Original Article

Longitudinal Changes in Physical Fitness Performance in Youth: A Multilevel Latent Growth Curve Modeling Approach

C. K. John Wang[1]

Do Young Pyun[1]

Woon Chia Liu[1]

B. S. Coral Lim[1]

&

Fuzhong Li[2]

[1] National Institute of Education, Nanyang Technological University, Singapore

[2] Oregon Research Institute, USA

Date of Initial Manuscript Submission: 20 April 2011

Date of Revision 3 Submission: 3 April 2013

Correspondence:
Professor C.K. John Wang, Ph.D.
Physical Education and Sports Science
National Institute of Education
Blk 5 #03-12
1 Nanyang Walk
Singapore 637616
Tel: (65) 67903690
Fax:(65) 68969260
Email: john.wang@nie.edu.sg

Ms 51 12

Longitudinal changes in physical fitness performance in youth: A multilevel latent growth curve modeling approach

Abstract

Using a multilevel latent growth curve modeling (LGCM) approach, this study examined longitudinal change in levels of physical fitness performance over time (i.e., 4 years) in young adolescents aged from 12 to 13 years old. The sample consisted of 6,622 students from 138 secondary schools in Singapore. Initial analyses found between-school variation on fitness test scores with intraclass correlations ranging from .02 to .19. Subsequent multilevel growth curve analyses revealed a quadratic trend of the longitudinal data across five stations of performance tests, with significant within-school student variability in change over time. The result of the multilevel LGCM showed that the there were strong school effects on all the physical fitness performances, in addition to inter- and intra-individual differences.

Keywords

Multilevel latent growth curve, longitudinal change, fitness tracking, fitness testing, LGCM

Longitudinal Changes in Physical Fitness Performance in Youth: A Multilevel Latent Growth Curve Modeling Approach

Introduction

The global increase in the prevalence of obesity and associated diseases has fostered a multi-pronged approach to reduce this epidemic in many countries. One area of focus is to promote lifetime physical activity and reduce physical inactivity among youth through physical education (PE) in school (McKenzie, 2001). Specific recommendations in PE are provided by the government in the United States. For example, the Healthy People 2020 document in the United States recommends that children should be active at least 50% of class time during a PE class (Park et al., 2008).

With the increased attention on PE in promotion of physical activity and reduction of obesity rate, it is not surprising that schools in many countries are using physical fitness testing as the primary form of assessment (Cale and Harris, 2009). However, it is known that other than habitual physical activity, factors such as physical growth, biological maturation, and other developments can influence young people's physical fitness performance (Physical Activity Guidelines Advisory Committee, 2008). Some researchers view fitness testing as invalid and believe that it should be discontinued for various reasons (Cale and Harris, 2009; Corbin et al., 1995; Rowland, 1995). Cale and Harris (2009: 89) pointed out that there has been lack of evidence to support positive roles of fitness testing to 'promote healthy lifestyles and physical activity, motivate young people, or develop the knowledge and skills that are important to a sustained engagement in an active lifestyle.' In contrast, Corbin and his colleagues (1995) has pointed out that fitness testing should play roles of learning about personal fitness and teaching students how to use information derived from its results. However, there are also researchers who endorse the use of physical fitness testing for formative evaluation to further educational goals (Silverman et al., 2008), promotion of physical activity (National Association for Sport & Physical Education [NASPE], 2005), and enhancement of motivation (Wiersma and Sherman, 2008). While there have been debates surrounding fitness testing in schools, a better approach would be to provide empirical evidence on whether there are any effects on the physical fitness across years of study, different schools, and students.

In the literature related to physical fitness testing, there is lack of research on the rate of changes in physical fitness among youth. With repeated measures in the same individual over an extended period of time, the change over time can be investigated. For example, questions like how much does each person change overtime (within-individual change), what are the differences between individuals in change (inter-individual change), and what is the process of change (shape of growth trajectory) can be answered (Singer and Willet, 2003). In addition to the time effect, contextual influences such as classroom, family, and schools can be also modeled using a multilevel approach (Card et al., 2007).

Latent growth curve modeling

The classic method of analysis for longitudinal repeated measures has been the repeated measures ANOVA (analysis of variance). This approach to model change requires more stringent data assumptions such as sphericity and normality (Fan and Fan, 2005). Regarding this, Fan and Fan (2005: 122) noted that unlike the traditional approach, the LGCM provides 'more flexibility in testing different research hypothesis related to the developmental trend, such as heteroscedastic residuals and nonlinear growth patterns.' That is, the repeated measures ANOVA focuses on mean changes, with linear growth determined by the mean changes over repeated measurement at the group level (Fan and Fan, 2005). However, the LGCM approach allows for individual growth curves to be analysed, taking into consideration of the means and covariances of the repeated measures (McArdle, 1988; Meredith and Tisak, 1990; Stoolmiller, 1995) by using factor analytic techniques (Rao, 1958; Tucker, 1958). Rovine and Molenaar (1998: 338) highlighted that 'by considering both simultaneously, one can assess the relative importance of fitting the mean structure and the covariance structure to the overall fit of the model.' In addition, the repeated measures ANOVA considers within/between individual changes as error variance (Voelkle, 2007). However, this measurement error can be modeled and estimated in LGCM. There are many more advantages of LGCM over the repeated ANOVA (for review, see Tomarken and Waller, 2005).

In modeling the factor variances and covariances, the LGCM allows researchers to test a linear, quadratic, cubic, or spline growth trajectory by including additional latent factors to the linear model. For example, given the loadings of the linear slope factor (e.g., 1, 2, 3, 4), the factor loadings in a quadratic function correspond to the squared values of the loadings of the linear slope factor (e.g., 1, 4, 9, 16), and the cubic model corresponds to the cubic values of the linear slope factor (e.g., 1, 8, 27, 64). Figure 1 shows a representation of a three-factor polynomial LGCM (Duncan et al., 2006).

###Insert Figure 1 here###

Multilevel LGCM

It is possible to have two or higher levels of analysis in the LGCM. The goal of Level-1 analysis is to describe the shape of each student's individual growth trajectory. Level-2 analysis focuses on inter-individual differences in change. Here, it can be asked if different students would show different patterns of within-individual change. The rationale for this is that the growth trajectory may vary across individuals as a result of different intercepts and slopes (Byrne, 2006).

In addition, it is also possible to test a level-3 (school) model in the LGCM within the SEM framework (Muthén, 1994). The population of the current study consisted of schools and students within these schools; therefore it is possible that there may be more similarities among students from the same school in general (Duncan et al., 1997). Duncan and his colleagues (2006) outlined a Limited Information Multilevel LGCM analysis (Muthén, 1991, 1994). This method uses an ad hoc estimator within the SEM framework for data with many groups of different sizes. This estimator provides similar results to the Maximum Likelihood estimator.

Purposes of the current study

The purpose of the present study was to use a latent growth curve modeling (LGCM) approach to examine the longitudinal change of physical fitness among Singaporean youth from secondary 1 to 4 (equivalent to Grade 7 to 10). As there were no previous studies in this area and there are multiple indicators of physical fitness performance, hypotheses were not stated. The following research questions were examined instead:

(a) What were effects across years of study in physical fitness among Singaporean youth?

(b) Were there effects across schools in physical fitness among Singaporean youth?

(c) Did the rates of change differ across individuals in physical fitness among Singaporean youth?

Method

Participants and procedures

A total of 6622 students' (3428 boys and 3194 girls) with complete physical fitness data for the four continuous year period were abstracted from 138 secondary schools in Singapore, registered in the Ministry of Education (MOE), the Singapore's school cockpit system. The given number of the participants was randomly selected from each secondary school using a simple random sampling technique. Each student must have had a complete record of the fitness test results for all the four year period from the same school. The participants were from 12 to 13 years old in the first year of secondary school and from 16 to 17 years old in the fourth year of study. Permission for accessing the data was given by the Ministry of Education Educational Research Committee. Research procedures for the study were approved by the Ethical Review Committee of the university.

Measures

***National physical fitness award scheme*.** The National Physical Fitness Award (NAPFA) is an annual test of physical fitness for Singapore students and is part of Singapore's Sports for Life programme. All healthy students from Primary 4 onwards are required to participate in the NAPFA test annually. It is a high stake test, and the protocols of conducting the tests must be followed. All testers must go through two-day training and be certified 'NAPFA Testers' by the Singapore Sports Council. All schools are required to inform MOE PE unit in advance of the test dates, and the officials from MOE perform random inspection during the test days.

The NAPFA is a battery of assessments used to identify cardiovascular fitness, muscular strength, joint flexibility, muscular endurance, muscular power, coordination and speed. These various components of fitness testing are reliable and valid indicators of physical fitness (Giam, 1981; Jackson, 2006). The NAPFA involves a series of six stations:

(a) Maximum number of bent-knee sit-ups in one minute, as a measure of anterior abdominal muscular endurance and strength;

(b) Better of two standing broad jump (SBJ) distances, as a measure of lower limb extensor muscular power;

(c) Better of two sit-and-reach (SnR; floating zero point) forward distances, as a measure of forward trunk flexibility, hip flexion and hamstring muscle stretch;

(d) Maximum number of overhand-grasp regular pull-ups in half a minute for males over 14 years old and overhand-grasp inclined pull-ups in half a minute for males younger than 14 years old and females, as a measure of upper limb muscular endurance and strength;

(e) Faster of two attempts to complete a 4 X 10 m shuttle-run, as a measure of general speed and coordination;

(f) Minimum time taken to 2.4 km run-walk, on a firm and level surface, as a measure of cardiorespiratory endurance (or aerobic fitness) and lower limb muscular endurance.

All of the station tests except 2.4 km run-walk are attempted on the same day, with a 2–5 minute rest period permitted between stations. The 2.4 km run-walk item may be attempted on a different day, although sometimes a 2-week window limit is set. The standards for these six test items vary for the different ages and sex groups. A minimum standard in all of these six tests is required before they can qualify for the gold, silver or bronze awards. This is to ensure that those who qualify will have a desirable level of fitness. Table 1 shows the standards for boys and girls from 13 to 16 years old for each test. The Gold award requires a minimum of 3 points in all six tests with at least 21 points. The Silver award requires at least a total of 15 points with a minimum 2 points in all six test items. The Bronze is awarded with at least 1 point in each test (minimum 6 points). The researchers excluded the pull-ups station from the analysis as there were two different methods for the boys. In addition, as the norms for boys and girls are different, the focus of this study was to identify separate trends for each gender. As such, the analyses for the two genders were separated.

###Table 1 near here###

Data analysis

The descriptive statistics (e.g., means and standard deviations) of the students' physical performance across time were tabulated. Next, the authors followed the four analytic steps outlined by Muthén (1994) who suggests that the four steps should be conducted prior to a full multilevel LGCM:

Step 1: Conventional confirmatory factor analysis of the total covariance structure to determine a rough fit for the proposed model. We fitted linear and quadratic curve to the data and compared the fit. This is a test of fit in terms of the general growth trend. The chi-square test of the model fit is usually inflated. The results of the test are not correct as it has ignored the hierarchical nature of the data. However, this step is still helpful in providing a rough fit of the data to the proposed model.

Step 2: Estimation of between structure variations (intraclass correlation coefficient; ICC) was performed to provide an indication of whether a multilevel analysis is warranted, that is, whether there are school effects. The larger the ICC, the more similar the scores in the same school are, relative to another. If ICC is close to zero, multilevel analysis may not be necessary as the scores are independent of one another. The ICCs were computed using EQS 6.1 for Windows in which the school is the clustering variable, and the repeated measures of the physical fitness scores are the dependent variables. Hox (2002) suggests that ICC coefficients of .05, .10, and .15 be considered as small, medium, and large values, respectively.

Step 3: The pooled within-sample covariance matrix provided in the EQS 6.1 in the analysis of the ICCs was used to estimate the within structure $S_{W.}$ This provides an estimation of individual-level parameters (within-subject). As this estimate is not distorted by the between covariation, it will provide a better model fit compared to S_T (total covariance matrix).

Step 4: An estimation of between structure S_B was conducted using the same model as S_W but with models across school covariation rather than individual-level data. This analysis focused on the differences between schools. The individual-level growth structure should be different from the between components. After the four analytic steps were conducted, the decision was made as to whether a multilevel LGCM was warranted.

Evaluating fit statistics in LGCM

The LGCM analyses were conducted with EQS for Windows 6.1 (Bentler and Wu, 1998). Multiple indices of fit provided by EQS were examined to evaluate the adequacy of the models such as chi-square statistic, the normed fit index (NFI), the comparative fit index (CFI), and the root mean square error of approximation (RMSEA) and its confidence intervals. For NFI and CFI, values less than .95 are taken to reflect acceptable fit (Hu and Bentler, 1999). The RMSEA is based on the analysis of residuals and compensates for the effects of model complexity. For this index, values less than .06 are taken to indicate acceptable fit (Hu and Bentler, 1999).

Results

Descriptive statistics

The means, standard deviations and distributions of the repeated measures of physical fitness of the sample are shown in Table 2. In general, the mean scores of all the stations showed improvement over time.

###Table 2 near here###

Step 1: Conventional confirmatory factor analysis of total covariance structure

In this step, the authors compared the linear and the quadratic models using several absolute and comparative fit indices. The fit indices of the fitted models are presented in Table 3. The higher the fit indices (NFI, CFI) and lower the error residuals (RMSEA), the better the fit of the model to the data. Across the five physical fitness tests, the quadratic models were found to be better fitting than the linear models for most of the stations, except for boys' sit-ups and girls' 2.4 km run-walk, when compared to the linear models. The quadratic trends implied the rate of change of the trajectory tended to peak and then decline or plateau over the four-year period. The linear trajectory reflects a constant change (either increasing or decreasing, or no change) across time. That is, the slope is a straight line over the four-year period. However, the limitation of this conventional technique is that it ignores the hierarchical nature of the data.

###Table 3 near here###

Step 2: Estimation of ICC (effects across schools)

To examine if there are effects across schools on the physical performances, ICCs were computed using EQS 6.1 for Windows. ICCs reflect the degree of association between the school-level and student-level measures. A high correlation suggests that variation between schools in that variable is high in comparison to the variation of that same variable within school. High ICCs may be due to different training programmes for the fitness tests, experience of teachers, and/or length of training of the students in different schools. The authors assumed that the testing protocols would have minimum effect as all testers were certified. As shown in Table 4, all the ICCs ranged from .02 to .19. According to the Hox's (2002) guideline, boys showed small effects for SBJ, shuttle run, and 2.4 km run-walk and medium effects for sit-ups and SnR; girls showed small effects for SBJ and shuttle run, medium effect for sit-ups and 2.4 km run-walk, and large effects for SnR. Since there were some levels of between-school variations, therefore, it was reasonable to proceed to examine the within-structure (step 3).

###Table 4 near here###

Step 3: Estimation of Pooled Within-Sample Covariance (S_W)

The pooled within-level matrices for the five physical performance tests were examined for the boys and the girls separately to examine the within individual change. The S_W analysis is independent of the between-level covariation. It examines the variation in the individual-level parameters only. Therefore, almost perfect fit was obtained for all of the five stations (See Table 5). All the values of NFI and CFI were 1.00.

Step 4: Estimation of Between-Sample Covariance (S_B)

In the fourth step, the authors tested a similar growth structure using the between-level covariance matrices. The focus is on the school-level variation only, that is, differences between schools. The between-level structure is not expected to be similar to the within-level structure as the between components may have a different meaning. The results are presented in Table 5. From the table, it seems that the quadratic curves fitted most of the data adequately according to the NFI and CFI indices. The information gathered from the four initial steps was used to specify a sequence of multilevel covariance structure analyses.

###Table 5 near here###

Limited information multilevel LGCM (effects across years and subjects)

The four initial analytical steps suggested that a Limited Information Multilevel LGCM model should be fitted for both within and between levels, as there were variations at both levels, for both genders. Specifically, Step 2 showed significant school effects, Steps 3 and 4 showed significant within- and between-level structures for boys and girls. Therefore, a Limited Information Multilevel LGCM was fitted to the data of the five physical performance tests for boys and girls separately (see Duncan et al., 2006: 111 for a graphical representation of the Limited Information Multilevel LGCM). In an analysis of each physical performance test, both the between and within-level structures were tested simultaneously using both S_B and S_W. Tables 6 and 7 present the fit indices and parameter estimates for the boys and the girls, respectively.

###Table 6 near here###

###Table 7 near here###

In terms of sit-up performance, the results of the present study showed that boys had higher sit-ups scores across the four years compared to the girls (see intercept scores in Tables 6 & 7). Boys also showed a steeper increase across the four years while the girls had a relatively flat curve (see slope values in Tables 6 & 7). The means for the intercept, slope, and quadratic were significant for the boys, indicating growth at the school levels in sit-up performance over time. The means for the intercept and slope for the girls were significant, but not in the quadratic trend ($t < 1.96$, indicating no significance). There were significant within-individual changes as well as between-individual changes in the intercept, slope, and quadratic trends among the boys and girls (see Tables 6 & 7).

Among the boys, while there was a steep increase in their SBJ performance across the four years, this increment slowed down over time (refer to the mean changes reported in Table 2). For both genders, the means of the intercept, slope and the quadratic trend were significant ($t > 1.96$). There were also significant individual differences and within individual differences in the variances in the slopes, intercepts and quadratic trends in both genders.

In terms of SnR performance, both genders (36.21 slope value for boys; 31.94 slope value for girls) improved their flexibility over time. The quadratic factors of both genders were significant ($t > 1.96$). However, there were no significant differences in the variances of the within-level slope and quadratic trend for the boys and within-level slope variance for the girls.

Both genders showed significant improvement in their shuttle run over the four year period. The means of the intercept, slope and quadratic means were significant for both genders

($t > 1.96$). There were significant variations at the between-level for intercept, slope, and quadratic for boys and girls. However, there were no significant difference in the variances in the slope and quadratic for both genders at the within-level ($t < 1.96$).

Finally, the 2.4 km run-walk mean coefficients of the intercept, slope and quadratic trend for both genders were significant ($t > 1.96$), indicating significant improvement in their 2.4 km run-walk. There were significant between-level differences for both genders. The within-individual slope variances for boys ($t = .00$) and girls ($t = .45$) were not significant.

In summary, at the within-individual level, the results of the multilevel LGCM showed that boys' and girls' had significant variances in the intercepts, slope and quadratic trends for sit-ups and SBJ. However, the within-level variances for slope and quadratic trends were not significant for boys' SnR, shuttle run, and 2.4 km run-walk. For girls, there were no significance in within-level slopes for SnR and 2.4 km run-walk and slope and quadratic trends of shuttle run.

Discussion

The purpose of the present study was to examine the changes in Singaporean youth physical fitness performance using a multilevel latent growth curve modelling approach. Overall, the present study tested the linear versus quadratic models in the initial analysis (Step 1). The results revealed that the shape of development of all five physical fitness performances (sit-ups, SBJ, SnR, shuttle run, and 2.4 km run-walk) followed a quadratic trend. The quadratic trend indicates a declining improvement over time. This study adds to the existing literature in that it documents the effect of school on physical fitness performance, in addition to between- and within-individual changes. Different schools may have different programmes for preparing students for the fitness tests, teachers' experience, length of time, or PE programmes, although the protocols of conducting the fitness tests remains fixed.

To our knowledge, there has been no study in the physical fitness domain that examined changes in fitness performance over a long period in a sample using LGCM apart from the introductory paper on LGCM written by Park and Schutz (2005). There are a few studies that examined the secular trends of fitness performance among Australian children aged 12 to 15 (Tomkinson et al., 2003) and 9-year-old Danish children (Wedderkopp et al., 2004); both studies were not longitudinal and only means and standard deviations were utilised in the analysis. The conclusions drawn from both studies indicate that aerobic fitness of Australian (both boys and girls using estimated VO_{2max} values) and Danish children (boys only using a maximal cycle ergometer test) has declined over the years. Most studies in the literature have ignored the fact that the participants were clustered in different classes, schools, geographical regions, and/or possibly social economic status. Since the data had a hierarchical (nested) structure, it was more appropriate to use a multilevel analysis. This current study contributed to the literature in terms of application of multilevel LGCM in analysing the growth trends of various physical fitness performances among youth.

To answer the first research question regarding the rates of change of physical fitness performance, the present study found that in general (ignoring the school effect), the rate of change can be represented by a quadratic trend for most of the five tests. In the quadratic trend, the physical fitness performance improved over time, particularly in secondary 1 to 3 and slowed down between secondary 3 to 4. There is a strong reason for the slowing down or plateau effects as indicated by the quadratic curves. The NAPFA test is criterion-referenced (see Table 1), and the students can score maximum points if they reach the performance standards. Therefore, there is no incentive for them to perform better. On the other hand, students could give less than their best effort if they might feel that a standard of a certain test is beyond their ability. For example, for ‘A’ grade (5 points) of standing broad jump, 15 years old students are required to jump further than 237 cm for boys and 182 cm for girls, according to the standards set by the Ministry of Education, Singapore. The descriptive statistics in Table 2 shows that the mean scores for secondary 3 boys and girls were only 206.11 cm and 161.29 cm, respectively, which were far below the cut-offs. In a linear trend, there is a constant improvement or plateau in the physical fitness performance. For the girls’ 2.4 km run-walk, it seems that the girls did not change over the four-year period; the mean scores from secondary 2 to 4 were similar. It could be deemed that the fitness activity levels of adolescent girls may not have improved. This is supported by various literature on girls activity levels that showed a significantly decrease, compared to that of adolescent boys (Butte et al., 2007; Lupker, 1999). Another plausible explanation would be girls’ lack of motivation in this high-intensity test which is relatively lengthy and boring (Harris and Cale, 2006). An activity programme conducted in more interesting and enjoyable manners could expect girls’ higher motivation and constant improvement in fitness (Kemp and Pienaar, 2009; Welk, 2008). In the case of the boys sit up, a constant improvement (39.31 for secondary 1; 42.77 for secondary 2; 44.16 for secondary 3) was demonstrated over the period. Unlike other tests, the standard for an ‘A’ grade (maximum points) required between ages of 13 to 18 years in boys is consistently 42 sit-ups per minutes. Students may not be motivated to perform any higher to attain as many as points from this test to secure a silver medal at easy which allows them to enjoy a one-month exemption from the national service. The non-significant within-individual variations across the four years in SnR, shuttle run, and 2.4 km run-walk could be because these fitness performances may be more difficult to improve.

The second research question was related to effects across schools on selected physical performance indicators. From the initial analyses of the intra-class correlation and the estimated latent variances of the Multilevel LGCM, the findings of the present study showed that there were strong school effects on all the physical fitness performances (see variances in Tables 6 and 7). A high proportion of the variances in the physical fitness performance scores could be attributed to the school effect. For example, considering the boys' sit-ups second-order-to-total variance ratios, approximately 61% of the total variance in boys' sit-ups intercept scores (computed from $S_B/ [S_B + S_w]$ = 60.66/[60.66 + 38.16]), 79% of the total variance in sit-ups slope scores (42.07/[42.07 + 10.87]), and 78% of the total variance in sit-ups quadratic scores (3.39/[3.39 + .94]) could not be accounted for by the individual level differences but could be accounted for by the school membership. This shows that the school effects are strong. Although they may not be direct evidence, different PE programmes, time for training for the physical fitness tests, and experience and competency of the PE teachers might cause such high variations among schools. There are strict guidelines to govern the way the NAPFA tests are conducted in Singapore. For example, all of the five tests (sit-ups, pull-ups, SBJ, SnR, and Shuttle run) must be attempted on the same day, and a two to five minute rest period is permitted between each of the first five items. The 2.4 km run-walk can be taken on a separate day. Each school is subjected to random checks by the Ministry officials as well. Therefore, there is a low probability that the variations among schools are due to the testing protocols. If there were errors in measurement or recording, the LGCM method would have modeled the errors in the analysis. That is one of the strengths of using LGCM whereby the estimated model parameters have been corrected for random measurement errors (Duncan et al., 2006).

The final research question was related to the change across individuals in terms of physical fitness performance. The evidence from the current study suggested that there were significant differences across individual in all of the physical fitness performance indices for both genders over the four year period. This is in line with the literature (e.g., Jones et al., 2000). The differences could be due to genetic or environmental factors (Rowland, 2005). The current study has several limitations. Firstly, it has been known that during pubertal years, biological maturation introduces a substantial inter-individual variability in timing and magnitude of the growth spurts in stature and body mass (Ellis et al., 2011). This study did not take into consideration of the pubertal timing and tempo. Secondly, although the testing protocol is standardized, there are many different testers in each school and the inter-rater reliability was not measured. Finally, the study took the objective data from the cockpit system. There is a need to triangulate the findings using other data from interviews or questionnaires.

Conclusions

The current study contributes significantly to the current literature on fitness performance among youth. This study demonstrated the use of the Multilevel LGCM methods to study change by taking into consideration of the hierarchical nature of the data. The study supported a quadratic trend of the longitudinal data across the five physical fitness tests with significant within-school variability in change over the period. The results provide useful information with practitioners to develop a new strategy, particularly for upper level secondary students. There are three major recommendations for practitioners from this study. Firstly, there is a need to review the standards of the physical fitness performance standards. The upper secondary students are not showing improvement after they reached a particular performance standard. Secondly, some of the tests are relatively easy for a particular gender (e.g., SnR for female), but some tests are quite tough (pull-ups for male older than 14 years old). However, all the six tests are given equal weighting to determine the standards. There is a need to review the test items as well. Finally, although there are variations within-individual, across-individual, and across schools, the year to year improvement in physical fitness performance was not substantial (intraclass correlation < .20). The results suggest that schools may not need to spend too much time, effort, and resources in getting students to prepare the physical fitness tests every year. Perhaps there is a need to delimit the testing of physical fitness during the schooling years to alternate years or during transition years in the Singapore context.

References

Bentler PM and Wu E (1998) EQS for Windows (Version 5.7). Encino, CA: Multivariate Software..

Butte NF, Puyay MR, Adolph AL, Vohra FA, and Zakeri I (2007) Physical activity in non-overweight and overweight Hispanic children and adolescent. *Medicine & Science in Sports and Exercise* 39: 1257-1266.

Byrne BM (2006) *Structural Equation Modeling with EQS: Basic Concepts, Applications and Programming* (2nd edn). Mahwah, NJ: Lawrence Erlbaum Associates..

Cale L and Harris J (2009) Fitness testing in physical education: A misdirected effort in promoting healthy lifestyles and physical activity? *Physical Education and Sport Pedagogy* 14: 89-108.

Card NA, Little TD, and Bovaird JA (2007) Modeling ecological and contextual effects in longitudinal studies of human development. In: Little TD, Bovaird JA, Card NA (eds). *Modeling Contextual Effects in Longitudinal Studies*. Mahwah, NJ: Lawrence Elbaum Asscoaites, Inc, pp. 1-12.

Corbin CB, Pangrazi RP, and Welk GJ (1995) A response to 'The horse is dead; let's dismount'. *Pediatric Exercise Science* 7: 347-351.

Duncan TE, Duncan SC, Alpert A, Hops H, Stoolmiller M, and Muthén B (1997) Latent variable modeling of longitudinal and multilevel substance use data. *Multivariate Behavioral Research* 32: 275-318.

Duncan TE, Duncan SC, and Strycker LA (2006) *An Introduction to Latent Variable GrowthCurve Modeling: Concepts, Issues, and Applications* (2nd edn). Mahwah, NJ: Lawrence Erlbaum Associates..

Ellis BJ, Shirtcliff EA, Boyce WT, Deardorff J, and Essex MJ (2011) Quality of early relationships and the timing and tempo of pubert: Effects depend on biological sensitivity to context *Developmental Psycholpathology* 23: 85-99.

Fan X and Fan X (2005) Power of latent growth modeling for detecting linear growth: Number of measurements and comoparison with other analytic approach. *Journal of Experimental Education* 73: 121-139.

Giam CK (1981) Physical fitness - Definition and assessment. *Singapore Medical Journal* 22: 176-181.

Harris J and Cale L (2006) A review of children's fitness testing. *European Physical Education Review* 12: 201-225.

Hox JJ (2002) *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hu L and Bentler PM (1999) Curoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6: 1-55.

Jackson AS (2006) The evolution and validity of health-related fitness. *Quest* 56: 160-175.

Jones MA, Hitchen PJ, and Stratton G (2000) The importance of considering biological maturity when assessing physical fitness measures in girls and boys aged 10 to 16 years. *Annals of Human Biology* 27: 57-65.

Kemp C and Pienaar AE (2009) Effect of aerobic-based physical activity programme on physical fitness of 10 - 15 year-old girls. *African Journal for Physical, Health Education, Recreation, and Dance* 14: 527-542.

Lupker RV (1999) How physically active are American children and what can we do about it?*International Journal of Obesity* 23: S12-S17.

McArdle JJ (1988) Dynamic but structural equation modeling of repeated measures data. In: Nesselroade JR and Cattel RB, (eds). *Handbook of Multivariate Experimental Psychology* (2nd edn). New York: Plenum, pp 561-614.

McKenzie TL (2001) Promoting physical activity in youth: focus on middle school environments. *Quest* 53: 326-334.

Meredith W and Tisak J (1990) Latent curve analysis. *Psychometrika* 55: 107-122.

Muthén B (1991) Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement* 28: 338-354.

Muthén B (1994) Multilevel covariance structure analysis. *Sociological Methods & Research* 22: 376-398.

National Association for Sport and Physical Education (2005). *Physical Education for Lifelong Fitness* (2nd edn). Champaign, IL: Human Kinetics.

Park I and Schutz RW (2005) An introduction to latent growth models: analysis of repeated measures physical performance data. *Research Quarterly for Exercise and Sport* 76: 176-192.

Park MJ, Brindis CD, Chang F, and Irwin CEJ (2008) A midcourse review of the Healthy People 2010: 21 critical health objectives for adolescents and young adults. *Journal of Adolescent Health* 42: 329-334.

Physical Activity Guidelines Advisory Committee. (2008). *Physical Activity Guidelines Advisory Committee Report 2008*. Washington, DC.

Rao CR (1958) Some statistical methods for comparison of growth curves. *Biometrics* 14: 1-17.

Rovine MJ and Molenaar PCM (1998) A LISREL model for the analysis of repeated measures with a patterned convariance matrix. *Structural Equation Modeling* 5: 318-343.

Rowland TW (1995) The horse is dead: let's dismount. *Pediatric Exercise Science* 7: 117-120.

Rowland TW (2005) *Children's Exercise Physiology* (2nd edn). Champaign, IL: Human Kinetics.,

Silverman S, Keating XD, and Phillips SR (2008) A lasting impression: a pedagogical perspective on youth fitness testing. *Measurement in Physical Education and Exercise Science* 12: 146-166.

Singer JD and Willet JB (2003) *Applied longitudinal data analysis*. New York: Oxford University Press.Stoolmiller M (1995) Using latent growth curve models to study developmental processes. In: Gottman JM (ed) *The Analysis of Change*. Mahwah, NJ: Lawrence Erlbaum, pp 103-138.

Tomarken AJ and Waller NG (2005) Structural equation modeling: strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology* 1: 31-65.

Tomkinson GR, Olds TS, and Gulbin J (2003) Secular trends in physical performance of Australian children. *Journal of Sports Medicine and Physical Fitness* 43: 90-98.

Tucker LR (1958) Determination of parameters of a function relation by factor analysis. *Psychometrika* 38: 1-10.

Voelkle MC (2007) Latent growth curve modeling as an integrative approach to the analysis of change. *Psychology Science* 49: 375-414.

Wedderkopp N, Froberg K, Hansen HS, and Andersen LB (2004) Secular trends in physical fitness and obesity in Danish 9-year-old girls and boys: odense school child study and Danish substudy of the European youth heart study. *Scandinavean Journal Medicine Science and Sports* 14: 150-155.

Welk GJ (2008) The role of physical activity assessments for school-based physical activity promotion. *Measurement in Physical Education and Exercise Science* 12: 184-206.

Wiersma LD and Sherman CP (2008) The responsible use of youth fitness testing to enhance student motivation, enjoyment, and performance. *Measurement in Physical Education and Exercise Science* 12: 167-183.

Table 1

NAPFA Standards for Boys and for Girls

Boys (Age)	Points	Sit-Ups (No)	SBJ (cm)	SnR (cm)	Inclined Pull-Ups (No)	Shuttle Run (sec)	2.4 Km Run-Walk (min:sec)
13	5	>42	>214	>41	>25	<10.3	<11:31
	4	38-42	202-214	38-41	22-25	10.3-10.7	11:31-12:30
	3	34-37	189-201	34-37	17-21	10.8-11.1	12:31-13:40
	2	29-33	176-188	30-33	12-16	11.2-11.5	13:41-14:50
	1	25-28	164-175	25-29	7-11	11.6-11.9	14:51-16:00
14	5	>41	>202	>39	>24	<10.4	<11:01
	4	36-41	189-202	36-39	21-24	10.4-10.9	11:01-12:00
	3	32-35	176-188	32-35	16-20	11.0-11.3	12:01-13:00
	2	27-31	163-175	28-31	11-15	11.4-11.7	13:01-14:10
	1	22-26	150-162	23-27	5-10	11.8-12.2	14:11-15:20
15	5	>42	>237	>45	>7	<10.2	<10:41
	4	40-42	228-237	42-45	6-7	10.2-10.3	10:41-11:40
	3	37-39	218-227	38-41	5	10.4-10.5	11:41-12:40
	2	34-36	208-217	34-37	3-4	10.6-10.9	12:41-13:40
	1	30-33	198-207	29-33	1-2	11.0-11.3	13:41-14:40
16	5	>42	>245	>47	>8	<10.2	<10:31
	4	40-42	236-245	44-47	7-8	10.2-10.3	10:31-11:30
	3	37-39	226-235	40-43	5-6	10.4-10.5	11:31-12:20
	2	34-36	216-225	36-39	3-4	10.6-10.7	12:21-13:20
	1	31-33	206-215	31-35	1-2	10.8-11.1	13:21-14:10
Girls (Age)	**Points**	**Sit-Ups (No)**	**SBJ (cm)**	**SnR (cm)**	**Inclined Pull-Ups (No)**	**Shuttle Run (sec)**	**2.4 Km Run-Walk (min:sec)**
13	5	>30	>170	>41	>16	<11.3	<14:31
	4	26-30	162-170	39-41	13-16	11.3-11.7	14:31-15:30
	3	22-25	153-161	36-38	10-12	11.8-12.2	15:31-16:30
	2	18-21	144-152	32-35	7-9	12.3-12.7	16:31-17:30
	1	14-17	135-143	27-31	3-6	12.8-13.2	17:31-18:30
14	5	>30	>177	>43	>16	<11.5	<14:21
	4	28-30	169-177	41-43	14-16	11.5-11.8	14:21-15:20
	3	24-27	160-168	38-40	10-13	11.9-12.2	15:21-16:20
	2	20-23	151-159	34-37	7-9	12.3-12.6	16:21-17:20
	1	16-19	142-150	29-33	3-6	12.7-13.0	17:21-18:20
15	5	>30	>182	>45	>16	<11.3	<14:11
	4	29-30	174-182	43-45	14-16	11.3-11.0	14:11-15:10
	3	25-28	165-173	39-42	10-13	11.7-12.0	15:11-16:10
	2	21-24	156-164	35-38	7-9	12.1-12.4	16:11-17:10
	1	17-20	147-155	30-34	3-6	12.5-12.8	17:11-18:10

	5	>30	>186	>46	>17	<11.3	<14:01
	4	29-30	178-186	44-46	14-17	11.3-11.5	14:01-15:00
16	3	26-28	169-177	40-43	11-13	11.6-11.8	15:01-16:00
	2	22-25	160-169	36-39	7-10	11.9-12.2	16:01-17:00
	1	18-21	151-159	31-35	3-6	12.3-12.6	17:01-18:00

Table 2

Descriptive Statistics for the Repeated Measures of Physical Fitness Test Data

	Year 1	Year 2	Year 3	Year 4
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Sit-Ups				
Male	39.31 (7.82)	42.77 (7.91)	44.16 (7.69)	44.64 (7.71)
Female	31.88 (6.37)	33.36 (6.75)	33.52 (6.85)	33.58 (6.71)
SBJ (cm)				
Male	180.26 (23.12)	194.88 (23.75)	206.11 (23.60)	213.21 (23.37)
Female	155.85 (18.16)	159.25 (18.65)	161.29 (18.02)	163.67 (17.63)
SnR (cm)				
Male	34.24 (7.11)	37.98 (7.51)	40.88 (8.32)	42.99 (8.30)
Female	36.23 (6.76)	38.43 (7.49)	40.08 (7.83)	41.80 (7.63)
Shuttle Run (sec)				
Male	11.07 (.84)	10.66 (.80)	0.43 (.78)	10.29 (.79)
Female	11.93 (.75)	11.77 (.75)	11.70 (.76)	11.64 (.73)
2.4 km Run-Walk (sec)				
Male	858.84 (145.63)	803.73 (141.47)	784.70 (151.38)	776.25 (152.40)
Female	1004.13 (124.92)	978.05 (123.31)	977.13 (127.02)	970.56 (126.19)

Note. SBJ = Standing Broad Jump, SnR = Sit & Reach, Shuttle Run = 4 x 10 m Shuttle Run

Table 3

Model Fit Indices for Boys and for Girls

	Growth Curve	NFI	CFI	RMSEA	90% of CI
Boys					
Sit Ups	Linear	.971	.985	.057	.041, .074
	Quadratic	.986	.987	.098	.079, .119
SBJ	Linear	.982	.982	.126	.110, .143
	Quadratic	.999	1.000	.000	.000, .044
SnR	Linear	.989	.989	.075	.059, .092
	Quadratic	1.000	1.000	.000	.000, .022
Shuttle Run	Linear	.994	.995	.051	.036, .069
	Quadratic	1.000	1.000	.047	.022, .078
2.4 km Run-Walk	Linear	.972	.986	.056	.040, .073
	Quadratic	1.000	1.000	.066	.040, .096
Girls					
Sit Ups	Linear	.986	.987	.081	.065, .098
	Quadratic	1.000	1.000	.067	.040, .098
SBJ	Linear	.986	.987	.090	.077, .103
	Quadratic	1.000	1.000	.042	.017, .075
SnR	Linear	.985	.985	.084	.071, .097
	Quadratic	1.000	1.000	.030	.000, .064
Shuttle Run	Linear	.996	.996	.056	.043, .070
	Quadratic	1.000	1.000	.030	.000, .064
2.4 km Run-Walk	Linear	.990	.990	.074	.057, .091
	Quadratic	1.000	1.000	.096	.068, .126

Note. SBJ = Standing Broad Jump (cm), SnR = Sit & Reach (cm), Shuttle Run = 4 x 10 m Shuttle Run (sec), 2.4 km Run-Walk (sec).

Table 4

Intraclass Correlations for the Repeated Measures

	Year 1 (T1)	Year 2 (T2)	Year 3 (T3)	Year 4 (T4)
Boys				
Sit-ups	.08	.11	.09	.08
SBJ	.03	.03	.03	.02
SnR	.07	.08	.12	.11
Shuttle Run	.03	.04	.05	.04
2.4 km Run-Walk	.04	.04	.04	.05
Girls				
Sit-ups	.10	.09	.11	.09
SBJ	.05	.07	.06	.07
SnR	.12	.16	.16	.19
Shuttle Run	.09	.07	.07	.06
2.4 km Run-Walk	.13	.11	.11	.11

Note. SBJ = Standing Broad Jump (cm), SnR = Sit & Reach (cm), Shuttle Run = 4 x 10 m Shuttle Run (sec), 2.4 km Run-Walk (sec).

Table 5

Model Fit Indices for Within-Level and Between-Level Structure for Boys and for Girls

		NFI	CFI	RMSEA	90% of CI
Within					
Sit Ups	Boys	1.000	1.000	.054	.029, .085
	Girls	1.000	1.000	.069	.043, .100
SBJ	Boys	1.000	1.000	.000	.000, .044
	Girls	1.000	1.000	.043	.019, .075
SnR	Boys	1.000	1.000	.000	.000, .022
	Girls	1.000	1.000	.031	.006, .064
Shuttle Run	Boys	1.000	1.000	.048	.023, .079
	Girls	1.000	1.000	.031	.006, .064
2.4 km Run-Walk	Boys	1.000	1.000	.067	.041, .098
	Girls	1.000	1.000	.098	.071, .127
Between					
Sit Ups	Boys	.999	.999	.073	.047, .103
	Girls	.994	.994	.132	.104, .160
SBJ	Boys	1.000	1.000	.000	.000, .041
	Girls	1.000	1.000	.029	.003, .062
SnR	Boys	.984	.984	.204	.176, .233
	Girls	1.000	1.000	.017	.000, .052
Shuttle Run	Boys	1.000	1.000	.028	.000, .061
	Girls	.999	.999	.041	.016, .072
2.4 km Run-Walk	Boys	1.000	1.000	.052	.027, .083
	Girls	1.000	1.000	.069	.043, .099

Note. SBJ = Standing Broad Jump (cm), SnR = Sit & Reach (cm), Shuttle Run = 4 x 10 m Shuttle Run (sec), 2.4 km Run-Walk (sec).

Table 6

Fit Indices and Parameter Estimates from the Limited Multilevel LGCM for Boys

Coefficient		Between-level		Within-level		Fit Indices
		Effect	t-value	Effect	t-value	
Sit-ups	**Means**					
	Intercept	27.83	34.65*			
	Slope	2.81	3.91*			NFI = .996
	Quadratic	- .53	-2.46*			CFI = .997
	Variances					RMSEA = .028
	Intercept	60.66	5.69*	38.16	14.06*	90% CI = (.013, .044)
	Slope	42.07	4.89*	10.87	3.39*	
	Quadratic	3.39	4.44*	.94	4.08*	
SBJ	**Means**					
	Intercept	632.74	355.37*			
	Slope	58.35	37.56*			NFI = .999
	Quadratic	- 6.57	-14.10*			CFI = .999
	Variances					RMSEA = .019
	Intercept	179.77	3.42*	428.20	20.94*	90% CI = (.000, .036)
	Slope	170.59	4.27*	81.56	3.94*	
	Quadratic	13.54	3.76*	6.55	5.09*	
SnR	**Means**					
	Intercept	120.21	172.17*			
	Slope	14.51	20.36*			NFI = .994
	Quadratic	-1.43	-7.30*			CFI = .994
	Variances					RMSEA = .046
	Intercept	43.79	5.43*	25.55	14.39*	90% CI = (.032, .061)
	Slope	36.21	4.40*	.00	.00	
	Quadratic	1.76	2.74*	.18	1.70	
Shuttle Run	**Means**					
	Intercept	38.87	579.36*			
	Slope	-1.68	- 21.83*			NFI = .993
	Quadratic	.26	10.74*			CFI = .994
	Variances					RMSEA = .049
	Intercept	.28	3.74*	.43	19.66*	90% CI = (.035, .064)
	Slope	.43	4.38*	.02	1.93	
	Quadratic	.04	4.56*	.00	.00	
2.4 km Run-Walk	**Means**					
	Intercept	3010.09	255.52*			
	Slope	-219.74	-16.74*			NFI = .997
	Quadratic	41.98	10.74*			CFI = .997
	Variances					RMSEA = .045
	Intercept	9204.52	4.00*	11953.00	12.95*	90% CI = (.031, .060)
	Slope	14456.12	5.07*	.00	.00	
	Quadratic	1190.62	4.70*	94.35	1.23	

Note. *Significant at the .05 probability level. SBJ = Standing Broad Jump (cm), SnR = Sit & Reach (cm), Shuttle Run = 4 x 10 m Shuttle Run (sec), 2.4 km Run-Walk (sec).

Table 7

Fit Indices and Parameter Estimates from the Limited Multilevel LGCM for Girls

Coefficient		Between-level		Within-level		Fit Indices
		Effect	t-value	Effect	t-value	
Sit-ups	**Means**					
	Intercept	22.56	33.78*			
	Slope	1.25	2.13*			NFI = .992
	Quadratic	- .29	-1.67			CFI = .994
	Variances					RMSEA = .053
	Intercept	43.28	5.86*	21.21	10.93*	90% CI = (.038, .068)
	Slope	20.34	4.12*	4.72	1.99*	
	Quadratic	1.43	3.40*	.64	3.71*	
SBJ	**Means**					
	Intercept	528.31	329.12*			
	Slope	11.26	9.97*			NFI = .996
	Quadratic	- .87	-2.53*			CFI = .996
	Variances					RMSEA = .049
	Intercept	201.76	4.71*	242.78	18.84*	90% CI = (.035, .064)
	Slope	74.55	3.51*	37.44	2.80*	
	Quadratic	5.87	2.99*	3.48	4.17*	
SnR	**Means**					
	Intercept	122.85	161.10*			
	Slope	7.50	12.15*			NFI = .993
	Quadratic	- .42	-2.14*			CFI = .994
	Variances					RMSEA = .056
	Intercept	60.38	6.26*	22.26	10.79*	90% CI = (.042, .070)
	Slope	31.94	5.05*	.08	.03	
	Quadratic	3.24	5.04*	.49	2.85*	
Shuttle Run	**Means**					
	Intercept	40.40	506.90*			
	Slope	- .51	- 6.91*			NFI = .997
	Quadratic	.06	3.00*			CFI = .997
	Variances					RMSEA = .028
	Intercept	.60	5.74*	.30	16.21*	90% CI = (.011, .045)
	Slope	.44	4.82*	.00	.00	
	Quadratic	.03	3.98*	.00	.83	
2.4 km Run-Walk	**Means**					
	Intercept	3396.33	238.42*			
	Slope	-85.16	-6.70*			NFI = .993
	Quadratic	16.70	4.54*			CFI = .994
	Variances					RMSEA = .064
	Intercept	21257.89	6.32*	7724.33	11.57*	90% CI = (.051, .078)
	Slope	15500.85	5.79*	362.12	.45	
	Quadratic	1191.26	5.31*	163.08	2.93*	

Note. *Significant at the .05 probability level. SBJ = Standing Broad Jump (cm), SnR = Sit & Reach (cm), Shuttle Run = 4 x 10 m Shuttle Run (sec), 2.4 km Run-Walk (sec).

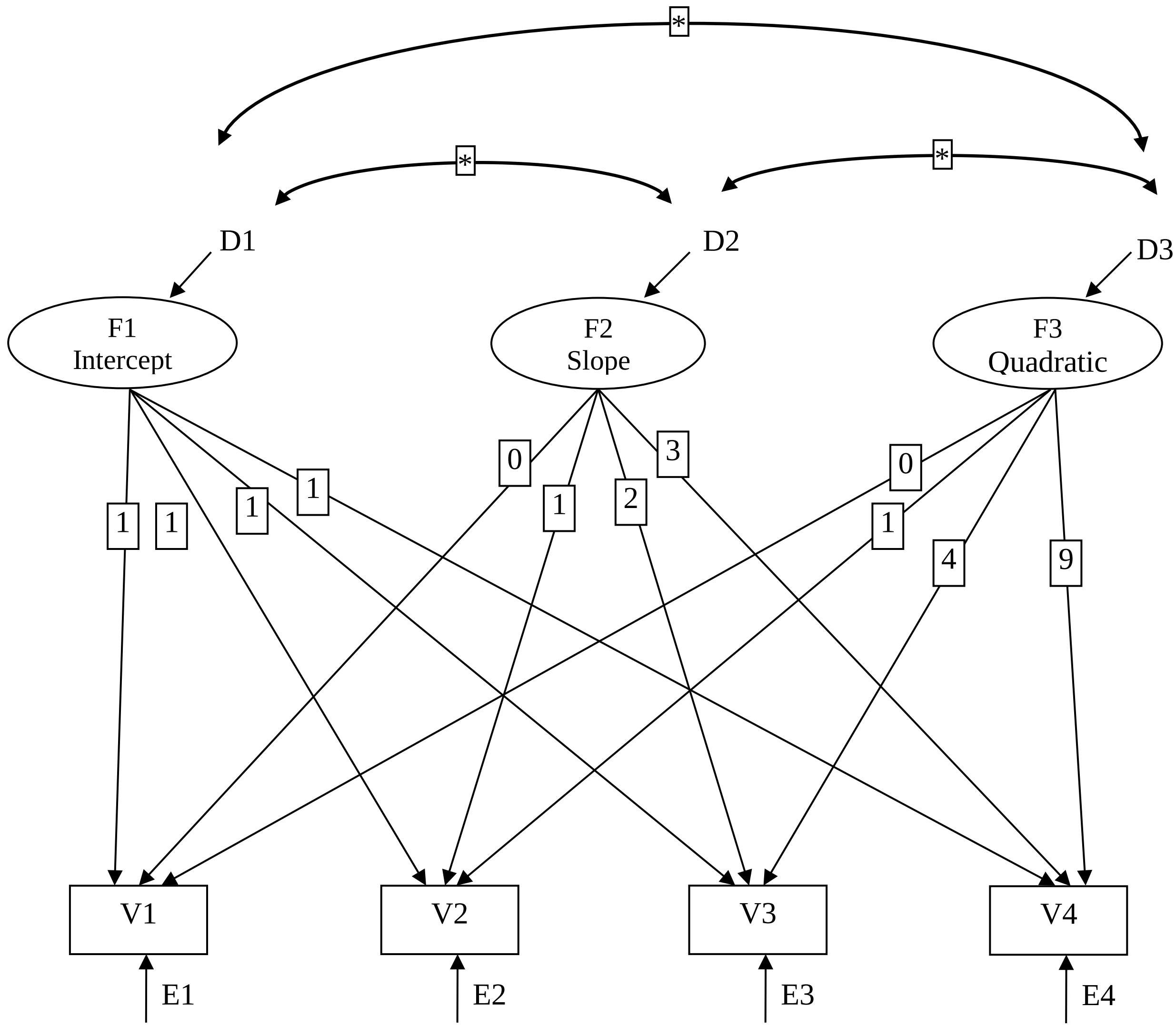

Figure 1. Representation of a three-factor polynomial LGCM.

* denotes factor covariances. V1 to V4 denote the same variable measured at four time points (Year 1 to Year 4), E1 to E4 represent the error variance terms. The three latent variables F1 to F3 represent the “intercept”, “slope” and “quadratic” factors for the latent growth curve model. The “intercept” is a constant for any given individual across time, hence the factor loadings are fixed at 1. The “slope” represents the slope of an individual’s trajectory or linear change rate of the variable over time. The “quadratic” carry the shape of the growth over time, the factor loadings for the quadratic factor (F3) are the squared loadings of the linear factor (F2).

Brief Biography

***John Wang, C. K.** is Professor in Sport and Exercise Psychology and Heads of the Physical Education and Sports Science in National Institute of Education, Nanyang Technological University, Singapore.

* **Pyun, Do Young** is an Assistant Professor in the Physical Education and Sports Science in National Institute of Education, Nanyang Technological University, Singapore.

***Liu, Woon Chia** is an Associate Professor in the Psychological Studies and Associate Dean (Practicum) in National Institute of Education, Nanyang Technological University, Singapore.

***Coral Lim, B. S.** is a Research Associate and a member of the Motivation in Educational Research Laboratory in National Institute of Education, Nanyang Technological University, Singapore.

#**Li, Fuzhong** is a Senior Research Scientist at the Oregon Research Institute, USA.

Address for correspondence: Dr John Wang, Professor, Physical Education and Sports Science, National Institute of Education, Blk 5 #03–12, 1 Nanyang Walk, Singapore 637616. [email: john.wang@nie.edu.sg]

*Nanyang Technological University, Singapore
[#]Oregon Research Institute