori correlated uniquenesses are required and result in a substantial improvement in fit (Table 6.1). For this reason, correlated uniquenesses are included in all subsequent models. Not surprisingly, given that the PBC scale is brief and has good psychometric properties, the goodness of fit for this model is exceptionally good (e.g., TLI = 1.000; Table 6.1). Inspection of the 48 factor loadings (3 items x 4 waves x 4 cohorts) indicates that all factor loadings are substantial, varying between .65 and .87 in a standardized metric. In summary, there is good support for configural invariance.

In Models Met-2A–MET-2C (Table 6.1; see Mplus syntax in Appendix 6.3) we demonstrate good support for the invariance of factor loadings over:

- A. four cohorts (between-group cross-sectional comparisons);
- B. four waves (within-person longitudinal comparisons);
- C. four cohorts and four waves (integrating both within- and between-comparisons; See syntax in Appendix 6.3 along with instructions as to how to alter the syntax to specify Models Met-2A and MET-2B).

Indeed, even constraining factor loadings to be invariant over both the four cohorts and the four waves resulted in almost no decrement of fit (e.g., TLI = .998, Model MET-2C). Hence, there is good support for metric invariance.

In Models Scl-3A—Scl-3C (Table 6.1; also see Mplus syntax in Appendix 6.4) we demonstrate good support for the corresponding tests of invariance of intercepts over four cohorts and four waves. Indeed, even constraining item intercepts to be invariant over both the four cohorts and the four waves resulted in almost no decrement of fit (e.g., TLI = .997). Hence, there is good support for metric invariance.

Model Scl-3D (Table 6.1; also see Mplus syntax in Appendix 6.4) is specific to the cohort-sequence design and critical to the interpretation of the results. In particular, it constrains the latent means to be the same with each of the six pairs of matching means (i.e., the matching means in boxes in Figure 6.1a). Thus, for example, the two estimates of M3 based on different cohorts (Wave 3, Year 7 cohort and Wave 1 Year 8 cohort) are constrained to be equal. The latent mean estimates based on Model Scl-3D are shown in Table 6.2 (with boxes representing the matching means constrained to be equal; Mplus syntax for this model, described in more detail in the results section, is included in Appendix 6.4).

In a supplemental model (Unq-4a) we also tested for the invariance of uniquenesses over cohorts and waves. However, unlike the corresponding tests of the invariance of factor loadings and intercepts, this model resulted in a moderately large decrement in fit (e.g., TLI = .971, a Δ TLI = .026 compared to Model Scl-3C). We note that this model is borderline in terms of traditional guidelines for testing invariance, but also that it might be possible to achieve a more acceptable fit with partial invariance based on ex-post-facto relaxation of invariance constraints on the uniquenesses (e.g., Byrne, Shavelson, & Muthén, 1989). However, instead we chose to reject this model, because the invariance of uniquenesses is not an assumption underlying the evaluation of latent means. However, strict invariance is a necessary condition for the appropriate evaluation of manifest means (i.e., means based scale scores formed from the average or sum of items and not corrected for measurement errors). Indeed, this is an important advantage of the latent-variable models, in that they do not require uniquenesses to be invariant over groups and occasions.

Finally, in Model M4-5A, we test a model in which the latent means are constrained to be equal across all four waves and four group groups—in addition to constraints on factor loadings and intercepts (Model Scl-3C). The fit for Model NGr-5A was very good (e.g., TLI = .995), indicating that latent means are very stable when averaged across gender and athlete groups. Were our primary interest in stability/change over this five-year period, it might be reasonable simply to stop at this point and to conclude that mean levels PBC are stable over this potentially turbulent adolescent period. However, because our primary interest is on stability/change in relation to gender, athlete group, and their interaction, we now extend the analyses to focus on these issues.

Cohort-Sequence Design of Multiple Indicators Multiple Causes (MIMIC) Models

A unique aspect of these data is that 38% of the adolescents are elite athletes selected through a highly competitive process, whereas the remaining 62% are non-athletes of similar ages attending the same school and taking mostly the same classes. From this perspective it is of substantive interest to compare developmental stability/change in PBC over time as a function of gender for these two groups of students.

In the first MIMIC model (Mim-6A, Table 6.1) we simply added the three MIMIC variables to the final cohort-sequence (Scl-3C with invariance of factor loadings and item intercepts, and matching means for overlapping time points), in which the latent factors (PBC) are predicted by the three covariates (gender: boys vs. girls, athlete group: athletes vs. non-athletes; and the gender-by-athlete-group interaction). In this model, no constraints were placed on the effects of the MIMIC variables, which were freely estimated, and the fit was very good (e.g., TLI = .984). In the next model (Mim-6B, Table 6.1; also Mplus syntax in

Appendix 6.7), the effects of the MIMIC variables were constrained to be to be the same within each of the six pairs of matching means (i.e., the matching means in boxes in Figure 6.1a). Thus, for example, the two estimates of gender differences for cell M3, based on different cohorts (e.g., Wave 3, Year 7 cohort and Wave 1 Year 8 cohort) were constrained to be equal. The imposition of this set of 18 constraints (3 MIMIC variables across 6 pairs of cells) resulted in nearly no decrement in fit (e.g., TLI = .988). In the next model (Mim-6C, Table 6.1) we constrained the effects of all the MIMIC variables to be zero (i.e., no differences associated with group or gender over the four waves and four cohorts). The results of this model showed a substantial decrement in fit (e.g., TLI = .960, Δ TLI = .028 compared to Model Mim-6B), indicating that there are gender and group differences. (Mplus syntax for this model, described in more detail in the results section, is included in Appendix 6.7).

An important feature of the MIMIC model is parsimony, especially when compared to the corresponding multi-group approach. Thus, for example, the 2 (athlete group) x 2 (gender) x 4 (cohort) design would require testing over 16 groups in the multiple group approach. In the 16 multigroup design it would be very complicated to disentangle invariance associated with each of the three facets (group, gender, and cohort) and their combinations. Furthermore, in the present investigation, the sample sizes of some of the groups would be so small as to make the analyses dubious. Nevertheless, there were critical drawbacks to the use of the MIMIC model, in that researchers cannot easily test the invariance of factor loadings. However, in the MIMIC model it is possible to test the invariance of items' intercepts by freely estimating the effects of the MIMIC variables on all items' intercepts and constraining their effects on the latent means to be zero. Comparison of this model (Min-6D in Table 1) with the corresponding model, with the effects of the MIMIC variables on the items' intercepts constrained to be zero and their effects on the latent means freely estimated (Mim-6A in Table 1; see syntax in Appendix 6.7), provides a test of intercept invariance across the set of MIMIC variables. If all intercept differences can be explained in terms of a much smaller number of latent means, then there is support for intercept invariance. Results support the invariance of the intercepts, in that Min-6A fits the data very well (e.g., CFI = .989), and that the difference between the two models is smaller than the typically recommended guidelines (ΔCFI ≤ .010; ΔRMSEA ≤ .015). Had there been a lack of invariance of intercepts over the MIMIC variables, it might have been appropriate to evaluate invariance for each of the MIMIC variables separately or, perhaps, to explore the possibility of partial invariance. We note, however, that the post hoc models of partial invariance are typically more defensible when there is a larger number of indicators per factor (there are only three in our example) and intercept invariance holds for all but one (or a small proportion) of a larger number of indicators.

Use of Model Constraint with Orthogonal Polynomial Contrasts to Evaluate Cohort-Sequence and MIMIC Latent Means

Based on the cohort-sequence model of these longitudinal data, we estimated 10 latent means (i.e., 2 means per year for each of the five years covered by this design). In order to evaluate the nature of change over time, we then fitted a latent growth model based on traditional orthogonal polynomial contrasts (e.g., Cohen, West & Aiken, 2014) to estimate polynomial (i.e., linear, quadratic, cubic etc.) components. Orthogonal polynomials have two defining characteristics: they sum to 0 for each component and are mutually independent for pairs of components (Cohen, West & Aiken, 2014). The coefficients used to define the polynomial contrasts (e.g., linear, quadratic, cubic components) will vary depending on the number of means (i.e., M1-M10) but are readily available from most textbooks that discuss contrast coding and can be generated from statistical packages such as R. (see Appendix 6.6 for the coefficients and Mplus syntax used here). Although familiar to most researchers who use contrasts in analyses of manifest models (e.g., polynomial contrasts with ANOVA or multiple regression) using standard statistical packages such as SPSS (where polynomial contrasts are the default for repeated measures ANOVAs), here the polynomial functions are fitted to latent means as part of the same analyses used to estimate the latent factor structure. In combination with grouping variables (gender: male vs. female; athlete group: elite-athlete vs. non-athlete) it is then possible to test interactions between each of the growth curve components, and various combinations of these grouping variables (e.g., does change over time differ for males and females). Although orthogonal polynomial contrast codes are often defined in terms of integer values, it is also possible to normalized them, such that the sum of squared coefficients for each contrast sums to 1.0.

We began with the total effects for each wave, averaged across gender and group. These are presented in the form of a table (Table 6.2) that highlights the cohort-sequence design underpinning these data. In Table 6.3 the same values are presented as a single column of means, along with tests of the orthogonal polynomial contrasts. These results show that there are few or no linear or quadratic effects in the longitudinal data represented by the 10 cells. Indeed, not even the mean squared differences between values of the 10 latent means (MSqDiff in Table 6.3) are statistically significant (see Mplus syntax in Appendix 6.7 for the computation of this value using model constraints). This suggests that PBC is highly stable over this adolescent period—at least when averaged across gender and athlete groups. These results are of course consistent with those based on Model Scl-3D. Again we note that because our primary interest is in stability/change in relation to gender, athlete group, and their interaction, we now look to extend the analyses to focus on these issues.

In Table 6.3, the next three columns represent the MIMIC variables: the main effects of gender and athletic group and gender-by-group interaction. The "grand mean" effects represent the main effects of each MIMIC variable averaged across the 10 cells in our longitudinal design. The results show that there are substantial main effects, due both to gender (higher values for boys) and group (higher values for athletes). However, the gender-by-group interaction is not statistically significant, suggesting that gender differences are similar across the two groups (or that group differences are similar for boys and girls) averaged over time and cohort (see Mplus syntax in Appendix 6.7 for the computation of this value using model constraints).

Next we evaluate developmental stability/change in these main effects; whether gender and group effects are longitudinally consistent across the 10 cells of our cohort-sequence design. In the language of ANOVA we are testing the time-by-gender, time-by-group, and time-by-gender-by-group interactions. Here we evaluate these effects using traditional polynomial contrasts typically used in ANOVAs, but keeping in mind that these contrasts are applied to latent means based on our cohort-sequence design.

For gender there is a significant negative linear effect, but a significantly positive quadratic effect; cubic and quartic components are non-significant. Inspection of the means (Table 6.3) demonstrates that gender differences in favor of boys were smallest at the start of high school (-.257 for M1; Wave 1, cohort Year 7), increase over time, level out, and then became smaller (-.349 for M10; Wave 4, cohort Year 10).

For the athlete group there is a significant negative linear effect: quadratic, cubic, and quartic components are non-significant. Inspection of the 10 cells across our cohort-sequence design suggests that group differences in favour of elite athletes are substantial at the start of high school (-.740 for M1) but decline rapidly, so that group differences are not even statistically significant for the last four cells (Years 10 and 11).

However, there are no significant differences over time for the gender-by-group interaction. The time-by-gender-by-group interactions are non-significant for all polynomial components.

Use of Latent Growth Curve Models to Evaluate Stability/Change Over Time

When researchers have longitudinal data, growth curve models aim to model trait trajectories in a variable or set of variables over time (see Chapter 7 in this book). Typically, the aim is to find a smoothed line through noisy longitudinal data that provides a test of a hypothesized trajectory. Example hypotheses may include the prediction that (a) a variable shows linear improvement (or decline) across the course of a training program; (b) a variable has a U-shaped or inverted U-shaped trajectory as individuals transition from a lower to higher competitive levels in football; or (c) a given training program dramatically lifts performance from the previous trend, prior to introduction of the intervention, and that the improvement is maintained over subsequent time waves. Growth curves can be estimated within a number of frameworks, including structural equation models, multilevel models, and hierarchical Bayes models (Diallo, Morin, & Parker, in press). In many cases, where data are balanced, trends are estimated from complete observed data and the Bayes priors are uninformative (See Chapter 8 in this book), the models provide very similar information. In all cases, latent growth curve (LGC) models provide considerable flexibility in estimating a range of trends, and a means of estimating models with missing data treated via full information maximum likelihood or multiple imputation, and allow for the estimation of growth trends from latent variables with various invariance constraints (see Diallo & Morin, in press; Diallo, Morin, & Parker, in press; Ram & Grimm, 2007; Ram & Gerstorf, 2009).

LGC models decompose variance in the repeated measures of a variable into intercept and slope components reflecting the initial level and growth trajectories (see Duncan, Duncan, & Strycker, 2006; also see chapter 7 in this book). Typically, tests of significance for both the mean and the variance of the intercept and slope components are provided as an indication of the significance of the observed growth component (mean) and of the presence of inter-individual variability on this growth component (variance) present within the sample. Additionally, comparisons of trajectories can be tested via either multigroup models or within a MIMIC framework. Here we apply this LGC approach to our cohort sequence design, as shown in Figure 6.1a (see Brodbeck et al., 2012, for an example; also see Appendix 6.8: Mplus Syntax for Latent Growth Curve Model LGC-7b1)

Coding the Intercept in LGC Models. A potential complication in applications of latent growth modeling is the question of how to code time. Historically, orthogonal polynomial contrasts (like those presented earlier) have been used, but these complicate interpretations in that it is difficult to graph the results in relation to a metric that is common to different growth components. However, in LGC models, estimates of the growth components and tests of their statistical significance are idiosyncratic to the placement of the intercept, which is a function of how time is coded. Thus for example, the intercept (and all but the highest order growth component) will vary according to the point in the temporal sequence (or outside the sequence) coded to be zero (i.e., the intercept). Importantly, however, the different codings result in equivalent models. In particular, coefficients based on models that differ only in terms of the intercept are merely transformations of each other so that they all result in the same graphs (Biesanz, et al., 2004; also see Mehta & West, 2000). For this reason it is critical to actually plot the polynomial function as the basis of interpretation of the results.

In some applications there might be a "logical" coding that is based on the experimental design and on research questions. Thus, for example, in experimental designs where there is an intervention introduced within the temporal sequence, it might be logical to code the intercept to be zero immediately prior to the intervention. In the present investigation we are interested in stability/change over the high school years, so it is reasonable to code the intercept to be at the start of high school (Yr7, Wave 1) or, perhaps, the end of the temporal sequence (Yr10, Wave 4). However, in some applications, which do not involve an intervention (or a critical transition experienced by all participants, such as moving from primary to secondary school), the placement of the intercept might be arbitrary. It is also important to emphasize that typically, interpretation of the growth coefficients (for all but the highest order component) is highly dependent upon the placement of the intercept. For the present purposes we illustrate this by considering three different codings of the data, placing the intercept at the:

- start of the temporal sequence: e.g., M1 (Yr7, Wave 1) = 0; M2 (Yr7, Wave 2) = 1; … M10 (Yr10, Wave 4) = 9; Model LGC-7b1a in Table 4 (see Appendix 6.7: Mplus Syntax for Latent Growth Curve Model LGC-7b1)
- mid-point of the temporal sequence: (e.g., M1 (Yr7, 1) = -4.5;… M4 (Yr7, Wave 4 and Yr8, Wave 2) = -0.5; M5 (Yr8, Wave 3 and Yr9, Wave1) = +0.5; … M10 (Yr10, Wave 4) = +4.5; Model LGCMb2 in Table 4
- end of the temporal sequence: e.g., M1 (Yr7, Wave 1) = -9; M2 (Yr7, Wave 2) = -8; … M10 (Yr10, Wave 4) = 0; Model LGCMb3 in Table 4).

Inspection of the results based on these three intercepts results in highly different growth coefficients (Table 4), even though the three models are equivalent and merely represent transformations of each other. Thus, for example, the intercept for the group contrast is highly significant (in favour of athletes) when the intercept is placed at the start of the time sequence, much smaller but still significant when the intercept is placed in the middle of the time sequence, and not statistically significant when placed at the end of the time sequence. These results demonstrate that the group differences vary systematically in relation to time (A similar pattern of results was demonstrated with the estimated means using the polynomial contrast approach in Table 3, in which group differences started out large in the first wave, gradually declined, and were non-significant for the last waves). Because the growth coefficients are so heavily dependent on the placement of the intercept, it is absolutely essential that interpretations are based on a graph of the LGC results.

Latent Growth Curve (LGC) Results

Preliminary results presented for the model-constraint approach showed significant linear and quadratic effects (but non-significant cubic or quartic components; see tests of orthogonal polynomial contrasts in Table 6.3). Hence, for LGC models, we limited consideration to linear and quadratic growth components. The fit for this model was very good (e.g., RMSEA = .031, CFI = .981, TLI = .978;). As noted earlier, all three codings (intercept at the start, middle, or end of the temporal sequence) necessarily result in the same goodness of fit and graph of the data, but different growth coefficient estimates for all but the highest-order growth component (quadratic in this application). Hence, the graph of the results is essential for interpretation.

When the intercept is placed at the start of the temporal sequence, overall there are small but significant linear (negative) and quadratic (positive) trends. Although there is substantial variation in the intercepts for different participants (residual variance component = .924), these residual variance components are not statistically significant for the linear and quadratic growth components. Whilst boys start off with much higher scores than girls (gender effect = -.273), there is a substantial linear decline in the size of the gender differences over time, as well as a small quadratic component. The athlete/non-athlete group difference is substantial at the start of the time sequence (group effect = .672), but linear and quadratic trends in this difference are not significant.

However, these estimates were systematically different for different codings of time. Thus, for example, compared to the start of the temporal sequence, the gender difference was substantially larger (-.756) at the midpoint and, to a lesser extent, at the end of the sequence (-.451). The athlete/non-athlete differences were even more dramatic, in that the very large difference at the start of high school (.672) were much smaller at the middle of the time sequence (.334), but not even statistically significant at the end of the time sequence (.046, SE = .134).

Summary, Implications, and Further Directions

In this chapter we have outlined a substantive-methodological synergy applying new and evolving statistical methodology to address substantively important issues with theoretical and practical implications. Our primary focus has been on methodological tests of invariance based on a cohort-sequence design (Figure 6.1a,b). Substantively, we have evaluated stability/change in PBC over the potentially turbulent adolescent period as a function of gender, group, and their interaction.

Methodological Implications, Limitations and Further Directions

We have presented a multi-group approach to testing invariance in a cohort sequence. We began by demonstrating good support for the invariance of factor loadings and item intercepts across the four cohort groups and four occasions, spanning five school years. Critically, we then demonstrated support for invariance constraints in latent means across these multiple groups for the overlapping time points (the cells in boxes in Figure 6.1a), resulting in a full model covering five years of high school, even though any one participant only completed data across two years.

Next, we extended the typical cohort-sequence design by integrating it with a MIMIC model of the effects of three covariates (gender, athlete group, and their interaction). The effects of these covariates and their interaction with time were explored by two complementary approaches, one based on the traditional orthogonal polynomial contrasts traditionally applied in ANOVAs (but here based on latent rather than manifest means), the other based on latent growth curve models. Not surprisingly, both approaches resulted in similar interpretations, but each has potentially complementary advantages. In order to illustrate the differences, we present graphs of the results based on both approaches. The LGC graph is smoothed, in that estimated points are only based on the linear and quadratic growth components. In the ANOVA approach, the means are freely estimated so that the graph does not represent smoothed polynomial plots. Nevertheless, the interpretation of both graphs is similar, as will typically be the case.

A potentially important advantage of the orthogonal polynomial approach is that it results in reasonably independent estimates of each polynomial component so that the convergence problems that plague LGC models are not so likely to occur—particularly for models estimating more than 2 or 3 growth components. Here, for example, we were able to show, with the orthogonal polynomial approach, that cubic and quartic growth components—and their interactions with the three covariates—were all nonsignificant.

This provided a good rationale for limiting LCG models to linear and quadratic growth components. However, the estimated trajectories based on LGC models that are so critical in the interpretation of the results, are not easily derived from estimates based on the orthogonal model. Also, there was considerable recent development in the application of LGC models, which is likely to provide new and evolving advantages to this approach (e.g., Biesanz, et al., 2004; Bollen & Curran, 2006; Morin, Maïano, et al., 2011, 2013). Although a detailed juxtaposition of the two approaches is beyond the scope of this chapter, we suggest that they should be viewed as complementary rather than alternative approaches to describing growth trajectories and how they vary as a function of other variables.

In this chapter we relied heavily on MIMIC models that implicitly assume that factor loadings and intercepts are invariant over these grouping variables as well as over cohort and occasion. Although it is possible to test the invariance of intercepts with the MIMIC model, the invariance of factor loadings is not readily testable. An alternative in the present investigation might have been to test invariance over 16 groups (2 gender x 2 athlete groups x 4 cohorts) and four occasions, but sample sizes in many of the groups were too small for this to be reasonable. Furthermore, the multiple-group approach to testing the invariance of estimates based on the MIMIC variable is only viable in applications such as the present one, in which the covariates are categorical, with relatively few discrete categories. For reasonably continuous covariates, Marsh, Nagengast, and Morin (2013) proposed a hybrid MIMIC model in which the same continuous variable (e.g., age) is included as a multi-group variable that allows more rigorous tests of invariance, and a continuous variable to determine how much explained variance is lost by categorizing a continuous model. Interestingly, Mehta and West (2000) had previously proposed a similar solution to evaluating age effects in a cohort-sequence design. We could have included age as a covariate, in addition to the implicit effect of age represented by the age-cohort. Although this was beyond the scope of the present analysis, these possibilities offer potentially important areas for further development.

In this chapter the factor structure was very simple, based on only three items with strong psychometric properties, and the fit of invariance models was always very good. However, this will not always be the case, particularly for models based on a large number of items for each construct and, perhaps, multiple latent constructs. When misfit is due to lack of invariance, it might be possible to fit partial invariance models. However, these should be used with appropriate caution, particularly when based on ex-post-facto model modifications. Nevertheless, when there is substantial misfit, even in the configural model with no invariance constraints, the use of growth curve models is problematic. In the past, applied researchers have sometimes resorted to typically dubious practices such as the use of item parcels, which camouflages misfit rather than resolving the problem (see discussion by Marsh, Lüdtke, Nagengast, Morin, & Von Davier 2013). Nevertheless, there are new and evolving approaches that do not require applied researchers to evaluate potentially overly restrictive models that preclude items loading on more than one factor. These include exploratory structural equation models that integrate most of the advantages of CFA with the flexibility of EFA structures and evolving Bayesian approaches (e.g., Marsh, Morin, Parker & Kaur, 2014; Morin, Marsh & Nagengast, 2013; see also Chapter 8 in this book). Although these approaches typically have not been applied to growth modeling, their application provides an important direction for further research.

References

Bell, R. Q. (1953). Convergence: An accelerated longitudinal approach. *Child Development*, *24*, 145–152.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological methods*, *9*(1), 30.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective.* New Jersey: Wiley.

Brodbeck, J., Bachmann, M. S., Croudace, T. J., & Brown, A. (2012). Comparing Growth Trajectories of Risk Behaviors From Late Adolescence Through Young Adulthood: An Accelerated Design. *Developmental Psychology*, *49*, 1732–1738.

Brown, A, Croudace, T. & Heron, J. (2011). Longitudinal data modelling: Sequential cohort design [Course handout]. Retrieved from http://www.restore.ac.uk/appliedpsychometrics/Tutorial.materials/Intro.to.longitudinal.modelling/

Byrne, B. M. (1996). *Measuring self-concept across the life span: Issues and instrumentation.* Washington, DC: American Psychological Association.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin, 105, 456–466.

Chen, F. F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of–fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255. doi: 10.1207/S15328007SEM0902_5

Diallo, T. M. O., & Morin, A. J. S. (in press). Power of Latent Growth Curve Models to Detect Piecewise Linear Trajectories. *Structural Equation Modeling.*

Diallo, T. M. O., Morin, A. J. S., & Parker, P. D. (in press). Statistical Power of Latent Growth Curve Models to Detect Quadratic Growth. *Behavioural Research Methods*.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications.* New York: Lawrence Erlbaum.

Dusek, J. B., & Flaherty, J. F. (1981). The development of self–concept during adolescent years. *Monographs of the Society for Research in Child Development*, *46*(4, Serial No. 191).

Eccles, J., Wigfield, A., Harold, R. D., & Blumfeld, P. (1993). Age and gender differences in children's self– and task–perceptions during elementary school. *Child Development*, *64*, 830–847.

Enders, C. (2010). *Applied missing data analysis.* New York: Guilford Press.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, *116*, 429–456.

Fleishman, F. A. (1964). *The structure and measurement of physical fitness.* Englewood. Cliffs, NJ: Prentice–Hall.

Graham, J. W. (2012). *Missing data: Analysis and design.* New York: Springer.

Hu, L., & Bentler, P. M. (1999). Cut–off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Jöreskog, 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., 1993. *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Chicago, IL: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago, Illinois: SPSS.

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non–arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling*, *13*, 59–72.

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self–concept: Preadolescence to early-adulthood. *Journal of Educational Psychology*, 81, 417–430.

Marsh, H. W. (1990). A multidimensional, hierarchical self–concept: Theoretical and empirical justification. *Educational Psychology Review*, *2*, 77–172.

Marsh, H.W. (1991). The failure of high ability high schools to deliver academic benefits: The importance of academic self–concept and educational aspirations. *American Educational Research Journal*, *28*, 445–480.

Marsh, H. W. (l993). The multidimensional structure of physical fitness: Invariance over gender and age. *Research Quarterly for Exercise and Sport*, *64*, 256–273.

Marsh, H. W. (1994). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. *Structural Equation Modeling, 1*, 5–34.

Marsh, H. W. (1996a). Construct validity of Physical Self–Description Questionnaire responses: Relations to external criteria. *Journal of Sport & Exercise Psychology*, *18*, 111–131.

Marsh, H. W. (1996b). Physical Self Description Questionnaire: Stability and discriminant validity. *Research Quarterly for Exercise and Sport*, *67*, 249–264.

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. (1998). Age and gender effects in physical self–concepts for adolescent elite–athletes and non–athletes: A multi-cohort-multi–occasion design. *Journal of Sport & Exercise Psychology*, *20*(3), 237–259.

Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (3rd ed., pp. 774–798). New York: Wiley.

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., & Nagengast, B. (2014). The Big-Fish-Little-Pond effect in mathematics: A cross-cultural comparison of U.S. and Saudi Arabian TIMSS responses. Journal of Cross-Cultural Psychology, 45(5), 777-804. doi:10.1177/0022022113519858

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J., Abdelfattah, F., & Nagengast, B. (2014). The internal/external frame of reference model of self-concept and achievement relations age-cohort and cross-cultural differences. American Educational Research Journal. (on-line version)

Marsh, H. W., Balla, J., & McDonald, R. P. (1988). Goodness of fit in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

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: Issues and techniques* (pp. 315–353). Hillsdale, NJ: Erlbaum.

Marsh, H. W., Chanal, J. P., Sarrazin, P. G. & Bois, J. E. (2006). Self–Belief does make a difference: A reciprocal effects model of the causal ordering of physical self–concept and gymnastics performance. *Journal of Sport Sciences*, *24*, 101–111.

Marsh, H. W., & Cheng, J. H. S. (2012). Physical self–concept. In G. Tenenbaum, R. C. Eklund, & A. Kamata (Eds.), *Measurement in sport and exercise psychology* (pp. 215–226). Champaign, IL: Human Kinetics.

Marsh, H. W., Craven, R. G., & Debus, R. (1991). Self–concepts of young children aged 5 to 8: Their measurement and multidimensional structure. *Journal of Educational Psychology*, *83*, 377–392.

Marsh, H. W., Craven, R. G., & Debus, R. (1998). Structure, stability, and development of young children's self–concepts: A multicohort–multioccasion study. *Child Development*, *69*(4), 1030–1053.

Marsh, H. W., Craven, R. G., & Debus, R. (1999). Separation of competency and affect components of multiple dimensions of academic self–concept: A developmental perspective. *Merrill–Palmer Quarterly–Journal of Developmental Psychology, 45*, 567–601.

Marsh, H. W., & Grayson, D. (1994). Longitudinal confirmatory factor analysis: Common, time–specific, item–specific, and residual–error components of variance. *Structural Equation Modeling, 1,* 116–146.

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. (2007). Applications of latent–variable models in educational psychology: The need for methodological–substantive synergies. *Contemporary Educational Psychology*, *32*, 151–170.

Marsh, H. W., Hau, K-T., Balla, J. R., & Grayson, D. (1998) Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33,* 181–220.

Marsh, H. W., Hau, K–T. & Grayson, D. (2005). Goodness of Fit Evaluation in Structural Equation Modeling. In A. Maydeu–Olivares & J. McArdle (Eds.), *Contemporary Psychometrics. A Festschrift for Roderick P. McDonald* (pp. 275–340). Mahwah 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.

Marsh, H. W., Kuyper, H., Morin, A. J. S., Parker, P. D. & Seaton, M. (2014). Big-Fish-Little-Pond Social Comparison and Local Dominance Effects: Integrating New Statistical Models, Methodology, Design, Theory and Substantive Implications. Learning and Instruction, 33, 50-66 http://dx.doi.org/10.1016/j.learninstruc.2014.04.0020959-4752/

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S., & Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): New approaches to key substantive issues in motivation and engagement. Journal of Psychoeducational Assessment, 29(4), 322-346. doi:10.1177/0734282911406657

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U. & Nagengast, B. (2010). A new look at the big–five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*, 471–491. doi: 10.1037/a0019227

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & Von Davier, M. (2013). Why Item Parcels Are (Almost) Never Appropriate: Two Wrongs Do Not Make a Right—Camouflaging Misspecification With Item Parcels in CFA Models. *Psychological Methods. Advance online publication*. doi:10.1037/a0032773

Marsh, H. W., Martin, A. J. & Jackson, S. (2010). Introducing a short version of the Physical Self Description Questionnaire: New strategies, short–form evaluative criteria, and applications of factor analyses. *Journal of Sport & Exercise Psychology*, *32*, 438–482.

Marsh, H. W., Morin, A. J. S., Parker, P., & Kaur, G. (2014). Exploratory Structural Equation Modeling: An Integration of the best Features of Exploratory and Confirmatory Factor Analysis. *Annual Review of Clinical Psychology*, *10*. Online pre–publication (ISI 2011: 12.422)

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16*(3), 439–476. doi:10.1080/10705510903008220

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of big–five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*, *49*(6), 1194–1218. doi:10.1037/a0026913

Marsh, H. W., Perry, C., Horsely, C., & Roche, L. (1995). Multidimensional self–concepts of elite athletes: How do they differ from the general population? *Journal of Sport & Exercise Psychology*, *17*, 70–83.

Marsh, H. W., Richards, G.E., Johnson, S., Roche, L., & Tremayne, P. (1994). Physical Self Description Questionnaire: Psychometric properties and a multitrait–multimethod analysis of relations to existing instruments. *Sport and Exercise Psychology*, *16*, 270–305.

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008).

The big–fish–little–pond–effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, *20*, 319–350. doi:10.1007/s10648–008–9075–6

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, *5*(1), 23. doi: I0.1037//10S2–989X.5.I.23

Meredith, W. (1964). Rotation to achieve factorial invariance. *Psychometrika*, *29*, 187–206.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.

Millsap RE. 2011. *Statistical Approaches to Measurement Invariance*. New York: Routledge

Morin, A. J., Maïano, C., Marsh, H. W., Janosz, M., & Nagengast, B. (2011). The Longitudinal interplay of adolescents' self–esteem and body image: A conditional autoregressive latent trajectory analysis. *Multivariate Behavioral Research*, *46*, 157–201.

Morin, A. J. S., Maïano, C., Marsh, H. W., Nagengast, B., & Janosz, M. (2013). School life and adolescents' self–esteem trajectories. *Child Development*, *84*, 1967–1988.

Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Exploratory Structural Equation Modeling: An Introduction. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (2nd ed.), pp. 395–436. Greewich, Connecticut: IAP.

Nesselroade, J. R., & Baltes, P. B. (1979). *Longitudinal research in the study of behavior and development.* San Diego, CA: Academic Press.

Parker, P. D., Marsh, H. W., Seaton, M., and Van Zanden, B (in press). If one goes up the other must come down: Examining ipsative relationships between mathematics and English self-concept trajectories across high school. *British Journal of Educational Psychology*.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Erlbaum.

Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. Psychology & Aging, 24, 778-791.

Ram, N., & Grimm, K. J. (2007). Using simple and complex growth models to articulate developmental change: Matching theory to method. *International Journal of Behavioral Development*, *31*, 303–316.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman and Hall.

Schutz, R. W., & Gessaroli, M. E. (1993). Use, misuse and disuse of psychometrics in sport psychology research. In R. N. Singer, M. Murphy, & L. K. Tennant (Eds.), *Handbook of research on sport psychology* (pp. 901–917). New York: Macmillan.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Validation of construct interpretations. *Review of Educational Research*, *46*, 407–441.

Tabachnick, B. G. & Fidell, L. S. (1996). *Using multivariate statistics* (3rd Ed.). New York: Harper Collins.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–69.

Wylie, R. C. (1979). The self-concept (vol. 2). Lincoln: University of Nebraska Press. Wylie, R.C. (1989). *Measures of self-concept.* Lincoln: University of Nebraska Press.

	Year 7		Year 8		Year 9		Year 10		Year 11	
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Year 7 Cohort 1	W1	W2	W3	W4						
Year 8 Cohort 2			W1	W2	W3	W4				
Year 9 Cohort 3					W1	W2	W3	W4		
Year 10 Cohort 4							W1	W2	W3	W4

Figure 6.1A. Cohort sequential design with four cohorts and four waves (W1-W4) for each cohort. Light Grey squares = Collected data. White square = Missing by design. M1–M10 are 10 latent means that span the five-year period. Estimates of M3–M8 are each based on results from two cohorts (i.e., M1 is based on wave 3 from year 7 cohort 1 and wave 1 from year 8 cohort 2), whilst those for M1–M2 and M9–M10 are based on a single cohort.

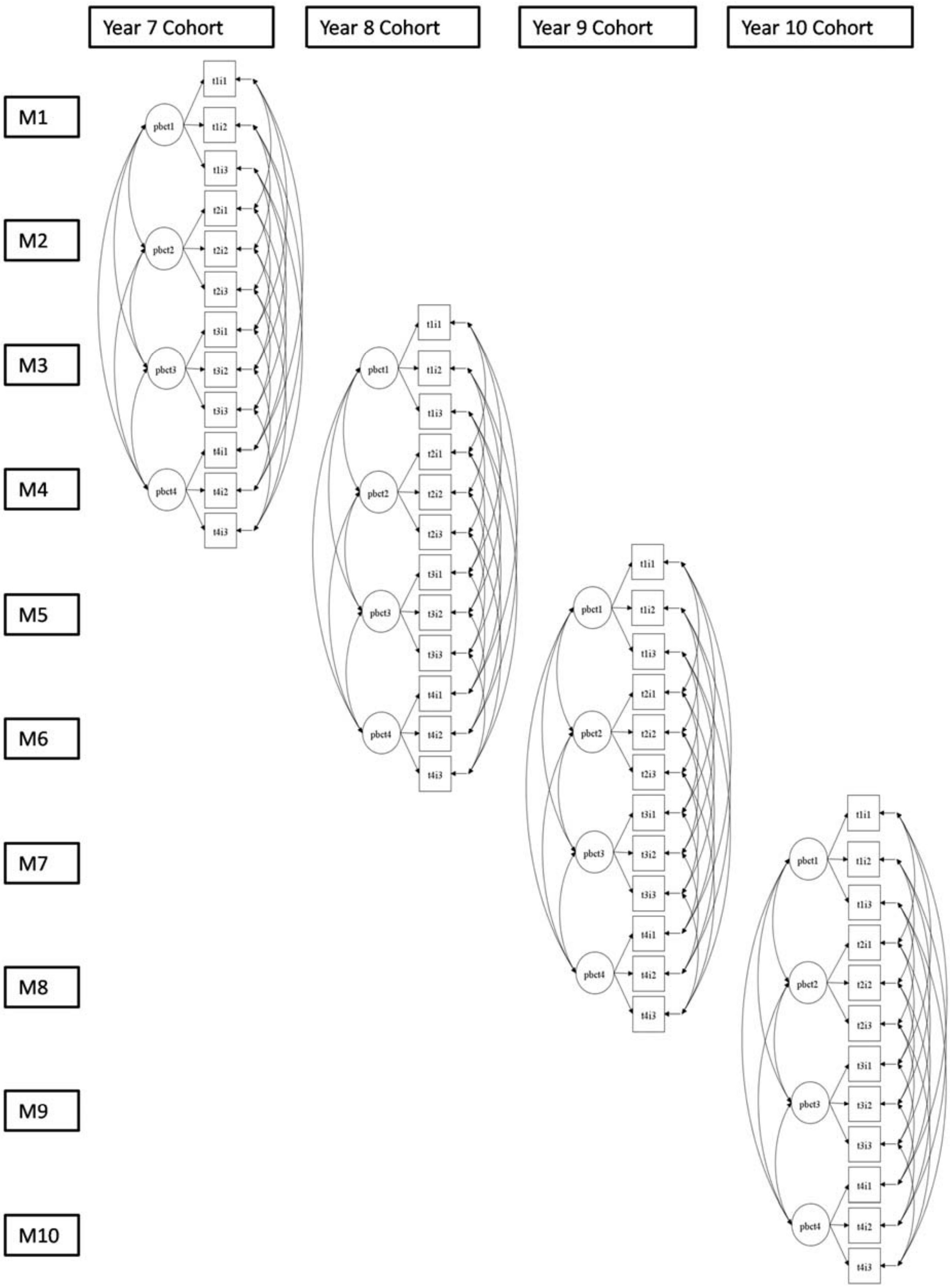

Figure 6.1B. Cohort sequential design with four cohorts and four waves of data for each cohort. M1–M10 represent the 10 latent means that span the five-year period. *M*s in boxes are matching means based on two different cohorts. White square = Missing by design. Estimates of M3–M8 are each based on results from two cohorts, whilst those for M1–M2 and M9–M10 are based on a single cohort.

Figure 6.2. Graphs of results for
(A) Latent Growth Curve Models based on Model M4-7b1 (see Tables 1 and 4)
(B) Observed Means based on model of orthogonal polynomial contrasts Model M4-6b (Tables 1 and 3). Time waves 0-9 represent waves 1 to 10 (see Figure 6.1a), spanning the five-year period (with two waves per year) from the start of high school (Year 7) to near the end of Year 11. The dependent variable is body composition, in which low scores reflect an overly fat body.

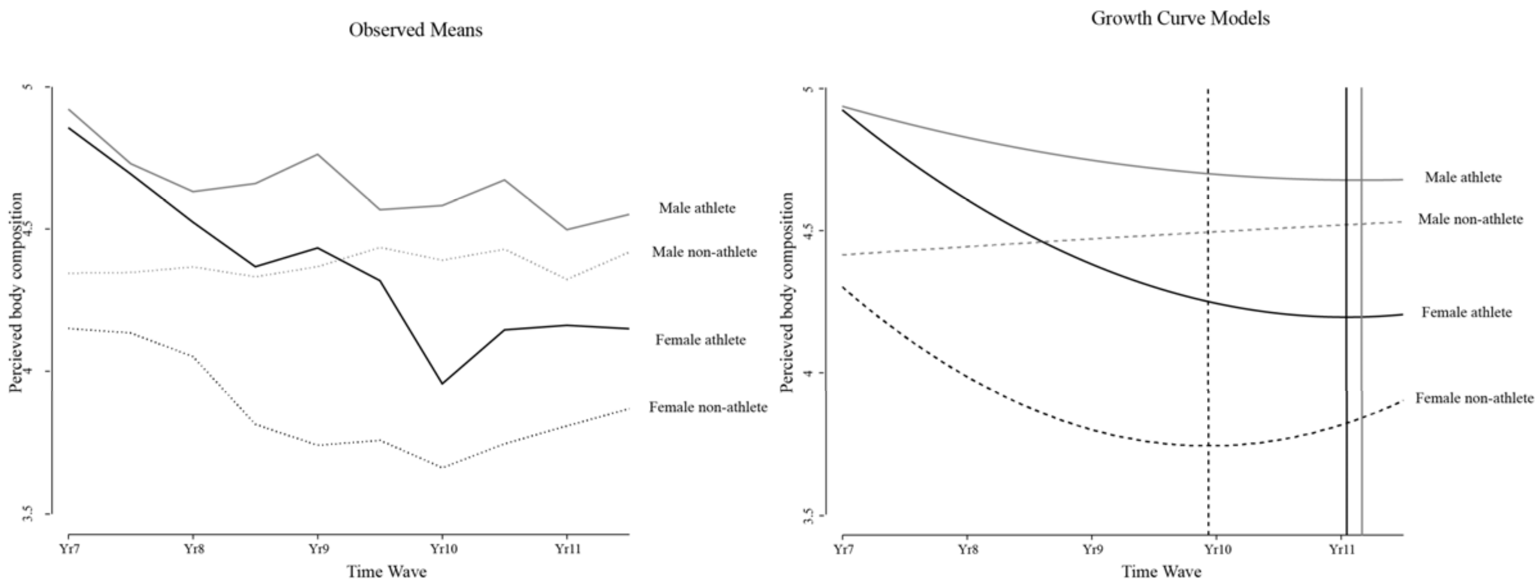

Table 6.1

Goodness of Fit for Alternative Cohort-Sequence Models

Model	ChiSq	DF	RMSEA	CFI	TLI	Description
Comparison of Models with No Invariance (Configural)						
Con-1a	291	192	.040	.979	.971	No Invariance, No Correlated Uniquenesses.
Con-1b[a]	120	120	.008	.999	1.000	Con-1a With Correlated Uniquenesses.
Comparison of Models with Cohort Invariance Of Factor Loadings (Metric)						
Met-2a	138	136	.010	.999	.999	Con-1b With Fls Invar Over Cohorts
Met-2b	147	144	.012	.998	.999	Con-1b With Fl Invar Over Waves
Met-2c[b]	155	150	.013	.998	.998	Con-1b With Fl Invar Over Waves And Cohorts
Comparison of Models with Cohort Invariance Factor Loadings & Intercepts (Scalar)						
Scl-3a	177	174	.011	.998	.999	Met-2c With Intercepts (In) Invar Over Cohorts
Scl-3b[c]	180	174	.012	.998	.998	Met-2c With Int Invar Over Waves
Scl-3c	188	180	.013	.997	.997	Met-2c With Int Invar Over Waves And Cohorts
Scl-3d[d]	197	186	.014	.997	.997	Met-3c With Int Invar Over Matching Latent Means
Comparison of Models with Cohort Invariance Factor Loadings, Intercepts Uniquenesses						
Unq-4a[e]	225	342	.040	.975	.971	Met-3c With Uniq invar Over Waves& Cohorts
Comparison of Models with Cohort Invariance Factor Loadings, Intercepts No Growth						
Ngr-5a	210	195	.015	.996	.995	Met-3c With Means = 0
Comparison Of Mimic Models Of Differences Associated With Gender, Athlete Group And Their Interaction						
Mim-6a	343	282	.026	.989	.984	Met-3c With Mimic=Free
Mim-6b[f]	368	300	.026	.988	.983	Met-3c With Mimic Inv Over Matching Cells
Mim-6c	548	330	.046	.960	.951	Met-3c With Mimic = 0
Mim-6d	199	186	.015	.997	.994	Met-3c Mimic Mns = 0 Mimic Intercepts=Free
Latent Growth Models With Mimic Variables						
LGC-7a	503	353	.036	.973	.969	Met-3c + Mns=0, Latent Growth
LGC-7b1[g]	448	346	.031	.981	.978	LGC-7a + Quadratic Growth, Intercept = M1

LGC-7b2	448	346	.031	.981	.978	LGC-7b1 + With Intercept = Zero Centered

Note. ChiSq = chi-square; df = degrees of freedom ratio; CFI = Comparative fit index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation. CUs = a priori correlated uniquenesses based on the negatively worded items. FL= factor loadings. IN = intercepts.

[a] See syntax in Appendix 6.2: Mplus Syntax for Configural Model Con-1b and how to specify correlated uniqueness

[b] See syntax in Appendix 6.3: Mplus Syntax for Metric Invariance Model Met-2c and how to specify factor loading invariance over within-group constructs (waves) and between-group constructs (cohort group);

[c] See syntax Appendix 6.4: Mplus Syntax for Scalar Invariance Model Scl-3b; Specification of intercept invariance over waves (a within-person construct)

[d] See syntax Appendix 6.5: Mplus Syntax for Scalar Invariance Model Scl-3d.

[e] See syntax Appendix 6.6: Mplus Syntax for Scalar Invariance Model unq-4a.

[f] See syntax Appendix 6.7: Mplus Syntax for MIMIC Model with Orthogonal Polynomial Growth Components Model Mim-6b

[g] See syntax Appendix 6.8: Mplus Syntax for Model LGC-7b1 (see Tables 6.1 and 6.4 in main text). Latent growth curve model with MIMIC variables.

Table 6.2

Latent Factor Means (SEs in parentheses) for Model M4-6b (Table 6.1): Complete Invariance of Factor Loadings and Intercepts over Time and Cohort, Invariance of Matching Latent Means over Cohort (shaded)

	Cohort 7	Cohort 8	Cohort 9	Cohort 10
M1	W1 4.602(.125)			
M2	W2 4.448(.125)			
M3	W3 4.477(.099)	W1 4.477(.099)		
M4	W4 4.422(.101)	W2 4.422(.101)		
M5		W3 4.433(.094)	W1 4.433(.094)	
M6		W4 4.453(.103)	W2 4.453(.103)	
M7			W3 4.462(.089)	W1 4.462(.089)
M8			W4 4.395(.093)	W2 4.395(.093)
M9				W3 4.395(.115)
M10				W4 4.485(.111)

Note. Cohort is the school year group for the first data collection. Wave 1–Wave 4 (W1-W4) are the four waves within each cohort. M1-M10 are the estimated latent means (matching means within boxes are constrained to be equal over cohort). Standard errors (SEs) in parentheses provide a test of significance for each estimate (if the ratio of the estimate/SE is greater than 1.96, the difference is significant at $p < .05$).

Table 6.3

Development as a Function of Athlete/Non-Athlete Group and Gender: Multi-Group (4 cohorts) Cohort-Sequence Analysis with Gender, Group and their Interaction as MIMIC Variables with Orthogonal Contrasts of Time Evaluated with Orthogonal (Model Constraint) Contrasts

	Main Effects			Interaction
Group	Total	Gender (M-F)	Group(E-N)	Gender-by-group
M1	4.602(0.125)	-0.257(0.142)	0.740(0.128)	-0.055(0.264)
M2	4.448(0.125)	-0.563(0.148)	0.491(0.137)	0.274(0.305)
M3	4.477(0.099)	-0.462(0.102)	0.351(0.099)	-0.055(0.206)
M4	4.422(0.101)	-0.619(0.104)	0.461(0.101)	0.218(0.203)
M5	4.433(0.094)	-0.701(0.099)	0.340(0.096)	0.184(0.201)
M6	4.453(0.103)	-0.731(0.105)	0.293(0.099)	0.292(0.211)
M7	4.462(0.089)	-0.752(0.105)	0.146(0.105)	0.009(0.211)
M8	4.395(0.093)	-0.753(0.120)	0.099(0.112)	-0.148(0.225)
M9	4.395(0.115)	-0.569(0.167)	0.024(0.158)	0.179(0.288)
M10	4.485(0.111)	-0.349(0.144)	0.010(0.155)	-0.018(0.336)
Summary				
Model Contrast (orthogonal polynomial contrasts)				
Grand Mean	4.457(0.090)	-0.576(0.064)	0.296(0.057)	0.088(0.130)
Linear	0.916(1.205)	-2.759(3.330)	-12.092(3.061)	-1.309(6.297)
Quadratic	1.143(0.886)	5.157(1.451)	0.729(1.430)	-1.910(2.949)

Cubic	-2.798(5.640)	10.864(9.541)	-4.948(10.176)	11.292(19.448)
Quartic	4.811(3.124)	4.770(5.411)	7.739(6.504)	1.414(11.789)
MSqDiff	0.003(0.002)	0.029(0.013)	0.050(0.023)	0.026(0.023)

Note. M1–M10 are the estimated latent means (see Figure 6.1a). Standard errors (SEs) in parentheses provide a test of significance for each estimate (if the ratio of the estimate/SE is greater than 1.96 the difference is significant at $p < .05$). For each of the covariates (MIMIC variables: female, elite, and their interaction), the corresponding tests evaluate the grand mean and the nature of growth. The Grand Means are the mean of M1–M10 and provide an overall test of each covariate. MSqDiff is the Mean Squared Difference among M1–M10; this provides a test of no growth (i.e., no significant differences between M1–M10). Polynomial components test the nature of growth. Results are based on Model_Mim-6b in Table 6.1 (also see Appendix 6.7 for Mplus syntax).

Table 6.4

Adolescent Development as a Function of Athlete Group (Elite vs. Non-Elite Athletes) and Gender (Male vs. Female): Multi-Group (4 Cohorts) Cohort-Sequence Analysis with Gender, Group and their Interaction as MIMIC Variables

		Main Effects		Interaction	Residual	Covariances	
Group	Total	Gender (M-F)	Group(E-N)	GenderxGroup	Variances	Intercept	Linear
Latent Growth Modeling (intercept at first time point)							
Intercept	4.889(0.076)	-0.273(0.133)	0.672(0.125)	-0.040(0.259)	0.924(0.189)		
Linear	-0.084(0.034)	-0.199(0.055)	-0.081(0.053)	0.103(0.110)	0.037(0.059)	0.065(0.111)	
Quadratic	0.007(0.003)	0.020(0.006)	0.001(0.006)	-0.011(0.012)	0.000(0.001)	-0.032(0.020)	.000(0.008)
Latent Growth Modeling (intercept at mid-point)							
Intercept	4.656(0.038)	-0.756(0.133)	0.334(0.073)	0.204(0.152)	1.021(0.079)		
Linear	-0.020(0.010)	-0.020(0.021)	-0.070(0.019)	0.006(0.041)	0.044(0.020)	-0.034(0.015)	
Quadratic	0.007(0.003)	0.020(0.006)	0.001(0.006)	-0.011(0.012)	0.000(0.001)	-0.030(0.007)	.001(0.001)
Latent Growth Modeling (intercept at last time point)							
Intercept	4.710(0.080)	-0.451(0.149)	0.046(0.134)	0.011(0.310)	0.489(0.206)		
Linear	-0.044(0.034)	-0.159(0.058)	0.058(0.057)	0.091(0.121)	0.055(0.067)	0.072(0.135)	
Quadratic	0.007(0.003)	0.020(0.006)	0.001(0.006)	-0.011(0.012)	0.000(0.001)	-0.027(0.022)	-0.001(0.008)

Note. Results based on three equivalent latent growth models (intercept, linear and quadratic components) that differ in terms of the placement of the intercept. For each of the covariates (MIMIC variables: female, elite, and their interaction), the corresponding tests evaluate the effect at the intercept and the linear and quadratic trend components (i.e., the extent to which the effect of the covariate varies with time). Standard errors (SEs) in parentheses provide a test of significance for each estimate (see Appendix 6.8: Mplus Syntax for Latent Growth Curve Model LGC-7b1).

Appendices

Appendix 6.1: Taxonomy of multigroup tests of invariance

Appendix 6.2: Mplus Syntax for Configural Model Con-1b and how to specify correlated uniqueness (see Table 6.1 in main text)

Appendix 6.3: Mplus Syntax for Metric Invariance Model Met-2c and how to specify factor loading invariance over within-group constructs (waves) and between-group constructs (cohort group); (also see Table 6.1 in main text)

Appendix 6.4: Mplus Syntax for Scalar Invariance Model Scl-3b (see Table 1 in main text). Specification of intercept invariance over waves (a within-person construct)

Appendix 6.5: Mplus Syntax for Scalar Invariance Model Scl-3d (see Table 1 in main text).

Appendix 6.6: Mplus Syntax for Uniqueness Invariance Model Unq-4a (see Table 1 in main text).

Appendix 6.7: Mplus Syntax for MIMIC Model with Orthogonal Polynomial Growth Components Model Mim-6b (see Table 1 in main text)

Appendix 6.8: Mplus Syntax for Model LGC-7b1 (see Tables 6.1 and 6.4 in main text). Latent growth curve model with MIMIC variables. (see Table 1 in main text)

Appendix 6.1

Taxonomy of multigroup tests of invariance

Model	Parameters constrained to be invariant
Model 1	None (configural invariance)
Model 2	Factor loadings (FL) [1] (Metric; weak measurement invariance)
Model 3	FL uniquenesses (uniq) [1, 2]
Model 4	FL, factor variance-covariances (FVCV) [1, 2]
Model 5	FL, intercepts (inter) [1, 2] (Scalar; strong factorial/measurement invariance)
Model 6	FL, uniq, FVCV [1, 2, 3, 4]
Model 7	FL, uniq, inter [1, 2, 3, 5] (strict factorial/measurement invariance)
Model 8	FL, FVCV, inter [1, 2, 4, 5]
Model 9	FL, uniq, FVCV, inter [1–8]
Model 10	FL, inter, factor means (FMn) [1, 2, 5] (latent mean invariance)
Model 11	FL, uniq, inter, FMn [1, 2, 3, 5, 7, 10] (manifest mean invariance)
Model 12	FL, FVCV, inter, FMn [1, 2, 4, 5, 6, 8, 10]
Model 13	FL, uniq, FVCV, inter, FMn [1–12] (complete factorial invariance)

Note. Bracketed values represent nesting relations in which the estimated parameters of the less general model are a subset of the parameters estimated in the more general model under which it is nested. All models are nested under Model 1 (with no invariance constraints), whereas Model 13 (complete invariance) is nested under all other models.

Appendix 6.2: Mplus Syntax for Configural Model Con-1b (see Table 6.1 in main text)

```
TITLE: Model Con-1Bc (see Table 1)
        DATA:
            FILE = wft1234ZBFATIMPlist.DAT;
            TYPE = IMPUTATION;
        VARIABLE:
        NAMES =
               t1bf1 t1I1 t1I2 t1I3 t1bf5 t1bf6
               t2bf1 t2I1 t2I2 t2I3 t2bf5 t2bf6
               t3bf1 t3I1 t3I2 t3I3 t3bf5 t3bf6
               t4bf1 t4I1 t4I2 t4I3 t4bf5 t4bf6
               elite year sex AGE;
        MISSING ARE ALL (-9);
        usevariables are
          t1I1 t1I2 t1I3
          t2I1 t2I2 t2I3
          t3I1 t3I2 t3I3
          t4I1 t4I2 t4I3
          ;
        USEOBSERVATIONS YEAR NE 6;
    GROUPING IS YEAR (7=yR7,8=Yr8,9=Yr9,10=Yr10);
         ANALYSIS: ESTIMATOR = MLR;
  MODEL:
  !BODY COMPOSITION Factor Loadings;
        PBCT1 by T1I1-T1I3;
        PBCT2 by T2I1-T2I3;
        PBCT3 by T3I1-T3I3;
        PBCT4 by T4I1-T4I3;
 !Latent means default to zero
        [PBCT1@0];
        [PBCT2@0];
        [PBCT3@0];
        [PBCT4@0];
 !Latent variances--free;
   !Latent variances--free;
       PBCT1; PBCT2 ; PBCT3 ; PBCT4 ;
 ! Correlataed residuals free
     T2I1-T2I3 pwith T1I1-T1I3 ;
     T3I1-T3I3 pwith T1I1-T1I3 ;
     T4I1-T4I3 pwith T1I1-T1I3 ;
     T3I1-T3I3 pwith T2I1-T2I3 ;
     T4I1-T4I3 pwith T2I1-T2I3 ;
     T4I1-T4I3 pwith T3I1-T3I3 ;
 ! INTERCEPTS free
     [T1I1-T1I3];
     [T2I1-T2I3];
     [T3I1-T3I3];
     [T4I1-T4I3];
 MODEL YR7:
  !BODY COMPOSITION Factor Loadings;
        PBCT1 by T1I1@1 T1I2-T1I3 ;
        PBCT2 by T2I1@1 T2I2-T2I3 ;
        PBCT3 by T3I1@1 T3I2-T3I3 ;
        PBCT4 by T4I1@1 T4I2-T4I3 ;
            !Latent variances--free;
              PBCT1-PBCT4;
  !BODY COMPOSITION INTERCEPTS
               [T1I1-T1I3] ;
               [T2I1-T2I3] ;
               [T3I1-T3I3] ;
               [T4I1-T4I3] ;
MODEL YR8:
  !BODY COMPOSITION Factor Loadings;
        PBCT1 by T1I1@1 T1I2-T1I3 ;
        PBCT2 by T2I1@1 T2I2-T2I3 ;
        PBCT3 by T3I1@1 T3I2-T3I3 ;
        PBCT4 by T4I1@1 T4I2-T4I3 ;
  !BODY COMPOSITION INTERCEPTS
               [T1I1-T1I3] ;
               [T2I1-T2I3] ;
               [T3I1-T3I3] ;
               [T4I1-T4I3] ;
MODEL YR9:
  !BODY COMPOSITION Factor Loadings;
        PBCT1 by T1I1@1 T1I2-T1I3 ;
        PBCT2 by T2I1@1 T2I2-T2I3 ;
```

```
        PBCT3 by T3I1@1 T3I2-T3I3 ;
        PBCT4 by T4I1@1 T4I2-T4I3 ;
            !Latent variances--free;
              PBCT1-PBCT4;
  !BODY COMPOSITION INTERCEPTS
              [T1I1-T1I3] ;
              [T2I1-T2I3] ;
              [T3I1-T3I3] ;
              [T4I1-T4I3] ;
MODEL YR10:
  !BODY COMPOSITION Factor Loadings;
        PBCT1 by T1I1@1 T1I2-T1I3 ;
        PBCT2 by T2I1@1 T2I2-T2I3 ;
        PBCT3 by T3I1@1 T3I2-T3I3 ;
        PBCT4 by T4I1@1 T4I2-T4I3 ;
            !Latent variances--free;
              PBCT1-PBCT4;
  !BODY COMPOSITION INTERCEPTS
              [T1I1-T1I3] ;
              [T2I1-T2I3] ;
              [T3I1-T3I3] ;
              [T4I1-T4I3] ;
    OUTPUT: stdyx tech4 SVALUES;
```

Notes to Appendix 6.2A:

1. Correlated Uniquenesses.

As noted in the chapter, one distinctive feature of longitudinal analyses is that they should normally include correlated uniquenesses between responses to the same item on different occasions. Here this was tested by comparing the basic measurement model with and without correlated uniquenesses. Shown here is the syntax for Model Con-1b (with correlated uniquenesses). Model Con-1a (without correlated unquenesses differs only in that the highlighted section (in yellow) is excluded. All subsequent models included the correlated uniquenesses.

2. Note: Modification indices and sample statistics are not available for multiple imputation, but can be obtained when the model is fit to any one of the imputation files separately

Appendix 6.3: Mplus Syntax for Metric Invariance Model Met-2c (see Table 6.1 in main text)

```
TITLE: Model Met-2c (see Table 1)
        DATA:
            FILE = wft1234ZBFATIMPlist.DAT;
            TYPE = IMPUTATION;
        VARIABLE:
        NAMES =
               t1bf1 t1I1 t1I2 t1I3 t1bf5 t1bf6
               t2bf1 t2I1 t2I2 t2I3 t2bf5 t2bf6
               t3bf1 t3I1 t3I2 t3I3 t3bf5 t3bf6
               t4bf1 t4I1 t4I2 t4I3 t4bf5 t4bf6
               elite year sex AGE;
        MISSING ARE ALL (-9);
        usevariables are
          t1I1 t1I2 t1I3
          t2I1 t2I2 t2I3
          t3I1 t3I2 t3I3
          t4I1 t4I2 t4I3
          ;
        USEOBSERVATIONS YEAR NE 6;
    GROUPING IS YEAR (7=yR7,8=Yr8,9=Yr9,10=Yr10);
         ANALYSIS: ESTIMATOR = MLR;
        MODEL:
        !BODY COMPOSITION
        PBCT1 by T1I1-T1I3 (L2-L4);
        PBCT2 by T2I1-T2I3 (L2-L4);
        PBCT3 by T3I1-T3I3 (L2-L4);
        PBCT4 by T4I1-T4I3 (L2-L4);
  !Latent means default to zero
        [ PBCT1@0]; [ PBCT2@0]; [ PBCT3@0]; [ PBCT4@0];
  !Latent variances--free;
        PBCT1 ; PBCT2 ; PBCT3 ; PBCT4 ;
  !BODY COMPOSITION Correlated residuals
        T2I1-T2I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T1I1-T1I3 ;
        T4I1-T4I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T3I1-T3I3 ;
  !BODY COMPOSITION INTERCEPTS
        [T1I1-T1I3]; [T2I1-T2I3]; [T3I1-T3I3]; [T4I1-T4I3];
    MODEL YR7:
        PBCT1 by T1I1-T1I3 (L2-L4);
        PBCT2 by T2I1-T2I3 (L2-L4);
        PBCT3 by T3I1-T3I3 (L2-L4);
        PBCT4 by T4I1-T4I3 (L2-L4);
    MODEL YR8:
        PBCT1 by T1I1-T1I3 (L2-L4);
        PBCT2 by T2I1-T2I3 (L2-L4);
        PBCT3 by T3I1-T3I3 (L2-L4);
        PBCT4 by T4I1-T4I3 (L2-L4);
  !Latent means default to zero
        [ PBCT1@0]; [ PBCT2@0]; [ PBCT3@0]; [ PBCT4@0];
  !Latent variances--free;
        PBCT1 ; PBCT2 ; PBCT3 ; PBCT4 ;
  !BODY COMPOSITION Correlated residuals
        T2I1-T2I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T1I1-T1I3 ;
        T4I1-T4I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T3I1-T3I3 ;
  !BODY COMPOSITION INTERCEPTS
        [T1I1-T1I3]; [T2I1-T2I3]; [T3I1-T3I3]; [T4I1-T4I3];
    MODEL YR9:
        PBCT1 by T1I1-T1I3 (L2-L4);
        PBCT2 by T2I1-T2I3 (L2-L4);
        PBCT3 by T3I1-T3I3 (L2-L4);
        PBCT4 by T4I1-T4I3 (L2-L4);
  !Latent means default to zero
        [ PBCT1@0]; [ PBCT2@0]; [ PBCT3@0]; [ PBCT4@0];
  !Latent variances--free;
        PBCT1 ; PBCT2 ; PBCT3 ; PBCT4 ;
  !BODY COMPOSITION Correlataed residuals
        T2I1-T2I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T1I1-T1I3 ;
```

```
       T4I1-T4I3 pwith T1I1-T1I3 ;
       T3I1-T3I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T3I1-T3I3 ;
 !BODY COMPOSITION INTERCEPTS
       [T1I1-T1I3]; [T2I1-T2I3]; [T3I1-T3I3]; [T4I1-T4I3];

   MODEL YR10:
       PBCT1 by T1I1-T1I3 (L2-L4);
       PBCT2 by T2I1-T2I3 (L2-L4);
       PBCT3 by T3I1-T3I3 (L2-L4);
       PBCT4 by T4I1-T4I3 (L2-L4);
 !Latent means default to zero
       [ PBCT1@0]; [ PBCT2@0]; [ PBCT3@0]; [ PBCT4@0];
 !Latent variances--free;
       PBCT1 ; PBCT2 ; PBCT3 ; PBCT4 ;
 !BODY COMPOSITION Correlataed residuals
       T2I1-T2I3 pwith T1I1-T1I3 ;
       T3I1-T3I3 pwith T1I1-T1I3 ;
       T4I1-T4I3 pwith T1I1-T1I3 ;
       T3I1-T3I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T3I1-T3I3 ;
 !BODY COMPOSITION INTERCEPTS
       [T1I1-T1I3]; [T2I1-T2I3]; [T3I1-T3I3]; [T4I1-T4I3];
 MODEL CONSTRAINT:
  L2 = 3 - L3 - L4; !mean unstandardized FL = 1
   OUTPUT: stdyx tech4 SVALUES;
```

Notes to Appendix Appendix 6.2B:

1. Factor loading invariance.

As noted in the chapter, In Models Met-2A–MET-2C (Table 6.1) we demonstrate good support for the invariance of factor loadings over:

- A. four cohorts (between-group);
- B. four waves (within-person);
- C. four cohorts and four waves (integrating both within- and between-comparisons).

Shown here is the syntax for Model Met-2C. Model Met-2A differs only in that with-group factor loadings (over time) are freely estimated (as in Model Con-1B in Appendix 1B). Model Met-2B differs only in that between-group factor loadings (over cohort) are freely estimated (as in Model Con-1B in Appendix 1B).

Appendix 6.4: Mplus Syntax for Scalar Invariance Model Scl-3b (see Table 1 in main text)

```
TITLE: Model Scl-3d (see Table 1)
        DATA:
            FILE = wft1234ZBFATIMPlist.DAT;
            TYPE = IMPUTATION;
        VARIABLE:
        NAMES =
               t1PBC1 t1I1 t1I2 t1I3 t1PBC5 t1PBC6
               t2PBC1 t2I1 t2I2 t2I3 t2PBC5 t2PBC6
               t3PBC1 t3I1 t3I2 t3I3 t3PBC5 t3PBC6
               t4PBC1 t4I1 t4I2 t4I3 t4PBC5 t4PBC6
               elite year sex AGE;
        MISSING ARE ALL (-9);
        usevariables are
          t1I1 t1I2 t1I3
          t2I1 t2I2 t2I3
          t3I1 t3I2 t3I3
          t4I1 t4I2 t4I3
          ;
        USEOBSERVATIONS YEAR NE 6;
    GROUPING IS YEAR (7=yR7,8=Yr8,9=Yr9,10=Yr10);
         ANALYSIS: ESTIMATOR = MLR;
        MODEL:
        !BODY COMPOSITION
        PBCT1 by T1I1-T1I3 (L2-L4);
        PBCT2 by T2I1-T2I3 (L2-L4);
        PBCT3 by T3I1-T3I3 (L2-L4);
        PBCT4 by T4I1-T4I3 (L2-L4);
!Latent means default to zero
          [ PBCt1@0];  [ PBCt2*0];  [ PBCt3*0]; [ PBCt4*0];
!Latent variances--free;
    PBCt1 ;  PBCt2 ;  PBCt3 ;  PBCt4 ;
! Correlataed residuals
    T2I1-T2I3 pwith T1I1-T1I3 ;
    T3I1-T3I3 pwith T1I1-T1I3 ;
    T4I1-T4I3 pwith T1I1-T1I3 ;
    T3I1-T3I3 pwith T2I1-T2I3 ;
    T4I1-T4I3 pwith T2I1-T2I3 ;
    T4I1-T4I3 pwith T3I1-T3I3 ;
! INTERCEPTS invariant over waves
    [T1I1-T1I3](INT2-INT4);
    [T2I1-T2I3](INT2-INT4);
    [T3I1-T3I3](INT2-INT4);
    [T4I1-T4I3](INT2-INT4);
MODEL YR7:
          PBCT1 by T1I1-T1I3 (L2-L4);
          PBCT2 by T2I1-T2I3 (L2-L4);
          PBCT3 by T3I1-T3I3 (L2-L4);
          PBCT4 by T4I1-T4I3 (L2-L4);
!Latent means default to zero
          [ PBCt1@0];  [ PBCt2*0];  [ PBCt3*0]; [ PBCt4*0];
!Latent variances--free;
    PBCt1 ;  PBCt2 ;  PBCt3 ;  PBCt4 ;
! INTERCEPTS
    [T1I1-T1I3](aINT2-aINT4);
    [T2I1-T2I3](aINT2-aINT4);
    [T3I1-T3I3](aINT2-aINT4);
    [T4I1-T4I3](aINT2-aINT4);
      MODEL YR8:
          PBCT1 by T1I1-T1I3 (L2-L4);
          PBCT2 by T2I1-T2I3 (L2-L4);
          PBCT3 by T3I1-T3I3 (L2-L4);
          PBCT4 by T4I1-T4I3 (L2-L4);
!Latent means default to zero
          [ PBCt1@0];  [ PBCt2*0];  [ PBCt3*0]; [ PBCt4*0];
!Latent variances--free;
    PBCt1 ;  PBCt2 ;  PBCt3 ;  PBCt4 ;
! INTERCEPTS
          [T1I1-T1I3](bINT2-bINT4);
          [T2I1-T2I3](bINT2-bINT4);
          [T3I1-T3I3](bINT2-bINT4);
          [T4I1-T4I3](bINT2-bINT4);
      MODEL YR9:
          PBCT1 by T1I1-T1I3 (L2-L4);
          PBCT2 by T2I1-T2I3 (L2-L4);
          PBCT3 by T3I1-T3I3 (L2-L4);
```

```
          PBCT4 by T4I1-T4I3 (L2-L4);
!Latent means default to zero
          [ PBCt1@0];  [ PBCt2*0];  [ PBCt3*0]; [ PBCt4*0];
!Latent variances--free;
    PBCt1 ;  PBCt2 ;  PBCt3 ;  PBCt4 ;
         ! INTERCEPTS
          [T1I1-T1I3](cINT2-cINT4);
          [T2I1-T2I3](cINT2-cINT4);
          [T3I1-T3I3](cINT2-cINT4);
          [T4I1-T4I3](cINT2-cINT4);
      MODEL YR10:
          PBCT1 by T1I1-T1I3 (L2-L4);
          PBCT2 by T2I1-T2I3 (L2-L4);
          PBCT3 by T3I1-T3I3 (L2-L4);
          PBCT4 by T4I1-T4I3 (L2-L4);
!Latent means default to zero
          [ PBCt1@0];  [ PBCt2*0];  [ PBCt3*0]; [ PBCt4*0];
!Latent variances--free;
    PBCt1 ;  PBCt2 ;  PBCt3 ;  PBCt4 ;
         ! INTERCEPTS
          [T1I1-T1I3](dINT2-dINT4); !  ; !(iNT1); !
          [T2I1-T2I3](dINT2-dINT4); ! ; !(iNT1); !
          [T3I1-T3I3](dINT2-dINT4); ! ; !(iNT1); !
          [T4I1-T4I3](dINT2-dINT4); !  ; !(iNT1); !
  MODEL CONSTRAINT:
    L2 = 1.0; ! traditional approach--FL of first item fixed to 1.0;
      OUTPUT: stdyx tech1 tech4 SVALUES;
```

Note. Intercepts for the indicators of the latent factors are specified by be invariant over waves (a within-person construct) but not age cohorts (a between-group construct). Thus, for example, for 'MODEL YR7:' the intercepts (highlighted in yellow) are constrained to be the same (aINT2-aINT4) across the four waves within this Year 7 age cohort. However, they are different from those specified for 'MODEL YR8:' (bINT2-bINT4) and subsequent age cohorts.

Appendix 6.5: Mplus Syntax for Scalar Invariance Model Scl-3d (see Table 1 in main text)

```
TITLE: Model Scl-3d (see Table 1)
        DATA:
            FILE = wft1234ZBFATIMPlist.DAT;
            TYPE = IMPUTATION;
        VARIABLE:
        NAMES =
               t1PBC1 t1I1 t1I2 t1I3 t1PBC5 t1PBC6
               t2PBC1 t2I1 t2I2 t2I3 t2PBC5 t2PBC6
               t3PBC1 t3I1 t3I2 t3I3 t3PBC5 t3PBC6
               t4PBC1 t4I1 t4I2 t4I3 t4PBC5 t4PBC6
               elite year sex AGE;
        MISSING ARE ALL (-9);
        usevariables are
          t1I1 t1I2 t1I3
          t2I1 t2I2 t2I3
          t3I1 t3I2 t3I3
          t4I1 t4I2 t4I3
          ;
        USEOBSERVATIONS YEAR NE 6;
    GROUPING IS YEAR (7=yR7,8=Yr8,9=Yr9,10=Yr10);
         ANALYSIS: ESTIMATOR = MLR;
        MODEL:
        !BODY COMPOSITION
        PBCT1 by T1I1-T1I3 (L2-L4);
        PBCT2 by T2I1-T2I3 (L2-L4);
        PBCT3 by T3I1-T3I3 (L2-L4);
        PBCT4 by T4I1-T4I3 (L2-L4);
!Latent means default to zero (default)
          [ PBCt1@0];  [ PBCt2*0];  [ PBCt3*0]; [ PBCt4*0];
!Latent variances--free (default);
    PBCt1 ;  PBCt2 ;  PBCt3 ;  PBCt4 ;
! Correlataed residuals free
    T2I1-T2I3 pwith T1I1-T1I3 ;
    T3I1-T3I3 pwith T1I1-T1I3 ;
    T4I1-T4I3 pwith T1I1-T1I3 ;
    T3I1-T3I3 pwith T2I1-T2I3 ;
    T4I1-T4I3 pwith T2I1-T2I3 ;
    T4I1-T4I3 pwith T3I1-T3I3 ;
    !INTERCEPTS constrained equal over waves)
    [T1I1-T1I3](INT2-INT4);
    [T2I1-T2I3](INT2-INT4);
    [T3I1-T3I3](INT2-INT4);
    [T4I1-T4I3](INT2-INT4);
MODEL YR7:
          PBCT1 by T1I1-T1I3 (L2-L4);
          PBCT2 by T2I1-T2I3 (L2-L4);
          PBCT3 by T3I1-T3I3 (L2-L4);
          PBCT4 by T4I1-T4I3 (L2-L4);
!Latent means default to zero
         [PBCT1-PBCT4*] (LMM1-LMM4);
!Latent variances--free;
    PBCt1 ;  PBCt2 ;  PBCt3 ;  PBCt4 ;
!INTERCEPTS constrained
    [T1I1-T1I3](INT2-INT4);
    [T2I1-T2I3](INT2-INT4);
    [T3I1-T3I3](INT2-INT4);
    [T4I1-T4I3](INT2-INT4);
      MODEL YR8:
          PBCT1 by T1I1-T1I3 (L2-L4);
          PBCT2 by T2I1-T2I3 (L2-L4);
          PBCT3 by T3I1-T3I3 (L2-L4);
          PBCT4 by T4I1-T4I3 (L2-L4);
!Latent means default to zero
   [PBCT1-PBCT4*] (LMM3-LMM6);
!Latent variances--free;
    PBCt1 ;  PBCt2 ;  PBCt3 ;  PBCt4 ;
! INTERCEPTS
          [T1I1-T1I3](INT2-INT4);
          [T2I1-T2I3](INT2-INT4);
          [T3I1-T3I3](INT2-INT4);
          [T4I1-T4I3](INT2-INT4);
      MODEL YR9:
          PBCT1 by T1I1-T1I3 (L2-L4);
          PBCT2 by T2I1-T2I3 (L2-L4);
```

```
          PBCT3 by T3I1-T3I3 (L2-L4);
          PBCT4 by T4I1-T4I3 (L2-L4);
!Latent means default to zero
    [PBCT1-PBCT4*] (LMM5-LMM8);
!Latent variances--free;
    PBCt1 ;  PBCt2 ;  PBCt3 ;  PBCt4 ;
         ! INTERCEPTS
          [T1I1-T1I3](INT2-INT4);
          [T2I1-T2I3](INT2-INT4);
          [T3I1-T3I3](INT2-INT4);
          [T4I1-T4I3](INT2-INT4);
      MODEL YR10:
          PBCT1 by T1I1-T1I3 (L2-L4);
          PBCT2 by T2I1-T2I3 (L2-L4);
          PBCT3 by T3I1-T3I3 (L2-L4);
          PBCT4 by T4I1-T4I3 (L2-L4);
!Latent means default to zero
    [PBCT1-PBCT4*] (LMM7-LMM10);
    !Latent variances--free;
    PBCt1 ;  PBCt2 ;  PBCt3 ;  PBCt4 ;
         ! INTERCEPTS
          [T1I1-T1I3](INT2-INT4);
          [T2I1-T2I3](INT2-INT4);
          [T3I1-T3I3](INT2-INT4);
          [T4I1-T4I3](INT2-INT4);
  MODEL CONSTRAINT:
   !Non-arbitary intercepts
    INT2 = 0 - INT3 - INT4 ;
      L2 = 1.0; ! traditional approach--FL of first item fixed to 1.0;
      OUTPUT: stdyx tech1 tech4 SVALUES;
```

Note. Intercepts for the indicators of the latent factors are specified by be invariant over waves (a within-person construct) and over age cohorts (a between-group construct). Thus, this model differs from Scalar Invariance Model Scl-3b (appendix 6.4) in that the intercepts are constrained to be the same (INT2-INT4) across the four waves within each of the four age cohort groups as well as the four waves.

Appendix 6.6 Mplus Syntax for Model Unq-4a (see Table 1 in main text)

```
        TITLE: Crosslag model ignoring multiple cohorts
        DATA:
         !   FILE = wft1234ZpbcATIMP1.DAT;
            FILE = wft1234ZbfATIMPlist.DAT;
            TYPE = IMPUTATION;

        VARIABLE:
        NAMES =
               t1pbc1 t1I1 t1I2 t1I3 t1pbc5 t1pbc6
               t2pbc1 t2I1 t2I2 t2I3 t2pbc5 t2pbc6
               t3pbc1 t3I1 t3I2 t3I3 t3pbc5 t3pbc6
               t4pbc1 t4I1 t4I2 t4I3 t4pbc5 t4pbc6
               elite year sex AGE;
        MISSING ARE ALL (-9);
        usevariables are
        !pbc
          t1I1 t1I2 t1I3
          t2I1 t2I2 t2I3
          t3I1 t3I2 t3I3
          t4I1 t4I2 t4I3
             ;
        USEOBSERVATIONS YEAR NE 6;
    GROUPING IS YEAR (7=yR7,8=Yr8,9=Yr9,10=Yr10);

        ANALYSIS: ESTIMATOR = MLR; !ML; TYPE = BASIC;
  MODEL:
        !factor loadings
        pbcT1 by T1I1-T1I3 (L2-L4);
        pbcT2 by T2I1-T2I3 (L2-L4);
        pbcT3 by T3I1-T3I3 (L2-L4);
        pbcT4 by T4I1-T4I3 (L2-L4);
 !Latent means default to zero
        [ pbct1*];
        [ pbct2*];
        [ pbct3*];
        [ pbct4*];
 !Latent means default to zero
          [ pbct1@0];
          [ pbct2*0];
          [ pbct3*0];
          [ pbct4*0];
 !Latent variances--free;
          pbct1 ;pbct2 ; pbct3 ; pbct4 ;
 !Correlataed residuals
       T2I1-T2I3 pwith T1I1-T1I3 ;
       T3I1-T3I3 pwith T1I1-T1I3 ;
       T4I1-T4I3 pwith T1I1-T1I3 ;
       T3I1-T3I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T3I1-T3I3 ;
        ! INTERCEPTS
        [T1I1-T1I3] (INT2-INT4);
        [T2I1-T2I3] (INT2-INT4);
        [T3I1-T3I3] (INT2-INT4);
        [T4I1-T4I3] (INT2-INT4);

  MODEL YR7:
        pbcT1 by T1I1-T1I3 (L2-L4); !(L1); !
        pbcT2 by T2I1-T2I3 (L2-L4); !(L1); !
        pbcT3 by T3I1-T3I3 (L2-L4); !(L1); !
        pbcT4 by T4I1-T4I3 (L2-L4); !(L1); !

        ! INTERCEPTS
     [ t1i1*4.66096 ] (int2);
     [ t1i2*4.56109 ] (int3);
     [ t1i3*4.86253 ] (int4);
     [ t2i1*4.66096 ] (int2);
     [ t2i2*4.56109 ] (int3);
     [ t2i3*4.86253 ] (int4);
     [ t3i1*4.66096 ] (int2);
     [ t3i2*4.56109 ] (int3);
     [ t3i3*4.86253 ] (int4);
     [ t4i1*4.66096 ] (int2);
     [ t4i2*4.56109 ] (int3);
     [ t4i3*4.86253 ] (int4);
```

```
        ! uniq invariant
      t1i1*(unq2);
      t1i2*(unq3);
      t1i3*(unq4);
      t2i1*(unq2);
      t2i2*(unq3);
      t2i3*(unq4);
      t3i1*(unq2);
      t3i2*(unq3);
      t3i3*(unq4);
      t4i1*(unq2);
      t4i2*(unq3);
      t4i3*(unq4);

        !Latent means default to zero
        [ pbct1*];
        [ pbct2*];
        [ pbct3*];
        [ pbct4*];

   MODEL YR8:
        pbcT1 by T1I1-T1I3 (L2-L4); !(L1); !
        pbcT2 by T2I1-T2I3 (L2-L4); !(L1); !
        pbcT3 by T3I1-T3I3 (L2-L4); !(L1); !
        pbcT4 by T4I1-T4I3 (L2-L4); !(L1); !
      [ t1i1] (int2);
      [ t1i2] (int3);
      [ t1i3] (int4);
      [ t2i1] (int2);
      [ t2i2] (int3);
      [ t2i3] (int4);
      [ t3i1] (int2);
      [ t3i2] (int3);
      [ t3i3] (int4);
      [ t4i1] (int2);
      [ t4i2] (int3);
      [ t4i3] (int4);
      [ pbct1];
      [ pbct2];
      [ pbct3];
      [ pbct4];
! uniq invariant
      t1i1*(unq2);
      t1i2*(unq3);
      t1i3*(unq4);
      t2i1*(unq2);
      t2i2*(unq3);
      t2i3*(unq4);
      t3i1*(unq2);
      t3i2*(unq3);
      t3i3*(unq4);
      t4i1*(unq2);
      t4i2*(unq3);
      t4i3*(unq4);

   MODEL YR9:
        pbcT1 by T1I1-T1I3 (L2-L4); !(L1); !
        pbcT2 by T2I1-T2I3 (L2-L4); !(L1); !
        pbcT3 by T3I1-T3I3 (L2-L4); !(L1); !
        pbcT4 by T4I1-T4I3 (L2-L4); !(L1); !

      [ t1i1] (int2);
      [ t1i2] (int3);
      [ t1i3] (int4);
      [ t2i1] (int2);
      [ t2i2] (int3);
      [ t2i3] (int4);
      [ t3i1] (int2);
      [ t3i2] (int3);
      [ t3i3] (int4);
      [ t4i1] (int2);
      [ t4i2] (int3);
      [ t4i3] (int4);
      [ pbct1 ];
      [ pbct2 ];
      [ pbct3 ];
      [ pbct4 ];

      pbct1*;
```

```
    pbct2*;
    pbct3*;
    pbct4*;
       ! uniq invariant
    t1i1*(unq2);
    t1i2*(unq3);
    t1i3*(unq4);
    t2i1*(unq2);
    t2i2*(unq3);
    t2i3*(unq4);
    t3i1*(unq2);
    t3i2*(unq3);
    t3i3*(unq4);
    t4i1*(unq2);
    t4i2*(unq3);
    t4i3*(unq4);

 MODEL YR10:
       pbcT1 by T1I1-T1I3 (L2-L4); !(L1); !
       pbcT2 by T2I1-T2I3 (L2-L4); !(L1); !
       pbcT3 by T3I1-T3I3 (L2-L4); !(L1); !
       pbcT4 by T4I1-T4I3 (L2-L4); !(L1); !

    [ t1i1] (int2);
    [ t1i2] (int3);
    [ t1i3] (int4);
    [ t2i1] (int2);
    [ t2i2] (int3);
    [ t2i3] (int4);
    [ t3i1] (int2);
    [ t3i2] (int3);
    [ t3i3] (int4);
    [ t4i1] (int2);
    [ t4i2] (int3);
    [ t4i3] (int4);
    [ pbct1 ];
    [ pbct2 ];
    [ pbct3 ];
    [ pbct4];

    pbct1*;
    pbct2*;
    pbct3*;
    pbct4*;
       ! uniq invariant
    t1i1*(unq2);
    t1i2*(unq3);
    t1i3*(unq4);
    t2i1*(unq2);
    t2i2*(unq3);
    t2i3*(unq4);
    t3i1*(unq2);
    t3i2*(unq3);
    t3i3*(unq4);
    t4i1*(unq2);
    t4i2*(unq3);
    t4i3*(unq4);

    MODEL CONSTRAINT:
       L2 = 3 - L3 - L4;
     OUTPUT: stdyx tech1 tech4 SVALUES;
```

Appendix 6.7: Mplus Syntax for MIMIC Model with Orthogonal Polynomial Growth Components Model Mim-6b (see Table 1 in main text)

```
TITLE: Model Mim-6b (see Table 1 and 3 in main text)
          DATA:
             FILE = wft1234ZBFATIMPlist.DAT;
             TYPE = IMPUTATION;

          VARIABLE:
          NAMES =
                 t1bf1 t1I1 t1I2 t1I3 t1bf5 t1bf6
                 t2bf1 t2I1 t2I2 t2I3 t2bf5 t2bf6
                 t3bf1 t3I1 t3I2 t3I3 t3bf5 t3bf6
                 t4bf1 t4I1 t4I2 t4I3 t4bf5 t4bf6
                 elite year sex AGE;

          MISSING ARE ALL (-9);

          usevariables are
            t1I1 t1I2 t1I3
            t2I1 t2I2 t2I3
            t3I1 t3I2 t3I3
            t4I1 t4I2 t4I3
            elite SEX SX_EL
            ;

          USEOBSERVATIONS YEAR NE 6;

      GROUPING IS YEAR (7=yR7,8=Yr8,9=Yr9,10=Yr10);

          define:
          center ELITE SEX (GRANDMEAN);
         SX_EL = ELITE * SEX;
            ANALYSIS: ESTIMATOR = MLR; !ML; TYPE = BASIC;

            MODEL:
            !BODY COMPOSITION
          PBCT1 by T1I1-T1I3 (L2-L4);
          PBCT2 by T2I1-T2I3 (L2-L4);
          PBCT3 by T3I1-T3I3 (L2-L4);
          PBCT4 by T4I1-T4I3 (L2-L4);
            !Latent variances--free;
              PBCT1 ;
              PBCT2 ;
              PBCT3 ;
              PBCT4 ;
              !BODY COMPOSITION Correlated residuals
             T2I1-T2I3 pwith T1I1-T1I3 ;
             T3I1-T3I3 pwith T1I1-T1I3 ;
             T4I1-T4I3 pwith T1I1-T1I3 ;
             T3I1-T3I3 pwith T2I1-T2I3 ;
             T4I1-T4I3 pwith T2I1-T2I3 ;
             T4I1-T4I3 pwith T3I1-T3I3 ;
              !BODY COMPOSITION INTERCEPTS
              [T1I1-T1I3] (INT2-INT4);
              [T2I1-T2I3] (INT2-INT4);
              [T3I1-T3I3] (INT2-INT4);
              [T4I1-T4I3] (INT2-INT4);

        PBCT1-PBCT4  on   sex   (sx)    ;
        PBCT1-PBCT4  on   elite (EL)    ;
        PBCT1-PBCT4  on   SX_EL (intS_E)    ;



        MODEL YR7:
            PBCT1 by T1I1-T1I3 (L2-L4);
          PBCT2 by T2I1-T2I3 (L2-L4);
          PBCT3 by T3I1-T3I3 (L2-L4);
          PBCT4 by T4I1-T4I3 (L2-L4);
        !Latent variances--free;
          PBCT1-PBCT4*;
          !BODY COMPOSITION Correlated residuals
         T2I1-T2I3 pwith T1I1-T1I3 ;
```

```
        T3I1-T3I3 pwith T1I1-T1I3 ;
        T4I1-T4I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T3I1-T3I3 ;
         !BODY COMPOSITION INTERCEPTS
         [T1I1-T1I3] (INT2-INT4);
         [T2I1-T2I3] (INT2-INT4);
         [T3I1-T3I3] (INT2-INT4);
         [T4I1-T4I3] (INT2-INT4);
!MEANS
  [PBCT1-PBCT4*] (LMM1-LMM4);

      PBCT1  ON Sex        (sx1);
      PBCT1  ON ELite      (el1);
      PBCT1  ON Sx_el      (ints_e1);
      PBCT2  ON Sex        (sx2);
      PBCT2  ON ELite      (el2);
      PBCT2  ON Sx_el      (ints_e2);
      PBCT3  ON Sex        (sx3);
      PBCT3  ON ELite      (el3);
      PBCT3  ON Sx_el      (ints_e3);
      PBCT4  ON Sex        (sx4);
      PBCT4  ON ELite      (el4);
      PBCT4  ON Sx_el      (ints_e4);

 MODEL YR8:
         PBCT1 by T1I1-T1I3 (L2-L4);
         PBCT2 by T2I1-T2I3 (L2-L4);
         PBCT3 by T3I1-T3I3 (L2-L4);
         PBCT4 by T4I1-T4I3 (L2-L4);
       !Latent variances--free;
         PBCT1-PBCT4*;
         !BODY COMPOSITION Correlated residuals
        T2I1-T2I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T1I1-T1I3 ;
        T4I1-T4I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T3I1-T3I3 ;
         !BODY COMPOSITION INTERCEPTS
         [T1I1-T1I3] (INT2-INT4);
         [T2I1-T2I3] (INT2-INT4);
         [T3I1-T3I3] (INT2-INT4);
         [T4I1-T4I3] (INT2-INT4);
!MEANS
  [PBCT1-PBCT4*] (LMM3-LMM6);


      PBCT1  ON Sex        (sx3);
      PBCT1  ON ELite      (el3);
      PBCT1  ON Sx_el      (ints_e3);
      PBCT2  ON Sex        (sx4);
      PBCT2  ON ELite      (el4);
      PBCT2  ON Sx_el      (ints_e4);
      PBCT3  ON Sex        (sx5);
      PBCT3  ON ELite      (el5);
      PBCT3  ON Sx_el      (ints_e5);
      PBCT4  ON Sex        (sx6);
      PBCT4  ON ELite      (el6);
      PBCT4  ON Sx_el      (ints_e6);
  MODEL YR9:
        PBCT1 by T1I1-T1I3 (L2-L4);
         PBCT2 by T2I1-T2I3 (L2-L4);
         PBCT3 by T3I1-T3I3 (L2-L4);
         PBCT4 by T4I1-T4I3 (L2-L4);
       !Latent variances--free;
         PBCT1-PBCT4*;
         !BODY COMPOSITION Correlated residuals
        T2I1-T2I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T1I1-T1I3 ;
        T4I1-T4I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T3I1-T3I3 ;
         !BODY COMPOSITION INTERCEPTS
         [T1I1-T1I3] (INT2-INT4);
         [T2I1-T2I3] (INT2-INT4);
         [T3I1-T3I3] (INT2-INT4);
         [T4I1-T4I3] (INT2-INT4);
```

```
!MEANS
  [PBCT1-PBCT4*] (LMM5-LMM8);
     PBCT1  ON Sex         (sx5);
     PBCT1  ON ELite       (el5);
     PBCT1  ON Sx_el       (ints_e5);
     PBCT2  ON Sex         (sx6);
     PBCT2  ON ELite       (el6);
     PBCT2  ON Sx_el       (ints_e6);
     PBCT3  ON Sex         (sx7);
     PBCT3  ON ELite       (el7);
     PBCT3  ON Sx_el       (ints_e7);
     PBCT4  ON Sex         (sx8);
     PBCT4  ON ELite       (el8);
     PBCT4  ON Sx_el       (ints_e8);
  MODEL YR10:
         PBCT1 by T1I1-T1I3 (L2-L4);
        PBCT2 by T2I1-T2I3 (L2-L4);
        PBCT3 by T3I1-T3I3 (L2-L4);
        PBCT4 by T4I1-T4I3 (L2-L4);
      !Latent variances--free;
        PBCT1-PBCT4*;
        !BODY COMPOSITION Correlated residuals
       T2I1-T2I3 pwith T1I1-T1I3 ;
       T3I1-T3I3 pwith T1I1-T1I3 ;
       T4I1-T4I3 pwith T1I1-T1I3 ;
       T3I1-T3I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T3I1-T3I3 ;
        !BODY COMPOSITION INTERCEPTS
        [T1I1-T1I3] (INT2-INT4);
        [T2I1-T2I3] (INT2-INT4);
        [T3I1-T3I3] (INT2-INT4);
        [T4I1-T4I3] (INT2-INT4);
!MEANS
  [PBCT1-PBCT4*] (LMM7-LMM10);
       PBCT1  ON Sex       (sx7);
       PBCT1  ON ELite     (el7);
       PBCT1  ON Sx_el     (ints_e7);
       PBCT2  ON Sex       (sx8);
       PBCT2  ON ELite     (el8);
       PBCT2  ON Sx_el     (ints_e8);
       PBCT3  ON Sex       (sx9);
       PBCT3  ON ELite     (el9);
       PBCT3  ON Sx_el     (ints_e9);
       PBCT4  ON Sex       (sx10);
       PBCT4  ON ELite     (el10);
       PBCT4  ON Sx_el     (ints_e10);

 MODEL CONSTRAINT:
new (G1-G10,lin, quad, cubic,quart,GROW,GROW_MSQ,female,elite,FExEL);
G1=LMM1;G2=LMM2;G3=LMM3;G4=LMM4;G5=LMM5;
G6=LMM6;G7=LMM7;G8=LMM8;G9=LMM9;G10=LMM10;

GROW=(G1+G2+G3+G4+G5+G6+G7+G8+G9+G10)/10;

GROW_MSQ=((G1-GROW)**2+(G2-GROW)**2+(G3-GROW)**2+(G4-GROW)**2+
(G5-GROW)**2+(G6-GROW)**2+(G7-GROW)**2+(G8-GROW)**2+(G9-GROW)**2+
(G10-GROW)**2)/10;

lin = (-9 * G1)+(-7 * G2)+(-5 * G3)+(-3 * G4)+(-1 * G5)+
    (1 * G6)+(3 * G7)+(5 * G8)+(7 * G9)+(9 * G10);
quad = (6 * G1)+(2 * G2)+(-1 * G3)+(-3 * G4)+(-4 * G5)+
    (-4 * G6)+(-3 * G7)+(-1 * G8)+(2 * G9)+(6 * G10);
cubic = (-42 * G1)+(14 * G2)+(35 * G3)+(31 * G4)+(12 * G5)+
    (-12 * G6)+(-31 * G7)+(-35 * G8)+(-14 * G9)+(42 * G10);
quart = (18 * G1)+(-22 * G2)+(-17 * G3)+(3 * G4)+(18 * G5)+
    (18 * G6)+(3 * G7)+(-17 * G8)+(-22 * G9)+(18 * G10);

new (F1-F10,Flin, Fquad, Fcubic,Fquart,F_MSQ);
  F1=SX1;F2=SX2;F3=SX3;F4=SX4;F5=SX5;
  F6=SX6;F7=SX7;F8=SX8;F9=SX9;F10=SX10;
  female=(SX1+SX2+SX3+SX4+SX5+SX6+SX7+SX8+SX9+SX10)/10;
   Flin = (-9 * F1)+(-7 * F2)+(-5 * F3)+(-3 * F4)+(-1 * F5)+
      (1 * F6)+(3 * F7)+(5 * F8)+(7 * F9)+(9 * F10);
  Fquad = (6 * F1)+(2 * F2)+(-1 * F3)+(-3 * F4)+(-4 * F5)+
      (-4 * F6)+(-3 * F7)+(-1 * F8)+(2 * F9)+(6 * F10);
  Fcubic = (-42 * F1)+(14 * F2)+(35 * F3)+(31 * F4)+(12 * F5)+
      (-12 * F6)+(-31 * F7)+(-35 * F8)+(-14 * F9)+(42 * F10);
  Fquart = (18 * F1)+(-22 * F2)+(-17 * F3)+(3 * F4)+(18 * F5)+
      (18 * F6)+(3 * F7)+(-17 * F8)+(-22 * F9)+(18 * F10);
```

```
F_MSQ=((SX1-FEMALE)**2+(SX2-FEMALE)**2+(SX3-FEMALE)**2+(SX4-FEMALE)**2+
(SX5-FEMALE)**2+(SX6-FEMALE)**2+(SX7-FEMALE)**2+(SX8-FEMALE)**2+(SX9-FEMALE)**2+
(SX10-FEMALE)**2)/10;


   new (E1-E10,Elin, Equad, Ecubic,Equart,EL_MSQ);
  E1=EL1;E2=EL2;E3=EL3;E4=EL4;E5=EL5;
  E6=EL6;E7=EL7;E8=EL8;E9=EL9;E10=EL10;
 elite=(EL1+EL2+EL3+EL4+EL5+EL6+EL7+EL8+EL9+EL10)/10;
  Elin = (-9 * E1)+(-7 * E2)+(-5 * E3)+(-3 * E4)+(-1 * E5)+
      (1 * E6)+(3 * E7)+(5 * E8)+(7 * E9)+(9 * E10);
  Equad = (6 * E1)+(2 * E2)+(-1 * E3)+(-3 * E4)+(-4 * E5)+
      (-4 * E6)+(-3 * E7)+(-1 * E8)+(2 * E9)+(6 * E10);
  Ecubic = (-42 * E1)+(14 * E2)+(35 * E3)+(31 * E4)+(12 * E5)+
      (-12 * E6)+(-31 * E7)+(-35 * E8)+(-14 * E9)+(42 * E10);
  Equart = (18 * E1)+(-22 * E2)+(-17 * E3)+(3 * E4)+(18 * E5)+
      (18 * E6)+(3 * E7)+(-17 * E8)+(-22 * E9)+(18 * E10);
  EL_MSQ=((EL1-ELITE)**2+(EL2-ELITE)**2+(EL3-ELITE)**2+(EL4-ELITE)**2+
(EL5-ELITE)**2+(EL6-ELITE)**2+(EL7-ELITE)**2+(EL8-ELITE)**2+(EL9-ELITE)**2+
(EL10-ELITE)**2)/10;

  new (I1-I10,Ilin, Iquad, Icubic,Iquart,FEXI_MSQ);
  I1=INTS_E1;I2=INTS_E2;I3=INTS_E3;I4=INTS_E4;I5=INTS_E5;
  I6=INTS_E6;I7=INTS_E7;I8=INTS_E8;I9=INTS_E9;I10=INTS_E10;
 FExEL=(I1+I2+I3+I4+I5+I6+I7+I8+I9+I10)/10;
  Ilin = (-9 * I1)+(-7 * I2)+(-5 * I3)+(-3 * I4)+(-1 * I5)+
      (1 * I6)+(3 * I7)+(5 * I8)+(7 * I9)+(9 * I10);
  Iquad = (6 * I1)+(2 * I2)+(-1 * I3)+(-3 * I4)+(-4 * I5)+
      (-4 * I6)+(-3 * I7)+(-1 * I8)+(2 * I9)+(6 * I10);
  Icubic = (-42 * I1)+(14 * I2)+(35 * I3)+(31 * I4)+(12 * I5)+
      (-12 * I6)+(-31 * I7)+(-35 * I8)+(-14 * I9)+(42 * I10);
  Iquart = (18 * I1)+(-22 * I2)+(-17 * I3)+(3 * I4)+(18 * I5)+
      (18 * I6)+(3 * I7)+(-17 * I8)+(-22 * I9)+(18 * I10);
FExI_MSQ=((I1-FExEL)**2+(I2-FExEL)**2+(I3-FExEL)**2+(I4-FExEL)**2+
(I5-FExEL)**2+(I6-FExEL)**2+(I7-FExEL)**2+(I8-FExEL)**2+(I9-FExEL)**2+
(I10-FExEL)**2)/10;

  !Non-arbitrary intercepts
  INT2 = 0 - INT3 - INT4 ;
  L2 = 1.0;
   OUTPUT: stdyx tech1 tech4 SVALUES;
```

Note.

1. The contrast coefficients used to define the polynomial contrast coefficients used to define the polynomial components (highlighted in yellow) will vary depending on the number of means (i.e., M1-M10) but are readily available from most textbooks the discuss contrast coding.

2. The interaction between gender and athlete group is defined as the cross-product of these two variables (SX_EL = ELITE * SEX;). The main and interaction effects (highlighted in green for the Year 7 cohort) is constrained to be equal for overlapping cells in the cohort sequence design (see Figure 6.1a). In different models, these estimates were freely estimated (Model Mim-6a in Table 6.1) or constrained to be zero (in Models Mim-6c and Mim-6d in Table 6.11) .

Appendix 6.8 Mplus Syntax for Latent Growth Curve Model LGC-7b1 (see Tables 6.1 and 6.4 in main text)

```
TITLE: Model LGC-7b1 (see Table 1 and 4 in main text)
           DATA:
              FILE = wft1234ZBFATIMPlist.DAT;
              TYPE = IMPUTATION;

           VARIABLE:
           NAMES =
                  t1bf1 t1I1 t1I2 t1I3 t1bf5 t1bf6
                  t2bf1 t2I1 t2I2 t2I3 t2bf5 t2bf6
                  t3bf1 t3I1 t3I2 t3I3 t3bf5 t3bf6
                  t4bf1 t4I1 t4I2 t4I3 t4bf5 t4bf6
                  elite year sex AGE;

           MISSING ARE ALL (-9);

           usevariables are
           !bf
             t1I1 t1I2 t1I3
             t2I1 t2I2 t2I3
             t3I1 t3I2 t3I3
             t4I1 t4I2 t4I3
            elite SEX SX_EL
            ;

           USEOBSERVATIONS YEAR NE 6;

       GROUPING IS YEAR (7=yR7,8=Yr8,9=Yr9,10=Yr10);

           define:
        sex = sex - 1.5;
        elite = elite -.5;

          SX_EL = ELITE * SEX;
             ANALYSIS: ESTIMATOR = MLR; ITERATIONS = 1000000; PROCESSORS=3;
             MODEL:
             !BODY COMPOSITION
           PBCT1 by T1I1-T1I3 (L2-L4);
           PBCT2 by T2I1-T2I3 (L2-L4);
           PBCT3 by T3I1-T3I3 (L2-L4);
           PBCT4 by T4I1-T4I3 (L2-L4);
             !Latent variances--free;
               PBCT1-PBCT4*;
               !BODY COMPOSITION Correlated residuals
              T2I1-T2I3 pwith T1I1-T1I3 ;
              T3I1-T3I3 pwith T1I1-T1I3 ;
              T4I1-T4I3 pwith T1I1-T1I3 ;
              T3I1-T3I3 pwith T2I1-T2I3 ;
              T4I1-T4I3 pwith T2I1-T2I3 ;
              T4I1-T4I3 pwith T3I1-T3I3 ;
               !BODY COMPOSITION INTERCEPTS
               [T1I1-T1I3] (INT2-INT4);
               [T2I1-T2I3] (INT2-INT4);
               [T3I1-T3I3] (INT2-INT4);
               [T4I1-T4I3] (INT2-INT4);

       !MEANS - Constrained to Identify growth component
       [PBCT1-PBCT4@0];

   !----------------------------------------------------------------------
   !Growth Components
   !----------------------------------------------------------------------
       !MEANS - Constrained to Identify growth component
       [PBCT1-PBCT4@0];
   !basic growth model
          int Lin Quad | PBCT1@0  PBCT2@1  PBCT3@2  PBCT4@3 ;
     !Means and Variances of growth components
       !Variances
       int*(VIM1);
       LIN*(VLM1);
       QUAD*(VQM1);
       !Means
       [int*](MI1);
       [Lin*](ML1);
      [QUAD*](MQ1);
```

```
    !Correlations between slopes and growth
    int with LIN (IL1) ;
    int with QUAD (IQ1) ;
    QUAD with LIN (LQ1) ;

    INT LIN QUAD on   sex   (sx1-SX3)    ;
    INT LIN QUAD on   elite (EL1-EL3)    ;
    INT LIN QUAD on   SX_EL (ELxSX1-ELxSX3)    ;

      MODEL YR7:
          PBCT1 by T1I1-T1I3 (L2-L4);
        PBCT2 by T2I1-T2I3 (L2-L4);
        PBCT3 by T3I1-T3I3 (L2-L4);
        PBCT4 by T4I1-T4I3 (L2-L4);
      !Latent variances--free;
        PBCT1-PBCT4*;
        !BODY COMPOSITION Correlated residuals
       T2I1-T2I3 pwith T1I1-T1I3 ;
       T3I1-T3I3 pwith T1I1-T1I3 ;
       T4I1-T4I3 pwith T1I1-T1I3 ;
       T3I1-T3I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T3I1-T3I3 ;
        !BODY COMPOSITION INTERCEPTS
        [T1I1-T1I3] (INT2-INT4);
        [T2I1-T2I3] (INT2-INT4);
        [T3I1-T3I3] (INT2-INT4);
        [T4I1-T4I3] (INT2-INT4);
!MEANS
!Growth Components
    !MEANS - Constrained to Identify growth component
    [PBCT1-PBCT4@0];
!basic growth model
         int Lin Quad | PBCT1@0  PBCT2@1  PBCT3@2  PBCT4@3 ;

  !Means and Variances of growth components
    !Variances
     int*(VIM1);
     LIN*(VLM1);
    QUAD*(VQM1);
    !Means
    [int*](MI1);
    [Lin*](ML1);
   [QUAD*](MQ1);

    !Correlations between slopes and growth
    int with LIN (IL1) ;
    int with QUAD (IQ1) ;
    QUAD with LIN (LQ1) ;!Growth Components: basic growth model

 MODEL YR8:
        PBCT1 by T1I1-T1I3 (L2-L4);
        PBCT2 by T2I1-T2I3 (L2-L4);
        PBCT3 by T3I1-T3I3 (L2-L4);
        PBCT4 by T4I1-T4I3 (L2-L4);
      !Latent variances--free;
        PBCT1-PBCT4*;
        !BODY COMPOSITION Correlated residuals
       T2I1-T2I3 pwith T1I1-T1I3 ;
       T3I1-T3I3 pwith T1I1-T1I3 ;
       T4I1-T4I3 pwith T1I1-T1I3 ;
       T3I1-T3I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T2I1-T2I3 ;
       T4I1-T4I3 pwith T3I1-T3I3 ;
        !BODY COMPOSITION INTERCEPTS
        [T1I1-T1I3] (INT2-INT4);
        [T2I1-T2I3] (INT2-INT4);
        [T3I1-T3I3] (INT2-INT4);
        [T4I1-T4I3] (INT2-INT4);
!MEANS
!  [PBCT1-PBCT4*] (LMM3-LMM6);
!Growth Components
    !MEANS - Constrained to Identify growth component
    [PBCT1-PBCT4@0];
!basic growth model
         int Lin Quad | PBCT1@2  PBCT2@3  PBCT3@4  PBCT4@5 ;

  !Means and Variances of growth components
    !Variances
```

```
      int*(VIM1);
      LIN*(VLM1);
     QUAD*(VQM1);
     !Means
     [int*](MI1);
     [Lin*](ML1);
    [QUAD*](MQ1);

     !Correlations between slopes and growth
     int with LIN (IL1) ;
     int with QUAD (IQ1) ;
     QUAD with LIN (LQ1) ;!Growth Components: basic growth model

   MODEL YR9:
        PBCT1 by T1I1-T1I3 (L2-L4);
         PBCT2 by T2I1-T2I3 (L2-L4);
         PBCT3 by T3I1-T3I3 (L2-L4);
         PBCT4 by T4I1-T4I3 (L2-L4);
       !Latent variances--free;
         PBCT1-PBCT4*;
         !BODY COMPOSITION Correlated residuals
        T2I1-T2I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T1I1-T1I3 ;
        T4I1-T4I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T3I1-T3I3 ;
         !BODY COMPOSITION INTERCEPTS
         [T1I1-T1I3] (INT2-INT4);
         [T2I1-T2I3] (INT2-INT4);
         [T3I1-T3I3] (INT2-INT4);
         [T4I1-T4I3] (INT2-INT4);
 !Growth Components
     !MEANS - Constrained to Identify growth component
     [PBCT1-PBCT4@0];
 !basic growth model
        int Lin Quad | PBCT1@4  PBCT2@5 PBCT3@6 PBCT4@7;
   !Means and Variances of growth components
     !Variances
      int*(VIM1);
      LIN*(VLM1);
     QUAD*(VQM1);
     !Means
     [int*](MI1);
     [Lin*](ML1);
    [QUAD*](MQ1);

     !Correlations between slopes and growth
     int with LIN (IL1) ;
     int with QUAD (IQ1) ;
     QUAD with LIN (LQ1) ;!Growth Components: basic growth model

   MODEL YR10:
         PBCT1 by T1I1-T1I3 (L2-L4);
         PBCT2 by T2I1-T2I3 (L2-L4);
         PBCT3 by T3I1-T3I3 (L2-L4);
         PBCT4 by T4I1-T4I3 (L2-L4);
       !Latent variances--free;
         PBCT1-PBCT4*;
         !BODY COMPOSITION Correlated residuals
        T2I1-T2I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T1I1-T1I3 ;
        T4I1-T4I3 pwith T1I1-T1I3 ;
        T3I1-T3I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T2I1-T2I3 ;
        T4I1-T4I3 pwith T3I1-T3I3 ;
         !BODY COMPOSITION INTERCEPTS
         [T1I1-T1I3] (INT2-INT4);
         [T2I1-T2I3] (INT2-INT4);
         [T3I1-T3I3] (INT2-INT4);
         [T4I1-T4I3] (INT2-INT4);
 !MEANS
  ![PBCT1-PBCT4*] (LMM7-LMM10);
 !Growth Components
     !MEANS - Constrained to Identify growth component
     [PBCT1-PBCT4@0];
 !basic growth model
        int Lin Quad | PBCT1@6 PBCT2@7 PBCT3@8 PBCT4@9;
   !Means and Variances of growth components
     !Variances
```

```
 int*(VIM1);
 LIN*(VLM1);
QUAD*(VQM1);
!Means
[int*](MI1);
[Lin*](ML1);
[QUAD*](MQ1);

!Correlations between slopes and growth
int with LIN (IL1) ;
int with QUAD (IQ1) ;

MODEL CONSTRAINT:

 !Non-arbitrary intercepts
INT2 = 0 - INT3 - INT4 ;
     !Non-arbitrary loadings
  L2 = 3 -  L3 - L4;

  OUTPUT: stdyx tech1 tech4 SVALUES;
```