Wu, A. D., Liu, Y., Gadermann, A. M., & Zumbo, B.D. (2010). Multiple-indicator Multilevel Growth Model: A Solution to Multiple Methodological Challenges in Longitudinal Studies. *Social Indicators Research: International Interdisciplinary Journal for Quality of Life Measurement, 97*, 123-142.

Best wishes,
Bruno

Professor Bruno D. Zumbo, Ph.D.
University of British Columbia
Scarfe Building, 2125 Main Mall
Department of ECPS
(Program Area: Measurement, Evaluation, & Research Methodology)
Vancouver, B.C.
CANADA V6T 1Z4
e-mail: bruno.zumbo@ubc.ca
web page: http://www.educ.ubc.ca/faculty/zumbo/
Phone: (604) 822-1931
Fax: (604) 822-3302
****

Soc Indic Res (2010) 97:123–142
DOI 10.1007/s11205-009-9496-8

Multiple-Indicator Multilevel Growth Model: A Solution to Multiple Methodological Challenges in Longitudinal Studies

Amery D. Wu · Yan Liu · Anne M. Gadermann · Bruno D. Zumbo

Accepted: 9 February 2009 / Published online: 4 June 2009

Abstract This paper described the versatility of the multiple-indicator multilevel (MIML) model in helping to resolve four common challenges in studying growth using longitudinal data. These challenges are (1) how to deal with changes in measurement over time and investigate temporal measurement invariance, (2) how to model residual dependence due to the nested nature of longitudinal data, (3) how to model observed trajectories that do not follow well-known functions commonly discussed in the methodology literature (e.g., a linear or quadratic curve), and (4) how to decide which predictors are relatively more important in explaining individuals' change over time. With an example of psychological well-being from the Wisconsin Longitudinal Study, we illustrated how the four methodological challenges can be resolved using the 3-phase MIML procedures and the Pratt's importance measures.

Keywords Growth and change · Quality of life · Latent growth modeling · Measurement invariance · Pratt's measures · Psychological well-being · Longitudinal studies

In the past several decades longitudinal designs for studying individuals' growth and change have slowly become popular in the area of psychological well-being (e.g., Mroczek and Spiro 2005). However, statistical techniques such as hierarchical linear modelling (HLM) and structural equation modelling (SEM), which have been developed to analyse repeated measures and longitudinal data, are still underutilized for studying psychological well-being. The advantage of using these techniques is that one can investigate individual and average trends in data collected over days, weeks, months or years. For example, in a 15 year longitudinal study, Lucas et al. (2003) report that on average people adapted back to their initial level of well-being after experiencing marital transitions; however, there were individual differences, and many individuals showed no adaptation at all. In a later

A. D. Wu · Y. Liu · A. M. Gadermann · B. D. Zumbo (✉)
Department of ECPS, University of British Columbia,
2125 Main Mall, Vancouver, BC V6T 1Z4, Canada
e-mail: bruno.zumbo@ubc.ca

study, Lucas et al. (2004) investigated the effect of unemployment on individuals' well-being and found that after a strong reaction to unemployment, individuals' life satisfaction shifted back toward their former baseline level, yet on average the shift was not complete (even after finding work again). Clearly, studies of the predictors and correlates of change using HLM and SEM techniques are producing interesting and important findings. We believe that a number of commonly-encountered methodological challenges have impeded the uptake of these methods in day-to-day research.

In the context of an example from psychological well-being, this paper illustrates how the multiple-indicator multilevel (MIML) growth model that incorporates Pratt's importance measures can help to resolve four commonly-encountered methodological challenges with longitudinal data. These challenges are (1) how to deal with changes in measurement over time and investigate temporal measurement invariance, (2) how to model residual dependence due to the nested nature of longitudinal data, (3) how to model observed trajectories that do not follow the well-known functions discussed in the methodology literature (i.e., linear, quadratic, or cubic curves), and (4) how to decide which predictors are relatively more important in explaining individuals' change over time.

This paper is organized as follows. First, we provided a detailed description of the MIML model (Chan 1998; Muthén and Muthén 1998–2007) with a focus on describing its versatility in solving data-analytical challenges resulting from common pitfalls in longitudinal designs. Second, the procedures for modeling MIML were demonstrated with an outcome measure of psychological well-being from the Wisconsin Longitudinal Study (WLS, WLS Handbook 2007). This dataset provides real examples for many data-analytical challenges in modeling change. Next, we described and applied a very useful method, Pratt's importance measures, to answer an often-asked question: which variable is relatively more important in predicting individuals' change over time. Finally, strengths of the MIML model were summarized, and the limitations were addressed in the Discussion section.

1 The MIML Model

The assessment of individual change requires, by definition, repeated measurements of the construct of interest across multiple time points. From a statistical point of view, these repeated measures are, by design, nested within individuals. Over the past several decades, HLM techniques have become an increasingly popular method for modeling nested data (Goldstein 1995; Hox 2002; Raudenbush and Bryk 2002; Singer and Willett 2003). In a conventional HLM analysis of change, the level-1 analysis directly models the intra-individual's change in the observed outcome over time, and the level-2 analysis models the inter-individuals' differences in their growth. Unlike HLM, the MIML model studies the growth curve of the "latent variable" created from multiple observed indicators via SEM techniques, entailing an extra level (a measurement model) at the foundation of the model. In this section, we describe the structure of the MIML model with an eye to its versatility.

Conceptually, the structure of the MIML model consists of three levels. Level-1, the *measurement model*, defines the scaling relationship between the latent variable, of which the change over time is studied, and the multiple observed indicators. Level-2, the *latent growth model*, captures the intra-individual change in the latent variables over time. Level-3, the *inter-individual model*, predicts the inter-individual differences in their growth.

Table 1 The equations for the 3-level MIML model

Level-1 measurement model	
$Y_{ijt} = \tau_{jt} + \lambda_{jt}F_{it} + r_{ijt}$	(1)
i = individuals, j = observed indicators, t = time points	
Y_{ijt} = observed response variables/indicators	
τ_{jt} = intercept of indicators	
λ_{jt} = factor loadings	
F_{it} = latent factor score across time points	
r_{ijt} = residual for Y_{ijt}	
Level-2 latent growth model (intra-individual model)	
$F_{it} = \eta_{0i} + b_t\eta_{1i} + \varepsilon_{it}$	(2)
η_{0i} = intercept growth factor	
η_{1i} = slope growth factor	
b_t = time score	
ε_{it} = residual for F_{it}	
Level-3 growth prediction model (inter-individual model)	
$\eta_{0i} = \alpha_0 + \gamma_0 X_i + \varsigma_{0i}$	(3)
$\eta_{1i} = \alpha_1 + \gamma_1 X_i + \varsigma_{1i}$	(4)
X_i = time-invariant predictors	
γ_0 and γ_1 are regression coefficient of predictors	
ς_{01} = residual for η_{0i}	
ς_{02} = residual for η_{0i}	

Using the notation and approaches developed by Bengt Muthén and Linda Muthén, and implemented in the Mplus software (Muthén and Muthén 1998–2007), Table 1 provides the mathematical equations for the MIML model and Fig. 1 depicts the equations graphically.

1.1 Level-1: The Measurement Model

The addition of a measurement model to a traditional multilevel growth model brings several advantages (see Eq. 1; Fig. 1). First, it simultaneously incorporates multiple indicators into one model, which, as a result, allows inferences to be made about the latent variable rather than the observed variables. The latent outcome variables, F_{T1}, F_{T2}, F_{T3}, and F_{T4} are depicted as ovals in Fig. 1. A latent variable is a mathematically created variable that is constructed by accounting for the interrelationships among the observed indicators. The observed indicators define the operational meaning of the latent variable. The measurement model allows one to partition random variance and systematic measurement variance (i.e., a second dimension which is unintentionally measured) from the true score variance. In other words, incorporating multiple indicators enables researchers to model growth and change in the true scores by accounting for the random and systematic measurement errors.

Furthermore, embedding a measurement model allows for the investigation and calibration of measurement invariance over time. The issue of *temporal measurement invariance* is often neglected by applied researchers who carry out longitudinal studies. Investigation of temporal measurement invariance examines whether the metric of the outcome variable, measured at multiple time points on the same group of individuals, remains the same. The purpose is to examine whether the scores of the outcome measure are confounded by a temporal change in the scale of measurement. In the MIML model, the

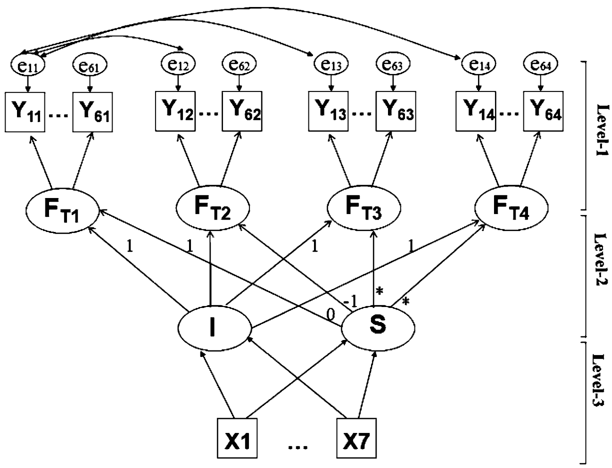

Fig. 1 A conceptual path diagram for the 3-level MIML model. *Note* Each of the latent outcome variables F_{T1} to F_{T4} at each time point was measured by six observed variables. e_{11} to e_{64} are residuals for the six indicators at four time points. Some correlations among the residuals were omitted for graphical simplicity. The * sign indicates that the loading for the slope growth factor were freely estimated (i.e., free time scores). I and S are intercept and slope growth factors. X_1–X_7 are predictors for I and S

outcome measures of interest are the latent variables. Thus, measurement invariance is investigated at the latent variable level through its measurement model. Establishing temporal measurement invariance is the prerequisite for analyzing change in the latent growth curve at level-2. Temporal measurement invariance provides evidence that the construct is measured on the same metric over time; hence cross-time comparison in the latent score is warranted so that the results and interpretations are not, at least, biased by lack of measurement invariance.

To date, measurement invariance is widely studied for cross-sectional data in the framework of multi-group confirmatory factor analysis (MG-CFA). The MG-CFA technique is, however, hardly used for testing measurement invariance across temporal groups prior to a growth study being carried out. It is worth noting that, even for cross-sectional MG-CFA, there is not yet full consensus on the necessary conditions for ensuring measurement invariance (see reviews in Cheung and Rensvold 2002; Vandenberg and Lance 2000; Wu et al. 2007). Recent developments, however, have come to an agreement that investigation of measurement invariance should, at the very least, (1) be based on the mean and covariance structure (MACS) of the observed indicators (Little 1997; Meredith 1993; Wu et al. 2007) and (2) meet the condition of *strong invariance*, i.e., cross-group equality of the indicator intercepts and the factor loadings (e.g., Brown 2006; Jöreskog and Sörbom 1989; Little 1997; Meredith 1993; Wu et al. 2007).

It should be noted that Meredith (1993) and Wu et al. (2007) argued that strict invariance is a necessary condition for examining measurement invariance, which requires, in addition to the indicator intercepts and factor loadings, equality of the residual variances across groups. However, a longitudinal design, in which the same subject is measured at multiple time points, usually leads to unequal residual variances across time. Brown (2006) pointed out that the fan-spread shape of indicator variances across time results in failure in testing equal residual variances. Thus, strict invariance appears rather unrealistic for establishing measurement invariance in the case of longitudinal studies. Hence, the present study omitted the discussion and investigation of the equality of the residual variances.

Testing equality of the indicator intercepts and factor loadings across longitudinal samples has somewhat different interpretations from those of cross-sectional samples. An indicator intercept is the value of an estimated observed variable when the latent variable is zero; it is the scaling constant (i.e., location) of the latent variable for the observed indicator. In a cross-sectional MG-CFA, testing the equality of the indicator intercepts is an investigation of whether the scaling constant remains the same for the observed variables *for all groups*. In a longitudinal study, however, testing equality of the indicator intercepts is an investigation of whether the scaling constant remains the same for the observed indicators *across all time points*.

In a cross-sectional MG-CFA, a loading represents the expected change in an observed variable per unit change in the latent variable. Testing of equality of the factor loadings is an investigation of whether the measurement unit (i.e., scale) of the latent variable remains the same for the observed indicators *for all groups* (see Wu et al. (2007) for a detailed explanation). In a longitudinal study, however, testing equality of the factor loadings is an investigation of whether the scaling unit of the latent variable remains the same for the observed variables *across all time points*.

Changing the observed indicators over time is one of the apparent and frequent sources of lack of measurement invariance. Researchers often modify their instrument by changing the contents and/or the response format. An example of such practice might be using a 4-point Likert response format at time one and a 6-point response format at time two, in addition to dropping or adding items to the questionnaire. Therefore, a growth study based on the observed total score of the modified indicators would be meaningless because cross-time total scores are clearly on different metrics (i.e., the range of the total score would be different across time). Fortunately, embedding a measurement model may resolve this problem through modeling change in the latent variable. Given a good model-data fit, temporal measurement invariance may still hold for the latent variable if the components for scaling the latent variable, the intercepts and loadings, are constrained to be equal. To achieve this, a transformation of the indicator scores is necessary prior to the invariance constraint. This technique will be explained in Sect. 2.

Another advantage of using the MIML model in a growth study is its capacity for dealing with residual dependence. In a longitudinal study, the design of repeated measures of the same individuals may lead to some residuals being dependent across temporal groups. The residual is that part of the observed score that is not accounted for by a statistical model (i.e., the difference between the observed score and the model predicted score). For the MIML measurement model, the residual of an observed indicator at a specific time point, r_{ijt}in Eq. 1, is the difference between the observed score and the predicted factor score for that specific time point, i.e., $r_{\text{ijt}} = Y_{\text{ijt}} - (\tau_{\text{jt}} + \lambda_{\text{jt}} F_{\text{it}})$.

Temporal residual dependence can occur in multiple ways. For example, it may occur among the same observed indicators measuring the same latent variable across time. To be specific, suppose the first indicator in Fig. 1 measures autonomy (one of the six observed indicators for psychological well-being) and its score was observed at four time points denoted as Y_{11}, Y_{12}, Y_{13}, and Y_{14}. Their corresponding residuals, denoted as e_{11}, e_{12}, e_{13}, and e_{14}, are very likely to correlate with one another even after controlling for their corresponding latent scores of psychological wellbeing (i.e., F_{T1}, F_{T2}, F_{T3}, and F_{T4}, respectively). Researchers should allow for the possibility of the residual dependence and model them appropriately to avoid model misspecification. Embedding a measurement model allows the researchers to specify various patterns of residual covariances according to their theory or specific research design.

1.2 Level-2: The Latent Growth Model (Intra-Individual Model)

The latent variables that define individuals' change over time are referred to as the *intercept growth factor* and the *slope growth factor*[1] (Muthén and Muthén 1998–2007). The intercept growth factor (denoted as η_{0i} in Eq. 2 and depicted as the "letter I in an oval" in Fig. 1) represents the estimated status of individuals' growth curve at the time point when the time score is zero (i.e., b_t equals zero in Eq. 2), which is usually specified to be at the first time point (time when the data were first collected) to indicate the estimated initial status of the latent outcome variable. The loadings (i.e., weights) of the intercept growth factor for each of the latent outcome variables (F_{T1}–F_{T4}) are fixed to be one to indicate that the prediction of F_{T1}–F_{T4} would begin with the constant of "1 multiplied by the score of the intercept growth factor", which will return to the score of intercept growth factor itself. The variance of the intercept growth factor indicates how diverse individuals are in their estimated initial value of the latent outcome variables. A small variance indicates that individuals are homogeneous in their initial status. Likewise, the mean of the intercept growth factor is the average of the latent outcome variable across individuals when the time score is zero.

When a linear curve is modeled, the slope growth factor (denoted as η_{1i} in Eq. 2 and depicted as the "letter S in an oval" in Fig. 1) represents the increase in the latent outcome variable for a time score increase of one unit (i.e., the constant growth rate of each individual over all time points). The mean of the linear slope growth factor indicates the average growth rate over individuals. The variance of the slope growth factor shows how diverse the individuals are in terms of their growth rate. The covariance between the intercept and growth factors indicates how the initial status and the growth rate are related. For example, a negative covariance indicates that the higher the individual's initial status, the slower the individual's growth rate is.

Time scores denoted as b_t in Eq. 2 (i.e., loadings for the slope growth factor), in essence, are the weights assigned to the slope growth factor in order to predict the latent outcome variable at a specific time point. They are parameters in the MIML model and can either be fixed or freely estimated to determine the shape of the growth curve. Note that a minimum of four time points is recommended for using free time score to capture the nonlinear growth,[2] and two of the time scores need to be fixed for the purpose of model identification; typically, the scores are set to be 0 and 1 for the first two time points ($b_1 = 0$ and $b_2 = 1$). Fixing the first two time scores to 0 and 1 specifies the time elapsed between the first and second time points to be the unit time interval for interpretation.

For example, time scores can be fixed to 0, 1, 2, and 3 to specify a linear growth curve for data collected every 6 months at four time points. Time scores of 0 and 1 are fixed to specify that the time unit of interpretation is 6-months, and aid in model identification.

[1] Other growth factors can also be modeled. For example, with sufficient time points, a higher order polynomial curve (such as a quadratic growth factor) can be incorporated to capture the non-linear trend in the observed trajectories.

[2] A minimum of four time points is recommended for growth models for two reasons. It is inflexible to make the model identify enough parameters in the growth model with less than four time points. Also, data with four time points give more power. See Muthén (1999) at http://www.statmodel.com/discussion/messages/14/20.html#POST16727.

Time scores of 2 and 3 are fixed to indicate a constant change across time points with equal time intervals in between. Following the same logic, time scores can be fixed to 0, 1, 2, and 6 to specify a linear growth curve for data collected at four time points: at the inception, 6, 12, and 36th months. Note that the choice of unit of time does not always have to occur between the first two time points. The choice should be guided by the researcher's substantive interest, longitudinal research design, or interpretational convenience.

Studying growth via a SEM technique allows a researcher to better capture the growth curve of the observed data. That is, freeing the time scores allows the model to better trace the data curve rather than imposing a pre-specified curve that may turn out to be poor-fitting when the observed data show no clear patterns or follow no familiar curves. As aforementioned, two time scores need to be fixed to 0 and 1 to set the unit of time. As an example, let us consider data collected at four time points with 6-month equal intervals. If the first two time scores are fixed to 0 and 1 and the last two are freely estimated to be 3.5 and 0.2, they can be interpreted as follows; if the amount of change during the first 6 months was scaled to be 1, the expected change at the third time point would be 3.5 and the expected change at the fourth time point would be 0.5. In other words, compared to the first interval of growth rate of 1, we expect a growth rate 2.5 times faster than the first interval during the second interval ($3.5 - 1 = 2.5$), but an even faster decrease in the growth rate of -3.3 during the third interval ($0.2 - 3.5 = -3.3$). These two free time scores reveal that the growth curve may not fit well to a particular known function such as a linear or quadratic; it reached a high peak at the third time point but showed a sharp decline at the fourth time point.

It is very important not to interpret the mean of the slope growth factor as a constant rate of change over all time points or over the study period, but as the rate of change for a time score change of one. Namely, the slope growth factor mean is the change in the latent outcome variable for a one unit change in the time score. So, if the unit time interval is scaled to occur between the first and second time points (6 months in our example) by fixing the first two time scores to 0 and 1 and freeing the last two, the growth factor mean is the change in the outcome variable for the first 6 months. Thus, the growth factor mean is the estimated mean difference between the latent outcome variables at the first and second time points. If the unit time interval is scaled to occur between the first and last time points (24 months in our example) by fixing the first and last time scores to 0 and 1 and freeing the middle two time scores, the growth factor mean is the change in the outcome variable for the 24 months.

Along these lines, the growth factor mean can be used to predict the means of the latent outcome variables at different time points. If the time score for the third time point is estimated to be 0.5, and the growth factor mean to be 1.8 while fixing the first two time scores to 0 and 1, using Eq. 2, $F_{\mathrm{it}} = \eta_{0\mathrm{i}} + b_{\mathrm{t}}\eta_{1\mathrm{i}} + \varepsilon_{\mathrm{it}}$, the predicted latent outcome mean at the third time point would be $0.5 \times 1.8 = 0.9$ (assuming that the estimated mean of the intercept growth factor $\eta_{0\mathrm{i}}$ is zero).

1.3 Level-3: The Growth Prediction Model (Inter-Individual Model)

The growth prediction model is formulated by including time varying and/or time-invariant predictors into the model to examine the relationship between the predictors and the intercept growth factor, as well as between the predictors and the slope growth factor (see Eqs. 3 and 4 and Fig. 1). Time-invariant predictors depict individuals' static status (e.g., gender) whereas time-varying predictors are variables of which the values vary across time hence may have different prediction on the growth factors (e.g., health condition). In addition, the SEM approach allows the flexibility of incorporating both observed and latent predictors. For instance, a researcher may be interested in observed predictors such as gender and age as well as latent predictors, such as personality traits that are indicated by five subscales (e.g., the Big-five personality approach).

2 An Illustration of the 3-Phase MIML Application to Psychological Well-Being

In this section, we will use an example of psychological well-being to illustrate the application of the MIML technique in analyzing longitudinal data. First, we will briefly describe the design of the Wisconsin Longitudinal Study (WLS). Next, we will highlight, with the WLS data as an example, data analytical challenges that are often encountered when analyzing longitudinal data. Lastly, along the way to illustrating the 3-phase MIML modeling strategy, we will provide solutions for each methodological challenge.

2.1 Wisconsin Longitudinal Study (WLS) Design

The Wisconsin Longitudinal Study (WLS) sample includes individuals who graduated from Wisconsin high schools in 1957. This sample is broadly representative of white, non-Hispanic American men and women who have completed at least a high school education. The four waves of psychological well-being data used in this study were collected in the following ways. In 1992, a telephone interview started the data collection and was followed by a mail survey 6 months later. The same data collection procedures were repeated in 2002 leading to four waves of data collection. The sample sizes for both the 1992 telephone interview and mail survey are around 8,500 respondents. The 2002 telephone survey only interviewed a subsample of 545 respondents. The sample size for the 2002 mail survey is 6,297 respondents. Most respondents were around 52 or 53 years old during the first telephone interview in 1992.

The latent outcome variable, psychological well-being was indicated by the six adapted subscales of Ryff's Scales of Psychological Well-Being (RPWB): (1) autonomy, (2) environmental mastery, (3) personal growth, (4) positive relations with others, (5) purpose in life, and (6) self-acceptance (Ryff 1989; Ryff and Keyes 1995). The original RPWB has a total of 120 items, with 20 items for each subscale. The WLS study adapted the RPWB such that the number of items and response scale points varied across data collections. Table 2 provides a description of the changes in the scales used across the four data collections. These scaling changes in the number of items and the response format have

Table 2 Description of changes in the RPWB across four waves of data collection

	1992 phone	1992 mail	2002 phone	2002 mail
Number of items per subscale	2	7	3	5 or 6
Total number of items across the six subscales	12	42	18	36
Number of scale points per item	7	6	6	6
Subscale total range	0–12	0–35	0–15	0–30

Note: *RPWB* Ryff's scales of psychological well-being raised some challenges in using the scores to study change. The last row of Table 2 shows that the amount of psychological well-being was measured on different metrics at each time point, making the comparison impossible without further statistical adjustment.

The predictors for the growth factors included the respondent's sex, their overall high school rank, and their scores on the Big Five personality traits (McCrae and John 1992) of extraversion (EXTRAV), agreeableness (AGREEABL), conscientiousness (CONSCO), neuroticism (NEURO), and openness (OPEN). For the present study, we used each personality trait as individual predictors for the growth factors. If preferred, a latent predictor variable of "personality" can be created and used as a single predictor instead.

In this section, we will describe the 3-phase MIML modeling procedures. The first phase includes the level-1 measurement model only. The second phase includes level-1 and level-2 models, i.e., adding the latent growth model to the measurement model. The third phase includes all three levels of models, i.e., adding the growth prediction model to the level-2 model. Figure 1 presents the path diagram of the 3-level MIML model. Note that the commonly used Chi-squared ratio test was not chosen for examining model fit because this index is affected by large sample size, such as that in the WLS, and easily becomes statistically significant when even trivial misfit occurs (Brannick 1995; Brown 2006; Cheung and Rensvold 2002; Kelloway 1995; Wu et al. 2007). Another disadvantage of the Chi-squared test in comparing model fit is that it always decreases when more parameters are added. Therefore, there is a possibility to choose a model with more parameters that are really unnecessary. Instead, we used the comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean residual (SRMR) (cf., Hancock and Lawrence 2005; Wu et al. 2007). Although the optimal cut-offs for good fit depend on a variety of factors such as model complexity (Browne and Cudeck 1992; Hu and Bentler 1999; Marsh et al. 2004), in broad strokes, RMSEA ≤ 0.08, SRMR ≤ 0.05, and CFI ≥ 0.90 are considered as good fit. Mplus 4.0 (Muthén and Muthén 1998–2007) was used for the data analyses. The appendix provides the Mplus syntax for the final model with all three levels—Mplus code for the level 1 and level 2 models can be created by selecting the appropriate code from the three-level model.

2.2 Phase One: Level-1 Model Only

As shown in Fig. 1, *Y* represents the observed indicators, and F_{T1} to F_{T4} are the latent variables representing psychological well-being at four time points. Psychological well-being at each time point was measured by six indicators (i.e., the six subscales of RPWB). In our example, the loadings of the first observed variable were fixed to 1 to set the scale of the latent variable (quantifying the latent variable) and, as explained above, the remaining five were constrained to be equal across the four time points. The intercepts for each indicator (τ_{ijt}) were all constrained to be equal over four time points. Two of the methodological challenges and their solutions will be addressed in this phase.

2.2.1 Solution to Scaling Change

The problem of scaling change can be solved by measurement invariance constraints based on the mean and covariance structure of the Z scores. In the WLS, the six indicators were modified across waves of data collection. Items were dropped from or added to the subscales from time to time (see the 'Number of items per subscale' of Table 2). The number of scale points for the Likert response was also modified (see Table 2). As a result, the ranges of the six subscale totals vary greatly across time points (i.e., 0–12 to 0–35, see Table 2). In a traditional growth study, the subscales scores are summed up to a total score representing individuals' amount of psychological well-being. The large difference in the range of subscales makes the cross-time comparison of the total scores impossible. It is not surprising if researchers find a poor model fit when examining strong invariance based on these observed scores with large range differences.

A normative transformation (i.e., standardization to Z scores)[3] is one appealing way to solve this problem when used with a latent growth model. The normative transformation converts the total score of each subscale into a Z score with a mean of zero and standard deviation of one. Z scores make the subscale totals over time "appear" to be on the same metric (−3 to 3) and seemingly comparable over time. However, this solution is not appropriate for studying observed trajectories because standardization may change the shape of the observed trajectories of the original raw scores (Willett et al. 1998). This is because, as mentioned earlier, unequal cross-time variances, hence SD, of the outcome variable is very common. Standardizing the subscale totals by dividing unequal SDs may distort the relative ranking of the cross-time raw subtotals, leading to a change in the shape of the observed trajectories. Although standardization transforms the cross-time raw subtotals into a "seemingly common" Z-scale and makes the comparison of the grand total "appear" feasible, it may greatly distort the temporal pattern of the observed trajectories.

Nonetheless, the problem of standardization in distorting observed trajectories will not occur if the study of the growth trajectory is at the latent variable level. This is because standardizing does not alter the overall distribution of the raw subscale totals. That is, the overall distributions of the raw subscale totals and of the Z scores will remain the same. For a latent growth model like the MIML model, the actual data for the measurement model is the mean and covariance structure (i.e., MACS) among the indicators (i.e., subscale totals) rather than the raw scores per se. Despite the fact that the means and covariances may change in magnitude, the structure will remain identical whether it was calculated based on the raw scores or the Z scores. Namely, individuals' scores on the latent outcome variable created based on the mean and covariance matrix of the Z scores would remain consistent to those of the raw scores (Cronbach 1990, p. 121; Gorsuch 1983, p. 299).

The use of the variance and covariance structure of the Z scores makes the investigation of strong invariance sensible without the price of distorting the observed trajectory. Since the observed indicators are now on the same metric of Z scores across time, testing the equality of the intercepts and loadings is now, at least, possible in terms of the face values of the observed metric.

[3] Readers should not confuse "normative" transformation with "normalized" transformation. Normative transformation is a linear transformation; it standardizes the raw scores into Z scores by subtracting the raw score from the mean then dividing by the standard deviation. Normalized transformation, on the other hand, stretches a distribution to make it nearly normal and spreads the data points in both tails of distribution, which is usually a non-linear transformation (Gorsuch 1983, p. 299). The normalized scores will affect factor analysis because the overall distribution of transformed scores will be different from that of the original raw scores. Therefore, normalized transformation is not recommended as a solution to the problem of changes in the observed score scaling.

2.2.2 Solution to Dependent Residuals

The problem of residual dependence can be solved by specifying correlated residuals. The SEM nature of the MIML model allows researchers to specify the residual relationships according to the data collection design and the substantive theory. In the present study, we specified the residuals of the same subscales to be correlated across the four time points. For example, the residual among the first indicators should all be correlated. That is, e_{11}, e_{12}, e_{13}, and e_{11} are mutually correlated, so are the four residuals among the rest of the five indicators. Figure 1 demonstrates these correlated relationships at the level-1 model. The double-headed arrows between residuals of the observed Y variables indicate the residual correlations. Note that in Fig. 1, for the simplicity of illustration, we only graphed correlations between the residual of the first indicator of the first time point with those of the other three time points.

The fit indices of the level-1 measurement model (with measurement invariance constraint and residual covariances as we specified) revealed that psychological well-being was aptly measured by the six indicator subscales and was measured invariantly across time, CFI = 0.959, RMSEA = 0.042, SRMR = 0.036. The establishment of measurement invariance qualifies the score comparison across the latent outcome variable at level-2 in the next phase. Note that the residual correlations as we specified were all statistically significant with sizes ranging from 0.278 to 0.832. If the residual dependence had not been modeled, the model would have fitted the data poorly with CFI = 0.869, RMSEA = 0.07, SRMR = 0.047 and been disqualified for phase two.

2.3 Phase Two: Level-1 and -2 Model

Two growth factors, the intercept and growth factors as represented by I and S in an oval, respectively in Fig. 1, are used to model the growth curve of the latent variable of psychological well-being. The growth factors are treated as latent variables/factors. For the present data, results show that phase two model fit the data well, CFI = 0.959, RMSEA = 0.042, SRMR = 0.037. However, the fit indices were almost identical to those of phase-one model indicating that adding the growth factors to model the change in psychological well-being did not help explain the data pattern much.

2.3.1 Solution to Observed Trajectories Failing to Follow a Known Function

This problem can be solved by using free time scores. As described earlier, the loadings of the slope growth factor, which are also called time scores, are important parameters of a latent growth model because they function to capture the shape of the growth curve. In the MIML model, the free time scores are recommended if they better capture the observed growth trajectories and yield better fit to the data. For the WLS data, we used the proportion scores to give us a rough guide for the shape of the observed growth curve (but not the actual "amount" of change across time, see Fig. 2). The proportion scores were calculated by taking the ratio of individuals' total scores at one time point to the maximum score of that time point. As we explained earlier, due to scaling change, it would be nonsensical to graph the mean curve based on the observed total scores, so would the Z scores that may distort the shape of the original curve. Figure 2 shows that the growth curve declines to the second time point, rises to the third time point, and declines again to the last time point.

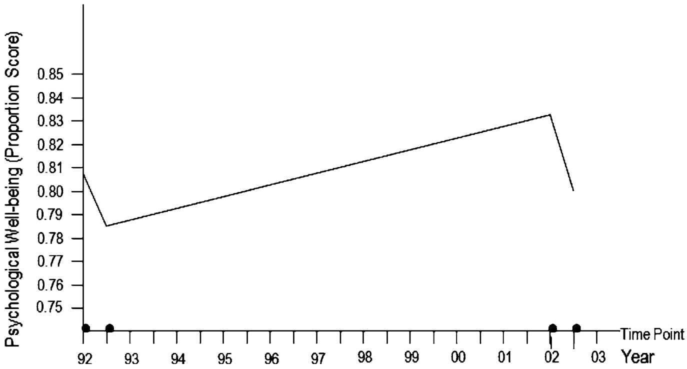

Fig. 2 The observed mean growth trajectory of psychological well-being. *Note* The observed trajectory was graphed based on the mean proportion score for four time points

It should be noted that the proportion score is an approximate alternative for sketching the unknown observed trajectories as if they had been measured on the same metric across time. First, when the proportion scores are compared among the individuals at one specific time point, the relative ranking remains the same. Second, when the proportion scores are compared across time points, they can capture the trend of the change for a specific individual given that the difficulty of endorsing or getting the items right remains similar across time. For example, if Johnny scores 7 points on a mathematics test with a maximum of 10 at grade one and 17 points on a test with a maximum of 20 at grade two and the two tests are of the same difficulty level with regard to his age, the increase of the proportion score from 0.7 of grade one to 0.85 of grade two could be an sufficient indication of Johnny's progress. Such cross-time comparison is meaningless if the age-appropriate difficulty has changed because it may distort the unknown observed trajectories.

Although the proportion score growth curve in Fig. 2 may appear somewhat volatile due to the graphical scaling on the Y axis, the increase/decrease is, in fact, pretty trivial (difference less than 0.05 in proportion). This indicates that individuals' psychological well-being remained considerably stable during the study course. The slight drops from the first to the second and from the third to the fourth time point may simply be a measurement artefact- a reflection of the drop in the social desirability from the telephone survey to the mail survey. In real research settings the small change may not be of substantive interest; however, for demonstration purposes, we used free time scores to model the somewhat zigzag trajectory. Given the observed pattern, we fixed the first two time scores to be 0 and -1 and freed the last two to be estimated. In particular, we found that setting the second time score to be -1 helped the estimation to converge. A fixed value of -1 (instead of 1) captured the drop at the second time point as shown in the proportion score growth curve in Fig. 2. Specifying the sign helped the model to converge. This goes to show that free time scores should not be used entirely thoughtlessly. Also, to help the estimation converge, researchers can specify the starting values for the free time score in terms of the size and sign based on their theory and/or the observed growth curve.

Table 3 shows the selected Mplus output for the phase two model. Note that all the loadings of the intercept growth factor were fixed at 1 and the first two loadings of the slope growth factor (i.e., the first two time scores) were fixed at 0 and -1. The following describes the freely estimated parameters of the level-2 latent growth model.

Table 3 Selected Mplus output for level-2 MIML model

	Estimates	SE	Est./SE
I			
F1	1.000	0.000	0.000
F2	1.000	0.000	0.000
F3	1.000	0.000	0.000
F4	1.000	0.000	0.000
S			
F1	0.000	0.000	0.000
F2	−1.000	0.000	0.000
F3	0.773	0.056	13.739
F4	−0.581	0.046	−12.553
Intercept			
F1	0.000	0.000	0.000
F2	0.000	0.000	0.000
F3	0.000	0.000	0.000
F4	0.000	0.000	0.000
Means			
I	0.000	0.000	0.000
S	0.000	0.003	0.023
Variances			
I	0.033	0.002	17.599
S	0.044	0.004	10.066
S with I	−0.036	0.002	−23.384
Residual variances			
F1	0.117	0.003	38.165
F2	0.086	0.006	15.255
F3	0.005	0.001	5.643
F4	0.119	0.004	28.769
Estimated means for the latent variables			
F1 0.000; F2 0.000; F3 0.000; F4 0.000; I 0.000; S 0.000			

First, the mean of the intercept growth factor (on the metric of −3 to 3) was estimated to be zero showing that the average score of psychological well-being at the outset is zero. This result makes sense because the indicators were transformed to Z-scores, which were all, by definition, centered at zero. Second, the mean of the slope growth factor was estimated to be close to zero, too, indicating little to no change during the unit time interval of the first 6 months. Third, the variances of the intercept growth factor and the slope growth factor were estimated to be 0.033 and 0.044, respectively. Despite being very small, they are statistically significant suggesting that individuals are diverse in their initial status and the first 6-month change. Fourth, the covariance of the intercept and slope growth factors was estimated to be −0.036 (i.e., correlation of −0.927), indicating that the initial status is highly negatively correlated with the first 6-month change in psychological well-being. Fifth, the last two loadings of the slope growth factor (time scores) were estimated to be 0.773, and −0.581, which correspond to the zigzag growth pattern shown in the observed mean growth trajectory. Finally, the residuals for the four well-being latent factors F_{t1}–F_{t4}, (ε_{it}in Eq. 2, not shown in Fig. 1), were all significant indicating that over and above the intercept and growth factors, other factors are in the play and needed for explaining individuals' psychological well-being over time. These results correspond to the fit indices showing no to little improvement after adding the growth factors.

For the present data, both the estimated means of the intercept growth factor and slope growth factor are extremely close to zero (values smaller than 0.001), and were unable to be shown in the Mplus output that only displayed up to the third decimal). This suggests that although the growth curve of psychological well-being appeared somewhat nonlinear on a graph, it remained very stable showing no considerable change. Our finding is in accordance with the notion that well-being is relatively stable as postulated, for example, in the adaptation model (cf. Frederick and Loewenstein 1999).

2.4 Phase Three: Level-1, -2 and -3 Model

In the third phase of MIML, we added *standardized* predictors for the growth factors. Hence, the MIML is a 3-level model with level-1 measurement model, level-2 latent growth model, and level-3 growth prediction model. At level-3, the MIML model investigates the effects of individuals' background variables on individuals' growth curve. This is shown by the arrows going from X1–X7 to I and S, where Xs are the time-invariant predictors measured before or at the first time point. The inter-individual predictors in the present study are sex, overall high school rank (HSRANK), and the Big Five Personality Traits (McCrae and John 1992). These time-invariant predictors vary across individuals, but not across time.

All of the fixed and free estimated parameters are the same as described in the second phase except for the newly added regression coefficients for the predictors. The phase three model still fit the data well, although slightly less satisfactory than the phase two model, with CFI = 0.912, RMSEA = 0.05, and SRMR = 0.039. Table 4 reports the selected Mplus output. The results show that except for agreeableness, all the other predictors were statistically significant, indicating that they played a role in explaining the variation in individuals' intercept growth factor (i.e., the initial status) and slope growth factor (i.e., the unit metric in the outcome variable for interpretation; change in the outcome per one unit change in time score; the first 6-month change). Except for neuroticism that had a negative effect on the intercept growth factor, all others had positive effects. The results for predicting the slope growth factor were reversed; except for neuroticism that had a positive effect, all others had negative effects.

2.4.1 Solution to Problem of Ordering the Relative Importance of the Predictors

It is often of theoretical and practical interest to learn which predictor in a regression model is relatively more important. Pratt's relative importance measures, *d,* were developed to order the relative importance of the explanatory variables (Thomas et al. 1998). The equation for calculating the d is given as $(\beta \times r)/R^2$, where β denotes the standardized partial regression coefficient, r is the simple correlation between the outcome variable and predictors, and R^2 is the variance of outcome variables explained by all the predictors. In order to obtain the β weights, the Mplus output command "STANDARDIZED" has to be used. Pearson correlation was obtained by correlating the intercept and growth factors with

Table 4 Selected Mplus output for level-3 MIML model

	Estimates	SE	Est./SE
I			
F2	1.000	0.000	0.000
F1	1.000	0.000	0.000
F3	1.000	0.000	0.000
F4	1.000	0.000	0.000
S			
F2	−1.000	0.000	0.000
F1	0.000	0.000	0.000
F3	0.647	0.034	18.993
F4	−0.440	0.024	−18.457
I ON			
SEX	0.021	0.004	5.563
HSRANK	0.000	0.000	5.698
EXTRAV	0.008	0.000	17.345
AGREEABL	−0.001	0.000	−1.793
CONSCO	0.005	0.001	9.713
NEURO	−0.022	0.001	−30.464
OPEN	0.003	0.000	6.628
S ON			
SEX	−0.036	0.005	−6.873
HSRANK	0.000	0.000	−3.582
EXTRAV	−0.011	0.001	−16.810
AGREEABL	0.001	0.001	1.781
CONSCO	−0.008	0.001	−10.927
NEURO	0.031	0.001	33.013
OPEN	−0.004	0.001	−6.018
S with I	−0.022	0.001	−24.641
Intercepts			
F1	0.000	0.000	0.000
F2	0.000	0.000	0.000
F3	0.000	0.000	0.000
F4	0.000	0.000	0.000
I	0.000	0.000	0.000
S	0.096	0.009	10.678
Residual variances			
F1	0.120	0.003	39.433
F2	0.065	0.003	19.382
F3	0.004	0.001	5.819
F4	0.131	0.003	39.178
I	0.019	0.001	18.953
S	0.035	0.002	14.976
Estimated means for the latent variables			
F1 0.0687; F2 0.0684; F3 0.0686 F4 0.0684 I 0.0685; S 0.0001			

Table 5 Ordering the importance of predictors using Pratt's measures

	Intercept growth factor				Slope growth factor			
	β	r	d	Order	β	r	d	Order
SEX	0.018	0.039	0.0015	–	−0.033	−0.049	0.0032	–
HSRANK	0.061	0.142	0.0181	–	−0.031	−0.117	0.0072	–
EXTRAV	0.258	0.496	**0.2664**	**2**	−0.255	−0.503	**0.2546**	**3**
AGREEABL	0.121	0.403	**0.1016**	**5**	−0.127	−0.422	**0.1063**	**5**
CONSCO	0.233	0.491	**0.2381**	**3**	−0.255	−0.515	**0.2606**	**2**
NEURO	−0.320	−0.486	**0.3239**	**1**	0.331	0.499	**0.3277**	**1**
OPEN	0.147	0.415	**0.1270**	**4**	−0.144	−0.415	**0.1186**	**4**
	$R^2 = 0.480$				$R^2 = 0.504$			

Note: β denotes the standardized beta-weight; r denotes the Pearson correlation. The Pratt's measures d is given by $(\beta \times r)/R^2$. A Pratt's measure less than $1/(2p)$ is considered unimportant where p refers to the number of predictors. Predictors that meet this criterion of importance are in bold font

predictors using the "WITH" command.[4] R-square values are the proportion of variance of the intercept and slope growth factors that were explained by predictors, they are automatically outputted in Mplus.

Using the criteria described by Thomas et al. (1998), variables with a Pratt's measure less than $1/(2p)$, where p is the number of predictors, were regarded as unimportant. Table 5 reports the Pratt's measures for the two growth factors, respectively. The results show that SEX and HSRANK failed to play an important role in prediction individual's growth variation in psychological well-being, whereas NEURO, EXTRAV, CONSCO, AGREEABL and OPEN were important variables for both intercept and slope growth factors. Among the important variables, neuroticism is the most important predictor for both growth factors. The results echoed previous findings that personality variables like neuroticism and extraversion are strong predictors for psychological well-being (e.g., Bostic and Ptacek 2001; DeNeve and Cooper 1998).

3 Closing Remarks

Elaborated from Golembiewski et al.'s study (1976), Chan (1998) pointed out there are three types of change that may occur in longitudinal studies—*alpha, beta, and gamma* change. Alpha change refers to the true score change in a construct given the same construct is measured on the same metric over time. This is an ideal condition for a meaningful study of quantitative change in a construct. Beta change refers to the change in the measurement metric though the same construct is measured over time. This is the

[4] Note that missing data may lead to the estimation of the beta-weights using the ON command and simple correlations using the WITH command being based on different subjects in Mplus. This problem could distort the calculation of the Pratt's measures; i.e., the sum of the Pratt's measures would not add up to one. Instead of using the WITH command, one possible solution is to save the scores of the growth factors I and S and use the simple bivariate correlations obtained in SPSS. To save the scores of the growth factors I and S in Mplus, use the SAVEDATA and SAVE = FSCRES commands for the OUTPUT.

assumption in the present study. Gamma change refers to qualitative change in the conceptual domain of the construct. That is, the meaning of the construct changes over time.

In this illustration, we did not find considerable quantitative change in the WLS cohort's psychological well-being under the premise of no qualitative change in the construct of psychological-well-being. Our study, however, did not rule out the possibility of qualitative change in the score meaning of psychological well-being; that is, the issue of temporal measurement validity was assumed and untouched in this study. Although embedding measurement may sidestep the problems of scaling change and allows for investigation of measurement invariance, it does not examine whether the same theoretical construct has been measured across data collections. In fact, in the present study, the correlations among the four latent psychological well-being scores are fairly low (ranging from 0.133 to 0.522). If the same construct had been measured, these scores should, at least, correlate fairly highly.

Dropping or adding items, as the WLS did, may lead to some content domains of the theoretical construct being underrepresented, overrepresented, or simply misrepresented. Any modification of the instrument may change the meaning of the test scores and result in invalidity in cross-time score interpretation in a growth study. Test scores may have different meanings across time even if the MIML invariance constraint is able to recalibrate them on the same metric at the latent variable level. Our best advice is to select one instrument that would work well over time. An extensive psychometric review of the instrument and a pilot study would certainly help the selection of the instrument. Otherwise, careful attention should be paid to ensure that the same content domains are represented when modifying the instrument. Also, additional evidence should be provided to show that the same theoretical construct has been measured across time.

Note that one cannot rule out gamma change even if one uses the same instrument over time. Respondents may have changed their thinking, emotions, or attitude towards the construct during the studied period. For example, the concept of psychological well-being may have different meanings to the cohort as they grow older. To reiterate, the MIML model is an insufficient tool if gamma change has occurred; in that case, other methods for studying qualitative change should be applied—e.g., a special case of latent class analysis called latent transition analysis (see Kaplan 2008).

In closing, the embedding of a measurement model into a growth model provides a very versatile framework for studying growth and change. In this paper, we described the capacity of the MIML model in solving many practical data analytical problems that often occur in growth studies. The problems of scaling change can be easily solved by modeling growth of the latent variable created by the mean and covariance structure of the Z scores. Built-in measurement invariance constraints investigate whether cross-time scores are warranted for a growth study. Making use of the flexibility of SEM, the concern over residuals dependence can be easily modeled according to the researcher's theory and study design. The issue of unknown observed growth curve can be aptly modeled using free time score of the slope growth factor. The use of Pratt's measures enables researchers to order the importance of the predictors for the growth factors. Despite all these strengths, to date, the MIML model is hardly used in day-to-day data analyses. This paper serves to motivate and promote future use of the MIML model.

Appendix: Mplus syntax for 3-level MIML growth model

```
TITLE: Latent Growth Model for Multiple Indicators Observed over Four Time Points.

DATA: FILE IS WLS_gc_fullZscore.dat;
      FORMAT is 385.F11;

VARIABLE:
     NAMES ARE  Za92p Ze92p Zpg92p Zpr92p Zpu92p Zs92p
                Za92m Ze92m Zpg92m Zpr92m Zpu92m Zs92m
                Za02p Ze02p Zpg02p Zpr02p Zpu02p Zs02p
                Za02m Ze02m Zpg02m Zpr02m Zpu02m Zs02m
                sex hsrank edusuc wksuc finsuc
                famsuc extr92 agre92 cons92 neu92 open92;

      USEVAR = Za92p Ze92p Zpg92p Zpr92p Zpu92p Zs92p
               Za92m Ze92m Zpg92m Zpr92m Zpu92m Zs92m
               Za02p Ze02p Zpg02p Zpr02p Zpu02p Zs02p
               Za02m Ze02m Zpg02m Zpr02m Zpu02m Zs02m
              sex hsrank extr92 agre92 cons92 neu92 open92;

ANALYSIS: TYPE IS MEANSTRUCTURE;

MODEL:

!! Level-1 Measurement Model
  !! Specifying Intercept Equality
           [Za92p Za92m Za02p Za02m] (1);
           [Ze92p Ze92m Ze02p Ze02m] (2);
           [Zpg92p Zpg92m Zpg02p Zpg02m] (3);
           [Zpr92p Zpr92m Zpr02p Zpr02m] (4);
           [Zpu92p Zpu92m Zpu02p Zpu02m] (5);
           [Zs92p Zs92m Zs02p Zs02m] (6);

  !! Specifying Loading Equality
        F1 by Za92p@1
          Ze92p-Zs92p (7-11);

        F2 by Za92m@1
          Ze92m-Zs92m (7-11);

        F3 by Za02p@1
          Ze02p-Zs02p (7-11);

        F4 by Za02m@1
          Ze02m-Zs02m (7-11);

  !! Correlated Residual Errors
        Za92p with Za92m Za02p Za02m;
        Za92m with Za02p Za02m;
        Za02p with Za02m;
        Ze92p with Ze92m Ze02p Ze02m;
        Ze92m with Ze02p Ze02m;
        Ze02p with Ze02m;
        Zpg92p with Zpg92m Zpg02p Zpg02m;
        Zpg92m with Zpg02p Zpg02m;
        Zpg02p with Zpg02m;
        Zpr92p with Zpr92m Zpr02p Zpr02m;
        Zpr92m with Zpr02p Zpr02m;
        Zpr02p with Zpr02m;
        Zpu92p with Zpu92m Zpu02p Zpu02m;
        Zpu92m with Zpu02p Zpu02m;
        Zpu02p with Zpu02m;
        Zs92p with Zs92m Zs02p Zs02m;
        Zs92m with Zs02p Zs02m;
        Zs02p with Zs02m;
!!------------------------------------------------------------------------------
!! Level-2 Latent Growth Model

   !! Specifying Growth Factor I and S and Growth Function
    I  S | F1@0 F2@ -1 F3* F4*;
!!------------------------------------------------------------------------------
!! Level-3 Inter-individual Model
   !! Predicting Growth Factors i and s (should be inactive while using the WITH command)
    I  S ON sex hsrank extr92 agre92 cons92 neu92 open92;

   !! Obtaining Simple Correlations (should be inactive while using the ON command)
    I  S WITH sex hsrank extr92 agre92 cons92 neu92 open92;

OUTPUT:
STANDARDIZED TECH4;
SAVEDATA: File is MIML.dat;
            SAVE = fscores;
```

References

Bostic, T. J., & Ptacek, J. T. (2001). Personality factors and the short-term variability in subjective well-being. *Journal of Happiness Studies, 2*, 355–373.

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior, 16*, 201–213.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Publications, Inc.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods, 1*, 421–483.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing MI. *Structural Equation Modeling, 9*, 235–255.

Cronbach, L. J. (1990). *Essentials of Psychological Testing* (5th ed.). New York, NY: Harper Collins Publishers, Inc.

DeNeve, K. M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin, 124*, 197–229.

Frederick, S., & Loewenstein, G. (1999). Hedonic adaptation. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 302–329). New York: Russell Sage Foundation.

Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD design. *Journal of Applied Behavioral Science, 12*, 133–157.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Hancock, G. R., & Lawrence, F. R. (2005). Using latent growth models to evaluate longitudinal change. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 171–196). Greenwich, CT: Information Age Publishing, Inc.

Hox, J. (2002). *Multilevel analysis techniques and application*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Jöreskog, K. G., & Sörbom, D. (1989). *LISERL 7: A guide to the program and applications* (2nd ed.). Chicago, IL: SPSS.

Kaplan, D. (2008). An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology, 44*, 457–467.

Kelloway, E. K. (1995). Structural equation modeling in perspective. *Journal of Organizational Behavior, 16*, 215–224.

Little, T. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53–76.

Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2003). Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status. *Journal of Personality and Social Psychology, 84*, 527–539.

Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2004). Unemployment alters the set point for life satisfaction. *Psychological Science, 15*, 8–13.

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 overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320–341.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*, 175–215.

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

Mroczek, D. K., & Spiro, A., I. I. I. (2005). Change in life satisfaction during adulthood: Findings from the veterans affairs normative aging study. *Journal of Personality and Social Psychology, 88*, 189–202.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Ryff, C. (1989). Happiness is everything, or is it? *Journal of Personality and Social Psychology, 6*, 1069–1081.

Ryff, C. D., & Keyes, D. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69*, 719–727.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and even occurrence*. London: Oxford University Press.

Thomas, D. R., Hughes, E., & Zumbo, B. D. (1998). On variable importance in linear regression. *Social Indicators Research, 45*, 253–275.

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*, 4–70.

Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology, 10*, 395–426.

Wisconsin Longitudinal Study Handbook. (2007). Edited by Erica Wollmering with Contributions by Wisconsin Longitudinal Study Staff.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation, 12*, 1–26.