

**LATENT GROWTH MODELS AND RELIABILITY ESTIMATION OF
LONGITUDINAL PHYSICAL PERFORMANCES**

by

Il Hyeok Park

B.P.E., Seoul National University, 1991

M.P.E., Seoul National University, 1993

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF**

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(School of Human Kinetics)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December 2001

© Il Hyeok Park 2001

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Human Kinetics

The University of British Columbia
Vancouver, Canada

Date December 13, 2001

Abstract

There are four purposes to this study. The first is to introduce Latent Growth Models (LGM) to Human Kinetics researchers. The second is to examine the merits and practical problems of LGM in the analysis of longitudinal physical performance data. The third purpose is to examine the developmental patterns of children's physical performances. The fourth purpose is to compare the capacity of the two most widely used longitudinal factor models, LGM and a quasi-simplex model, to accurately estimate reliability for longitudinal data under various conditions. In study 1, the first, second and third purposes of the study were accomplished, and in study 2, the fourth purpose was accomplished.

In study 1, two longitudinal data sets were obtained, however, only one set was deemed appropriate for subsequent analyses. The data included seven physical performance variables, measured at five time points, from 210 children aged eight to twelve years, and five predictor variables of physical performances. The univariate LGM analyses revealed that the children's individual development over a 5-year period was adequately explained by either a Linear (jump-and-reach and sit-and-reach), Quadratic (flexed-arm hang), Cubic (standing long jump) or Unspecified Curve model (agility shuttle run, endurance shuttle run and 30-yard dash). The children improved in their physical performances between ages 8 and 12 except for flexibility, in which children's performance declined over time. Children showed considerable variations in the developmental rate and patterns of physical performances. Among the predictor variables, the test practice (the number of previous testing sessions) and age in months showed positive effects on the children's performance at the initial time point. A negative test practice effect on the development in physical performance was also found. The effect of other predictor variables varied for different performance variables. The multivariate analyses showed that the factor structure of three hypothesized factors, "Run", "Power" and "Motor Ability", holds at all five time points. However, only the change in the "Run" factor was adequately explained by the Unspecified Curve model. There were significant test practice, age, measured season and measured year effects on the performance at the initial time of testing, and significant test practice and measured year effects on the curve factor. The cross-validation procedure generally supported these findings. It was concluded that a LGM has several merits over traditional methods in the analysis of change in that a LGM provides an individual level of analysis, and thus allows one to test various research questions regarding the predictors of change, measurement error, and multivariate change. Additionally, it requires less strict statistical assumptions than traditional methods. Because of the merits of the LGM analysis used here, this study provided some interesting findings regarding children's development of physical performances-- findings that were not detectable in previous studies because of the use of traditional statistical analyses. The difficulty in comparing non-nested models, and the unknown relationship between the change in indicator variables and the change in the factor in the analysis of multivariate "curve-of-factors" model were discussed as practical problems in the application of LGM.

In study 2, several longitudinal developmental data sets with known parameters under various

conditions were generated by computer. The conditions were varied by the magnitude of correlations between initial status and change, the magnitude of reliability, and the magnitude of correlated errors between time points. The data were analyzed using two models, a LGM and a simplex model, and the estimated reliability coefficients were compared. The simplex model overestimated the reliability in all conditions, while the LGM provided relatively accurate reliability estimates in almost all conditions. Neither the magnitude of correlation between the initial status and change nor the magnitude of reliability affected the reliability estimation, while the correlated errors led to an overestimation of reliability for both models. On the other hand, the magnitude of reliability showed a negative effect on the goodness-of-fit of the simplex model. It was concluded that a LGM, rather than the often used simplex model, be used for reliability analyses of longitudinal data.

TABLE OF CONTENTS

Abstract	ii
Table of Contents	iv
List of Tables	vii
List of Figures	ix
Glossary of Abbreviations	x
Acknowledgment	xi
 CHAPTER I. INTRODUCTION	 1
Introduction	1
The Purposes of the Study	5
 STUDY 1. THE ANALYSIS OF LONGITUDINAL PHYSICAL PERFORMANCE DATA	 7
 STUDY 1-CHAPTER II. LITERATURE REVIEW	 8
Analysis of Change and Latent Growth Models	8
Relative Methods and Limitations	8
Latent Growth Model	11
Development of Physical Performance	19
Physical Performance Tests	21
The Development of Children's Physical Performance	24
The Factor Structure of Physical Performance	26
 STUDY 1-CHAPTER III. METHODOLOGY	 29
The Data	29
Data Analyses	33
Univariate LGM	33
Descriptive statistics	33
Identification of the best fitting growth curve	33
Predictor effects	36
Pseudo cross-validation	37
Multivariate LGM	37
Descriptive statistics	38
Verification of the factor structure	38
Identification of the best growth curve	38
Predictor effects	38
Pseudo cross-validation	38
Estimation of LGMs	41
Model Evaluation	41

STUDY 1-CHAPTER IV. RESULTS	42
Univariate Latent Growth Models for Motor Performances	42
Flexed-Arm Hang (FAH)	42
Descriptive Statistics	42
Identification of the Best Fitting Growth Curve	42
Predictor Effects	46
Six Other Physical Performance Variables	48
Descriptive Statistics	48
Identification of the Best Fitting Growth Curve	50
Predictor Effects	54
Pseudo Cross-validation	56
Descriptive Statistics	56
Identification of the Best Fitting Growth Curve	58
Parameter Estimates of the Best Fitting Growth Models	60
Predictor Effects	61
Discussion of the Development of Physical Performance	63
Multivariate Latent Growth Models for Physical Performances	65
Run	65
Descriptive Statistics	65
Verification of the Factor Structure	65
Identification of the Best Fitting Growth Curve	68
Predictor Effects	70
Power	72
Descriptive Statistics	72
Verification of the Factor Structure	72
Identification of the Best Fitting Growth Curve	74
Motor Ability	76
Descriptive Statistics	76
Verification of the Factor Structure	76
Identification of the Best Fitting Growth Curve	79
Pseudo Cross-validation	79
Descriptive Statistics	79
Verification of Factor Structure	79
Identification of the Best Fitting Growth Curve	82
and Predictor Effects	82
Discussion of the Multivariate Development of Physical Performance	84
 STUDY 1-CHAPTER V. DISCUSSION	 88
Merits of Latent Growth Models	88
Problems of Using Latent Growth Models	92
 STUDY 2. COMPARING THE LATENT GROWTH MODEL AND QUASI-SIMPLEX MODEL IN THE ESTIMATION OF LONGITUDINAL RELIABILITY	 94

STUDY 2-CHAPTER II. LITERATURE REVIEW	95
Concepts of Reliability and Traditional Estimation Methods	95
Estimation of Longitudinal Reliability	97
Comparing the Quasi-simplex and Latent Growth Models	101
STUDY 2-CHAPTER III. METHODOLOGY	104
Data and Conditions	104
Condition A: The Magnitude of the Correlation Between the Intercept and Change	104
Condition B: The magnitude of reliability	105
Condition C: The Correlation Between Errors	105
Data Generation Procedure	105
Generating Initial Status, Linear Change and Errors	107
Computing True Scores at Each Time Point	107
Changing the Variance of Errors	107
Computing Observed Scores	107
Model Fitting and Evaluation	109
STUDY 2-CHAPTER IV. RESULTS	110
The Effect of Correlation Between Initial Status and Linear Change	110
The Effect of the Magnitude of Reliability	112
The Effect of Correlated Errors	115
STUDY 2-CHAPTER V. DISCUSSION	120
CHAPTER VI. SUMMARY AND CONCLUSION	124
REFERENCES	127
APPENDICES	139
Appendix A: Example Data Records for Five Selected Subjects (Michigan Data Set 1)	139
Appendix B: Program Commands for Latent Growth Models	140
Appendix C: Descriptive Statistics and Parameter Estimates of Latent Growth Models	148
Appendix D: Descriptive Statistics and Parameter Estimates for Generated Data Sets	172

LIST OF TABLES

Table		
1.3.1	Examples of unreliable measurement – Standing long jump	30
1.3.2	Descriptions of variables that were used in the study (Michigan data)	32
1.3.3	Factor loadings of intercept and change factors for four LGMs	35
1.4.1	Correlation coefficients and descriptive statistics for flexed-arm hang	43
1.4.2	Fit indices for latent growth models for flexed-arm hang	43
1.4.3	Estimated parameters (standard errors) of the Quadratic model for flexed-arm hang	45
1.4.4	Fit indices of the Quadratic models with predictors for flexed-arm hang	47
1.4.5	Parameter estimates of predictor variables' effects on growth factors	47
1.4.6	Means and standard deviations for six physical performance variables	49
1.4.7	Best fitting growth curve models and goodness-of-fit indices for the six physical performance variables	51
1.4.8	Predictors' effects on growth factors for six physical performance variables	55
1.4.9	Means and standard deviations for seven physical performance variables of data set 2	57
1.4.10	Identification of the best fitting model: The comparison between data set 1 and cross-validation data (data set 2)	59
1.4.11	Predictors' effects on growth factors for six physical performance variables	62
1.4.12	Descriptive statistics for the ASR, ESR and DASH across all time points	66
1.4.13	Fit indices of the 5-factor models for the verification of the factor structure of "Run"	67
1.4.14	Fit indices of latent growth models for "Run"	69
1.4.15	Fit indices of models for the verification of the factor structure of "Power"	73
1.4.16	Parameter estimates of the 5-factor model with correlated errors and the equality of factor loadings over time for "Power"	75
1.4.17	Fit indices of latent growth models for "Power"	75
1.4.18	Fit indices of models for the verification of the factor structure of "Motor Ability"	77
1.4.19	Parameter estimates of the 5-factor model with correlated errors and the equality of factor loadings over time for "Motor Ability"	78
1.4.20	Fit indices of latent growth models for "Motor Ability"	80
1.4.21	Comparison of multivariate analyses results between data set 1 and data set 2	81

1.4.22	Fit indices of latent growth models for "Run" factor (data set 2)	83
1.4.23	Goodness-of-fit indices of growth models for two factors	85
1.5.1	The results of ANOVA analysis with polynomial contrasts for FAH, data set 1	89
2.3.1	Summary of conditions of generated data	106
2.3.2	Descriptive statistics of true and error scores for condition A1	108
2.4.1	Fit indices and estimated reliability coefficients of models with various correlations between initial status and linear change	111
2.4.2	The true and estimated parameters (standard errors) of the Linear model for condition A1	113
2.4.3	Fit indices and estimated reliability coefficients of models with various magnitudes of reliability	114
2.4.4	Fit indices and estimated reliability coefficients of models with various magnitudes of correlated errors	116
2.4.5	The true and estimated variances of Linear model for condition Cs	118

LIST OF FIGURES

Figure

1.2.1.	Linear LGM	15
1.2.2.	Quadratic LGM	17
1.2.3.	Linear LGM with a time-invariant predictor	17
1.2.4.	A Factor-of-curves model: 3 variables and 4 time points with linear change	18
1.2.5.	A curve-of-factors model: A linear model with 3 variables and 4 time points	20
1.3.1.	Unspecified Curve LGM	34
1.3.2.	The procedure for the verification of factor structure	39
1.3.3.	5-factor CFA model with a factor at each time point	40
1.3.4.	5-factor CFA model with correlated errors	40
1.4.1.	Linear and quadratic components of change in FAH	45
1.4.2.	Curve-of-factors model for "Run"	71
2.2.1.	A quasi-simplex model with five time points	98
2.2.2.	Two-factor LGM	100
2.4.1.	Parameter estimates (standard errors) of the Simplex 2 model for condition A1	113

GLOSSARY OF ABBREVIATIONS

AAHPER: American Alliance for Health, Physical Education and Recreation
AAHPERD: American Alliance for Health, Physical Education, Recreation and Dance
AIC: Akaike's Information Criteria
ANOVA: Analysis of Variance
ANCOVA: Analysis of Covariance
ASR: Agility Shuttle Run
CFA: Confirmatory Factor Analysis
DASH: 30-yard Dash
DM MANOVA: Doubly Multivariate Analysis of Variance
df: Degrees of Freedom
ECVI: Expected Cross-Validation Index
ESR: Endurance Shuttle Run
FAH: Flexed-Arm Hang
HLM: Hierarchical Linear Model
JAR: Jump-And-Reach
LGM: Latent Growth Model
MANOVA: Multivariate Analysis of Variance
ML: Maximum Likelihood
NNFI: Non-Normed Fit Index
p: probability
PCPFS: President's Council on Physical Fitness and Sports
PPMC: Pearson Product-Moment Correlation
RM ANOVA: Repeated Measures Analysis of Variance
RMSEA: Root Mean Square Error of Approximation
SAR: Sit-And-Reach
SD: Standard Deviation
SEM: Structural Equation Modelling
SLJ: Standing Long Jump
SRMR: Standardized Root Mean Square Residual

Acknowledgment

My special appreciation should go to my supervisor, Dr. Robert Schutz. I learned not only a tremendous amount of knowledge but also the spirit of research and teaching from him. The spirit has made me confident in what I am doing. He is an ideal supervisor, teacher and researcher. I will certainly follow his steps, and will try my best to contribute my area, measurement and statistics in Human Kinetics.

I appreciate the help and support from my committee members and university examiners, Dr. Seong-Soo Lee, Dr. Edward Rhodes, Dr. Bruno Zumbo and Dr. Heather McKay. Their comments and inputs were invaluable, and refined my research project in several ways.

I would like to thank Dr. John Haubenstricker and Dr. Vern Seefeldt for providing their valuable data set. Without the data set, this research project could not be possible.

I also would like to thank Dr. Jong Taek Kim, my former supervisor, who encouraged me to go to University of British Columbia (UBC) to continue studying in this area.

I have many friends who deserve my appreciation. Among them, Dr. Terry Wood helped and encouraged me to come to UBC; Dr. Hanjoo Eom gave me a part of his research spirit; Dr. Yuanlong Liu supported my work a lot; and Jaehan In provided his computer programming expertise in managing irregularly structured data.

During those years, my family were always with me. It is almost impossible to express my appreciation to my parents who have fully supported my education and work in every possible way. I will never be able to repay everything I owe them. My special thanks go to my parents-in-law who also fully supported my work.

Finally, my wife, Hyun Kyoung Lee, deserves the half of the celebration for becoming a "doctor". She has been with me for all those years in Vancouver, has brought our two children, and has encouraged me to finish up this research project.

CHAPTER I. INTRODUCTION

Introduction

The development of physical performance during childhood and adolescence has been, and continues to be, one of the most researched domains in human kinetics. Examples include: (a) attempts to depict or identify the trajectory of change over time in various physical performances of children and adolescents (e.g., Cearley, 1957; Clarke & Wickens, 1962; Haubenstricker & Seefeldt, 1986; Malina & Bouchard, 1991; Mirward & Bailey, 1986; Montoye, 1984; Morrow, Jackson & Bell, 1978; Shuleva, Hunter, Hester, & Dunway, 1990; Thomas & French, 1985), (b) comparisons of the change in performance between groups such as males and females or athletes and nonathletes (e.g., Erbaugh, 1984; Espenschade, 1947; Halverson & Williams, 1985; Pangrazi & Corbin, 1990; Smoll & Schutz, 1990), and (c) investigations of the relationship between physical performance and anthropometrical growth (e.g., McCloy, 1935; Nelson, Thomas & Nelson, 1991; Rowe, 1933; Sellis, 1951; Solley, 1960; Teeple & Massey, 1976). Most of these researchers focused on describing how children's performance changes over time. However, quantitative descriptions of development in physical performance are limited because these studies used primarily group statistics while ignoring individual developmental patterns. In addition, attempts to explain the reason why children show inter-individual differences in development have not been adequately made, although there have been some studies in which the difference between groups in development or the relationship between different variables in development was examined. The lack of adequacy in the studies of children's performance development is due, in some part, to the lack of a valid statistical model that enables one to adequately describe and explain (predict) change. For a more adequate description and explanation of change, one needs to use a method that enables one to analyze the change at both the individual and the group level. Traditional approaches are based either only on individual level analysis, or only on group level analysis, and thus are limited in adequately describing and explaining change.

The most widely used traditional method for analysis of change is based on the differences in mean scores. The simplest method may be the paired t-test in which the difference between two repeatedly measured mean scores is statistically tested. The more general form of this test is the repeated measures analysis of variance (RM ANOVA) that allows one to compare mean scores that are measured on more than two occasions. Additionally, in RM ANOVA, one can identify the shape of a variable's change by using preplanned orthogonal polynomials (Winer, Brown & Michels, 1991). The ANOVA procedure has some utility in describing change, but has limited utility in explaining change. At the most, the ANOVA procedure allows one to examine only the differences among groups in change. The ANOVA procedure can be misleading because it is mainly based on the group level statistics, and thus may not properly represent the individual level of change. In addition, this method is a univariate technique and requires the assumptions of "sphericity" that are frequently violated in

practice. Although multivariate analysis of variance (MANOVA) or doubly multivariate analysis of variance (DM MANOVA) for multiple indicator variables can be used under the violations of sphericity (Schutz & Gessaroli, 1987; Stevens, 1996), these methods also have limitations in that they are based solely on the comparisons of mean scores. In general, the covariances between repeated measurements, an important component of information, are not adequately accounted for in these statistical models (Labouvie, 1982).

There have been other approaches to the analysis of change, such as the application of stochastic models, time series analysis and growth curve fitting (Bock & Tissen, 1976, 1980; Cromwell, Labys & Terraza, 1994; Crosbie, 1995; Frederiksen & Rotondo, 1979; Rogosa, Brandt & Zimowski, 1982; Tissen & Bock, 1990). These methods are based on the individual level of change, thus have some merits in describing change. However, these approaches are limited in that they may include only one variable in an analysis, depend too much on approximations, or require large numbers of repeated measurements. In addition, these procedures require more than one step of analysis to obtain the group level of statistics in change as well as the individual level of statistics, and thus have limitations in explaining change.

Recent developments in factor analytic solutions for repeated measures data have received a significant amount of interest. Based on the formative work of Rao (1958) and Tucker (1958), Meredith and Tisak (1984) proposed a 'latent growth model (LGM)' approach for repeated measures data analysis formulated within the framework of structural equation modelling (SEM). The basic idea of LGM is that change is an unobservable latent trait. Thus, in a LGM, initial status and change are represented by latent factors. For example, in the linear LGM, the intercept and the slope of a growth line form latent factors. The basic linear model can be extended to a curvilinear model by adjusting the loadings of the slope factor or by adding one or more change factors (McArdle, 1988; Meredith & Tisak, 1990). This statistical analysis method is especially useful when one has an a priori hypothesis about the change of measures over time. The unique feature of LGM that is distinguished from a usual SEM is that one takes into account both the means and covariances of repeatedly measured variables in the analysis (McArdle, 1988; Meredith & Tisak, 1990; Stoolmiller, 1995). Thus, one may statistically examine the hypothesized change of means of variables and the covariances among variables at the same time in a LGM analysis, while one can examine only the hypothesized covariances among variables in the usual SEM. This eventually leads one to be able to examine the change at both individual and group level at the same time. The LGM approach offers several other important features. First, individual change can be represented by either a straight line or a curvilinear trajectory. Second, occasions of measurement need not be equally spaced. Third, measurement errors can be accounted for by the statistical model. Fourth, multiple predictors or correlates of change can be easily included in the model. Fifth, as in general SEM analysis, statistical models are very flexible, allowing one to extend the basic idea in several ways in order to test various hypotheses (Willett & Sayer, 1994).

McArdle (1988) extended the basic model and suggested two more complex models, which he called a 'factor-of-curves LGM' and a 'curve-of-factors LGM', that are more appropriate for multivariate data. In a 'factor-of-curves LGM', several first-order intercept and change factors explain the trajectories of several variables over time, and the correlations among intercept and among change factors are explained by second-order intercept and change factors. In a 'curve-of-factors LGM', on the other hand, several measures at a single time point form a latent construct and the curve of this latent construct over time is represented by second-order intercept and change factors (McArdle, 1988). Duncan and Duncan (1996) applied these two models in a growth study of adolescent substance use over time, and recommended more use of these models in longitudinal research. Other extensions have also been made. Muthen (1994, 1997) and Muthen and Curran (1997) extended and applied the LGM idea to clinical trial data and compared trajectories of change between groups. A number of studies have employed a cohort-sequential design with missing data (Duncan & Duncan, 1994, 1995; Duncan, Duncan & Li, 1998; McArdle & Hamagami, 1991). Autoregressive models and stability analyses are very closely related to LGM (Kenny & Campbell, 1989; Marsh & Grayson, 1994b; Meredith & Tisak, 1990).

Despite the many strengths of using LGM in a longitudinal study, there are problems that prevent practitioners from employing this approach in the study of development of physical performance. First, although there exist many introductory publications in which the merits of LGM are summarized, the specific strengths of LGM over traditional approaches have not been adequately shown, especially in the human kinetics field. For example, nowhere in the literature, to the current researcher's knowledge, is there a detailed discussion and comparison between an ANOVA procedure and LGM, with examples. Meredith and Tisak's (1990) presentation was too mathematically sophisticated for most practitioners, and Duncan, Duncan, Strycker, Li & Alpert's (1999) example was to show that the ANOVA model is a special case of LGM. There has been a lack of presentations that showed the specific strengths of a LGM approach over traditional procedures for the analysis of longitudinal data.

Second, there may be practical problems in the application of LGM to the longitudinal analysis of physical performance data. One such problem is that choosing between the unspecified curve model and the specified curve model (e.g., quadratic or cubic model) is not clear in some situations because these models are not nested to each other, and thus a statistical test that compares these models is not available. Another problem is related to the application of multivariate LGM to physical performance data. In most applications, multivariate LGM has been used with psychological variables, variables that are different from physical performance variables. The developmental curve of each subtest in a physical performance test battery may be very different from all other subtests in terms of both the rate of change and the nature (linear or curvilinear) of change across measures and time. For example, in a physical fitness test battery, a person's level of strength generally improves until late adolescence while the level of flexibility starts to decrease at early adolescence (Haubenstricker & Seefeldt, 1986;

Haywood, 1993). In general, some potentially important multivariate characteristics of longitudinal physical performance data are not well known. Additionally, there may be other practical problems in the application of LGM to longitudinal physical performance data.

In addition to the information of change, LGM provides estimates of reliability for a repeatedly measured variable (McArdle & Epstein, 1987; Tisak & Tisak, 1996). In a LGM, the variance of the observed variable is decomposed into two parts: the true score variance that is explained by the growth factors and the measurement error variance that is not explained by the growth factors. This way of estimating reliability is especially useful in a longitudinal study where the estimation of reliability is not feasible unless there is more than one measurement at each time point. Traditional methods that are based on the test-retest method and internal consistency have shortcomings in that these require more than one measurement at each time point. Another longitudinal path analytic model, a quasi-simplex model, has also been used for the estimation of reliability in a longitudinal study (e.g., Blalock, 1963; Heise, 1969; Siegel & Hodge, 1968). This model was initially suggested to separate the temporal instability of true scores from the measurement error. In this model, the true score at a certain time point is explained by the true score of the immediately preceding time point. The unexplained part of the observed variable at each time point is regarded as an error component. The basic idea has been extended and widely used by others (e.g., Jöreskog, 1970; Werts, Jöreskog & Linn, 1971; Wheaton, Muthen, Alwin & Summers, 1977; Wiley & Wiley, 1970).

While these two models, LGM and a quasi-simplex model (more generally, an autoregressive model), are the most widely used factor models for the analysis of longitudinal data, which one of the two models provides more accurate reliability estimates for repeatedly measured variables is not known. Although the implications and underlying assumptions regarding change of these two models are different, choosing one model over the other is not feasible in practice, because these two models are empirically difficult to distinguish (McArdle & Epstein, 1987; Rogosa & Willett, 1985a). Although there have been a few studies in which these two models are compared (e.g., Kenny & Campbell, 1989; Mandys, Dolan & Molenaar, 1994; Rogosa & Willett, 1985a), most of these studies focused more on the rationales, strengths and weaknesses of applying these two models in the analysis of longitudinal data rather than on the accuracy of reliability estimation. The capability of these two longitudinal models in the estimation of longitudinal reliability needs to be examined. This is especially important in the longitudinal study of physical performance because the repeated testing of physical performance is costly in terms of time and money. If the estimation of reliability for longitudinal data can be achieved analytically, it will be of a considerable benefit in the study of longitudinal physical performance.

The LGM method has seldom been utilized in human kinetics research. Duncan and Duncan (1991), in their introductory study, applied LGM to children's perception of physical competence. There have been other related works in which the SEM methodology was applied to repeated measures data. Marsh (1996) used confirmatory factor analysis with multitrait-multimethod data to examine the

stability of physical self-description, and Duncan and Stoolmiller (1993) used autoregressive models to examine social and exercise behaviour. However, all these studies used psychological variables that have different characteristics from physical performance variables. Schutz (1995, 1998) examined the stability of performances in sports, but focused more on the stability of professional players in terms of their relative positions on several performance records and the stability of factor structures. Although there are clear benefits of using LGM for longitudinal data, there has been an obvious lack of studies using LGM in physical performance research. This may be due to the lack of researchers' knowledge and the lack of proper guidance about LGM methodology, as well as the practical problems of applying LGM to longitudinal and multivariate physical performance data previously discussed.

The Purposes of the Study

There are four purposes of this study. First, there is an inadequate body of literature, especially in the Human Kinetics area, in which LGM is presented in sufficient detail to allow practitioners to easily follow and apply this statistical model in a longitudinal study. Thus, the first purpose of the present study is to introduce LGM to Human Kinetics researchers. More specifically, by analyzing real data, the present study includes the examination and presentation of: (a) how an individual level of developmental change is examined, (b) how predictors of change are implemented in the LGM statistical model, and (c) how the change of a multivariate latent factor can be examined.

The second purpose is to examine the merits and practical problems of LGM in the analysis of longitudinal physical performance data. Based upon the findings of the first purpose, the merits and practical problems of LGM are examined and compared with those of traditional analysis models such as ANOVA. In the present study, the examination of the merits and problems of LGM were made from a practical rather than a theoretical point of view. Many published LGM studies have shown the theoretical merits of LGM, but inadequate attention has been given to the practical problems of using this statistical model.

The third purpose is to examine the developmental patterns of children's physical performances. Although developmental patterns of children's physical performance have been studied, most previous researchers based their conclusions solely on group statistics, and thus the children's development and the variations in development of physical performance were not adequately examined. By using LGM, the present study provided more informative results regarding the children's development in physical performances than previous studies. Specifically, the present study includes the investigations of: (a) the individual level of developmental patterns in physical performances in childhood (between ages 8 and 12), (b) the variables that explain (predict) the between-person variations in the development of physical performance, and (c) the validity of multivariate latent factors as measures of longitudinal development in physical performances as well as the children's developmental patterns in multivariate latent physical performance factors. Because the current study had to use already existing data due to the difficulty of

obtaining a new longitudinal data set, the research questions regarding the development of specific physical performances were established based on the available variables.

The fourth purpose is to compare the capacity of the two most widely used longitudinal factor models, LGM and a quasi-simplex model, to accurately estimate reliability for longitudinal data under various conditions. This is important, especially with longitudinal studies of physical performance in which the measurements are costly. If valid reliability estimation for longitudinal data can be obtained by means of an analysis, it will be a benefit for the longitudinal study of physical performance. The conditions were varied to examine the effects of: (a) the magnitude of correlation between the initial status and change, (b) the magnitude of reliability, and (c) the magnitude of correlated errors on the estimation of reliability. The selected conditions of these three sub-purposes do not fully examine the effects of these variables on reliability estimation, but are expected to establish the basis for further research questions for future research on this topic.

Although all the purposes of the present study are closely related to each other, the fourth purpose is somewhat distinctive in that it requires computer simulated data sets to accomplish the purpose. Consequently, the present research endeavour was structured as two studies. In study 1, the first, second and third purposes of the study were accomplished. Study 1 includes the analyses of a longitudinal data set, the interpretation of the results, a discussion of the development of children's physical performance, and an elucidation of the merits and practical problems of using LGM in the analysis of longitudinal physical performance data. In study 2, the fourth purpose was accomplished. Study 2 includes the computer simulation of longitudinal data sets with known parameters, the analyses of these data sets, and the comparison and evaluation of the LGM and quasi-simplex models in estimating reliability.

STUDY 1. THE ANALYSIS OF LONGITUDINAL PHYSICAL PERFORMANCE DATA

STUDY 1-CHAPTER II. LITERATURE REVIEW

Analysis of Change and Latent Growth Models

The term "analysis of change" encompasses a vast amount of analysis issues and methods, from analyzing a simple treatment effect in an experimental study to analyzing a complex developmental change of an attribute. It is almost impossible to discuss all the issues and models of the analysis of change. The discussion in this section is limited to the analysis methods that are particularly relevant to the analysis of development (growth) or longitudinal data.

The analysis of change has long been an interest of researchers in almost all empirical sciences. Although systematic research about the analysis of change was initiated in the early 1900s, reports of research on growth can be dated back to the 18th century (Baltes & Nesselroade, 1979). However, it is believed that the study of change started much earlier in time. Recently, with the aid of computer development, a large number of research articles have been published on the statistical models and methods for the analysis of change (e.g., Collins & Horn, 1991; Gottman, 1995; Nesselroade & Baltes, 1979; von Eye, 1990). Selecting an appropriate analysis method from among these many available methods requires considerations of the research question and the theory behind the change of the variable used in the research (Rogosa, Brandt & Zimowski, 1982). In general, a researcher considers two major objectives of analyzing change in selecting an analysis method, description and explanation (Baltes & Nesselroade, 1979; Burr & Nesselroade, 1990). Description includes the direction, shape and amount of change, while explanation pertains to the predictor(s) of change, relationship of changes between two or more variables and what makes the differences among individuals in the rate of change in relation with other variable(s). Various analysis methods have different merits and limitations in accomplishing these two objectives.

Relative Methods and Limitations

The simplest but most restricted design for the analysis of change is the pre-post test design. In this design, gain scores (G scores) are calculated to represent change ($G \text{ score} = \text{Post test score} - \text{Pre test score}$), and then a statistical analysis is applied to these G scores (e.g., one sample t-test). However, the problems of the pre-post test design and G scores in analyzing change were detected early (e.g., Thorndike, 1924; Wilder, 1957; Zieve, 1940), and have been one of the major issues in the area of analysis of change (Schutz, 1989). The first problem is the ceiling-floor effect. Generally, the scores at the top end do not change upward and the scores at the bottom end do not change downward at post-test (Wilder, 1957). The ceiling-floor effect is related to the problem of "regression toward the mean", and causes a negative correlation between the initial score and the rate of change (Thorndike, 1924; Zieve, 1940). Second, the G scores are inherently unreliable (Lord & Novic, 1968). This has been one of the major drawbacks to the use of G scores as measures of change (Burr & Nesselroade, 1990). Third, the G

scores that are based on only two points in time do not adequately describe any nonlinear change over time (Rogosa et al., 1982). This is an especially serious limitation in a developmental study where the trajectory of a variable is often a major interest. Although Rogosa (1995) argued that the ceiling-floor effect and unreliability of G scores may not be problems in certain situations, the pre-post design and G scores have limitations in accomplishing both objectives of analysis of change, description and explanation, due to the above mentioned problems.

To overcome these problems of G scores, several alternatives have been proposed. Lord (1956) suggested a true change score, which is obtained after correcting for measurement errors in pre- and post-test scores. DuBois (1957) and Manning and DuBois (1962) suggested a residual gain score that is based on the difference between the predicted (via linear regression) post-test and raw post-test scores. There have been some other works on this issue (e.g., Tucker, Demarin & Messic, 1966). However, Cronbach and Furby (1970) pointed that none of these adjustments on G scores were satisfactory.

Other suggestions have been made in the perspective of research design. One suggestion was to include a control group in the design, and compare the change between the treatment and control groups. The traditional analysis of variance (ANOVA) procedure is used to analyze this type of data. The analysis of covariance (ANCOVA) was also frequently used when there exist differences among groups in pre-test scores. This is, however, most suitable for experimental research. Another simple suggestion was employing multiple time points in the measurement (Nesselroade, Stigler & Baltes, 1980; Rogosa & Willet, 1985b). This is especially important in studying development or growth, because this allows one to examine a nonlinear change of an attribute over time. With more than two points in time, various types of analysis methods can be employed, but traditionally, the ANOVA procedure has been mostly used. The ANOVA with trend analysis (polynomial contrasts) is useful because it allows one to examine the change in mean scores, to decompose the variance into linear, quadratic, cubic etc. components, and to examine interaction effects where there are multiple groups in the design. The ANCOVA procedure also provides more valid tests of differential change among groups with multiple testing periods (Richards, 1975).

Certainly, multiple time points of measurement provide a better opportunity in describing change in a longitudinal study. However, ANOVA has still limitations in the explanations of the change (i.e., in examining the causes of change, relationship of changes between two or more variables and what makes the differences among individuals in the rate of change in relation with other variables). Although a regression approach suggested by Hummel-Rossi and Weinberg (1975) may provide some insight into explaining change, at most, the ANOVA and ANCOVA procedures allow one to examine only the differences among groups in change. The ANOVA procedure can also be misleading. For example, Schutz and Park (in press) presented an example where ANOVA failed to detect important aspects of change (discussed in Study 1-Chapter V) if it is not properly used. The ANOVA and ANCOVA also suffer from quite restrictive assumptions underlying statistical models, such as

sphericity and random assignment of subjects to groups. Although multivariate analysis of variance (MANOVA) or doubly multivariate analysis of variance (DM MANOVA) for multiple indicator variables can be used under the violations of sphericity (Schutz & Gessaroli, 1987), these methods also have limitations in that they are based solely on the comparisons of mean scores. Fundamentally, these traditional methods do not fully use the information that longitudinal data provide. Although in MANOVA the covariances between different variables within a time point are partially used in obtaining the best linear combination of variables, in general, the covariances between repeated measurements are not adequately employed in the analysis (Labouvie, 1982). This leads to inadequate description and explanation of change with these methods.

Other approaches have been used for the analysis of change such as the application of stochastic models and time series analysis, growth curve fitting, qualitative analysis of change, the application of multi-level analysis (hierarchical linear model) and the application of factor analysis. The application of a stochastic model for the analysis of change is based on the probability of an individual achieving any one of a number of possible scores at some time in the future, given a current score (Schutz, 1970). A special case of the stochastic model approach is time series analysis. Time series analysis takes account of the change of each time interval of the data, and estimates a mathematical model that predicts the score at certain time point (Cromwell, Labys & Terraza, 1994; Crosbie, 1995; Frederiksen & Rotondo, 1979). This approach is often applied to a single subject (or any single measurement unit) who is measured at several time points, and is used extensively by econometricians in predicting economic indices (e.g., stock price). This procedure requires a large number of time points (more than 50), and this is not feasible in a typical developmental study done with human subjects. Also, this approach focuses mainly on the shape of the change, thus may be useful for the description of the change, but has very limited utility in the explanation of the change.

A similar method that has been used for a long time in biological and medical sciences is growth curve fitting. Some classify this method as a special case of time series analysis (von Eye, 1990). Essentially, this procedure involves finding the best fitting mathematical model and its parameters that explain the change of a variable as a function of time (Rogosa et al., 1982). The mathematical model ranges from a simple linear model or polynomial model that can be estimated by least squares method to a more complex triple-logistic model with marginal maximum likelihood estimation and multilevel statistical procedures (Bock & Tissen, 1976, 1980; Rogosa et al., 1982; Tissen & Bock, 1990). The measurement unit is often a single subject (or single measurement unit) as in other time series analysis, and the collection of parameters of the same curve model that are fitted to several subjects can be used in the second level of analysis to examine the relationship between change and other variable(s). This approach is different from traditional ANOVA procedure in that the change is described at an individual level. However, as Tissen and Bock (1990) noted, this approach requires approximations at many stages, thus leads to a questionable utility of this approach for the explanation of the change purpose. This area

still requires more development (Tissen & Bock, 1990).

Qualitative change models have also been used in several areas. In general, a qualitative change model is employed when the variable of interest represents the change in a discrete state (measured on a nominal or categorical scale). Various statistical models have been used to analyze such data, including the longitudinal Guttman simplex model (Collins & Cliff, 1985), log linear model (Goodman, 1972, 1978), logit or probit regression models (Goldberger, 1964), and hazards modeling (Allison, 1982). The regression models and hazards modeling have some merits in that these models may include both categorical and continuous independent variables that may explain the change (Burr & Nesselroade, 1990). Although these methods have been mostly limited to the univariate case, these are powerful tools in describing and explaining change. Additionally, these models are very useful in a criterion-referenced assessment context (Schutz, 1989).

Recently, several new statistical models for the analysis of change have been suggested. The application of a hierarchical linear model (HLM), suggested by Bryk and Raudenbush (1987), has some merits in that it describes the change at the individual level and one may include predictors of change in the model. In addition, this model does not require the same number of repeated measures for each individual, and the measurement intervals need not be the same for all individuals. Although this model is limited to the univariate case and lacks flexibility in modeling, it provides powerful methods to describe and explain change.

Application of factor analysis techniques to longitudinal data is also a relatively new approach. An auto-regressive model (quasi-simplex model) has been widely used (e.g., Jöreskog, 1970; Rogosa & Willet, 1985a) since it was introduced in the 1950s (Guttman, 1954). However, this model concentrates more on the stability, or change of relative positions of subjects within a group by using only the covariance matrix as data (Rogosa & Willet, 1985b). A more detailed description of this model is presented in Study 2-Chapter II, "Longitudinal Reliability" section. Another relatively new technique is employing means as well as covariances in the factor analysis model (Meredith & Tisak, 1984, 1990). This model has some merits in analyzing change over traditional methods. This model is called a "Latent Growth Model".

Latent Growth Model

Based on the formative work of Rao (1958) and Tucker (1958), latent growth model (LGM) was first suggested by Meredith and Tisak (1984, 1990) within the framework of structural equation modeling (SEM), and later extended by others (e.g., McArdle, 1988; McArdle & Epstein, 1987; Muthen, 1997). Although the name "Latent Growth Model" contains the term "growth", this statistical model can be applied to any repeated measures data. However, it may be most useful when one has an a priori hypothesis regarding the pattern or shape of the change of a variable. LGM has several merits in describing and explaining change. First, by using both the means and covariances of repeatedly

measured variables as data, a LGM allows one to take into account the individual level of change as well as the group level of change. Second, individual change can be represented by either a straight line or a curvilinear trajectory. Third, occasions of measurement need not be equally spaced. Fourth, measurement errors can be accounted for by the statistical model, and reliability estimates for the variables at each time point are available. Fifth, multiple predictors or correlates of change can be easily included in the model. Finally, as in general SEM analysis, statistical models are very flexible, allowing one to extend the basic idea in several ways in order to test various hypotheses, such as multivariate LGM, multi-group analysis and cohort sequential analysis (McArdle, 1988; McArdle & Epstein, 1987; Meredith & Tisak, 1990; Muthen, 1997).

The Basics of LGM

The basic idea of a LGM (Tisak & Meredith, 1990) is that the growth (change) of an attribute is an unobservable latent trait. Thus, in a LGM, change is described by one or more latent variables (factors). We may express the observed score for the i th individual at time t , $Y_i(t)$, as

$$Y_i(t) = \sum_{k=1}^d W_{ik} \lambda_k(t) + E_i(t) \quad (1.2.1)$$

where $\lambda_k(t)$ is the k th unspecified (or specified) longitudinal curve for all individuals and W_{ik} is the weight that the i th individual attaches to the $\lambda_k(t)$ curve, $i = 1, 2, \dots, N$. $E_i(t)$ is the error or residual of the i th individual. Let m be the number of repeated measurements and λ_d be the factor loadings of d th order curve factor, then we can express this in matrix form as follows;

$$\mathbf{y}'_i = (Y_i(t_1), Y_i(t_2), \dots, Y_i(t_m)) \quad (1.2.2)$$

$$\mathbf{e}'_i = (E_i(t_1), E_i(t_2), \dots, E_i(t_m)) \quad (1.2.3)$$

$$\Lambda = \begin{bmatrix} \lambda_1(t_1) & \lambda_2(t_1) & \dots & \lambda_d(t_1) \\ \lambda_1(t_2) & \lambda_2(t_2) & \dots & \lambda_d(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1(t_m) & \lambda_2(t_m) & \dots & \lambda_d(t_m) \end{bmatrix} \quad (1.2.4)$$

Then, equation (1.2.1) becomes,

$$\mathbf{y} = \Lambda \mathbf{w} + \mathbf{e} \quad (1.2.5)$$

The subscripts were omitted in equation 1.2.5 for simplicity. There are assumptions that are imposed in

the factor analysis model. The assumptions include the mean error scores to be zero ($E[e] = 0$), the covariances between errors to be zero ($E[ee'] = \Theta$, a diagonal matrix), and the covariance between the factor and the error to be zero ($E[we'] = 0$), where $E[.]$ is the expectation operator. In addition, let $E[w] = \alpha$, and $E[ww'] = \Psi$. Then the expected mean vector and the variance-covariance matrix can be expressed as;

$$E[y] = \Lambda\alpha = \mu \quad (1.2.6)$$

$$E[yy'] = \Lambda\Psi\Lambda' + \Theta = \Omega \quad (1.2.7)$$

These look essentially the same as the general factor analytic form, the difference being that $E[y] \neq 0$ and $E[w] \neq 0$. Thus, these are the basic equations of factor analysis with means (Harman, 1976; Meredith & Tisak, 1990; Mulaik, 1972; Tisak & Meredith, 1990).

With an additional assumption of joint multivariate normality of the y variables, maximum likelihood estimation and hypothesis testing is possible. Let 1 and 0 denote column vectors of ones and zeros, respectively, and π a vector of free parameters for the model. Then a partitioned matrix $\Sigma(\pi)$ is defined as,

$$\Sigma(\pi) = \begin{bmatrix} \Omega & \mu \\ \mu' & 1 \end{bmatrix} = \begin{bmatrix} \Lambda & 0 \\ 0' & 1 \end{bmatrix} \begin{bmatrix} \Psi & \alpha \\ \alpha' & 1 \end{bmatrix} \begin{bmatrix} \Lambda & 0' \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} \Theta & 0 \\ 0' & 0 \end{bmatrix} \quad (1.2.8)$$

and the partitioned matrix S consists of the covariance matrix and mean vector of y variables that are obtained from the data:

$$S = \begin{bmatrix} \hat{\Omega} & \hat{\mu} \\ \hat{\mu}' & 1 \end{bmatrix} \quad (1.2.9)$$

Maximum likelihood estimation minimizes the fitting function, F_{ML} , where

$$F_{ML} = \log|\Sigma(\pi)| + tr(S\Sigma^{-1}(\pi)) - \log|S| - t \quad (1.2.10)$$

and "log" is the natural logarithm, $|\cdot|$ is a determinant of a matrix, $tr(\cdot)$ is a trace of a matrix, and t is the number of measured variables (or the number of repeated measures of the same variable). Maximum likelihood estimation procedures require a relatively large sample size (i.e., 200), and can be done using any one of a number of commercially available SEM programs (e.g., LISREL, EQS, MPLUS or

SEPATH). These programs also provide several goodness-of-fit indices for the fitted model.

Another way of representing a LGM is through the use of a path diagram. This is common practice in a SEM, and is a conceptually easier way to represent and understand these rather complex relationships. Figure 1.2.1 shows the diagram for a linear LGM. Following the general rules of SEM, boxes represent observed (measured) variables, and ovals represent unobserved latent variables (factors). In Figure 1.2.1 there are five observed variables, labelled as “Time1 to Time5” (instead of y variables as used in the matrix equations), and two latent variables, named “Intercept” and “Slope” (these two latent variables are included in the matrix W in the matrix equations). Arrows represent the relationships among observed and latent variables. Single-headed arrows are used to show a causal relationship between variables where the variable at the tail of the arrow is hypothesized to cause (or explain) the variable at the head of the arrow. The magnitude of causal relationship between an observed variable and a latent variable is represented by a path coefficient (or factor loading, λ_s), and it is equivalent to a B coefficient (non-standardized slope coefficient) in a regression analysis. The path coefficients in Figure 1.2.1 are all fixed at 1s for the intercept factor and at 0, 1, 2, 3, and 4 for the slope factor. Thus, in Figure 1.2.1, the Time1 through Time5 variables are dependent (endogenous) variables, while the intercept and slope factors are independent (exogenous) variables. The relationship between observed and latent variables can be represented by a linear equation. For example, the Time1 variable is represented as; $\text{Time1} = (1) \times \text{Intercept} + (0) \times \text{Slope} + e_1$. The double-headed arrow shows the covariances (correlations in standardized units) between two variables.

Unlike a regression analysis or a usual SEM, all path coefficients in Figure 1.2.1 are fixed at certain values. Because of these fixed coefficients, the latent variables have specific meanings. The “Intercept” factor represents a true score at the first time point (initial status), and the “Slope” factor represents the true rate of linear change over time. Each subject has his or her own intercept and slope, and it is expected that there will be between-subject variation in the intercept and in the slope. The mean and the variance of the intercept factor are represented by α_I and ψ_{II} , respectively. The mean and the variance of the slope factor are represented by α_S and ψ_{SS} , respectively. The covariance between the intercept and slope factor is represented by ψ_{IS} . An error (e) represents that part of an observed variable that is not explained by the intercept and slope factors. Thus according to Figure 1.2.1, the score of each individual at each time point can be expressed as;

$$\text{Time1} = \text{Intercept} + (0) \times (\text{Slope}) + e_1$$

$$\text{Time2} = \text{Intercept} + (1) \times (\text{Slope}) + e_2$$

$$\text{Time3} = \text{Intercept} + (2) \times (\text{Slope}) + e_3$$

$$\text{Time4} = \text{Intercept} + (3) \times (\text{Slope}) + e_4$$

$$\text{Time5} = \text{Intercept} + (4) \times (\text{Slope}) + e_5$$

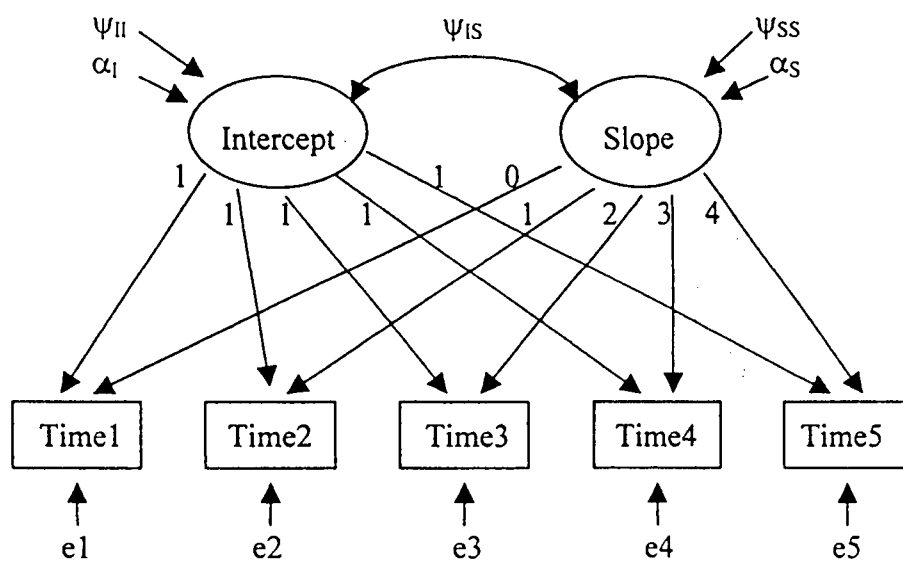


Figure 1.2.1. Linear LGM

Thus apart from error, which is unique at each time point, the difference between Time1 and Time2 is Slope, and the difference between Time1 and Time3 is $2 \times \text{Slope}$ and so on, implying a linear change over time. A score of an individual at any time point is a function of one's own intercept and slope. The means and variances of the observed variables and covariances between observed variables are used as data for the statistical analysis. The means and variances of the latent variables and covariance between the two latent variables are estimated by the model. The mean and the variance of the intercept factor are the true mean and the between-subject variance of the initial time point, respectively. The mean of the slope factor is the average linear change between adjacent time points, and the variance of the slope factor is the between-subject variation of the magnitude of the linear change over time. The covariance between the two factors shows the magnitude and the direction (positive or negative) of the relationship between the score at the initial time point (Time1) and the rate of the change. The variances of the errors are also estimated by the model (Lawrence & Hancock, 1998).

Extensions of LGM

The basic model can be extended in several ways. First, by adjusting some of the path coefficients of the slope factor, or by adding additional change factor(s), one can specify a model that describes a curvilinear change. In Figure 1.2.1, if the last three path coefficients of the slope factor are freely estimated rather than fixed at specific values, the model describes a less restricted type of change. This model is often called an "unspecified curve model" (McArdle, 1988). In this model, the second path coefficient should be still fixed at 1 to provide a scale to the change factor. On the other hand, if one add a third factor to a linear model (Figure 1.2.2), it becomes a quadratic model that describes a quadratic change over time. The higher order curvilinear model is formed in the same way. Another kind of curvilinear model is possible. For example, Willet and Sayer (1996), in their introductory paper, transformed the subjects' ages (time point) using a logarithmic function and applied a linear LGM.

Predictor(s) of the intercept and change can be easily included in a LGM. Figure 1.2.3 shows a linear LGM with a time-invariant predictor. The γ s represent the path coefficients from the predictor variable to the intercept and slope factors, and the magnitude of these coefficients shows the strength of the predictor variable in explaining these two factors. In addition, incorporation of time-varying predictor(s) in the model is also possible (Kaplan, 2000).

Tisak and Meredith (1986), based on the work of Tucker (1966), showed a multivariate generalization of the LGM. A multivariate LGM includes several variables that are repeatedly measured at multiple time points. McArdle (1988) described two types of multivariate LGM and named these two models a "factor-of-curves" model and a "curve-of-factors" model. In a 'factor-of-curves LGM' (Figure 1.2.4), several first-order intercept and change factors explain the trajectories of several variables over time, and the correlations among intercepts and among change factors are explained by the second-order intercept and slope factors. This model is more parsimonious than the model in which the second-order

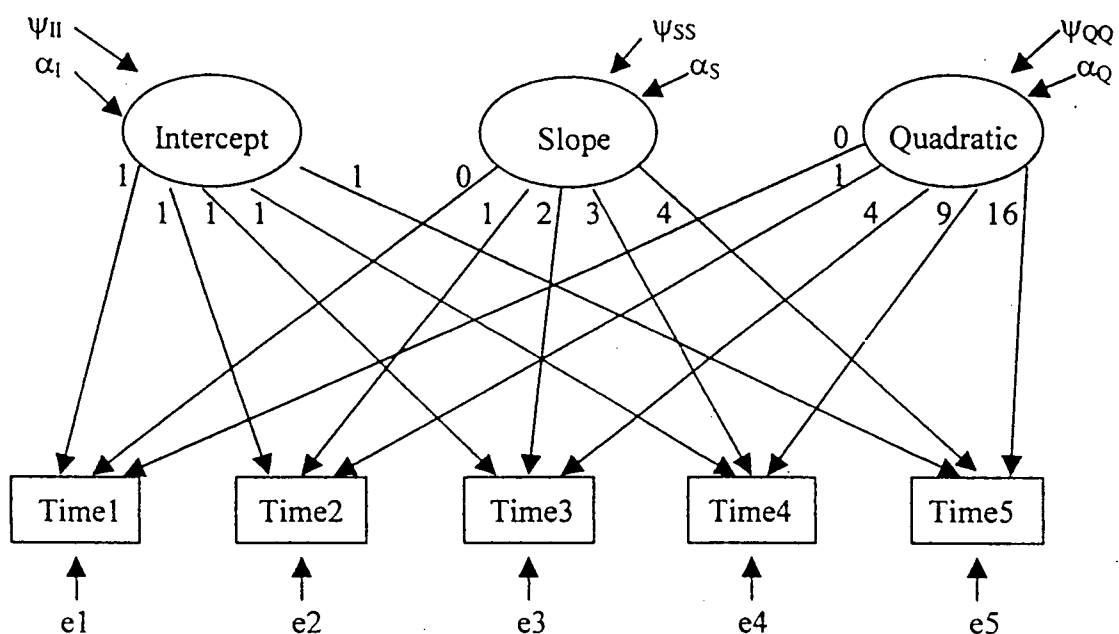


Figure 1.2.2. Quadratic LGM (covariances between latent factors are omitted)

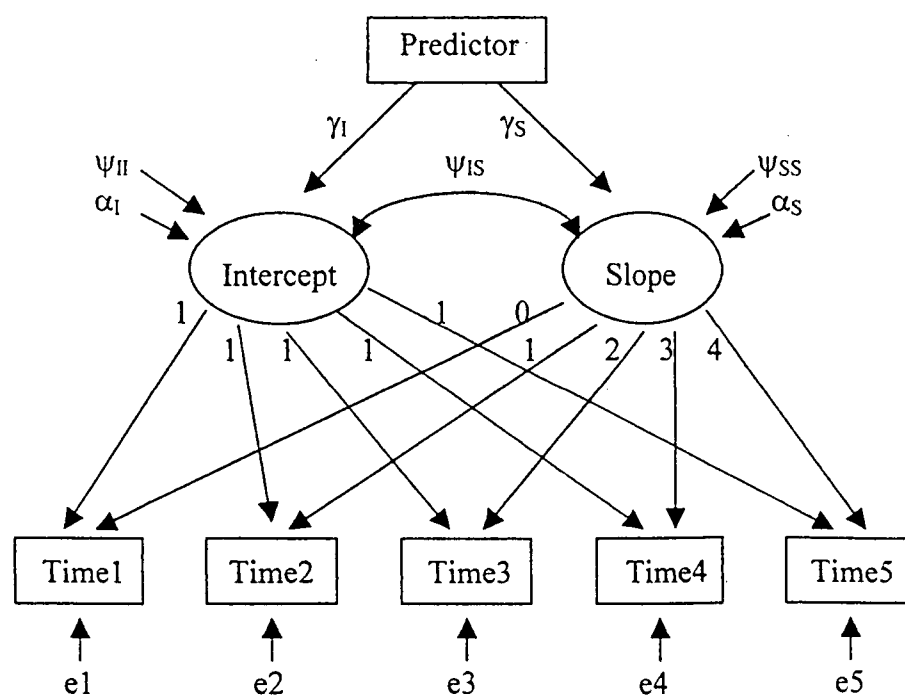


Figure 1.2.3. Linear LGM with a time-invariant predictor

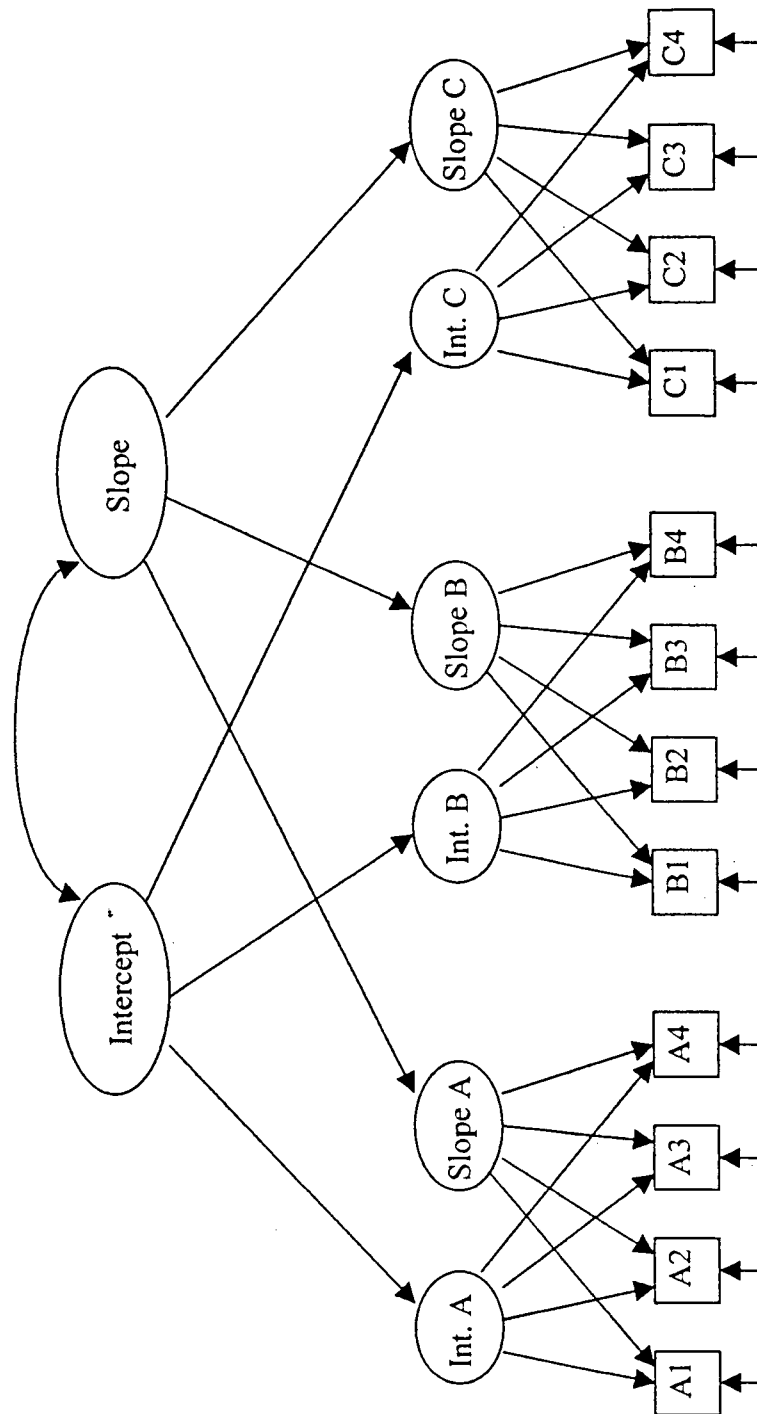


Figure 1.2.4. A Factor-of-curves model: 3 variables and 4 time points with linear change (parameters are omitted)

factors are not specified, and the correlations between the first-order factors are estimated; however, when two or more variables show different types of change, it is unclear how to specify the model.

In a 'curve-of-factors LGM' (Figure 1.2.5), on the other hand, several measures at a single time point form a latent construct, and the change of this latent construct over time is explained by second-order intercept and change factors (McArdle, 1988; Tisak & Meredith, 1990). A 'curve-of-factors LGM' requires a few extra steps in the analysis because before a 'curve-of-factors LGM' is applied to a data set, several conditions have to be satisfied. First, one has to examine if the hypothesized factor structure holds at each time point. That is, one has to test if the measured variables form a factor at each time point (i.e., examination of a measurement model). Second, once a factor is believed to be formed at all time points, one has to examine if the factor loadings for the same variable are equal over time. In other words, one should test if the same attribute (factor) is measured over time. Thus a 'curve-of-factors LGM' analysis includes the following steps: (a) test of a measurement model, (b) test of equality of factor loadings over time, (c) selecting the best growth model (i.e., linear, quadratic, cubic or curve etc. model), and (d) test of predictor effects, if necessary.

Other types of extensions and applications have also been made. One may employ a multi-group analysis model when the examination of the differences among groups in the change of an attribute is a main concern. Willet and Sayer (1996) applied this model to compare healthy children and non-healthy children in the growth of reading and mathematics ability. Muthen and Curran (1997) further extended this application to an experimental clinical trial and examined the treatment effect by comparing the treatment and control groups. They also included the interaction effect in the model and developed a procedure to obtain a statistical power to detect a significant treatment effect. Another extension of the LGM is the application to the cohort sequential design (Meredith & Tisak, 1990). As a matter of fact, this is another application of multi-group analysis of LGM to sequential data. In this model, several cohort groups are included (and treated as different groups), but the time (age) is specified as continuous across cohort groups. The extension of LGM to a binary outcome (dependent) variable has been also made (Muthen, 1996); however, as Muthen (1996) noted, this method requires a weighted least squares estimation that is computationally heavy and requires a large sample size.

Development of Physical Performance

The term "physical performance" encompasses a broad range of systematic human body movement. As well, there exist a huge number of tests that were developed to measure several types of physical performance. In this section, only the physical performance tests and corresponding physical performances that were employed in the data sets used in the present study are discussed.

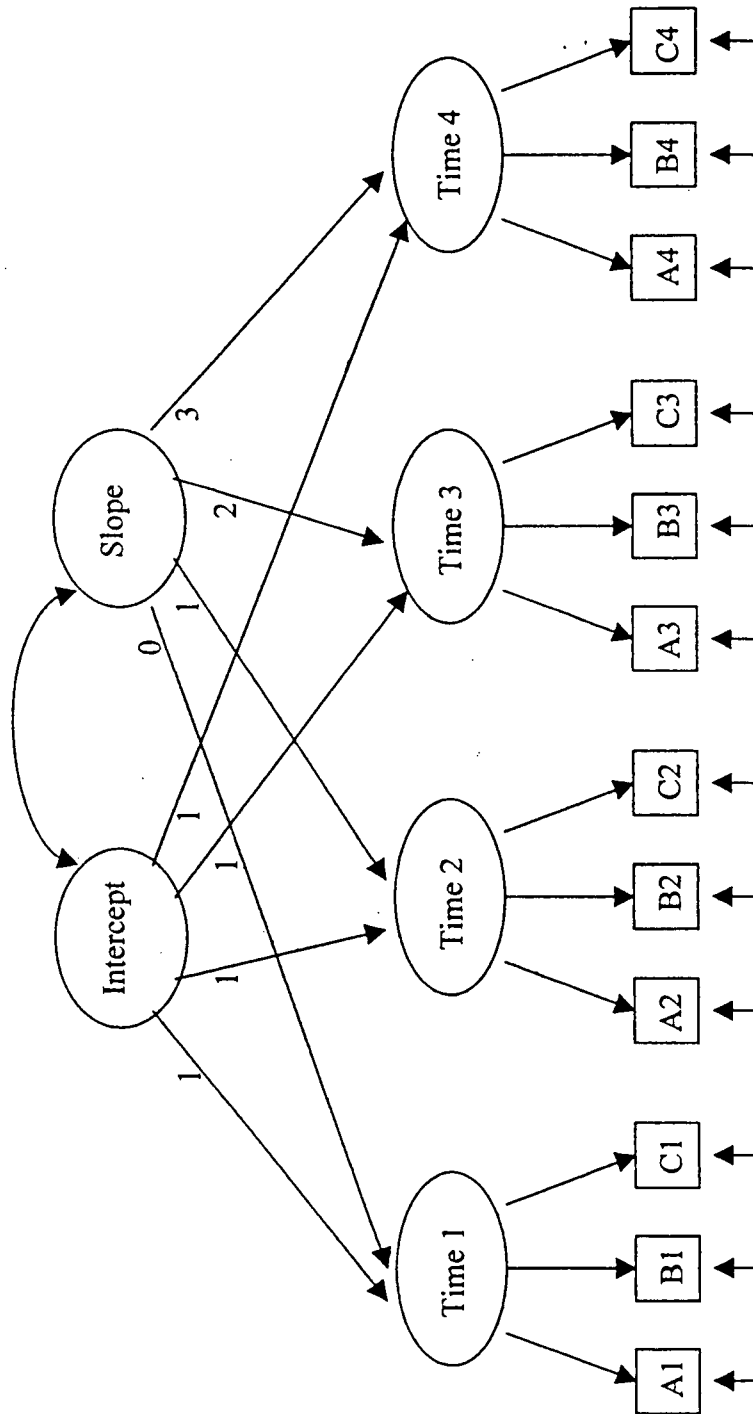


Figure 1.2.5. A curve-of-factors model: A linear model with 3 variables and 4 time points (parameters are omitted)

Physical Performance Tests

Flexed-arm-Hang (FAH)

The FAH is used to measure upper arm and shoulder girdle muscular strength and endurance (Corbin & Pangrazi, 1992). Upper body muscular strength and endurance are considered to be an important component of health-related physical fitness (AAHPERD, 1988; President's Council on Physical Fitness and Sports [PCPFS], 1987). Consequently, the FAH test is included in health-related physical fitness test batteries such as The Chrysler Fund-AAU Fitness Test (1987) and FITNESSGRAM Test (Institute for Aerobics Research, 1987). This test is often used instead of a pull-up test for females and younger boys. The FAH test has shown good reliability, but questionable validity (Cotton & Marwitz, 1971; Pate, Burgess, Woods, Ross & Baumgartner, 1993). Pate et al. (1993) reported that the concurrent validity of this test is relatively low (.50); however, they showed evidence of construct validity of the test. Pate et al. (1993) also noted that the subject's performance is confounded by body weight. Because the body weight of a subject affects the performance of this test significantly, this test should be regarded as a measure of strength and endurance relative to one's body weight. Another problem of the FAH is the frequent occurrence of a relatively large percentage of zero or near-zero scores, particularly among girls and young boys (Reiff, Dixon, Jacoby, Ye, Spain & Hunsicker, 1986; Ross & Gilbert, 1985).

Jump-and-Reach (JAR)

The JAR test is used to measure explosive power of the leg extensors (Safrit & Wood, 1995). Jumping tests such as JAR and "Standing Long Jump (SLJ)" have been described as tests of power and of explosive strength (Baumgartner & Jackson, 1999; Fleishman, 1964; McCloy & Young, 1954). The JAR test, first developed by Sargent (1921), is also referred to as the "Sargent Jump Test" or "Vertical Jump Test". This test is one of the most widely used tests to measure jumping ability and power, and is especially relevant for testing athletes such as volleyball and basketball players because jumping is an important part of those games (Baumgartner & Jackson, 1999). The reported validity (.78) is in an acceptable range and the reliability (.93) is relatively high (Safrit & Wood, 1995); however, others reported low correlations between jumping tests and mechanical measures of power (Barlow, 1970; Considine, 1970). Another concern is the negative correlation between jumping tests and body weight (Baumgartner & Jackson, 1999). Consequently, the JAR should be regarded as a test of lower leg power relative to body weight rather than as a test of absolute power.

Sit-and-Reach (SAR)

The SAR test is used to measure flexibility of the low back and posterior thigh, and has been applied to all age groups (Safrit & Wood, 1995). This test is often included as a test item in health-related physical fitness test batteries such as the AAHPERD Physical Best Test Battery (AAHPERD, 1988), Chrysler Fund-AAU Fitness Test (1987), Fit Youth Today Test (American Health and Fitness Foundation, 1986), FITNESSGRAM Test (Institute for Aerobics Research, 1987) and the Presidential

Physical Fitness Test Battery (PCPFS, 1987). Reported validity estimates of this test are varied, ranging from .60 to .90 (Jackson & Baker, 1986; Safrit & Wood, 1995). Jackson and Baker (1986) found that the SAR test had moderate validity (.60 to .73) in measuring hamstring flexibility; but low validity (.27 to .30) in measuring low back flexibility. The reliability of this test is relatively high: reported reliability coefficients ranged from .70 to .99 (Jackson & Baker, 1986; Safrit & Wood, 1995). Performance on the test, however, is somewhat dependent on the ratio of trunk length to lower body length (Safrit & Wood, 1995).

Agility Shuttle Run (ASR)

The ASR test is used to measure agility, running speed and change of direction (Corbin & Pangrazi, 1992). Agility is an attribute that is more strongly related to a specific sport (Safrit & Wood, 1995), thus, the ASR test is often included in performance-related physical fitness test batteries such as the AAHPER Youth Fitness Test (AAHPER, 1976) and Manitoba Physical Fitness Performance Test (Manitoba Department of Education, 1977). Interestingly this test has also been included in health-related physical fitness test batteries such as the FITNESSGRAM (Institute for Aerobics Research, 1987) and the Presidential Physical Fitness Test Battery (PCPFS, 1987). The ASR test has been widely used in various school settings and applied to all age groups from age 6 through adult. Although there have been a few studies that revealed an evidence of construct validity (e.g., Hilsendager, Stow & Ackerman, 1969), no studies that are directly related to the validity of the ASR test have been conducted. In relation to this, Safrit and Wood (1995) noted that "Agility" is highly specific to a task; thus there is no valid measure of overall agility. A task-specific measure of agility might be used as a measure of performance-related physical fitness. The reported reliability coefficients (.68 to .75) were in an acceptable range (Klesius, 1968). Several studies have shown that there exists a practice effect on ASR performance, and recommended several practice trials before the actual measurement or more than two trials in a measurement occasion (Baumgartner & Jackson, 1970; Hilsendager, Stow & Ackerman, 1969; Marmis, Montoye, Cunningham & Kozar, 1969).

Endurance Shuttle Run (ESR)

This test measures leg muscular endurance. Muscular endurance is defined as "the ability of the muscle to maintain submaximal force levels for extended periods" (Heyward, 1984), or "the ability to persist in physical activity or to resist muscular fatigue" (Baumgartner & Jackson, 1999). Muscular endurance is often measured by repetitions of a movement of a specific muscle group. Because the different muscle groups may show different levels of endurance, a specific muscle group must be selected and tested, given the purpose of the measurement (Safrit & Wood, 1995). Most field tests were developed to measure the endurance of the arms and shoulder girdle, (e.g., FAH), the endurance of abdominal muscles (e.g., sit-ups) and cardiorespiratory endurance (e.g., distance run). There are a few tests that measure the endurance of the leg muscle group (e.g., leg press). These tests generally require one to perform a movement to exhaustion or to perform as fast as one can during a specific time period

(usually the length of time is 1 min). The ESR test has seldom been used as a test of endurance, and to this researcher's knowledge no published studies on this test are available. This test, however, was an integral component of the test battery used in the longitudinal study that generated the data used in this study. The ESR test requires repetitions of leg movement, and the mean of the test scores found in the current study ranged from 43.93 to 37.64 seconds at age 8 to 12.5 (see Table 4.6 and Table 4.9 in chapter 4), indicating that this test measures leg muscular endurance (or anaerobic capacity of the leg muscle group). Because of the relatively short distance of running and completion time, this test may include the elements of speed and running efficiency as well. This test should also be regarded as a measure of relative endurance rather than absolute endurance because the body weight affects the performance. Information regarding the validity and reliability of this test is not available.

30-Yard Dash (DASH)

The DASH is used to measure running speed (Corbin & Pangrazi, 1992). Running speed is regarded as a performance-related attribute, and measured by the elapsed time required to run a specified distance or the distance the subject can run during a specified time period. Various running distances or time periods have been used depending on the purpose of the test and the subjects' ages: ranging from 10 to 60 yards or from 4 to 8 seconds, respectively (Baumgartner & Jackson, 1999; Fleishman, 1964; Haubenstricker & Seefeldt, 1986; Jackson, 1971; Jackson & Baumgartner, 1969; Seils, 1951). A fifty-yard dash test is most widely used, and this test is included in the AAHPER Youth Fitness Test battery (AAHPER, 1976). This test has been applied mainly to children and adolescents. Construct validity has been established for this test (Hastad & Lacy, 1994), but there has been lack of studies in which other types of validity of the DASH test were examined. Safrit and Wood (1995) noted that the 50-yard dash is a function of running efficiency as well as pure speed. In addition, this test has an element of explosive power, thus it shows a relatively high correlation with the performance of the JAR and SLJ tests (Costill, Miller, Myers, Kehoe & Hoffman, 1968; Marsh, 1993). The reported reliability coefficients of the 50-yard dash test are relatively high, ranging from .86 to .94 (Fleishman, 1964; Jackson & Baumgartner, 1969), and the reliability of the 30-yard dash test was also relatively high (Seils, 1951).

Standing Long Jump (SLJ)

The SLJ is used to measure explosive power of the lower limb extensors (Corbin & Pangrazi, 1992). The nature of this test is very similar to that of JAR test. Because of the ease of administration of the test, it has been widely used in school and nonschool settings, and included in several performance-related fitness test batteries such as the AAHPER Youth Fitness Test (AAHPER, 1976). The SLJ is generally accepted as an adequate measure of explosive power, although an element of timing (skill) exists in executing the jump that does not exist to the same extent for JAR (Safrit & Wood, 1995). The reported reliability coefficients are high, ranging from .83 to .99 (Klesius, 1968; Safrit & Wood, 1995). However, like the JAR test, the SLJ test has shown low correlations with mechanical measures of power

(Barlow, 1970; Considine, 1970), and a negative correlation with body weight (Baumgartner & Jackson, 1999).

The Development of Children's Physical Performance

The discussion about the development of children's physical performance in this section is limited mostly to the age range 8 to 13 years, which Haubenstricker and Seefeldt (1986) recognized as "middle and late childhood". In addition, only the development of males is discussed because the present study includes only males as subjects. The specific interests of the present study are the pattern (shape) of development. Although an important issue in the present study is the identification of individual levels of change as well as the group level of change, the discussion in this section is based largely on group level statistics, as this was the only available information in the literature.

Many studies concerning the development of physical performance have been conducted since the early 1900s. These studies focused on issues such as the developmental process of motor skill acquisition (e.g., Halverson, 1931; Haubenstricker & Seefeldt, 1986; Vilchkovsky, 1972), the impact of training and other environmental factors on the acquisition of motor performance (e.g., Dusenberry, 1952; Halverson, Robertson, Safrit & Roberts, 1977; Werner, 1974), the relationship between physical performance and physical growth (e.g., Cearley, 1957; Clarke & Wickens, 1962; Selis, 1951), and comparisons between males and females in development (e.g., Morris, Williams, Atwater & Wilmore, 1982; Smoll & Schutz, 1990). These studies provide limited information regarding the development of children because the majority of these studies are cross-sectional rather than longitudinal. The lack of longitudinal studies is due to the difficulties of conducting a longitudinal study.

Muscular Endurance (FAH and ESR)

Different studies employed different tests and different age groups in examining the development of children's muscular endurance. Many studies indicated that children's muscular endurance improves linearly at an early age, but the rate of the improvement may decrease or increase after age 11 or 12. Montoye and Lamphiear (1977), in their cross-sectional study, found that children improved linearly in their arm strength between the ages of 10 and 12, and the improvement was larger between age 12 and 13. Bischoff and Lewis (1987) used a sit-ups test in their cross-sectional study, and reported similar findings in a little different age range. They showed that children's improvement was approximately linear between age 7 and 11, and the improvement was accelerated between age 11 and 12. Jones (1946) also reported similar results. On the contrary, Baumgartner, East, Frye, Hensley, Knox and Norton (1984) showed children's upper arm muscular endurance improved rapidly and then the rate of the improvement decreased between ages 7 and 9. Milne, Seefeldt and Reuschlein (1976) obtained similar results using a 400-ft shuttle run test on children aged between 6 and 8. Conflicting results were also found in Smoll and Schutz (1990). In their cross-sectional study, children showed larger improvements in the measure of FAH between age 13 and 17 than between age 9 and 13, while sit-ups

showed opposite results. Although the Canada Fitness Survey (1985) revealed a little more complex change (e.g., cubic change) in muscular endurance measured by push-ups and sit-ups tests between age 8 and 12, in general, the improvement of children was approximately linear or quadratic with larger change at an early age. Different results that were reported from various studies may be due to the inclusion of different muscle groups, different tests, and the subjects with different characteristics.

Power (JAR and SLJ)

The development of explosive power of the lower leg muscle group or of jumping ability is similar to that of muscular endurance. Many studies have reported a linear change between ages 5 and 12 in jumping ability as measured by the JAR and SLJ (Bayley, 1935; Clarke & Wickens, 1962; Herkowitz, 1978; Milne et al., 1976). Hauebnstricker and Seefeldt's (1986) summary of several longitudinal studies also showed a linear change of jumping ability until the age of 16. A few different findings have been also reported. The cross-sectional study by Marmis, Montoye, Cunningham and Kozar (1969) showed a cubic change of SLJ between ages 9 and 13. They reported a faster improvement between ages 10 and 12 than between ages 9 and 10, and slower improvement between ages 12 and 13. Nonlinear changes were also reported by Caskey (1968), Selis (1951), and Smoll and Schutz (1990).

Flexibility (SAR)

Unlike other children's physical performances, flexibility declines with age. The rate of decrement in childhood is very small: many studies showed that flexibility remains at the same level or slightly declines between ages 8 and 12. In studies by Bischoff and Lewis (1987) and Gallahue (1982), the decrement in children's flexibility becomes evident from age 10 for males. However, in a large norm establishment study by Hastad, Marett and Plowman (1983), the level of flexibility remains about the same between ages 8 and 12. The fiftieth percentile norms that were suggested by PCPFS (President's Council on Physical Fitness and Sports, 1987) and National Children and Youth Fitness Studies [NCYFS] (Ross & Gilbert, 1985; Ross, Pate, Delpy, Gold & Svilar, 1987) also supported these findings. Because of the small rate of decrement within this age range, it is difficult to observe the general shape of the change. Although Bischoff and Lewis's (1987) study implied a cubic change, this study was based on a very small sample size. Because flexibility showed two types of change, no change and decrement, within the age range from 8 to 12, it is possible to infer that the change occurs in a quadratic form.

Agility (ASR)

Although there is some disagreement on the rate of change (e.g., Marmis et al., 1969; Milne et al., 1976), generally, studies indicated that agility improves constantly in childhood. Selis (1951) reported a linear improvement in children aged from 7 to 9 in the ASR test. AAHPER's youth fitness test norm (AAHPER, 1976) also implied a linear improvement from age 10 to 15. In addition, Clarke and Wickens (1962) and Marmis et al. (1969) showed that the level of children's agility improved

linearly from age 10 to 12, but they found that there was no change in children's agility from age 9 to 10. Thus, the possibility of a quadratic change should be considered within this age range.

Running Speed (DASH)

The development of running speed appears to be similar to that of explosive power (jumping ability) and agility. Haubenstricker and Seefeldt (1986) reported that children improved constantly in running speed up to age 17. Marmis et al. (1969) and Morris et al. (1982) also supported this for the age range of 3 to 12 years. There were some other reports that showed a nonlinear development of running speed within this age range. Cearley (1957) concluded that children's development in running ability is nonlinear from age 9 to 13. Although Selis (1951) noted that children's running ability improved constantly, his results implied a faster improvement before school age. Milne et al. (1976) also supported this finding. The tendency of faster development in running speed may exist in an early age range.

In summary, children improve rapidly in their physical performance between ages 8 and 12.5, except for flexibility which generally declines over time. However, there has been lack of agreement regarding the pattern of change. Two notable patterns were a constant (linear) change, and a faster change at an earlier age. The different results among studies might be due to the fact that they included different age groups and different performance variables. In addition, most of the previous studies employed cross-sectional research designs and small samples. There has been lack of physical performance development studies that employ a longitudinal design with large samples.

The Factor Structure of Physical Performance

Attempts to identify an underlying factor structure of physical performance have been made since the 1930s (e.g., Buxton, 1938; Coleman, 1937; Metheny, 1938; Rarick, 1937). The purpose of early studies was to determine the existence of a general motor ability factor or to extract a number of underlying latent factors given various performance variables. Most of these studies showed that the earlier concept of general motor ability does not exist, but rather physical performance is specific to particular muscle groups or particular types of movement (Coleman, 1937; Cumbee, 1954; Fleishman, 1964; Rarick, 1937; Seashore, 1942). Various studies identified somewhat different factors because different studies employed different physical performance variables. However, there exists some agreement about the factors that may encompass the domain of physical fitness of motor performance in identified factors. In general, these factors can be categorized as strength, explosive power, speed, endurance (muscular and cardiorespiratory), coordination, balance, and flexibility.

Strength, muscular endurance and explosive power are among the most dominant factors that have been identified in numerous studies (e.g., Barry & Cureton, 1961; Fleishman, 1964; Larson, 1941; Rarick & Dobbins, 1975). In these studies, the three factors were recognized as different elements of strength. Fleishman (1964) conducted an extensive study of the factor structure of physical performance,

and reported three primary strength factors, named dynamic strength, static strength and explosive strength (power). In his study, dynamic strength included several muscular endurance test items (e.g., pull-ups, FAH), and explosive strength included several explosive power test items (e.g., JAR, SLJ). Others (e.g., Larson, 1941; Marsh, 1993) supported these findings. Fleishman further showed some evidence of separate components that are specific to a particular muscle group (e.g., upper arm strength) or a particular type of movement (e.g., running).

Others disputed Fleishman's earlier categorization, and supported his latter notes that physical performance is specific to particular muscle groups or particular types of movement. Studies by Cousins (1955), Liba (1967) and Start, Gray, Glencross and Walsh (1966) suggested that dynamic, static and explosive strength are not unidimensional factors and that separate factors of arm and leg involvement exist in each of these three elements. Jackson (1971) also supported this and further showed that there exist distinctive factors of running, jumping and throwing. The specificity of physical performance to particular muscle groups is also evident in endurance. Baumgartner and Zuidema (1972), in their factor analysis of physical fitness tests, identified three main factors: upper body strength and endurance, leg strength and endurance and cardiorespiratory endurance.

One of the notable results from factor analytic studies of physical performance is that a speed run variable is included as an element of explosive power, and regarded as a measure similar to the JAR and SLJ. Start et al. (1966) found a high correlation between explosive power and speed, and concluded that explosive power is linked with speed rather than strength. Costill, Miller, Kehoe and Hoffman (1968) also reported relatively high correlations between JAR, SLJ and DASH (40-yards). Costill et al. (1968) described this factor that includes speed variables (e.g., DASH) as well as jumping ability variables as "explosive leg strength and power". These findings imply that a speed variable may share the same underlying construct with JAR and SLJ. On the contrary, studies by Fleishman (1964), Metheney (1938), and Phillips (1949) suggested that speed and explosive power are two distinctive factors. Overall, there is a lack of evidence that confirms whether speed and explosive power are two distinctive constructs or aspects of the same underlying construct.

Only a few studies included the measure(s) of agility (e.g., ASR) in a factor analysis. In studies in which an agility test is included, it often showed a high degree of relationship with gross body coordination, explosive power or running ability (Fleishman, 1964; Larson, 1941; Phillips, 1949; Ponthieux & Barker, 1963). Most notably, Phillips (1949) employed three agility tests in his factor analysis and concluded that there is no common factor to the three agility tests other than speed. These findings may be due to the fact that a typical agility test (e.g., ASR) involves running ability and speed, and/or coordination.

Unlike the literature on other factors, most studies are in agreement that there is a distinctive construct of flexibility (Fleishman, 1964; Harris, 1969; Hilsendager, Karnes & Spiritoso, 1969; Marsh, 1993). In these studies, a distinctive flexibility factor was extracted in factor analyses and/or flexibility

showed relatively low correlations with other performance factors. The SAR test has been most frequently involved in these studies. Other distinctive performance factors, such as cardiovascular endurance, balance and coordination, have been identified in many studies, but these are not the main interest of the present study.

Most of findings described above were based on samples of college students, but the degree of specificity is less clear for young children. Young children tend to show general motor ability, that is, a child who shows high level of performance in one type of physical task generally shows high level of performance in other types of task as well. However, Rarick's (1980) note that "it is in the early childhood and preschool years that there occurs a gradual transition from generality to gradually increasing specificity of motor functioning" (p. 179), implies that the specificity of physical performance may be achieved from age 6 or 7 years. Barry and Cureton (1961) employed children aged between 7 and 11 years in their study, and extracted factors that are specific to particular body parts and particular types of movement. Studies by Ismail and Cowell (1961) and Rarick and Dobbins (1975) used different performance variables but generally supported the specificity of children's physical performance. In addition, Marsh (1993) showed that the structure of physical fitness factors (i.e., the relationship between a factor and a test, and the relationship between factors) is invariant across age groups of 9, 12 and 15 years.

According to the literature, there seem to exist distinctive latent factors, such as strength and endurance, explosive power, speed, flexibility and agility. Further, these latent factors may be specific to particular muscle groups, such as the lower leg muscle group and upper arm muscle group, and/or to particular types of movement such as running and jumping. However, for young children, there is a possibility that there exists a general motor ability. More studies are required to determine if the specificity of latent physical performance depends on particular muscle groups or on particular types of movement, and if a general motor ability exist in early childhood.

STUDY 1-CHAPTER III. METHODOLOGY

The Data

Initially, two longitudinal data sets were obtained. The first data set was obtained from Korea (Korean data), and the second data set was obtained from Michigan State, US (Michigan data). For both data sets, initial data screening was conducted to examine the reliability of the data and to identify extreme values.

Korean Data

The Korean data set was obtained from four high schools that are located in metropolitan Seoul, Korea. The measurements were taken as a part of the annual report of students' physical growth and fitness level. These annual measurements are mandatory in Korea for all students, aged from 11 to 18 (grade 5 to grade 12). The obtained data set includes four body size variables and six performance variables that were measured from 706 cohort male students, aged from 13 to 17 (grade 7 to grade 11). The variables are weight, height, sit-height, chest girth, 100-m dash, standing long jump, pull-ups, sit-ups, softball throw and 1000-m endurance run. The measurements were taken from 1993 to 1997, once a year with approximately the same time intervals between assessments.

A close examination of the data set revealed that this data set was highly unreliable. A few examples of unreliable cases of standing long jump are presented in Table 1.3.1. In Table 1.3.1, for example, the record of the subject 245 (the second row) showed a decrease of 98 cm between age 14 and 15 and an increase of 100 cm between age 15 and 16. Considering that this subject did not show a considerable change (other than normal growth) in his height (163, 167 and 169 cm at age 14, 15 and 16, respectively) and weight (50, 54 and 59 kg at age 14, 15 and 16, respectively) during this period, this record is not reasonable. This kind of unreliable record was found in all six performance variables and throughout the data set. This may be due to different test procedures/criteria caused by different administrators, and insincere participation of the subjects on the measurements. This has been a problem in the measurement of physical and fitness growth in Korea. Because the extent of the unreliability of this data set was difficult to be determined, further analyses on this data set were not conducted.

Michigan Data

The Michigan data set was provided by Haubenstricker and Seefelt from Michigan State University. The data were collected as a part of a large research project, "The Motor Performance Study", and the original data set included seven demographic variables and nine motor performance variables that were repeatedly measured on 585 male children. The initial measurement was started in 1968 with 30 subjects, and approximately 30 subjects were recruited and repeatedly measured twice every year until they were dropped out of the study. The most recent measurements were taken in 1997.

Table 1.3.1

Examples of unreliable measurement -- Standing long jump (cm)

Subject ID	Age 13	Age 14	Age 15	Age 16	Age 17
28	140	117	210	170	195
245	235	248	150	250	240
531	215	210	253	242	170
550	165	270	270	236	262
598	198	170	264	195	238
664	196	200	235	130	190
690	167	167	157	241	205
707	264	177	190	260	210

The age at the initial measurement varied from 3 years old to 15 years old, and most of the subjects were measured between age 8 and 13 years. On average, each subject was measured every six months for about five years. Age was categorized by month, and the size of a category is six months. For the purpose of this latent growth model (LGM) application, subjects were matched based on this 6-month age category and then the data were analyzed as if these subjects were a cohort group. Thus, for example, one subject's age was 8 in 1970 while another's age was 8 in 1990, but they were treated as if they were a cohort group. The differences between subjects in measurement year were accounted for by means of a covariate (or predictor) in all statistical models, thus the effect of measurement year was controlled for in the LGM analyses. The usage of this variable, measurement year, as a predictor variable is explained later in this chapter.

A LGM requires a relatively large sample size with no missing values. To satisfy these conditions, only parts of the whole data set were selected based on listwise deletion (no missing value in all used variables), sample size, and the number of repeated measures (i.e., > 4). As a result of this procedure, two different data sets were obtained from the same pool of subjects (i.e., $N = 585$) with sample sizes of 218 (data set 1) and 212 (data set 2). These data sets have five time points, and the intervals between two adjacent time points were approximately one year. The subjects' ages in data set 1 and data set 2 were different. The subjects' ages at initial measurement were 8 years in data set 1, and 8.5 years in data set 2. Many of the subjects were included in both data set 1 and 2, but provided different performance records due to the differences in measurement time (age) between the two data sets. Thus, the two data sets were not completely mutually exclusive from each other; that is the same 168 subjects were included in both data sets, while the data set 1 had additional 50 subjects and the data set 2 had 44 additional subjects. The data set 1 was used as the main data for the analyses, and the data set 2 was used for a pseudo cross-validation.

Some of the demographic variables were excluded because there were too little between-person variations within a variable. For example, the variable "race" was not included in this study because more than 95% of the subjects were Caucasian. Some of the motor performance variables were also excluded because the measurement was stopped at the initial stage of the project. The resultant data sets include five demographic variables (used as predictor variables) and seven motor performance variables. These variables are not representative of all the important physical performance and predictor variables, since already existing data were used in the present study. The descriptions of the included variables in this study are presented in Table 1.3.2.

Both data sets, data set 1 and 2, were examined for extreme values because means and covariances that are used as data in a LGM analysis are highly sensitive to extreme values. In data set 1, eight subjects showed extreme values in ASR and ESR. These values were more than 4.00 standard deviation (SD) away from the mean of the variable, and more than 1.25 SD away from the adjacent values. These 8 subjects were excluded from the analyses. In data set 2, eight subjects showed extreme

Table 1.3.2

Descriptions of variables that were used in the study (Michigan data)

Variable name	Description
The number of pre-measurements	The number of measurements that were taken before the initial time point (age 8 and 8.5 in data set 1 and 2, respectively). This variable was used as a predictor variable.
Age	The subjects' age in months at the initial time point (age 8 and 8.5). Although the subjects were matched by age category, there still exist variations of age in months, and these variations may affect the level of the motor performance of young children. The maximum difference between any two subjects on this variable was 6 months within a data set. This variable was used as a predictor variable.
Grade	The subjects' grade at the initial time point (at age 8 and 8.5). The maximum difference between any two subjects was 2. This variable was used as a predictor variable.
Measurement season	The season that the measurement was taken. This variable has only two values, summer (coded as 0) and winter (coded as 1). This variable was used as a predictor variable.
Measurement year	The year when the subject's age was 8 and 8.5 in data set 1 and 2, respectively. The values ranged from 1968 to 1992. This variable was used as a predictor variable.
Flexed-arm-hang (FAH)	Measured in seconds. A larger score represents a better performance. This variable measures the muscular endurance of upper arm. A detailed test procedure and the characteristics of the test are given in Safrit and Wood (1995).
Jump-and-reach (JAR)	Measured in inches to the nearest half inch. A larger score represents a better performance. This variable measures dynamic leg power. A detailed test procedure and the characteristics of the test are given in Safrit and Wood (1995: Vertical jump test).
Sit-and-reach (SAR)	Measured in inches to the nearest half inch. A positive larger value represents a better flexibility of the hamstring and lower back. A detailed test procedure and the characteristics of the test are given in Safrit and Wood (1995).
Agility shuttle run (ASR)	Measured in seconds to the nearest one-tenth of a second. A smaller score represents a better performance. This variable measures agility and running ability. A detailed test procedure and the characteristics of the test are given in Safrit and Wood (1995).
Endurance shuttle run (ESR)	Measured in seconds to the nearest one-tenth of a second. A smaller score represents a better performance. Two laps of 300 feet each. This variable measures the muscular endurance of leg and running ability.
30-yard dash (DASH)	Measured in seconds to the nearest one-tenth of second. A smaller score represents a better performance. This variable measures power and running ability.
Standing long jump (SLJ)	Measured in inches to the nearest half inch. A larger score represents a better performance. This variable measures lower leg power. A detailed test procedure and the characteristics of the test are given in Safrit and Wood (1995).

values in ASR, ESR and DASH. These values were more than 3.90 SD away from the mean of the variable, and more than .66 SD away from the adjacent values. These eight subjects were also excluded from the analyses. The resulted sample sizes were 210 for the data set 1, and 204 for the data set 2. The full data records for five selected subjects are presented in Appendix A as an example.

Data Analyses

Data set 1 was used as the main data for all the analyses, and data set 2 was used for a cross-validation. Both data sets were analyzed by the following procedures.

Univariate LGM

Descriptive Statistics

Descriptive statistics for each variable at each time point were obtained using the SPSS windows program (SPSS Inc., 1997: Version 8.0). The descriptive statistics include mean, standard deviation, skewness and kurtosis. The Pearson product-moment correlation (PPMC) coefficients between time points within a variable were also calculated.

Identification of the Best Fitting Growth Curve (LGM)

To identify the best growth curve, four LGMs were fitted and compared for each of seven performance variables. These four LGMs were the “Linear”, “Quadratic”, “Cubic” and “Unspecified Curve” (Curve hereafter) models. The Linear model and the Quadratic model are shown in Figure 1.2.1 and Figure 1.2.2, respectively. The Cubic model has four factors, intercept, linear, quadratic and cubic factors, and describes a cubic change. The cubic factor has fixed factor loadings (coefficients) of 0, 1, 8, 27, and 64 (0^3 , 1^3 , 2^3 , 3^3 and 4^3). The Quartic model was not fitted because this model is under-identified unless some constraints are imposed in the model. The Curve model is a two-factor model with both an intercept and a curve factor (Figure 1.3.1). In a Curve model, the factor loadings for the last three time points of the curve factor were freely estimated (denoted by *). Thus, in this model, the different rates of change at each time interval are estimated. This model is similar to a Quartic model (with five time points) in that it estimates the changes of each interval, but is different from a Quartic model in that the only one change factor explains a between-subject variation in change, while in a Quartic model four change factors (linear, quadratic, cubic and quartic) explain a between-subjects variation in change. The factor loadings of factors for each model are presented in Table 1.3.3. The estimation of parameters and the model evaluation procedure are presented later in this chapter. Fit indices and parameter estimates were compared for the four models, and once the best model was selected, the validity of equality of error variances over time was examined. The model with equality of error variances over time provides a more parsimonious model, and thus would be the final model of choice should this equality assumption hold.

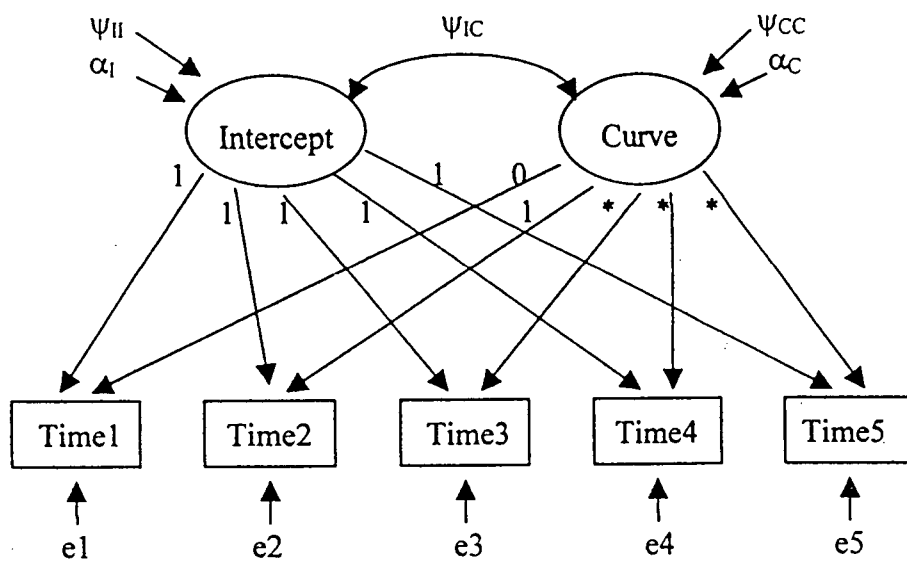


Figure 1.3.1. Unspecified Curve LGM

Table 1.3.3

Factor loadings of intercept and change factors for four LGMs

<u>Model</u>	<u>Factors</u>	<u>Factor loadings</u>				
		Time1	Time2	Time3	Time4	Time5
Linear	Intercept	1	1	1	1	1
	Linear	0	1	2	3	4
Quadratic	Intercept	1	1	1	1	1
	Linear	0	1	2	3	4
	Quadratic	0	1	4	9	16
Cubic	Intercept	1	1	1	1	1
	Linear	0	1	2	3	4
	Quadratic	0	1	4	9	16
	Cubic	0	1	8	27	64
Unspecified	Intercept	1	1	1	1	1
Curve	Curve	0	1	*	*	*

Note. * = free estimates.

Predictor Effects

Once the best growth model was selected, five predictors were sequentially included in the selected model to examine the effect of these predictors on the initial status and the change (the word "effect" is used as a statistical term implying a predictability, and does not necessarily imply a causal effect). The selection of the five predictors was largely dependent on the availability of the variables from the data. Other possibly important predictor variables such as activity level, injury, participation in a specific activity program, height, weight etc. were not available. The five predictors were the number of pre-measurements, age, grade, measurement season and measurement year. These predictor variables were expected to have effects on physical performance to some extent. The number of pre-measurements is the number of measurements taken before age 8. The children who were measured more frequently might show better performance than the children who were measured less frequently because of the practice effect. The age variable is the age in months at initial time point (at age 8 and 8.5 for data set 1 and 2, respectively). Children showed differences up to seven months in their age at the initial time point. The children who were older (within the same age group) might show better performance than the children who were younger. As well, the children who were within the same age group but were in a higher grade might show a better performance. The measurement season may affect the children's performance because the level of activity is generally lower during winter, and the temperature may also affect the physical conditions of children in performing a test. The measurement year might have an effect on the performance because the living environments that are related to the level of physical activity had been considerably changed from the 1970s to the 1990s.

The effects of these five predictor variables on the initial status and the development of physical performances were examined hierarchically. The order of variable input was as follows; (a) the number of pre-measurements, (b) age, (c) grade, (d) measurement season, and (e) measurement year. These predictor variables were included additively in the model one by one. If the effect of the predictor variable was not significant at α level of .05, this effect was fixed at zero in the examination of the next predictor effect. If a predictor variable was significant, the effect of the next predictor variable was examined after controlling for the previously examined predictor variable. For example, the age effect was examined after controlling for the test practice effect (the number of pre-measurement), the grade effect was examined after controlling for the test practice and age effects, and so on. The order of variable input was determined based on the rationale that the effect of a certain predictor variable should be examined after controlling for other variables. For example, the grade effect should be examined after controlling for the age effect because older (by a few months) children tend to be in a higher grade and show a higher level of physical performance than younger children. In this case, the better performance of older children is more likely to be an age effect rather than a grade effect. If the previously significant effect of a predictor became nonsignificant at an α level of .10 after including the

next predictor in the model, the previous variable was dropped out of the model. One exception was the effects of age and grade. If the effect of age became nonsignificant after grade was included in the model, the effect of grade was dropped out of the model because a significant grade effect is meaningful only after controlling for age effect.

Pseudo Cross-validation

The analyses for the cross-validation were conducted with data set 2 following the same procedure described above. The best growth model that was selected in data set 1 for each variable was fitted to data set 2, and compared with the results of data set 1. As well, the best growth models for data set 2 were also identified. The examination of the predictor effects was conducted with the same sequence, and the results were compared with those of data set 1.

Multivariate LGM

Two types of LGM, a curve-of-factors model and a factor-of-curves model, could be applied for the multivariate longitudinal data. In the present study, however, only the curve-of-factors model (Figure 1.2.5) was employed for the hypothesized factors. The factor-of-curves model (Figure 1.2.4) was not used because it was not appropriate for the physical performance data for two reasons. First, the observed variables showed different types of change over time. This resulted in a different number of the first-order change factors between variables. In this case, how to model the second-order change factors is not clear. Second, there is a lack of theoretical background that supports the common cause for the development of performance variables (i.e., endurance, power, flexibility, agility). Thus the factor-of-curves model that explains the change of several variables by the common (second-order) change factor(s) is not appropriate for the physical performance variables that were used in the present study.

For the applications of the curve-of-factors model, three sets of variables were hypothesized to form factors at each time point. These three sets of variables (factors) were (a) Run: ASR, ESR, DASH, (b) Power: JAR, SLJ, DASH, and (c) Motor Ability: FAH, SLJ, SAR, DASH, ESR. The "Run" factor represents a running ability and includes the same type of movement (e.g., Jackson, 1971). The "Power" factor includes performance variables that measures the explosive power of lower leg and involves the same muscle group (e.g., Cousins, 1955; Liba, 1967; Start, Gray, Glencross & Walsh, 1966). Although the DASH variable is a measure of a speed, many studies reported that the DASH might be included as a measure of explosive power (e.g., Miller et al., 1968; Start et. al., 1966). The "Motor Ability" factor represents a general motor ability, thus includes the variables that measure the muscular endurance of the upper arms, explosive power, flexibility, speed and the muscular endurance of the legs. Although a general motor ability factor has not been documented by the literature, most studies supporting the specificity of motor performance were based on a college student sample. The existence of a general motor ability can be demonstrated with young children, as noted by Rarick (1980). For each hypothesized factor, the following analyses were sequentially conducted.

Descriptive Statistics

In addition to the descriptive statistics that were obtained in the univariate analyses, the PPMC coefficients between different variables within a time point and between different variables between different time points were calculated and examined for each hypothesized factors.

Verification of the Factor Structure

Before LGMs are fitted to the hypothesized factors, the factor structure at each time point should be verified. Verification of the factor structure involved the sequential examination of several models. This procedure is summarized in the Figure 1.3.2. First, a 5-factor confirmatory factor analysis (CFA) model with one factor at each time point was fitted to each of the three hypothesized physical performance factors, Run, Power and Motor Ability (Figure 1.3.3). In this model, the factor loadings of the first observed variable were fixed at 1 to provide a scale to the factor at each time point, and the covariance of the factors between the time points were freely estimated. However, the correlations of errors between time points were fixed at zero. When the absolute goodness-of-fit of this model was not satisfied (i.e., when this model was rejected), the correlated errors of the same variable between the time points were included in the next model (correlated-errors model: Figure 1.3.4). This is a common practice in a multivariate longitudinal factor model (Marsh & Hau, 1996; Schutz, 1998). When one of these two models, the 5-factor CFA models without correlated errors and with correlated errors, fits the data well, the equality of factor loadings over time was examined in the next model. This model examines if the same construct was measured over time. When both models were statistically rejected, further analyses were not conducted because the existence of a factor (latent trait) at each time point was not verified. When the model with the equality of factor loadings over all five time points fit the data well, the LGMs were fitted to the verified factor. When the equality of factor loadings model was rejected, the equality of factor loadings between each time points was examined to determine at which interval the factor structure changed.

Identification of the Best Fitting Growth Curve

If a 5-factor model or correlated errors model with equality of factor loadings was acceptable, four multivariate LGMs (curve-of-factors models) were fitted to each hypothesized factor. These models were "Linear", "Quadratic", "Cubic" and "Curve" models as in the univariate LGM analyses. These four models are merely the extensions of the univariate LGMs. The linear "curve-of-factors model" is shown in Figure 1.2.5.

Predictor Effects

Once the best growth model was determined, predictors were included in the model to examine the effects of the predictors. The procedure and the sequence of the variable input were the same as in the univariate analyses.

Pseudo Cross-validation

Data set 2 was used for a cross-validation of all multivariate LGMs. All above-mentioned

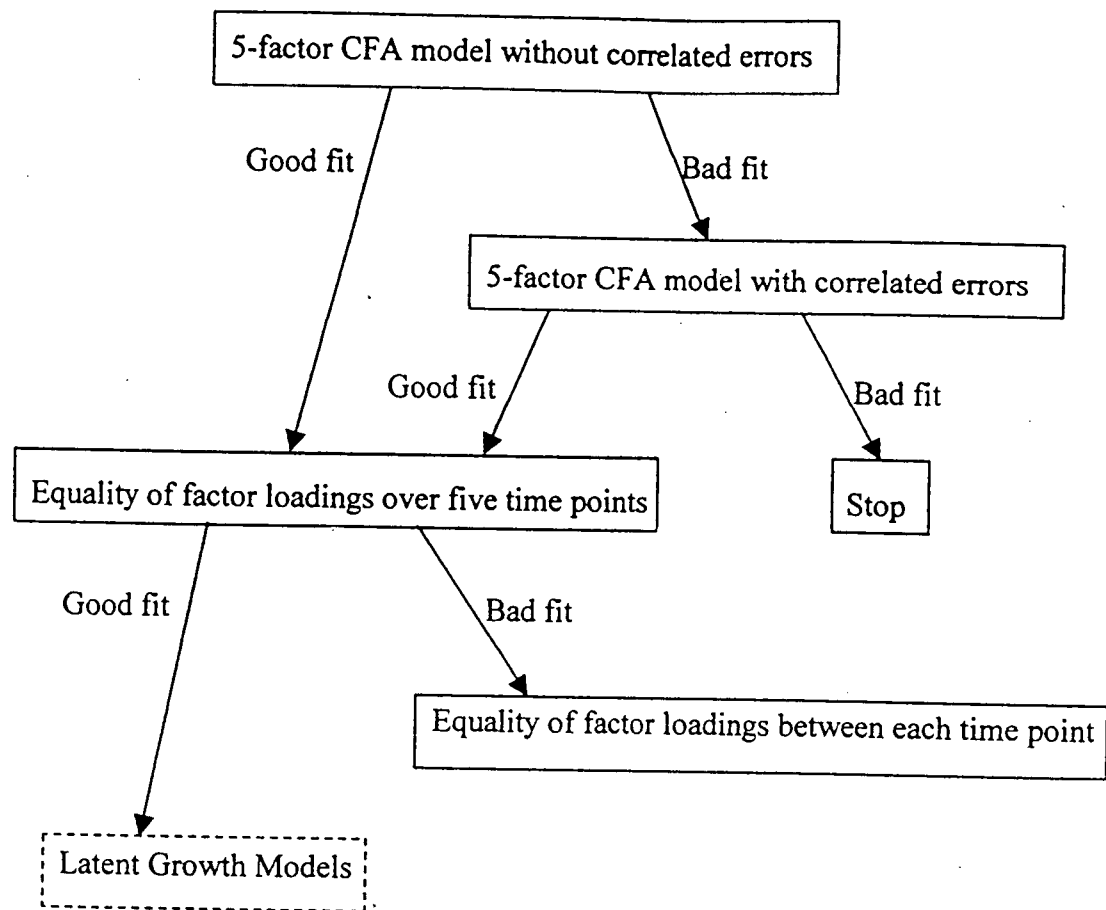


Figure 1.3.2. The procedure for the verification of factor structure

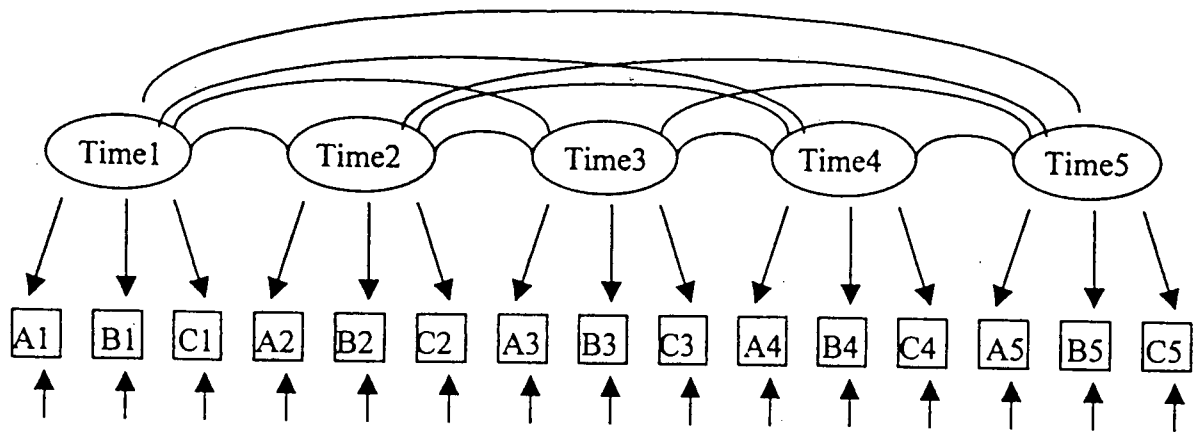


Figure 1.3.3. 5-factor CFA model with a factor at each time point (parameter estimates are omitted)

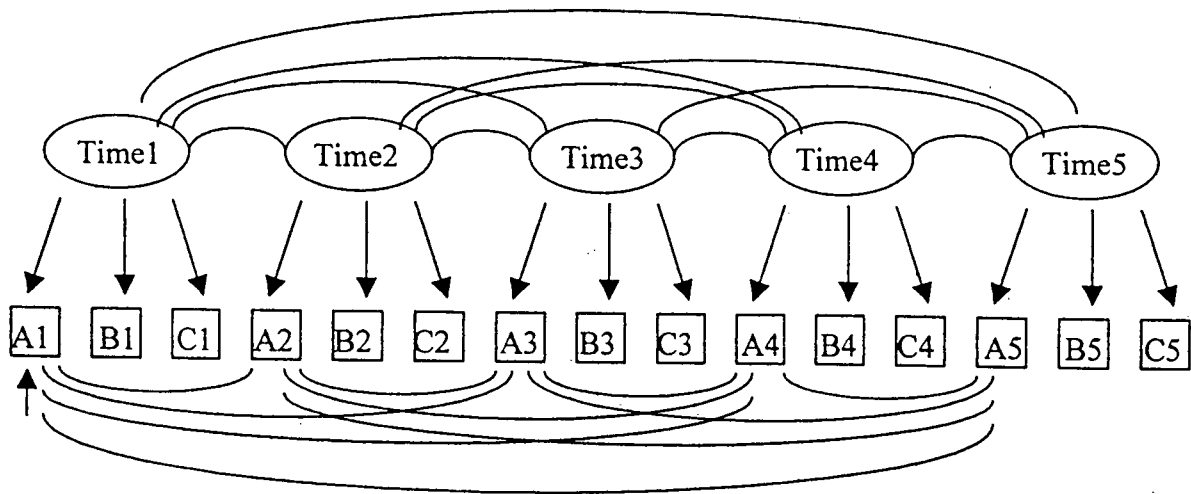


Figure 1.3.4. 5-factor CFA model with correlated errors (only the correlations between the error of variable A are shown)

procedures were equally applied and the results were compared with those of the data set 1.

Estimation of LGMs

Maximum Likelihood (ML) procedures were used for the estimation of parameters for all growth models. The LISREL (Joreskog & Sorbom, 1999: Version 8.3) program was used in the estimation of parameters and the calculations of goodness-of-fit indices for most LGMs. In addition, the MPLUS (Muthen & Muthen, 1998: Version 1.04) program was used for the curve-of-factors models with predictors, because the LISREL program (and LISREL model) does not allow one to estimate these models.

Model Evaluation

The evaluation of fitted models was conducted using several pieces of information from the analyses results. First, the results were examined for unacceptable parameter estimates (e.g., negative variance). If all estimated parameters were within acceptable ranges, several goodness-of-fit indices were used to evaluate a model. The absolute fit of a model was evaluated by the χ^2 statistic with the associated degrees of freedom (df), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR) and Non-normed Fit Index (NNFI). For the comparisons of two nested models, the χ^2 difference test was used. For the comparisons of non-nested models, RMSEA and the Expected Cross Validation Index (ECVI) were used as well as other fit indices. In general, criteria for evaluating a model using absolute fit indices (i.e., RMSEA, SRMR and NNFI) were based on Hu and Bentler's (1999) suggestions. They suggested that one accepts a model when the RMSEA is smaller than .06, the SRMR is smaller than .08, or the NNFI is larger than .95. The ECIV is meaningful only when it is compared to the ECVI of another model. A smaller absolute value of ECIV indicates a better fitting model to the data. More weight was given to the χ^2 statistic than other fit indices in the evaluation of a relatively small model (i.e., univariate LGM) than in the evaluation of a relatively large model (i.e., all multivariate models). In addition, if a model was accepted, residuals of the fitted covariance matrix and mean vector were examined for any extreme values.

STUDY 1-CHAPTER IV. RESULTS

Univariate Latent Growth Models for Motor Performances

Univariate Latent Growth Model (LGM) analyses results are presented in the following sections. The analyses results for only flexed-arm-hang, among seven physical performance variables, are presented in detail, because the results for flexed-arm-hang showed most of major aspects of LGM. The analyses results for six other variables are presented briefly.

Flexed-Arm Hang (FAH)

Descriptive Statistics

Table 1.4.1 shows the correlation coefficients between time points and other descriptive statistics for the FAH measure at each time point (histograms for FAH scores at each time point are presented in Appendix C, Figure C.1). As expected, the data were positively skewed, but the degrees of skewness were moderate to small (all less than 2.0: Cuttance, 1987; Muthen & Kaplan, 1985). Kurtosis values at ages 8, 9 and 10 were relatively high. However, these levels of skewness and kurtosis are regarded as a low to medium level of departure from a normal distribution, and have a relatively low impact on the maximum likelihood estimation in a structural equation modelling analysis (Cuttance, 1987; Muthen & Kaplan, 1985). Thus maximum likelihood estimation methods were used in the estimation of all latent growth models.

The mean of the 210 children's FAH scores increased over the 5-year period, indicating an improvement in their upper arm muscular endurance, relative to their body weight. The largest rate of increase occurred between ages 8 and 9, and generally, the rate of the change decreased in subsequent years. Relatively large magnitudes of standard deviations are partially due to a few extreme scores. The correlation coefficients between adjacent time points were high, indicating year-to-year consistency of relative positions of children in their FAH scores. However, the correlation matrix approximated a simplex pattern, with the correlation coefficients becoming smaller as a coefficient gets further away from the main diagonal. This pattern implies that the children changed over time at different rates. That is, there was a considerable between-person variation in both the rate and the year of maximum development rate of their muscular strength.

Identification of the Best Fitting Growth Curve

Four growth models were fitted and compared to identify the best growth curve model. Table 1.4.2 shows the goodness-of-fit indices of these four fitted latent growth models. Although the non-normed fit indices were high, the χ^2 and root mean square error of approximation statistics indicated that the Linear and the Unspecified Curve (Curve in Table 1.4.2) models should be rejected. The Quadratic model (model 2), however, showed a very good fit to the data with six degrees of freedom. A test for the

Table 1.4.1

Correlation coefficients and descriptive statistics for flexed-arm hang (FAH)

	Age 8	Age 9	Age 10	Age 11	Age 12
Age 9	.780				
Age 10	.696	.840			
Age 11	.626	.768	.785		
Age 12	.652	.724	.711	.797	
Mean (sec.)	16.64	19.29	21.35	22.45	23.91
SD	13.40	14.61	15.97	15.22	15.73
Skewness	1.87	1.51	1.63	1.14	1.12
Kurtosis	4.08	2.35	3.20	1.27	1.64
Range	0 - 72	1 - 77	2 - 92	1 - 76	1 - 88

Table 1.4.2

Fit indices for latent growth models for flexed-arm hang (FAH)

Model	$\chi^2(df)$	p-value	RMSEA	ECVI	SRMR	NNFI
1. Linear	46.91 (10)	< .001	.138	.335	.084	.96
2. Quadratic	8.96 (6)	.176	.048	.177	.039	1.00
3. Quadratic, equal error variance	19.43 (10)	.035	.069	.192	.036	1.00
4. Curve	38.43 (7)	< .001	.152	.319	.064	.95

Note. Curve = Unspecified Curve model, df = degrees of freedom, RMSEA = root mean square error of approximation, ECVI = expected cross-validation index, SRMR = standardized root mean square residual, NNFI = non-normed fit index.

equality of error variances across time points (model 3) revealed that the error variances could not be considered equal at all five time points (χ^2 difference = 10.47, $df = 4$, $p = .033$).

The ECVI statistic was also smallest for the Quadratic model with unequal error variances. Thus, the Quadratic model with different magnitudes of error variances (model 2) is most favorable in explaining individual changes in FAH test scores over a 5-year period. The Cubic model was not fitted because given the difference in the degrees of freedom between the Quadratic and Cubic models, the Cubic model could not be significantly better than the Quadratic model (requires a χ^2 difference of 11.07 or larger with $df = 5$ at $p < .05$). Individual children's FAH developed in a quadratic fashion over time.

Parameter estimates for the Quadratic model are presented in Table 1.4.3. The mean of the intercept factor was very close to the actual mean of FAH (16.64) at age 8, while the variance of the intercept factor was considerably smaller than the actual variance (179.56). The mean and the variance of the intercept factor represent the true (error free) initial status and variation of FAH performance at age of 8 that were explained by the Quadratic growth model.

The mean of the linear factor was 2.75 ($p < .001$), indicating an average linear increase of 2.75 seconds between ages 8 and 9. The dotted line in Figure 1.4.1 shows the trajectory of mean scores if this linear increase had continued for the rest of the time period (i.e., between ages 9 and 12). However, the actual average improvement was smaller due to the negative mean of the quadratic factor. The score of the linear factor represents the linear component of the individual change over time. The significant variance of the linear factor implies that there were differences in this change rate among children in the population. The magnitude of this variance ($SD = 7.04$) suggests that some of these children may have actually declined in their FAH scores over the five testing periods. The raw data showed that 20% of the children actually declined in their FAH score. However, this does not necessarily mean that their muscular endurance declined over time because FAH measures relative strength and endurance (relative to their body weight).

The mean of the quadratic factor was $-.241$ ($p = .118$) meaning, on the average, the rate of improvement decelerated over 5 years. However, this nonsignificant quadratic factor mean but significant variance ($p < .001$) indicate that some of the children decelerated and some accelerated in their development, but averaging those decelerations and accelerations produced a mean value close to zero. As well, it is possible that some children showed zero scores for the quadratic factor, and changed linearly over time. The significant variance of the quadratic factor resulted in the much better model fit obtained for the Quadratic model than that for the Linear model. This means that the inter-individual variation of the development in FAH among children is not adequately explained by the linear factor only. The curved line in Figure 1.4.1 shows the change described by the model. The difference between the dotted straight line and the curved line become larger as age increases, and this difference reflects the quadratic component of FAH score change over time.

Table 1.4.3

Estimated parameters (standard errors) of the Quadratic model for flexed-arm hang (FAH)

	Intercept factor	Linear factor	Quadratic factor	Error Variance	
<u>Mean</u>	16.69 (.930) $p < .001$	2.75 (.655) $p < .001$	-.241 (.154) $p = .118$	<u>Age 8</u>	25.92 (11.20) $p = .021$
<u>Variance</u>	157.36 (19.93) $p < .001$	49.57 (11.81) $p < .001$	2.72 (.599) $p < .001$	<u>Age 9</u>	28.70 (4.84) $p < .001$
<u>Covariance</u>				<u>Age 10</u>	44.09 (6.46) $p < .001$
Linear factor	-5.57 (12.47) $p = .655$			<u>Age 11</u>	47.55 (6.58) $p < .001$
Quadratic factor	.158 (2.648) $p = .952$	-10.85 (2.53) $p < .001$		<u>Age 12</u>	24.26 (13.39) $p = .070$

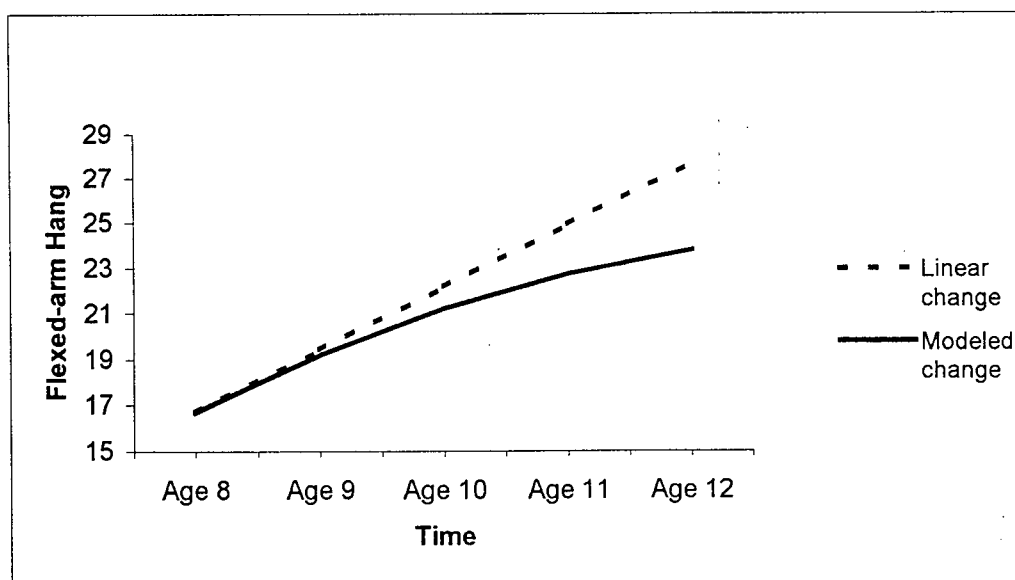


Figure 1.4.1. Linear and quadratic components of change in FAH

The covariances between the intercept and the linear and quadratic factors were not significant ($p = .655$ and $.952$, respectively). In standardized units these values were $-.063$ and $.008$, respectively. These small correlations were not expected because, in general, initial status and change show a low to medium level of negative correlation (Schutz, 1989). The covariance between the linear and quadratic factor was significant, the standardized value was $-.94$, indicating a very high negative correlation. This implies that the higher the linear development rate (the larger the change between ages 8 and 9), the faster the deceleration in development.

Error variances at the first four time points were significantly different from zero, while the error variance at the last time point was not. These error variances are the variances that were not explained by the quadratic growth model. The error variances at ages 10 and 11 were larger than those at other time points. Although four of five error variances were significantly different from zero, these were relatively small compared to the estimated total variances at each time point (i.e., 183.28, 205.82, 248.61, 254.79, and 242.55 at time 1, 2, 3, 4 and 5, respectively). This resulted in the relatively high estimated reliabilities for each time point of $.86$, $.86$, $.82$, $.81$ and $.90$, respectively.

Predictor Effects

Five predictors were sequentially included in the Quadratic model (model 2 in Table 1.4.2) in the next series of analyses (see Figure 1.2.3). The correlations between predictor variables and descriptive statistics are presented in Appendix C, Table C.1. The sequence of predictor models and goodness-of-fit indices of each model are presented in Table 1.4.4. All five models fit the data very well; the χ^2 statistics for all models were not significant, RMSEAs and SRMRs were low, and NNFIIs are all close to 1.00, indicating a very good fit for all models.

The parameter estimates of the predictor variables' effects on the intercept, linear and quadratic factors for each model are presented in Table 1.4.5. The test practice effects (the number of pre-measurements) on the intercept, linear and quadratic factors were not significant ($p = .523$, $.719$ and $.604$, respectively). Thus, this predictor variable was excluded in subsequent analyses.

The next predictor, age, had a positive effect on the intercept ($p = .022$) but no significant effects on the linear and quadratic factors ($p = .936$ and $.617$, respectively). This means that although there was only a small degree of variation in the age variable (maximum difference among children was seven months), older children had higher levels of muscular endurance. However, the standardized coefficient was very small ($.167$), and the variance of age at the initial time point explained only about 3 % of the variance of the initial status. As indicated by nonsignificant effects on both change factors, the rate of change was not influenced by age at the initial time point. Thus the age effects on the linear and quadratic factors were excluded in the following analyses while the age effect on the intercept factor was included.

After controlling for age (on the intercept only), none of grade, measurement season or

Table 1.4.4

Fit indices of the Quadratic models with predictors for flexed-arm hang (FAH)

Predictors	$\chi^2(df)$	p-value	RMSEA	ECVI	SRMR	NNFI
5. The number of pre-measurements	9.52 (8)	.300	.029	.227	.034	1.00
6. Age	10.28 (8)	.246	.036	.230	.034	1.00
7. Grade	14.77 (12)	.255	.035	.292	.035	1.00
8. Measurement season	14.18 (12)	.289	.032	.290	.035	1.00
9. Measurement year	13.90 (12)	.307	.030	.288	.035	1.00

Note. df = degrees of freedom, RMSEA = root mean square error of approximation, ECVI = expected cross-validation index, SRMR = standardized root mean square residual, NNFI = non-normed fit index.

Table 1.4.5

Parameter estimates of predictor variables' effects on growth factors

Predictors	Intercept	Linear	Quadratic
5. The number of pre-measurements	.297 (.465) p = .523	.118 (.328) p = .719	-.040 (.077) p = .604
6. Age	1.086 (.475) p = .022	-.027 (.339) p = .936	-.040 (.080) p = .617
7. Grade	.688 (1.904) p = .718	-.204 (1.308) p = .876	.074 (.308) p = .810
8. Measurement season	-1.863 (1.832) p = .309	.646 (1.307) p = .621	-.213 (.308) p = .489
9. Measurement year	-.134 (.167) p = .935	.115 (.119) p = .334	.024 (.028) p = .391

measurement year was significantly related to any of the three model parameters, intercept, linear or quadratic. The nonsignificant grade effect was due partly to a positive correlation between age and grade (.28). In general, measurement season (winter or summer) did not affect the performance on the muscular endurance test, and children's muscular endurance did not show differences across years, 1970s through 1990s.

In summary, the children's individual development in their relative upper arm muscular endurance and strength over a 5-year period was explained well by a Quadratic model with unequal error variances. On the average, their FAH scores increased, but the rate of the increase declined over the 5-year period. There were considerable inter-individual variations among children in the initial status (the FAH score at age 8), the linear development rate and the deceleration (acceleration) of the development. Five predictors were included in the Quadratic model to predict these variations, but only age had a significant positive effect on the initial status. The magnitude of the age effect on the initial status was very small, with only 3 % of the variance of the FAH initial status being explained by the variance in age. This was not unexpected, given the very small variance in age (a maximum difference between children of seven months).

Six Other Physical Performance Variables

Following is a summary of the results of the analyses for the remaining six physical performance variables; Jump-and-Reach (JAR), Sit-and-Reach (SAR), Agility Shuttle Run (ASR), 300-foot Endurance Shuttle Run (ESR), 30-yard Dash (DASH) and Standing-Long-Jump (SLJ). Detailed descriptive statistics and parameter estimates for each variable are presented in Appendix C, Table C.2 to Table C.13.

Descriptive Statistics

Means and standard deviations of all six physical performance tests at five time points are presented in Table 1.4.6. Children's physical performances improved over a 5-year period (from 8 to 12 years old) except for the SAR. The mean scores for JAR and SLJ, measured in inches, increased over time, and the mean scores for ASR, ESR and DASH, measured in seconds, decreased over time. However, children's flexibility measured by SAR (in inches) decreased over time. Children showed relatively large changes (in percent) on JAR (42.3%) and SLJ (22.9%) between ages 8 and 12. The other four variables showed an average change of 13.2% over a 5-year period. In general, the rates of improvements measured by mean scores were largest between ages 8 and 9, and the rate decreased in subsequent years, except for JAR and SAR. The JAR and SAR showed the largest change between ages 9 and 10. The standard deviations of SAR were relatively large compared to the magnitude of the mean scores, indicating large between-person variability in hamstring flexibility.

All six variables showed skewness and kurtosis values close to zero, indicating a small departure from a normal distribution. The largest absolute skewness value among all six variables across

Table 1.4.6

Means and standard deviations for six physical performance variables

Variable	Age 8	Age 9	Age 10	Age 11	Age 12	% improvement over a 5-year period
JAR	9.42 ± 1.78	10.34 ± 1.81	11.58 ± 1.89	12.33 ± 1.85	13.40 ± 2.07	42.3 %
SAR	7.89 ± 2.29	7.70 ± 2.20	7.37 ± 2.22	7.12 ± 2.34	6.89 ± 2.51	- 12.7 %
ASR	12.46 ± 1.05	11.92 ± .90	11.39 ± .82	11.06 ± .74	10.76 ± .75	13.6 %
ESR	43.93 ± 3.64	42.00 ± 3.04	40.55 ± 3.03	39.46 ± 2.68	38.32 ± 2.67	12.8 %
DASH	5.21 ± .44	4.93 ± .40	4.75 ± .36	4.65 ± .35	4.50 ± .33	13.6 %
SLJ	53.38 ± 7.56	57.46 ± 7.61	61.40 ± 7.21	64.48 ± 6.49	67.76 ± 6.91	26.9 %

Note. JAR = jump-and-reach (inches), SAR = sit-and-reach (inches), ASR = agility shuttle run (seconds),

ESR = 300-foot endurance shuttle run (seconds), SLJ = standing long jump (inches), DASH = 30-yard dash (seconds).

all time points was .907 on ASR at age 10. The largest absolute kurtosis value was 1.418 on ESR at age 10. Thus, maximum likelihood estimation methods were used in the estimation of all latent growth models.

The correlation coefficients between time points for JAR ($r = .49$ to $.69$), ASR ($r = .46$ to $.64$), ESR ($r = .52$ to $.67$) and DASH ($r = .55$ to $.69$) indicate that there were moderately high levels of year-to-year consistency of relative positions among children in these measures. The SAR ($r = .74$ to $.86$) and SLJ ($r = .66$ to $.83$) showed high levels of year-to-year-consistency. Similar to the Flexed-arm Hang (FAH) test, however, the correlation coefficients between time points approximated simplex patterns for all six variables. This implies that there were considerable between-person variations in the development rates for each of the six physical performances.

Identification of the Best Fitting Growth Curve

Best fitting growth models and goodness-of-fit indices for the six physical performance variables are presented in Table 1.4.7.

Jump-and-reach. A Linear model with equal error variances at each time point described the individual changes of JAR very well. The χ^2 statistic was not significant ($\chi^2(14) = 17.31$, $p = .240$), and all other fit indices indicated that the Linear model with equal error variances fit the data very well. These results imply that individuals improved linearly in jumping ability between ages 8 and 12. The true mean score at age 8 was 9.43 inches ($p < .001$), and on the average, the children improved .99 inches ($p < .001$) per year. The variances of the intercept (1.94, $p < .001$) and linear (.08, $p < .001$) factors were significantly different from zero, indicating that there was a considerable inter-individual variation in the rate of improvement as well as in the initial status. This implies that children improved in jumping ability at different rates. The covariance between the intercept and linear factors was not significant, as indicated by a standardized covariance (i.e., correlation r) of $-.06$ ($p = .656$). The initial status and the rate of change did not show a significant relationship, similar to the FAH results.

Sit-and-reach. The analyses results for SAR were very similar to those of JAR except that children showed decreasing scores in SAR over time. A Linear model with unequal error variances among time points described the individual changes of SAR very well. The χ^2 statistic was not significant ($\chi^2(10) = 10.54$, $p = .395$), and all other fit indices indicated a good model fit. Individuals linearly declined in lower back and hamstring flexibility between ages 8 and 12. The true mean score at age 8 was 7.91 inches ($p < .001$), and on the average, the children declined .26 inches ($p < .001$) per year. The variances of the intercept (4.127, $p < .001$) and linear (.05, $p = .006$) factors were significantly different from zero, but the covariance between these two factors was not significant ($r = -.03$, $p = .816$). Thus, as in JAR, there was a considerable variation in the rate of decrease in flexibility, but the relationship between the initial status and the rate of decrease was not significant.

Agility shuttle run. For ASR, both an Unspecified Curve (Curve hereafter) model ($\chi^2(7) = 7.78$,

Table 1.4.7

Best fitting growth curve models and goodness-of-fit indices for the six physical performance variables

Variable	Identified model	Error variances	χ^2 (df)	p-value	RMSEA	ECVI	SRMR	NNFI
JAR	Linear	Equal	17.31 (14)	.240	.030	.137	.038	1.00
SAR	Linear	Unequal	10.54 (10)	.395	.019	.147	.016	1.00
ASR	Curve	Unequal	7.78 (7)	.353	.018	.160	.030	1.00
ESR	Curve	Unequal	3.18 (7)	.868	< .001	.134	.023	1.01
DASH	Curve	Unequal	18.73 (7)	.009	.085	.208	.037	.97
SLJ	Cubic	Equal	4.04 (5)	.544	< .001	.144	.020	1.00

Note. Curve = Unspecified Curve model, JAR = jump-and-reach (inches), SAR = sit-and-reach (inches), ASR = agility shuttle run (seconds), ESR = 300-feet endurance shuttle run (seconds), SLJ = standing long jump (inches), DASH = 30-yard dash (seconds).

$p = .353$) and a Quadratic model ($\chi^2(6) = 4.75$, $p = .577$) fit the data very well, but the Quadratic model produced an improper solution (the correlation between the linear and quadratic factor was high ($-.92$), but was not significant ($p = .066$)). Thus, the Curve model with unequal error variances over time was selected as the best fitting model for ASR. The χ^2 statistic was not significant ($\chi^2(7) = 7.78$, $p = .353$), and all other indices indicated a good model fit. Thus, the individual changes in ASR scores and the variation in change was adequately explained by the curve factor. The true mean of the ASR at age 8 was 12.45 ($p < .001$) seconds, and on the average, the children improved in agility (decreasing means imply an improvement because this test was measured in time) over a 5-year period. The parameter estimates of the Curve model (more specifically, the factor loadings of the curve factor) provided the average improvement at each time interval. The improvement in agility was largest between ages 8 and 9 (.54) and the rate of the improvement decreased in subsequent years (.52, .34, and .29 seconds between ages 9 and 10, 10 and 11, and 11 and 12, respectively). The significant variances of the intercept (.637, $p < .001$) and curve (.03, $p = .016$) factors implied that children showed inter-individual differences in the rate of improvement as well as in the initial status. The negative correlation ($r = -.67$, $p = .002$) between the intercept and curve factors indicated that the children with higher performance levels at age 8 showed smaller rates of improvement.

Endurance shuttle run. The analyses results for ESR were very similar to those of ASR. A Curve model with unequal error variances described the individual changes of ESR very well. The χ^2 statistic was not significant ($\chi^2(7) = 3.18$, $p = .868$), and all other fit indices indicated a good model fit. The true mean at age 8 was 43.93 seconds ($p < .001$), and in general, the children's performance improved over time. As in ASR, the improvement in ESR was largest between ages 8 and 9 (1.92), and the rate of the improvement decreased in subsequent years (1.47, 1.09 and 1.12 seconds between ages 9 and 10, 10 and 11, and 11 and 12, respectively). The significant variances of the intercept (7.65, $p < .001$) and curve (.36 $p = .009$) factors implied that the children showed inter-individual variations in the initial status and the rate of improvement in endurance. The negative correlation ($r = -.62$, $p = .002$) between the intercept and curve factors indicate that children with higher performance levels at age 8 showed smaller rates of improvement.

30-yard dash. None of the four growth curve models, the Linear, Quadratic, Curve or Cubic model, adequately described the individual changes of DASH. The Linear model ($\chi^2(10) = 55.45$, $p < .001$; RMSEA = .144) and the Quadratic model ($\chi^2(6) = 18.199$, $p = .006$; RMSEA = .095; ECVI = .217) were rejected. Although the Cubic model showed a very good model fit ($\chi^2(1) = 1.54$, $p = .215$; RMSEA = .050), this model produced an improper solution (the standardized variances of both the slope and the quadratic factors were greater than 1.00). The Curve model with unequal error variances among time points was selected as the best fitting model, although the model fit was not satisfactory. The model fit was slightly better than that of the Quadratic model, and produced a proper solution. The

χ^2 statistic was significant ($\chi^2(7) = 18.73$, $p = .009$) and the RMSEA (.085) was greater than .06, but the SRMR (.037) and NNFI (.97) indicated a good model fit. Thus, the subsequent analyses for the predictors' effect that are presented in the next section were conducted based on this model. The true mean at age 8 was 5.21 seconds ($p < .001$), and on the average, children's performance in DASH improved over time. The average improvement was largest between ages 8 and 9 (.275), and the rate of the improvement decreased until the fourth time point (.188 and .097 seconds between ages 9 and 10, and 10 and 11, respectively). The rate of improvement became larger between ages 11 and 12 (.142). Although estimated mean scores implied a cubic change over time, the variation in the rate of change was explained moderately well by the only one change factor, the curve factor. The significant variances of the intercept and curve factors (.125, $p < .001$ and .005, $p = .038$, respectively) indicated that there was a considerable inter-individual variation in the initial status and in the rate of change. However, the inter-individual variation of change was relatively small (standard deviation of the curve factor = .07). The negative correlation ($-.66$, $p = .003$) between the intercept and curve factors indicated that the children with higher performance levels at age 8 showed lower rates of improvement.

Standing long jump. For SLJ, a Cubic model with equal error variances among time points described the individual changes very well. The χ^2 statistic was not significant ($\chi^2(5) = 4.04$, $p = .544$), and all other fit indices indicated a good model fit. The mean of the intercept (53.36, $p < .001$) and linear (4.52, $p < .001$) factors were significant, while the mean of the quadratic ($-.34$, $p = .385$) and cubic (.028, $p = .659$) factors were not. In general, children improved in SLJ score over time. The very small magnitude of the cubic factor mean resulted in the average improvement that is close to a quadratic change. The estimated true mean at age 8 was 53.36 inches and the improvement was highest between ages 8 and 9 (4.20), and the rate of the improvement decreased in subsequent years (3.68, 3.33 and 3.15 inches between ages 9 and 10, 10 and 11, and 11 and 12, respectively). The variances of the intercept (48.79, $p < .001$) and the linear (37.56, $p < .001$) factors were significant. The variances of the quadratic (12.44, $p = .001$) and the cubic (.30, $p = .002$) factors were also significant although the means of these two factors were not significant. This implies that there were considerable inter-individual variations in each component of the change, the linear, quadratic and cubic. The significant variances of the quadratic and cubic factors resulted in the very good model fit of the Cubic model. The correlation between the linear and quadratic factors was negative ($-.94$, $p < .001$), while the correlation between the linear and cubic factor was positive (.87, $p = .002$). The correlation between the quadratic and cubic factors was negative ($-.98$, $p = .002$). These pieces of information (i.e., large positive mean of the linear factor, negative mean of the quadratic factor, very small positive mean of the cubic factor, high negative correlations between the linear and quadratic factors, and between the quadratic and cubic factors and high positive correlation between the linear and the cubic factors) imply that the children who showed larger improvement at the first time interval showed larger decrease in the rate of the improvement in a

subsequent interval, and the rate of deceleration in the improvement decreased faster in subsequent years. It is possible that some showed decrease in the rate of the improvement at the beginning of the time period and then the rate of improvement accelerated in later years, or some accelerated in the rate of improvement at the beginning and then decelerated in later years. None of the change factors showed a significant correlation with the intercept factor.

Predictor Effects

The effects of the five predictors on the intercept and change factors for the six physical performance variables are summarized in Table 1.4.8. For simplicity, only the significant ($p < .05$) effects are presented, and the estimated values are reported in standardized units.

Jump-and-reach. There were positive test practice and age effects on the intercept, and a negative measurement year effect on the linear factor. This indicates that the children who were measured more frequently before age 8, and who were older (within the group) showed better performances at age 8. In addition, the children who were measured in the 1970s showed faster improvements than the children who were measured in 1990s. However, the magnitudes of these predictors' effects were small, each explained less than 12% of the variance of the intercept or linear factor. Although the grade effect on the intercept was significant, the age effect on the intercept became nonsignificant with the inclusion of a grade effect, due to the positive correlation between age and grade (.28). Thus, the grade was excluded in the subsequent analyses.

Sit-and-reach. There was a positive test practice effect on the intercept factor, and a negative test practice effect on the linear factor. This indicates that children who had been measured more frequently before age 8 showed higher levels of flexibility at the age of 8 but declined faster (or improved less) over a 5-year period. However, the magnitudes of the test practice effects on both factors were small, explaining only 4% and 10% of the variances of the intercept and the linear factors, respectively.

Agility shuttle run. There were negative test practice, age and measurement year effects on the intercept factor, indicating that the children who were measured more frequently before age 8 and were older, showed better performances at age 8. As well, the children who were measured in the 1990s showed better performances than children who were measured in the 1970s. The positive measurement season effect on the intercept factor indicates that the children who were measured during a winter season showed a lower level of performance. There was also a positive test practice effect on the curve factor, indicating that the children who were measured more frequently before age 8 showed lower rates of improvement in agility.

Endurance shuttle run. There were negative test practice, age, grade and measurement year effects on the intercept, and a positive test practice effect on the curve factor. Thus, the children who were measured more frequently before age 8, were older, and in a higher grade showed better endurance at age 8. As well, the children who were measured in the 1990s showed better performance at age 8 than

Table 1.4.8

Predictors' effects on growth factors for six physical performance variables

Variable	Factor	<u>Predictors</u>				
		Test practice	Age	Grade	Season	Year
JAR	Intercept	.339	.193	.118*		
	Linear					-.290
SAR	Intercept	.200				
	Linear	-.318				
ASR	Intercept	-.392	-.188	-.317*	.312	-.226
	Curve	.389				
ESR	Intercept	-.383	-.172	-.218		-.130
	Curve	.297				
DASH	Intercept	-.320	-.260		.215	-.194
	Curve	.438	.287			.428
SLJ	Intercept	.356	.144			.228
	Linear					
	Quadratic					
	Cubic					

Note. Only significant effects ($p < .05$) are shown in a standardized unit.

JAR = jump-and-reach (inches), SAR = sit-and-reach (inches), ASR = agility shuttle run (seconds), ESR = 300-foot endurance shuttle run (seconds), DASH = 30-yard dash (seconds), SLJ = standing long jump (inches), Test practice = the number of pre-measurements, Season = measurement season, Year = measurement year.

* = Grade variable was excluded in the subsequent analyses because the age effect became nonsignificant with the presence of this variable.

children who were measured in the 1970s. The children who were measured more frequently before age 8 showed slower improvement in endurance performance.

30-yard dash. There were negative test practice, age and measurement year effects and a positive measurement season effect on the intercept factor. The children who were measured more frequently before age 8, were older, and were measured in the summer season showed better performance in DASH at age 8. As well, the children who were measured in the 1990s showed better performances at age 8 than the children who were measured in the 1970s. There were positive test practice, age and measurement year effects on the curve factor. The children who were measured more frequently, were older and were measured in the 1990s showed smaller rates of improvement than the children who were measured less frequently, were younger and were measured in the 1970s.

Standing long jump. There were positive test practice, age and measurement year effects on the intercept factor. The children who were measured more frequently before age 8, were older, and were measured in the 1990s showed better jumping ability than the children who were measured less frequently before age 8, were younger and were measured in the 1970s.

In summary, the children improved in their motor performance in all variables over a 5-year period, except for the SAR where children's flexibility declined over time. The patterns of children's development were different across performance measures. The children showed linear development in JAR and SAR. For ASR, ESR, SLJ and DASH, the change was largest between ages 8 and 9, and in general, the rate of the change decreased in the subsequent years. A positive test practice effect on all performance measures at age 8 was observed, while the grade effect after controlling for age was observed in ESR only. There was also a positive age effect on all performance measures at age 8 except for SAR. A significant measurement season effect on the ASR and DASH at age 8 revealed that the children who were measured in the summer performed better than the children who were measured in the winter in agility and speed. Children who were measured in the 1990s showed better performance in ASR, ESR, SLJ and DASH at age 8 than the children who were measured in 1970s.

Pseudo Cross-validation

The analyses results for data set 2 are summarized and compared to those for data set 1 in this section. On the average, the children's age at each time point was six months older than that of data set 1. Thus, the age at each time point was 8.5, 9.5, 10.5, 11.5 and 12.5 years. Detailed descriptive statistics and parameter estimates are presented in Appendix C (Table C.14 to Table C.28).

Descriptive Statistics

Means and standard deviations of seven motor performance tests at five time points for data set 2 are presented in Table 1.4.9. In general, the descriptive statistics that were obtained from data set 2 were within the expected range. Children's physical performances at time 1 were slightly better (worse for SAR) in data set 2 than the performances in data set 1 due to the fact that they were six months older,

Table 1.4.9

Means and standard deviations for seven physical performance variables of data set 2

Variable	Age 8.5	Age 9.5	Age 10.5	Age 11.5	Age 12.5	% improvement over 4 years	
						Data set 2	Data set 1
FAH	17.60 ± 13.85	20.09 ± 14.86	22.46 ± 15.69	23.71 ± 16.42	24.31 ± 16.71	38.1 %	43.7 %
JAR	9.80 ± 1.90	10.89 ± 1.83	11.98 ± 1.85	12.81 ± 1.88	14.01 ± 2.31	42.9 %	42.3 %
SAR	7.96 ± 2.17	7.59 ± 2.26	7.38 ± 2.33	7.24 ± 2.48	7.18 ± 2.52	- 9.8 %	- 12.7 %
ASR	12.14 ± 1.00	11.59 ± .80	11.19 ± .86	10.88 ± .76	10.56 ± .72	13.0 %	13.6 %
ESR	42.86 ± 3.35	41.26 ± 3.22	39.97 ± 2.90	38.94 ± 2.81	37.64 ± 2.54	12.2 %	12.8 %
DASH	5.06 ± .41	4.88 ± .41	4.73 ± .37	4.56 ± .35	4.44 ± .34	12.3 %	13.6 %
SLJ	55.44 ± 7.10	59.29 ± 7.33	62.47 ± 7.16	66.01 ± 7.30	69.42 ± 7.17	25.2 %	26.9 %

Note. FAH = flexed-arm hang (seconds), JAR = jump-and-reach (inches), SAR = sit-and-reach (inches), ASR = agility shuttle run (seconds), ESR = 300-foot endurance shuttle run (seconds), SLJ = standing long jump (inches), DASH = 30-yard dash (seconds).

on the average. The magnitude of the change over a 5-year period was smaller than that of data set 1 for all the variables except for JAR -- an expected finding, given the observed deceleration in development after age 9 in data set 1. The children showed a slightly larger change in JAR in data set 2. However, differences in change between data sets 1 and 2 were small (average difference was 1.9%). The pattern of the change was also similar to that of the data set 1. In general, the change was largest between the first two time points, and the rate of the change decreased in subsequent years. As in data set 1, the standard deviations of FAH and SAR were relatively large compare to the magnitude of means, indicating large between-person variability on these performances.

The FAH showed relatively large kurtosis values at first two time points (3.82 and 2.98). However, the kurtosis at other time points and skewness at all time points of FAH were close to zero (smaller than 2.00), indicating small or medium departure from a normal distribution. For other variables, skewness and kurtosis at all time points were close to zero as in data set 1. The largest absolute skewness and kurtosis values across these six variables and all time points were 1.07 and 1.95 on ASR at age 8.5, respectively. Thus, the maximum likelihood estimation was also used for all LGMs for data set 2.

The correlation coefficients between time points of each variable showed a very similar pattern with that of data set 1. In general, correlation coefficients of data set 2 were larger than those of data set 1 except for SAR, although the difference between the two data sets was small. The range of correlation coefficients between time points across all six variables (excluding SAR) was .52 to .86 in data set 2, while it was .46 to .83 in data set 1. This indicated that for these six performance variables, children showed higher consistency over time in their relative positions in data set 2 where children were six months older. For SAR, the range of correlation coefficients among time points was .70 to .84 in data set 2 while it was .74 to .86 in data set 1. As in data set 1, the correlation coefficients between time points approximated simplex patterns for all seven variables. This implies that there were inter-individual variations in the rates of change in performances among children.

Identification of the Best Fitting Growth Curve

Table 1.4.10 shows the comparison between the two data sets in the goodness-of-fit of the selected best fitting models for each variable. In the first vertical column block, a summary of the results of the best fitting models for data set 1 are presented, and in the second column block, two goodness-of-fit indices for the best fitting models of data set 1 that were fitted to data set 2 (cross-validation models) are presented. For example, the quadratic model with unequal error variances that was selected as the best model for the FAH in data set 1 was fitted to FAH for data set 2, and the χ^2 and RMSEA statistics of this model are presented in the second column block. The third column block shows the selected best fitting models, and fit indices for data set 2.

In general, the results of the two data sets were comparable. The goodness-of-fits of the cross-

Table 1.4.10

Identification of the best fitting model: The comparison between data set 1 and cross-validation data (data set 2)

Variable	Data set 1				Data set 2			
	<u>Best fitting model</u>				<u>Cross-validation</u>			
	Identified model	Error variances	χ^2 (df)	RMSEA	Identified model	Error variances	χ^2 (df)	RMSEA
FAH	Quadratic	Unequal	8.96 (6)	.048	Quadratic	Equal	5.21 (10)	< .001
JAR	Linear	Equal	17.31 (14)	.030	Linear	Unequal	15.15 (10)	.044
SAR	Linear	Unequal	10.54 (10)	.019	Linear	Equal	19.99 (14)	.048
ASR	Curve	Unequal	7.78 (7)	.018	Curve	Unequal	21.32 (7)	.100
ESR	Curve	Unequal	3.18 (7)	< .001	Linear	Unequal	16.80 (10)	.062
DASH	Curve	Unequal	18.73 (7)	.085	Linear	Equal	14.60 (14)	.020
SLJ	Cubic	Equal	4.04 (5)	< .001	Linear	Unequal	15.26 (10)	.051

Note. FAH = flexed-arm hang (seconds), JAR = jump-and-reach (inches), SAR = sit-and-reach (inches), ASR = agility shuttle run (seconds), ESR = 300-foot endurance shuttle run (seconds), SLJ = standing long jump (inches), DASH = 30-yard dash (seconds), Cross-validation model = the best fitting model of data set 1.

validation models for data set 2 were worse in five variables, JAR, SAR, ASR, ESR and SLJ, but all of these five variables showed RMSEA statistics smaller than .10 which means an acceptable fit (Browne & Cudeck, 1993), except for ASR. The cross-validation models showed a better fit for the FAH and DASH. In terms of the best fitting models for data set 2 (third column block), the same growth curve models described the children's individual changes well for the FAH, JAR and SAR, but the equality of error variances over time were different between the two data sets. The error variances over time were equal for JAR and SLJ in data set 1, and for FAH, SAR and SLJ in data set 2. The more parsimonious models fitted the data well for the ESR, DASH and SLJ. Linear models fit the data well for these three variables in data set 2, while a Curve or Cubic model was the best fitting model for data set 1. For ASR, none of the Linear, Quadratic, Cubic or Curve models fit the data well in data set 2.

Parameter Estimates of the Best Fitting Growth Models

In the following comparisons, ASR is excluded because none of the growth models were selected as the best fitting model for this measure. The direct comparisons of parameter estimates of the best fitting models between two data sets were not possible because of the differences in the selected growth models. However, in general, the parameter estimates of the growth models for data set 2 were within the expected range considering the developmental trend and age differences between data set 1 and 2. For all variables, the mean of the intercept factor reflected slightly better performances except for SAR. For example, the estimated mean of the intercept factor of the FAH (Table B.16: 17.53, Standard Error (SE) = .96, $p < .001$) was slightly larger than that of data set 1 (Table 1.4.3: 16.69, SE = .93, $p < .001$), indicating a better performance at age 8.5 than at age 8. For SAR, the estimated mean of the intercept factor (Table B.20: 7.86, SE = .15, $p < .001$) was smaller than that of data set 1 (Table B.5: 7.91, SE = .15, $p < .001$), reflecting a slightly worse performance in flexibility at age 8.5 than at age 8. The estimated variances of the intercept factors for all the variables were similar to those of data set 1.

The estimated means of the change factors (linear, quadratic and/or curve factors) were also within the expected range. The mean of the growth factors showed that there were smaller changes in the performances over a 5-year period in data set 2, as compared to data set 1 for all the variables except for JAR. For example, the absolute magnitude of the estimated mean of the linear factor (Table B.24: -1.27, SE = .05, $p < .001$) for the ESR was smaller than the estimated average change of the four time intervals (-1.40) of data set 1. This reflects the decreasing rate of the change in the physical performances over years (two data sets combined, between ages 8 and 12.5). The JAR showed a slightly larger estimated linear factor mean (Table B.18: 1.02, SE = .03, $p < .001$) than that of data set 1 (Table B.3: .99, SE = .03, $p < .001$). All of the estimated variances of the growth factors were significant, implying that there were considerable between-person variations in children's development of physical performances, as in data set 1.

Unlike in data set 1, the correlation between the intercept and the linear factors was significant (-.27, $p = .049$) for FAH, indicating that children who showed better performance at age 8.5 showed

lower rate of improvement. However, the magnitude of the correlation was small. As well, the correlation between the intercept and the linear factors was significant ($- .43, p = .001$) for SLJ, while none of the correlations between the intercept and growth factors of the Cubic model was significant in data set 1. Other parameter estimates for the covariances among factors were similar to those in data set 1.

Predictor Effects

The effects of the predictors on the intercept and change factors for six physical performance variables of data set 2 are summarized in Table 1.4.11. As noted earlier, ASR is excluded from the analyses because none of the growth models were selected as the best fitting model for this measure. For simplicity, only the significant ($p < .05$) effects are presented and done so in standardized units. The statistical test of predictor effects for ASR was not conducted because there was no best fitting growth model for this variable. There were similarities and differences between the two data sets in the effect of the predictors. Where there was an agreement in the significance of an effect, the magnitude of the effect was similar and the direction (positive or negative) of the effect was the same.

The test practice and age effects on the intercept factor were similar between the two data sets. As in data set 1, a positive test practice effect on the performance at the first time point (intercept factor) was found for all variables except for FAH. As well, a positive age effect on the performance at the first time point was found for all the variables except for the SAR in both data sets. The significance of other predictor effects varied between the two data sets. A test practice effect on the change factor was found for the ESR, DASH and SLJ in data set 2, while it was found for the SAR, ASR, ESR and DASH in data set 1. The age effect on the change factor was found for the ESR and DASH in data set 2, while it was found for the DASH only in data set 1. The grade effect was significant on the intercept factor for FAH and DASH and on the change factor for DASH in data set 2, while the grade effect on only the intercept for ESR was significant in data set 1. There was no measurement season effect in data set 2, but there were on the intercept factor for ASR and DASH in data set 1. The measurement year effect was found only for JAR in data set 2, while it was found for the JAR, ASR, ESR, DASH and SLJ in data set 1.

In summary, children's development in physical performance was comparable between data set 1 and data set 2, except for ASR. The same growth models described the individual changes well for the FAH, JAR and SAR in both data sets. More parsimonious models described the change for the ESR, DASH and SLJ in data set 2 than in data set 1. However, for ASR, none of the growth models fit the data well in data set 2, while a Curve model fit the data very well in data set 1. Although the difference was small, the changes in the performances over a 5-year period were smaller for all variables in data set 2, where the children's ages at five time points ranged from 8.5 to 12.5 than in data set 1, where the children's ages at five time points ranged from 8 to 12. This implies that the change was larger at younger ages and the rate of the change decreased at subsequent years. As in data set 1, the test practice and age effects on the intercept and change factors were dominant in most of the variables. The effect of

Table 1.4.11

Predictors' effects on growth factors for six physical performance variables

Variable	Factor	<u>Predictors</u>				
		Test practice	Age	Grade	Season	Year
FAH	Intercept		.154	-.198		
	Linear					
	Quadratic					
JAR	Intercept	.394	.159			.170
	Linear					-.360
SAR	Intercept	.152				
	Linear					
ASR	Intercept	Not conducted				
	Curve					
ESR	Intercept	-.435	-.199			
	Linear	.274	.247			
DASH	Intercept	-.256	-.233	.181		
	Linear	.262*	.316	-.307		
SLJ	Intercept	.354	.155			
	Linear	-.237				

Note. Only significant effects ($p < .05$) are shown in a standardized unit.

FAH = flexed-arm hang (seconds), JAR = jump-and-reach (inches), SAR = sit-and-reach (inches), ASR = agility shuttle run (seconds), ESR = 300-foot endurance shuttle run (seconds), DASH = 30-yard dash (seconds), SLJ = standing long jump (inches), Leaning = the number of pre-measurements, Season = measurement season, Year: measurement year.

* = Test practice effect on the linear factor was excluded after the Grade variable was included because this effect became nonsignificant ($p > .10$).

other predictor variables varied by variable between the two data sets.

Discussion of the Development of Physical Performance

As mentioned in Chapter 1-III, the variables that were used in the present study are not representatives of the all important physical performance and predictor variables. More important physical performance variables such as the variables that measure cardiovascular endurance and/or strength of various body parts were not included since already existing data were used in the present study. As well, more important predictor variables such as height, weight, percent body fat and/or the level of physical activity were not included in the present study.

Children showed two dominant patterns of individual development in physical performances between ages 8 and 12.5 (data sets 1 and 2 combined). These were a constant (linear) change and a deceleration of the change (similar to a quadratic change) over time. Children showed linear changes in two of the seven performance variables, (JAR and SAR), and decelerations in change in the rest of the variables (FAH, ASR, ESR, DASH and SLJ) in data set 1. These two dominant patterns of change were generally supported by the cross-validation procedure. Thus, it is concluded that in early childhood an individual child shows a constant rate of development in physical performance or a faster development during early years and a subsequent decrease in developmental rate. Although the results were not directly comparable to those of previous studies because only group level statistics were available from previous studies, these two dominant patterns of development in physical performance variables during this age range generally agreed with the findings by Baumgartner, East, Frye, Hensley, Knox and Norton (1984), Bayley (1935), Clarke and Wickens (1962), Haubenstricker and Seefeldt (1986), Herkowitz (1978), Marmis, Montoye, Cunningham and Kozar (1969), Milne, Seefeldt and Reuschlein (1976), Morris, Williams, Atwater and Wilmore (1982), and Selis (1951).

The comparisons of overall change in physical performance measures between data set 1 and the cross-validation data (data set 2) indicated that the rate of the development in physical performances decreased between ages 8 and 12.5. Although differences were very small, the rate of change of all performance variables, except JAR, is larger in data set 1 where the children's age ranged from 8 to 12 than in the cross-validation data where the children's age ranged from 8.5 to 12.5. The percentages of change over a 5-year period were larger in data set 1 for all the variables except for JAR (see Table 1.4.9), and the average changes estimated by LGMs were also larger for data set 1. The JAR showed slightly larger average change in the cross-validation data than in data set 1. This implies that the deceleration in the development rate of physical performances started within this age range, from age 8 to age 12.5. The deceleration in the development rate of physical performances within this age range was reported in many studies (e.g., Baumgartner et al., 1984; Milne et al., 1976; Selis, 1951). However, these results do not agree with the aforementioned results for some variables that showed linear changes

within a data set (i.e., SAR in data set 1, and SAR, ESR, DASH and SLJ in data set 2).

The significance and the magnitude of the correlation between the initial status (performance at the first time point) and the rate of change varied by variable. The correlations between the initial status and the rate of change were not significant for FAH, JAR, SAR and SLJ in data set 1. These results disagreed with the general belief that the initial status negatively correlated with the rate of change, but supported Rogosa's (1995) arguments that the correlation between the initial status and the rate of change is not always present, but depends on the specific time interval that is selected in a study. For the variables that showed significant correlations between the initial status and the rate of change (ASR, ESR and DASH), the correlations were negative as noted in previous studies (e.g., Schutz, 1989). It is interesting to note that all the variables that involve running (ASR, ESR and DASH) showed significant correlations between the initial status and the rate of change, and the magnitudes of the correlations were similar across variables (ranged from $-.67$ to $-.62$). This implies that the correlation between the initial status and the rate of change depends not only on the specific time interval that is selected in a study but also on the specific performance measures. The analysis results of the cross-validation data generally supported these findings.

The effects of predictors on the initial status and change were varied by variables. Most notable predictors were the number of pre-measurement (test practice effect) and age in months within the same age group. The test practice effect on the intercept factor (initial status) was significant for all the variables except for FAH. Children who were measured more frequently before the initial time point (age 8) showed better performances at initial time point. This effect is a long-term test practice effect rather than a short-term test (memory or practice) effect, because the interval between any two measurements was six months on the average. The test practice effect was not significant on the performance level of FAH, because the element of skill is relatively small for this test. The effect of age on the intercept was significant for all the variables except for SAR. Children who were older than others (within the same age group) by the length of up to seven months showed higher level of performances at the initial time point. The magnitudes of these test practice and age effects were small to medium, each explained 2% to 15% of variation of the initial status. These effects of the number of pre-measurement and age on the initial status were also evident in the cross-validation procedure, showing similar effects on the performance variables. The effects of other predictors were varied by variables (and also by the data sets). The negative test practice effect on the change rate of a few variables was partially due to the negative relationship between the initial status and the rate of growth.

Multivariate Latent Growth Models for Physical Performances

The results of multivariate LGMs are presented in this section. As explained in Study 1-Chapter III, the models that examine the factor structure at each time point (Figure 1.3.3 and Figure 1.3.4) and the curve-of-factors models (Figure 1.2.5) were fitted to three hypothesized factors, “Run” (ASR, ESR, DASH), “Power” (JAR, SLJ, DASH) and “Motor Ability” (FAH, SLJ, SAR, DASH, ESR). Descriptive and related statistics for “Power” and “Motor Ability” factors for data set 1 and 2 are presented in Appendix C, Table C.29 to Table C.35.

Run (ASR, ESR, DASH)

Descriptive Statistics

Descriptive statistics for the ASR, ESR and DASH are presented in Table 1.4.12. The change in mean and standard deviation of each variable, and the correlations between time points within the same variable were discussed in previous sections. The magnitudes of correlations between different variables within a time point were medium to high (.54 to .80). Generally, the correlations between ASR and ESR were relatively high at all time points (.68 to .80), while other correlations were lower (.54 to .67). This indicates that there was a considerable amount of variation that was shared by these three variables at each time point, and the variation that was shared by ASR and ESR was larger than the variation that was shared by DASH and other variables. The magnitudes of correlations between different variables between different time points were medium, with the smallest coefficient (.37) being between ESR at age 8 and DASH at age 11, and the largest coefficient (.65) being between ESR at age 11 and ASR at age 12.

Verification of the Factor Structure

As explained in Study 1-Chapter III, the factor structure of “Run” should be verified before the multivariate LGM is examined. This requires sequential testing of several models. The goodness-of-fit indices of these sequential tests for the verification of the factor structure are presented in Table 1.4.13.

The first step was the confirmation of the factor structure at each time point. This was done using a 5-factor measurement model with one factor (representing the three performance measures, ASR, ESR and DASH) at each time point. The LISREL computer program commands for the analyses are shown in Appendix B. The 5-factor model (model 1) was rejected. Although the SRMR was in an acceptable range ($< .08$), the χ^2 statistic was almost four times the degrees of freedom, the RMSEA was unacceptably high ($> .06$), and the NNFI was low ($< .90$). In the next model (model 2), the errors of the same variable between time points were allowed to be correlated. This model, the 5-factor model with correlated errors, was not rejected. The RMSEA ($< .06$) and SRMR ($< .08$) were small and the NNFI was large ($> .95$). The ECVI was also much smaller than that of model 1. Thus, there were significant correlations of errors of the same variable between time points. This implies that there was an element

Table 1.4.12

Descriptive statistics for the ASR, ESR and DASH across all time points

Age 8			Age 9			Age 10			Age 11			Age 12			
ASR	ESR	DASH	ASR	ESR	DASH	ASR	ESR	DASH	ASR	ESR	DASH	ASR	ESR	DASH	
ASR	1.00														
ESR	.68	1.00													
DASH	.67	.65	1.00												
ASR	.59	.53	.51	1.00											
ESR	.60	.63	.54	.76	1.00										
DASH	.45	.46	.64	.57	.59	1.00									
ASR	.56	.57	.52	.63	.64	.42	1.00								
ESR	.52	.56	.51	.58	.63	.49	.74	1.00							
DASH	.41	.47	.60	.44	.50	.58	.55	.56	1.00						
ASR	.46	.52	.49	.58	.58	.49	.60	.57	.55	1.00					
ESR	.47	.52	.45	.62	.65	.53	.57	.62	.47	.71	1.00				
DASH	.38	.37	.56	.42	.47	.57	.40	.47	.68	.55	.54	1.00			
ASR	.52	.48	.52	.55	.56	.45	.58	.59	.53	.64	.65	.54	1.00		
ESR	.53	.52	.58	.57	.61	.55	.61	.66	.52	.60	.67	.57	.80	1.00	
DASH	.47	.43	.62	.39	.46	.55	.39	.46	.58	.48	.46	.69	.63	1.00	
Mean	12.46	43.93	5.21	11.92	42.00	4.93	11.39	40.55	4.75	11.06	39.46	4.65	10.76	38.32	4.50
SD	1.05	3.65	.44	.90	3.04	.40	.82	3.03	.36	.74	2.68	.35	.75	2.67	.33

Table 1.4.13

Fit indices of the 5-factor models for the verification of the factor structure of "Run"

Models	$\chi^2(df)$	p-value	RMSEA	ECVI	SRMR	NNFI
1. 5-factor model	313.60(80)	< .001	.140	2.33	.063	.87
2. 5-factor model with correlated errors	67.82(50)	.048	.037	.98	.032	.98
3. Equal factor loadings over time	77.66(58)	.043	.037	.95	.039	.99

Note. df = degrees of freedom, RMSEA = root mean square error of approximation, ECVI = expected cross-validation index, SRMR = standardized root mean square residual, NNFI = non-normed fit index.

in the variance of each observed variable that was not explained by the "Run" factor but that was lasting over time. However, these correlations between errors were relatively small, ranging from -.02 to .21, indicating the lasting part of the each variable's variance that was not explained by the "Run" factor was relatively small. In the next step, the equality of factor loadings over time was examined (model 3). In this model, the factor loadings across all five time points within the same variable were constrained to be equal, and the variance of the factor was allowed to change over time. The χ^2 difference test of this model against model 2 (χ^2 difference = 9.84, $df = 8$, $p > .05$) revealed that the fit of this model was not significantly worse than model 2. The RMSEA, SRMR and NNFI indices also revealed that this model fit the data very well, and the ECIV was smallest among three models (see Table 1.4.13). Thus, it was concluded that the factor structure of the latent variable "Run" did not change over time. The relative magnitudes of explanatory power of the "Run" factor for each observed variable did not change over a 5-year period. This means that conceptually the same latent construct, "Run", was measured over time.

The estimated factor loadings are discussed in the next section. The estimated correlations of factors between time points ranged from .70 to .85, indicating that children showed a relatively high year-to-year stability in the performance of the factor "Run".

Identification of the Best Fitting Growth Curve

Four different growth models (curve-of-factors models), the Linear, Quadratic, Cubic and Unspecified Curve (the Curve hereafter) models, were fitted and compared to examine the children's development in the "Run" performance over time. The factor loading of each variable was constrained to be equal over time, and errors of the same variable between time points were allowed to be correlated in these models. Once the best growth model was selected, the equality of error variances over time for each variable was examined.

The results of the model fit are presented in Table 1.4.14. In terms of the χ^2 statistic, RMSEA and SRMR, the Cubic model (model 6) fit the data best among three growth models, the Linear (model 4), Quadratic (model 5) and Cubic models. The χ^2 difference tests showed that the Cubic model was significantly better than both the Linear (χ^2 difference = 50.99, $df = 9$, $p < .001$) and the Quadratic (χ^2 difference = 12.6, $df = 5$, $p < .05$) models. However, the Cubic model showed very little differences from the Curve model (model 7(a)) in terms of the RMSEA, SRMR, NNFI and ECVI. In addition, the Curve model was more parsimonious (larger degrees of freedom) than the Cubic model. Thus, the Curve model was selected as the best fitting model in describing the children's development in the "Run" performance. Children's development in "Run" performance was well explained by one latent change factor, curve (see Figure 4.2). The test of the equality of error variances over time showed that only model 7(d) was not significantly worse than model 7(a) (χ^2 difference = 7.87, $df = 4$, $p > .05$). Thus, it is concluded that the error variances were equal over time only for the variable DASH.

The parameter estimates of model 7(d), the Curve model with equal error variances over time

Table 1.4.14

Fit indices of latent growth models for "Run"

Models	$\chi^2(df)$	p-value	RMSEA	ECVI	SRMR	NNFI
4. Linear	157.37(78)	< .001	.067	1.27	.066	.95
5. Quadratic	118.98(74)	< .001	.051	1.13	.063	.97
6. Cubic	106.38(69)	.003	.049	1.13	.062	.98
7(a) Curve	117.24(75)	.001	.050	1.12	.065	.98
Equal error variance over time for:						
(b) ASR	148.40(79)	< .001	.062	1.22	.067	.96
(c) ESR	152.30(79)	< .001	.063	1.23	.064	.96
(d) DASH	125.11(79)	< .001	.050	1.11	.066	.97

Note. df = degrees of freedom, RMSEA = root mean square error of approximation, ECVI = expected cross-validation index, SRMR = standardized root mean square residual, NNFI = non-normed fit index.

for DASH, are presented in Figure 1.4.2. In this figure, the path showing correlated errors are omitted for simplicity. Factor loadings for each observed variable at each time point and the correlation between the intercept and curve factors are standardized values (represented in *italic*).

Because the presented factor loadings are standardized values, the magnitudes of the loadings for a variable across time points are not identical although these factor loadings were constrained to be equal in raw values (in its own scale). All other estimates are raw values. All presented estimates were significant at $p < .05$.

The standardized factor loadings for the observed variables were relatively high (.71 to .89), indicating that these variables were well explained by one underlying latent factor "Run" at each time point. Among the three variables, the DASH showed the lowest standardized loadings at all time points. This indicates that the proportion of variance that was explained by the "Run" factor was smaller for DASH than those of ASR and ESR. The mean of the intercept factor was 12.44 (seconds). This was very close to the mean of ASR at age 8 because the "Run" factor was scaled by fixing the factor loading of this variable at 1.0 at each time point. This means that the scale of the "Run" factor was same as that of the ASR, and thus the interpretation of this factor may be based on this scale. The variance of the intercept factor was .57 ($p < .001$), indicating that there was a significant inter-individual variation among children in the performance of "Run" at age 8. The mean of the curve factor was -.58 ($p < .001$), implying that on average, children improved (by .58 seconds) in the "Run" performance between ages 8 and 9. By multiplying factor loading of each time point (i.e., 1.0, 1.81, 2.31 and 2.87) to this mean of the curve factor, one obtains the amount of change between the first time point and the specific time point. Thus, by age 10, the average score on the "Run" had decreased 1.05 seconds ($1.81 \times -.58$), indicating an improvement of .47 seconds ($1.05 - .58$ or $.81 \times .58$) from age 9 to 10. In general, the rate of the improvement decreased until age 11 (an improvement of .29 seconds between 10 and 11), and slightly increased between age 11 and 12 (.32 seconds). The significant variance of the curve factor ($p = .018$) indicated that there was a significant inter-individual variation in the development of "Run" performance. The correlation between the intercept and the curve factors was negative and relatively high ($r = -.66$, $p < .001$). This implies that the children who showed better performance at age 8 showed a slower improvement.

Predictor Effects

As in the univariate LGM, five predictors were sequentially included in the selected Curve model. All models with predictors showed a good model fit ($RMSEA < .06$). There were significant test practice ($-.410$, $p < .001$), age ($-.230$, $p = .001$), measurement season (.274, $p < .001$) and measurement year ($-.234$, $p = .003$) effects on the intercept factor. The children who were measured more frequently before age 8, older (by up to seven months), measured during a summer season (compared to a winter season) and measured in the 1990s (compared to the children who were measured in the 1970s) showed

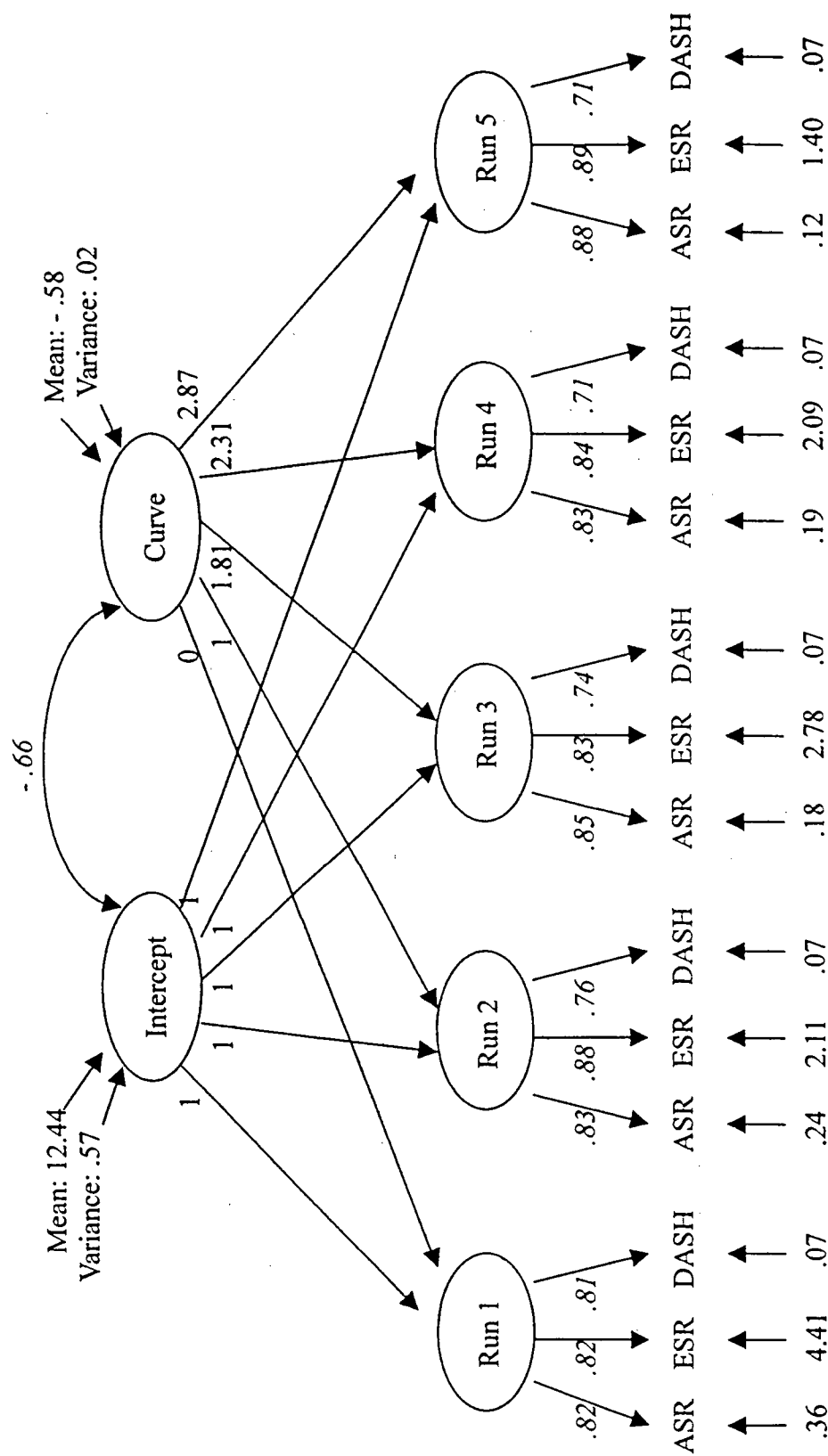


Figure 1.4.2. Curve-of-factors model for "Run"

better performances in “Run” at age 8. The significant test practice (.491, $p < .001$) and measurement year (.327, $p = .011$) effects on the curve factor indicated that the children who were measured more frequently and measured in the 1990s showed slower improvement in “Run” performance over time.

Power (JAR, SLJ, DASH)

Descriptive Statistics

The means and standard deviations of JAR, SLJ and DASH were presented and discussed in the univariate LGM sections. The correlations among these three variables across five time points are presented in Appendix C, Table C.29. The magnitudes (in absolute values) of correlations between different variables within a time point were of medium magnitude (.49 to .72). In general, the correlations between JAR and SLJ showed higher absolute values (.63 to .72) than other correlations (.49 to .67). These two variables that have the element of jumping have a relatively large amount of shared variance. The magnitudes of correlations between different variables between different time points were small to medium with the smallest (in an absolute value) coefficient of -.34 between JAR at age 8 and DASH at age 9, and the largest coefficient of .66 between JAR at age 9 and SLJ at age 10.

Verification of the Factor Structure

The goodness-of-fit indices of models for the verification of the factor structure for the “Power” factor are presented in Table 1.4.15. As in the “Run” factor the 5-factor model with correlated errors (model 2) showed a good model fit. All fit indices indicate a close fit of this model to the data. Thus, errors of each variable over time were significantly correlated. There was a part in the variance of each variable that was not explained by the “Power” factor and was lasting over time. The magnitudes of correlations between errors were relatively small, and ranged from .06 to .24.

In the next step, a 5-factor model with the equality of factor loadings across all time points as well as the correlated errors (model 3(a)) was examined. In this model, the variance of the factor was allowed to change over time (the factor loadings of JAR was fixed at 1.0 at each time point). This model showed a significantly worse model fit as compared to model 2 (χ^2 difference = 34.43, $df = 8$, $p < .05$). Although the absolute fit of this model was good (RMSEA < .06 and NNFI > .95), it is concluded that the factor structure changed over a 5-year period. To examine when the factor structure changed, four additional models were fitted and compared to model 2 using the χ^2 difference test. Model 3(b), in which the factor loadings between time 1 and 2 (age 8 and 9) were constrained to be equal, was significantly worse than model 2 (χ^2 difference = 15.38, $df = 2$, $p < .05$). However, model 3(c) (χ^2 difference = .42, $df = 2$, $p > .05$), model 3(d) (χ^2 difference = 3.32, $df = 4$, $p > .05$), and model 3(e) (χ^2 difference = 7.34, $df = 6$, $p > .05$) were not significantly worse than model 2. Thus, it is concluded that the factor structure changed between ages 8 and 9 (time 1 and 2), and then remained relatively stable through to age 12.

Table 1.4.15

Fit indices of models for the verification of the factor structure of "Power"

Models	$\chi^2(df)$	p-value	RMSEA	ECVI	SRMR	NNFI
1. 5-factor model	368.24(80)	< .001	.152	2.62	.061	.85
2. 5-factor model with correlated errors	44.09(50)	.708	< .001	.91	.020	1.00
3. Equal factor loadings over time						
(a) Time 1 = 2 = 3 = 4 = 5	78.52(58)	.038	.039	.96	.066	.99
(b) Time 1 = 2	59.47(52)	.222	.022	.92	.048	.99
(c) Time 2 = 3	44.51(52)	.760	< .001	.90	.022	1.01
(d) Time 2 = 3 = 4	47.41(54)	.725	< .001	.89	.027	1.01
(e) Time 2 = 3 = 4 = 5	51.43(56)	.648	< .001	.88	.034	1.00

Note. df = degrees of freedom, RMSEA = root mean square error of approximation, ECVI = expected cross-validation index, SRMR = standardized root mean square residual, NNFI = non-normed fit index.

According to the parameter estimates of model 2 in which the factor loading of the JAR was fixed at 1.0 at each time point, and the variance of the "Power" factor was allowed to change over time, the absolute magnitudes of the factor loadings of the SLJ (5.13) and DASH (-.26) at the first time point were considerably larger than those of other time points (the average factor loadings at the other four time points were 4.02 and -.16, respectively; these raw factor loadings are not presented, instead standardized factor loadings are shown in Table 1.4.16). Thus, the amount of variance that was explained by the "Power" factor for the SLJ and DASH, relative to that of JAR, was larger at the first time point than at the other time points.

The standardized parameter estimates of model 3(a), instead of model 3(e), are presented in Table 1.4.16 because this model was used as the base model for the subsequent LGM analyses. This model, in which the equality of factor loadings over all time points was imposed, showed an acceptable absolute fit (i.e. RMSEA < .06). The correlated errors are omitted in Table 1.4.16. Because the presented factor loadings are standardized values, these were not identical over time. The absolute magnitudes of the factor loadings were moderate to large (.62 to .89). The factor loading for SLJ was largest while the factor loading for DASH was smallest among three observed variables at each time point. This implies that the latent factor "Power" explained smaller proportion of the variance of DASH than those of the other two variables. The correlations of factors between time points were high (.80 to .96), indicating that children showed a relatively high level of year-to-year stability in the "Power" performance. In general, these correlations approximate a simplex pattern, with the correlation coefficients becoming smaller as a coefficient gets further away from the main diagonal. This indicates that there was an inter-individual variation in the development of the "Power" performance.

Identification of the Best Fitting Growth Curve

Although it is concluded that the factor structure changed between ages 8 and 9, the factor loadings of each observed variable across all time points were constrained to be equal for the multivariate LGM analysis because the multivariate latent growth model (a curve-of-factors model) requires the equality of factor structure (loadings) over time. Four growth models were fitted, and the results of the model fit are presented in Table 1.4.17.

The goodness-of-fit indices indicate that all four growth models were rejected. The χ^2 statistic was large, and the RMSEA, SRMR and NNFI were in unacceptable ranges. In addition, maximum likelihood estimation produced improper solutions (i.e., negative variances). Attempts to resolve this by using several different sets of starting values resulted in the convergence to the same solution. It is concluded that none of the growth models adequately explain the change in the "Power" factor. Thus, the interpretation of parameter estimates and further analyses for the predictors' effects were not conducted.

Table 1.4.16

Parameter estimates of the 5-factor model with correlated errors and the equality of factor loadings over time for "Power"

<u>Standardized factor loading</u>			<u>Correlations of factors between time points</u>				
Time	Variable	Loading	Age 8	Age 9	Age 10	Age 11	Age 12
<u>Age 8</u>	JAR	.79					
	SLJ	.87	1.00	.91	.87	.78	.80
	DASH	-.66					
<u>Age 9</u>	JAR	.81					
	SLJ	.81		1.00	.92	.86	.86
	DASH	-.62					
<u>Age 10</u>	JAR	.74					
	SLJ	.83			1.00	.96	.91
	DASH	-.67					
<u>Age 11</u>	JAR	.79					
	SLJ	.89				1.00	.89
	DASH	-.69					
<u>Age 12</u>	JAR	.74					
	SLJ	.89					1.00
	DASH	-.74					

Note. Correlated errors are omitted. All estimates were significant at an alpha level of .01.

Table 1.4.17

Fit indices of latent growth models for "Power"

Models	$\chi^2(df)$	p-value	RMSEA	ECVI	SRMR	NNFI
4. Linear	1285.11(78)	< .001	.181	3.49	.286	.35
5. Quadratic	1273.10(74)	< .001	.190	3.61	.337	.32
6. Cubic	1265.03(69)	< .001	.197	3.63	.337	.27
7. Curve	1274.43(75)	< .001	.182	3.41	.294	.32

Note. df = degrees of freedom, RMSEA = root mean square error of approximation, ECVI = expected cross-validation index, SRMR = standardized root mean square residual, NNFI = non-normed fit index.

Motor Ability (FAH, SLJ, SAR, DASH, ESR)

Descriptive Statistics

The means and standard deviations of each variable at each time point were presented and discussed in previous sections. The correlations between variables across five time points are presented in Appendix C, Table C.30. The magnitudes (in absolute values) of correlations between different variables within a time point range from small to medium (.19 to .67). Generally, correlations between SLJ and DASH (.51 to .67), between SLJ and ESR (.56 to .66) and between DASH and ESR (.54 to .65) showed larger values than correlations between other variables at each time point. This indicates that the amount of shared variation among the SLJ, DASH and ESR was relatively large while the FAH and SAR had smaller amount of shared variation with other variables. The magnitudes of correlations between different variables between different time points range from also small to medium (.11 to .59).

Verification of the Factor Structure

The results of the model fit for the verification of the factor structure for the “Motor Ability” (Table 1.4.18) factor were very similar to those of the “Power” factor. The 5-factor model with correlated errors showed a good model fit (model 2). The model with the equality of factor loadings over five time points (model 3) was rejected compared to model 2 (χ^2 difference = 41.11, $df = 16$, $p < .05$). Further analyses revealed that the factor structure changed between ages 8 and 9 (only model 3(a) was significantly worse than model 2 at an α level of .05).

According to the parameter estimates of model 2, the absolute magnitudes of factor loadings for all the variables except for the FAH, which was used as a scaling variable (factor loading of this variable was fixed at 1.0), were considerably larger at the first time point (1.15, .17, -.07 and -.53 for the SLJ, SAR, DASH and ESR, respectively) than those of the other time points (average factor loadings of four time points were .80, .10, -.04 and .31, respectively). For these variables, the amount of variation that was explained by the “Motor Ability” factor, relative to that of the FAH, was considerably larger at the first time point than at the other four time points. As was done for the “Power” factor, the standardized parameter estimates of model 3(a) are presented in Table 1.4.19. The absolute fit of this model was acceptable in terms of the χ^2 statistic, RMSEA, and NNFI.

The display of correlated errors is omitted in Table 1.4.19. The absolute magnitudes of factor loadings were small to large (.29 to .82). The factor loadings for SLJ (.76 to .80), DASH (.72 to .78) and ESR (.73 to .82) were relatively large, while the factor loadings for FAH (.36 to .47) and SAR (.29 to .38) were relatively small. This indicates that the factor “Motor Ability” was highly characterized by three variables, the SLJ, DASH and ESR. The amount of variance that was explained by the “Motor Ability” factor was relatively small for the FAH and SAR. The correlations of factors between time points were high (.81 to .93), indicating that children showed a high level of year-to-year stability in “Motor Ability” performance.

Table 1.4.18

Fit indices of models for the verification of the factor structure of "Motor Ability"

Models	$\chi^2(df)$	p-value	RMSEA	ECVI	SRMR	NNFI
1. 5-factor model	2307.67(265)	< .001	.228	15.67	.151	.49
2. 5-factor model with correlated errors	242.81(215)	.094	.022	2.18	.060	.99
3. Equal factor loadings over time						
(a) Time 1 = 2 = 3 = 4 = 5	283.92(231)	.010	.029	2.20	.085	.99
(b) Time 1 = 2	252.99(219)	.057	.024	2.19	.069	.99
(c) Time 2 = 3	246.31(219)	.099	.021	2.16	.062	.99
(d) Time 2 = 3 = 4	253.17(223)	.081	.023	2.16	.067	.99
(e) Time 2 = 3 = 4 = 5	260.78(227)	.061	.024	2.15	.067	.99

Note. df = degrees of freedom, RMSEA = root mean square error of approximation, ECVI = expected cross-validation index, SRMR = standardized root mean square residual, NNFI = non-normed fit index.

Table 1.4.19

Parameter estimates of the 5-factor model with correlated errors and the equality of factor loadings over time for "Motor Ability"

<u>Standardized factor loading</u>			<u>Correlations of factors between time points</u>				
Time	Variables	Loading	Age 8	Age 9	Age 10	Age 11	Age 12
<u>Age 8</u>	FAH	.47					
	SLJ	.82					
	SAR	.38	1.00	.92	.91	.81	.83
	DASH	-.77					
	ESR	-.73					
<u>Age 9</u>	FAH	.41					
	SLJ	.76					
	SAR	.35		1.00	.91	.87	.87
	DASH	-.72					
	ESR	-.77					
<u>Age 10</u>	FAH	.36					
	SLJ	.76					
	SAR	.33			1.00	.92	.87
	DASH	-.74					
	ESR	-.73					
<u>Age 11</u>	FAH	.37					
	SLJ	.80					
	SAR	.30				1.00	.93
	DASH	-.74					
	ESR	-.76					
<u>Age 12</u>	FAH	.37					
	SLJ	.80					
	SAR	.29					1.00
	DASH	-.78					
	ESR	-.82					

Note. Correlated errors are omitted. All estimates were significant at an alpha level of .01.

Identification of the Best Fitting Growth Curve

The goodness-of-fit indices of the four multivariate LGMs for the “Motor Ability” factor are presented in Table 1.4.20. As was shown for the “Power”, none of the four growth models fit the data well. It is concluded that none of the four growth models adequately explained the children’s development in “Motor Ability” performance over a 5-year period. Further analyses for the predictors’ effects were not conducted.

In summary, although all hypothesized factors showed good model fits for the factor structure, children’s development in the “Run” performance only, among three hypothesized factors, was explained well by a multivariate LGM, a Curve model. In this Curve model for the “Run” factor, three observed variables were well explained by one underlying latent factor at each time point. The children’s performance represented by the “Run” factor improved over a 5-year period, and the change was largest between ages 8 and 9. There were significant test practice, age, measurement season and measurement year effects on the intercept factor, and significant test practice and measurement year effects on the curve factor. Children showed change in factor structure in “Power” and “Motor Ability” performances between ages 8 and 9. None of the multivariate LGMs for these two factors adequately explained the children’s development in these latent traits.

Pseudo Cross-Validation

The series of all analyses and model testings presented in the previous sections dealing with multivariate LGMs were replicated on data set 2. The results are summarized and compared to those for data set 1 in Table 1.4.21. As noted previously, on average, the children’s age at each time point was six months older than that of data set 1. Thus, the mean age at each time point was 8.5, 9.5, 10.5, 11.5 and 12.5 years.

Descriptive Statistics

In general, the absolute magnitudes of correlations between variables within a time point and between different variables between time points were slightly larger in data set 2 than in data set 1 for all hypothesized factors with some exceptions (see the first column block in Table 1.4.21). The patterns of the correlations were very similar to those of data set 1. For example, for the “Run” factor, the correlations between ASR and ESR showed the highest values among correlations between different variables within a time point at each time point in both data set 1 (Table 1.4.12: .68 to .80), and data set 2 (Table B.31: .75 to .82).

Verification of Factor Structure

As in data set 1, the 5-factor model with correlated errors fit the data well in all three hypothesized factors ($RMSEA < .06$). The test of the equality of factor loadings over time for the “Run” factor also revealed that the factor structure of the “Run” factor did not change over time (χ^2 difference

Table 1.4.20

Fit indices of latent growth models for "Motor Ability"

Models	$\chi^2(df)$	p-value	RMSEA	ECVI	SRMR	NNFI
4. Linear	1578.75(261)	< .001	.131	6060	.294	.67
5. Quadratic	1545.33(257)	< .001	.126	6.18	.295	.67
6. Cubic	1530.05(252)	< .001	.125	6.10	.296	.67
7. Curve	1539.05(258)	< .001	.124	6.08	.295	.67

Note. df = degrees of freedom, RMSEA = root mean square error of approximation, ECVI = expected cross-validation index, SRMR = standardized root mean square residual, NNFI = non-normed fit index.

Table 1.4.2.1

Comparison of multivariate analyses results between data set 1 and data set 2*

Factor	Correlations between observed variables**		Equality of factor loadings over time		Factor loadings of observed variables***		Correlations of factors between time points***		The best growth model	
	Data 1	Data 2	Data 1	Data 2	Data 1	Data 2	Data 1	Data 2	Data 1	Data 2
Run	.37 ~ .80	.45 ~ .82	1 = 2 = 3 = 4 = 5	1 = 2 = 3 = 4 = 5	.71 ~ .89	.74 ~ .92	.70 ~ .85	.69 ~ .88	Curve	Curve
Power	.34 ~ .72	.49 ~ .77	1 ≠ 2 = 3 = 4 = 5	1 = 2 = 3 = 4 ≠ 5	.62 ~ .89	.68 ~ .90	.80 ~ .96	.77 ~ .94	None	None
Motor Ability	.11 ~ .67	.20 ~ .70	1 ≠ 2 = 3 = 4 = 5	1 = 2 = 3 = 4 ≠ 5	.29 ~ .82	.34 ~ .86	.81 ~ .93	.76 ~ .95	None	None

Note. *The range of estimated values are standardized absolute values.

**The ranges of correlations exclude the correlations of the same variable between different time points.

***The presented factor loadings and correlations of factors between time point are based on the model with correlated errors and equality of factor loadings over five time points.

= 11.06, $df = 8$, $p > .05$, as compared to the model with free factor loadings) as in data set 1. However, the other two factors ("Power", and "Motor Ability") showed that the factor structure changed between time 4 and 5 (ages 11.5 and 12.5), while the factor structure changed between time 1 and 2 (ages 8 and 9) in data set 1 (see the second column block in Table 1.4.21).

Although the model with equality of factor loadings over all five time points was rejected against the model with changing factor loadings in three hypothesized factors, the model with equality constraints still showed a good fit to the data for all hypothesized factors ($RMSEA < .06$). Thus, the range of factor loadings and factor correlations presented in Table 1.4.21 are based on this model with equality constraints. The factor loadings were slightly larger in data set 2 than in data set 1 for all factors. The average factor loading was .79 in data set 2 and it was .76 in data set 1.

Identification of the Best Fitting Growth Curve and Predictor Effects

As in data set 1, only the change of the "Run" factor was adequately explained by a multivariate LGM (Table 1.4.22). In terms of the $RMSEA$, $ECVI$ and $NNFI$, the Quadratic (model 5), Cubic (model 6) and Curve (model 7(a)) models showed very similar fits, with the Quadratic model showing a slightly better fit than other models in terms of the $RMSEA$ and $ECIV$. However, the Curve model was selected because this model differed only slightly from the Quadratic model in terms of model fit but it had more degrees of freedom. In addition, because a Curve model was selected as the best fitting model for the "Run" factor in data set 1, it provides easier comparisons between the two data sets. With this Curve model, only the DASH variable showed equal error variances over time (model 7(d): χ^2 difference = 4.67, $df = 4$, $p > .05$), as in data set 1.

The estimated mean of the intercept factor (12.06) was slightly smaller than that of the data set 1 (12.44), indicating a better average performance at age 8.5 than at age 8 in "Run". The variance of the intercept factor (.59, $p < .001$) was very close to that of data set 1 (.57). As in data set 1, the average change was largest between the first two time points (-.46) and the rate of the change decreased in subsequent years (-.38, -.34 and -.32 between time 2 and 3, 3 and 4, and 4 and 5, respectively). The variance of the Curve factor (.02, $p = .001$) was very close to that of data set 1 (.02). The correlation between the intercept and the curve factors (-.58) was negative and relatively high, as in data set 1 (-.66).

All of the LGMs for the "Run" factor with the predictors fit the data very well ($RMSEA < .06$). As in data set 1, there were significant test practice (-.426, $p < .001$) and age (-.207, $p = .003$) effects on the intercept factor. However, unlike data set 1, the effects of measurement season and measurement year on the intercept factor were not significant. There was a significant test practice effect (.338, $p = .001$) on the curve factor, as in data set 1, but unlike data set 1, the measurement year effect was not significant. In addition, the age effect (.298, $p = .003$) on the curve factor was significant in data set 2, while it was not significant in data set 1.

Table 1.4.22

Fit indices of latent growth models for "Run" factor (data set 2)

Models	$\chi^2(df)$	p-value	RMSEA	ECVI	SRMR	NNFI
4. Linear	123.82(78)	< .001	.053	1.17	.052	.98
5. Quadratic	108.16(74)	.006	.045	1.11	.047	.98
6. Cubic	105.59(69)	.003	.048	1.15	.046	.98
7(a) Curve	113.04(75)	.003	.046	1.12	.053	.98
Equal error variance over time (Curve model)						
(b) ASR	143.13(79)	< .001	.061	1.23	.059	.97
(c) ESR	153.19(79)	< .001	.064	1.27	.053	.96
(d) DASH	117.71(79)	.003	.045	1.10	.054	.98

Note. df = degrees of freedom, RMSEA = root mean square error of approximation, ECVI = expected cross-validation index, SRMR = standardized root mean square residual, NNFI = non-normed fit index.

The χ^2 statistic and RMSEA of four growth models of the other two factors, “Power”, and “Motor Ability” are presented in Table 1.4.23. The χ^2 statistics and RMSEAs indicate that none of the four growth models fit the data for these two hypothesized factors. In addition, the maximum likelihood estimation produced improper solutions (negative variances), as in data set 1. Thus, further analyses were not conducted for these factors.

In summary, the analysis results of data set 2 were very similar to those of data set 1. Children’s development in the “Run” performance was adequately explained by the Curve model. There were some similarities and differences between data sets 1 and 2 in the significance of the predictors’ effects on initial status and the change. However, none of the growth models adequately explained the change in the other two factors, “Power” and “Motor Ability”. For these two factors, the factor structure changed between time 4 and 5 (ages 11.5 and 12.5) while it changed between time 1 and 2 (ages 8 and 9) in data set 1.

Discussion of the Multivariate Development of Physical Performance

Multivariate analyses of the data included two main parts, an examination of the hypothesized factor structure and an examination of growth curves of latent factors. The examination of the factor structure for three hypothesized factors provided evidence that rejects the early concepts of general motor ability, and partially supports the specificity of the physical performance factors to the particular muscle groups or particular types of movement. The factor models for the examination of the factor structure at each time point for all three hypothesized factors fit the data well. However, the standardized factor loadings of the FAH and SAR for the “Motor Ability” factor that was hypothesized based on the earlier concept of general motor ability, were relatively small (.29 to .47), while the factor loadings of the other three observed variables (SLJ, DASH and ESR) were relatively large (.73 to .86). This implies that the “Motor Ability” was largely characterized by the three variables, SLJ, DASH and ESR, and did not explain well the variance of the FAH and SAR. Thus, the concept of general motor ability was rejected in this study as in many earlier studies (e.g., Cousins, 1955; Baumgartner & Zuidema, 1972; Jackson, 1971). It is unlikely that a single general motor ability factor explains all the physical performances even for boys in early childhood. On the contrary, the factor loadings of all observed variables for the “Run” (.71 to .92) and “Power” (.62 to .89) factors were relatively large, indicating that these factors explained the performances of running and explosive leg power fairly well. These two factors were characterized by a particular type of movement or by a particular muscle group. Thus, the findings by Baumgartner and Zuidema (1972), Cousins (1955), Jackson (1971), Liba (1967), and Start, Gray, Glencross and Walsh (1966) were not refuted. However, this is not a strong support for the specificity notion of physical performance latent variable(s), because not all possible sets of physical performance variables were included in the examination of factor structure in the present study. Because

Table 1.4.23

Goodness-of-fit indices of growth models for two factors

Model	<u>Power</u>		<u>Motor ability</u>	
	$\chi^2(df)$	RMSEA	$\chi^2(df)$	RMSEA
Linear	1265.39(78)	.190	1637.42(261)	.137
Quadratic	1262.43(74)	.194	1561.21(257)	.129
Cubic	1256.85(69)	.200	1558.69(252)	.131
Curve	1264.23(75)	.193	1630.36(258)	.137

Note. df: degrees of freedom, RMSEA: root mean square error of approximation.

of this limitation in this study, the specificity of factor structure is not clearly supported for the population of young children. Regarding the fact that most of the factor analytic studies were conducted on a college population, more studies are needed to verify the factor structure of physical performance for children.

The examination of the equality of factor loadings (the equality of relative contributions of indicator variables) over time for "Power" and "Motor Ability" revealed that the factor structure changed between ages 8 and 9 in data set 1 and between ages 11.5 and 12.5 in the cross-validation data (data set 2). However, the magnitude of the change in factor structure was small for both data sets, showing that the models with equality constraint between all time points fit the data fairly well. Although these results were consistent within each data set, it is difficult to draw a general conclusion regarding the time point of change in factor structure because of the disagreement between the two data sets: In the present study, the time point of the change in factor structure was specific to selected sample and time points. These results generally support the findings by Marsh (1993), in that factor structure is not equal over time but the difference is small, although his conclusions were based on the comparisons of independent age groups (ages of 9, 12 and 15) and different sets of performance variables. On the contrary, the "Run" factor showed the equality of factor loadings over time in both data sets. This implies that while each indicator variable showed different variations in individual change rate among children, the relationship among these three variables did not change over time. In other words, the same underlying latent trait explained the variations in three running ability variables, ASR, ESR and DASH, at each of five time points. Strictly speaking, the results of the present study indicate that only the "Run" factor would be a valid latent trait to employ as a measure in a longitudinal analysis (Marsh, 1993). However, this may be an overly cautious conclusion because the other two hypothesized factors showed only marginal differences in factor loadings over time. As noted by Marsh (1993), inadequate attention has been given to the issue of factorial invariance over time for physical performance variables, and thus more study is needed to investigate this issue, especially for the populations of children and youth.

The Curve model adequately explained the children's development in the "Run" performance. In general, the estimated parameters for development in "Run" performance over time were similar to those for ASR, because ASR was used as the scaling variable. While the parameters for change factors (i.e., intercept and curve factors) were similar to those for ASR, these two change factors adequately explained the changes in means and variances of "Run" factor that adequately explained the means and variances of the other two indicator variables, ESR and DASH at five time points. On average, children improved in "Run" performance over a 5-year period. The children's average change in the "Run" performance was largest between the first two time points and the rate of the change decreased in subsequent years. Positive test practice and age effects on the intercept, and a negative test practice effect on the curve factor were found in both data sets, but the effects of other predictors varied between

the two data sets.

Unlike for the "Run", none of the specified LGMs adequately explained the children's development in the other two factors, "Power" and "Motor Ability". This, however, was not because of the change in the factor structure over time of these two factors. The differences in factor loadings were marginal where the change in the factor structure occurred. Although the results are not presented, additional LGM analyses were conducted excluding the time point that showed a different factor structure from the rest of the time points (i.e., time 1 in data set 1, and time 5 in data set 2). These analyses did not produce acceptable model fits. This implies that the specified growth models were not adequate to explain the complex components of individual change in the latent trait of "Power" or "Motor Ability". Thus, results of the present study indicated that only the "Run" factor was a valid construct for the explanation of development in children's performance. The other two latent factors did not adequately represent the children's development in physical performance. It is noteworthy that the change of all observed variables for the "Run" factor were adequately explained by the Curve model in univariate analyses, while the observed variables for the other two factors showed different patterns of development (in terms of the selected best growth model). This issue can be viewed as a relationship between univariate change and multivariate change rather than a specific problem of the multivariate change of physical performance. This issue is discussed in Study 1-Chapter V.

STUDY 1-CHAPTER V. DISCUSSION

Merits of Latent Growth Models

In the present study, the latent growth model (LGM) approach was employed for the analyses of longitudinal physical performance data. The application of this statistical model to the physical performance data revealed several merits over traditional methods in describing and explaining the development of children's performance over time. One of the most notable merits of LGM in describing change was the capability of modelling and analyzing change at the individual level. This does not mean that a LGM estimates the change parameters of every subject, but rather that a LGM estimates the inter-individual variation as well as the mean of individual change. For example, in the present study, the Quadratic model adequately explained the children's individual change in FAH performances, and provided mean and variance estimates of linear and quadratic factors. That is, the Quadratic model decomposed the variations of change in FAH performances into two components, the linear and quadratic. The significant mean and variance of the linear factor indicated that children improved in their FAH scores over time, and there was considerable inter-individual variation in the rate of improvement among children. The significant variance but the non-significant mean of the quadratic factor indicated that, on the average, there was no quadratic effect, but some children accelerated and some children decelerated in the change rate of FAH performance. In addition, the variance and mean of the change factors implied that the development of some children in FAH score might be linear (i.e., a zero score for the quadratic factor).

These kinds of inferences can be made based on only an individual level of analysis. This merit has been emphasized by many (e.g., Meredith & Tisak, 1990; Willet & Sayer, 1994). In general, traditional methods do not directly provide the information regarding the individual level of change. To obtain a similar level of information, curve fitting methods and/or stochastic models require two steps of analysis, one at the individual level and one at the group level. However, for these models the same mathematical model has to be fitted to all the subjects at the individual level analysis in order to conduct the group level of analysis. This is a serious shortcoming of these approaches. Traditional ANOVA with polynomial contrast (trend analysis) provides information that is similar to that of LGM in terms of describing change. In an ANOVA procedure, the within-subjects variance is decomposed into linear, quadratic, cubic, etc. components.

An example of ANOVA results for FAH (data set 1) is presented in Table 1.5.1. The within-subjects sum of squares indicated that most of the change in mean scores of FAH is explained by the linear effect (97.1%). However, the within-subjects error sum of squares indicated that there is a considerable between-subjects variation (29.8%) in quadratic change. These two pieces of information together agreed with the results of the LGM analysis in that although individuals change in quadratic fashions, the inter-individual differences cancelled each other out and produced a non-significant group

Table 1.5.1

The results of ANOVA analysis with polynomial contrasts for FAH, data set 1

Source	Sum of Squares	%	df	Mean Square	F	p-value
<u>Between-subject</u>						
Intercept	451157.36		1	451157.36		
Error	185995.64		209	889.93		
<u>Within-subjects</u>						
Linear	6589.71	97.1%	1	6589.71	72.41	< .001
Quadratic	168.58	2.5%	1	168.58	2.38	.124
Cubic	18.11	0.3%	1	18.11	.49	.485
4th order	8.57	0.1%	1	8.57	.22	.640
Error						
Linear	19020.09	38.3%	209	91.01		
Quadratic	14789.28	29.8%	209	70.76		
Cubic	7716.59	15.5%	209	36.92		
4th order	8159.07	16.4%	209	39.04		

level quadratic effect. The information that is not available in ANOVA results, however, is the significance test for this inter-individual variation in quadratic change, as well as other types of change (e.g., linear, cubic). This is available in LGM by means of significance testing for the variances of change factors. In addition, in ANOVA, one is interested more in the mean scores and variances due to the within-subjects factor, thus usually the error terms of ANOVA results are not interpreted. The requirement of satisfying the sphericity assumption is another shortcomings of ANOVA procedure, because generally one may anticipate that the variance of a measure changes over time in a longitudinal study.

The LGM's capability for an individual level of analysis for change further enables one to extend the basic model for the description of change to various models for the explanation of change. One such extension is examining the effect of predictor(s) on change. This is not possible in the traditional ANOVA model. The inclusion of the predictor variables in a LGM is conceptually similar to multiple regression analysis in that the effects of several variables on the change can be examined, and is similar to the analysis of covariance (ANCOVA) in that the effects of several covariates (predictors) are controlled for. However, using a regression or an ANCOVA model requires one to estimate the individual change scores before the effect of a predictor is examined. Unlike traditional models, a LGM estimates the parameters of the change and predictor effects at the same time. In the present study, the effects of five predictor variables on the development of physical performance variables were examined. The effects of these five predictors were hierarchically included in the model based on a priori hypotheses, and the significance of a test as well as the magnitude of the effect was obtained. The examination of the effect of a single predictor variable on the change and the examination of the effect of a predictor variable after controlling for other predictor variables were made. Although the results varied by variables and by data sets, and only a small part of variation in change was explained, this analysis procedure for the examination of predictors' effect provided useful information in explaining inter-individual differences in children's development. This capability of examining predictors' effect on change is a notable merit of LGM, and consequently, has been emphasized and employed in many studies (e.g., Duncan & Duncan, 1995; Meredith & Tisak, 1990; Muthen & Curran, 1997; Willet & Sayer, 1994).

LGM, like the general structural equation modelling (SEM), allows one to decompose the variance of an observed variable into two components, the true score variance and the error variance. The intercept and change factors describe only the true score component of an observed variable, thus represent the true score at the first time point and the true change. The error component is the uniqueness that is not explained by the intercept and change factors. Thus, in a LGM, the error component of a variable is taken into account in the analysis, while it is not in traditional methods. This may lead one to make different conclusions regarding description and/or explanation of change. For example, the LGM revealed that the age effect on the intercept factor of FAH in data set 1 was

significant, while a randomized group ANOVA analysis showed a non-significant age effect ($F = 1.329$, $p = .238$) on FAH at the first time point (age 8).

In addition to this, LGM allows one to examine various research questions regarding the error component of observed variables. Generally, two types of research questions are examined. First, the equality of error variances over time can be examined. This is used not only to test theoretically based hypotheses about the equality of the error variances over time, but also to obtain a more parsimonious model. In this study, the equality of error variances over time was tested for the latter purpose. Two data sets showed different results in the univariate LGM analyses. The JAR and SLJ showed equal error variance over time in data set 1, while FAH, SAR and DASH showed equal error variance over time in data set 2. There is no theoretical base that supports these findings. It is rather unreasonable to expect the equality of error variances over time, because in a longitudinal study one expects that the observed variance as well as the true score variance of a variable changes over time. In the multivariate analyses with the "curve-of-factors" model, two data sets showed consistent results. Only DASH, among three variables that form the factor "Run" at each time point, showed equal error variances over time. The implication of the error variance in this model is different from that of the univariate model. In the "curve-of-factors" model, the error variance represents the component that is not explained by the "Run" factor. The equality of error variance over time for DASH implies that the magnitude of unexplained variance in DASH was the same over time.

The second type of research question that is related to the error component of observed variables is the correlation of errors between time points. The examination of the correlation of errors between time points has been a common practice in the factor analysis of repeated measures data especially in a multivariate model (e.g., Marsh, 1993; Marsh & Hau, 1996; Schutz, 1998). In a univariate LGM, the examination of correlated errors between all possible pairs of time points is not possible because of an identification problem (i.e., the number of free parameters is larger than the number of means and covariances that are used as data). Only some of the possible pairs of time points can be examined, and the extent of how many correlated errors can be examined depends on the number of time points and the model (e.g., linear, quadratic, cubic etc. and any constraints that are imposed in the model). Thus, one should be cautious when including correlated errors in a univariate model because of the identification problem. Including correlated errors in a univariate model should be done only when the reason for this can be justified by theory or by a specific research condition (e.g., different testers over time). In the present study, the correlation of errors between time points was not examined for the univariate LGM because there was no theoretical base or other research condition that supports this. In the multivariate LGM analyses (with "curve-of-factors" model), however, correlated errors were included in the model for the same variable between time points in order to obtain a better fitting measurement model. Significant correlated errors were found for all hypothesized factors in both data sets. This implies that there exists a lasting component over time within a variable that was not explained by the hypothesized

factor at each time point. This kind of correlated errors in a multivariate longitudinal model has been found in several studies (e.g., Marsh, 1993; Marsh & Grayson, 1994; Schutz, 1998).

There are other extensions that are based on the individual level of analysis for change. These are models in which the relationship between changes in two or more variables are examined, multi-group analysis models, cohort-sequential analysis models, etc. (Meredith & Tisak, 1990; Willet & Sayer, 1994). Although these models were not included in this study, the flexibility of LGM that allows one to examine various research questions is a strength of LGM. These merits of LGM, based on the individual level of analysis for change, are comparable to those of the hierarchical linear model (HLM) (Bryk & Raudenbush, 1992). A comparison between these two statistical models was not made in this study. HLM is also based on the individual level of analysis for change, thus allowing one to examine the predictors' effects on change, the relationship between intercept and change, the inter-individual variation of change. HLM is more efficient than LGM in the parameter estimation procedure, and does not require that all the subjects be measured at approximately the same time. However, LGM is generally more flexible in modelling and allows one to examine various research questions that are not possible in HLM (Chou, Bentler & Pentz, 1998; Willet & Sayer, 1994). For example, hypothesis testing with error variances, examining relationships between changes of different variables, cohort sequential analysis, and multivariate extensions are available only in LGM. A few of these were examined and presented in the present study.

Problems of Using Latent Growth Models

The application of LGM in the analysis of longitudinal data in this study not only showed several merits as discussed above, but also raised a few practical issues. First, selecting one model over another, such as between the Curve model and the Quadratic or Cubic models, was a somewhat arbitrary process at times. This is due to the fact that the χ^2 difference test is not available in the comparison of these models because the Curve model and Quadratic or Cubic model are not nested to each other. For the comparison of non-nested models, the usage of the expected cross-validation index (ECVI) and/or Akaike's information criteria (AIC) is recommended (Akaike, 1987; Browne & Cudeck, 1993; Cudeck & Browne, 1983). For both of these indices, a lower absolute value indicates a better fitting model. In the present study, the ESR in data set 1 produced results that revealed very small differences in model fit between the Curve and Quadratic models. Both models were not rejected in terms of the χ^2 statistic, and all other fit indices indicated that both models fit the data well (see Appendix B). Although the ECVI and AIC indicated that the Curve model is better, the differences between two models in these indices were very small (i.e., ECVI were .134 and .139 for the Curve and Quadratic models, respectively). Because both models fit the data very well and showed very small differences in model fit, it is difficult to select one model over another. In this study, the Curve model was selected because the

Curve model was more parsimonious than the Quadratic model. The degrees of freedom (df) was larger for the Curve model (df = 7 and 6 for the Curve and Quadratic models, respectively), and the inter-individual variation in change was explained by the only one change factor in the Curve model while it was explained by two change factors, the linear and quadratic, in the Quadratic model. However, this may not be the case in other situations. The parsimony of these models changes as the number of time points increases. For example, with seven time points, the Quadratic model is more parsimonious than the Curve model in terms of degrees of freedom (df = 18 and 19 for the Curve and Quadratic models, respectively). The difference between two models in the degrees of freedom becomes larger as the number of time points increases, and the Quadratic model becomes more and more parsimonious than the Curve model as the number of time points increases. This is due to the fact that in the Curve model the shape of change is not specified, thus the change parameter has to be estimated for each time interval. For this reason, the Curve model can be regarded as an exploratory, rather than confirmatory, way of finding the best-fitting curve compared to other models such as Quadratic and Cubic models. Thus, in comparing these models, one has to consider if the model fitting should be confirmatory (based on theory) or exploratory (unspecified curve) as well as the parsimony of the model.

The second practical issue is the relationship between the change of each indicator variable and the change of the latent factor in the "curve-of-factors" model. The application of the "curve-of-factors" model in this study revealed that the change in the "Run" factor was explained well by the Curve model in both data sets. Interestingly in the univariate analyses, the Curve model fit the data well for all three observed variables that were used as indicators of the "Run" factor (see Table 4.10). Thus, the change of the latent factor that was well explained by the Curve model explained well the three observed variables that were also well explained by the Curve model in the univariate analyses. The indicator variables for other hypothesized factors showed growth curves that were different from each other. For example, for "Power", the best fitting models for three indicator variables (JAR, DASH and SLJ) were the Linear, Curve and Cubic models. However, it is not conclusive if the different growth curves of indicator variables causes the poor fit for the "curve-of-factors" model, because the Cubic model should be able to take into account all the variance components that are based on the linear and the quadratic change. More research with different approaches on this relationship between the change in each indicator variable and the change in a latent factor is needed.

**STUDY 2. COMPARING THE LATENT GROWTH MODEL AND QUASI-SIMPLEX
MODEL IN THE ESTIMATION OF LONGITUDINAL RELIABILITY**

STUDY 2-CHAPTER II. LITERATURE REVIEW

Concepts of Reliability and Traditional Estimation Methods

Reliability is the extent to which a test or any measuring procedure yields the same results under the same conditions (Carmines & Zeller, 1979). It is sometimes represented as the consistency or reproducibility of measured scores. Ideally, a perfect measurement tool will produce the exact same scores for a group of individuals if it is repeatedly administered under identical conditions, assuming that there is no change in the subjects' true attribute. However, to a certain extent, all measurements that are taken from human subjects are unreliable (Crocker & Algina, 1986). In other words, it is almost impossible to perfectly measure an attribute even if there exists some true level of the attribute within a person.

An observed score resulting from the measurement of such an attribute includes two components, a true attribute component and an unreliable component. The unreliable component is called the measurement error. Based on classical test theory, the observed score is a composite of two components, a true (theoretical) score and error score. That is, $x = \tau + e$, where, x is the observed score; τ is the true score; and e is random error. Given the assumptions that the correlation between the true score and error score is zero and the mean of the error scores is zero, it can be shown that the variance of observed scores is the sum of the true and error score variances, $\sigma_x^2 = \sigma_\tau^2 + \sigma_e^2$, where σ_x^2 , σ_τ^2 , and σ_e^2 are the variance of the observed scores, true scores and error scores, respectively (Crocker & Algina, 1986). Given this, the reliability of variable x , ρ_x , is defined by the following equation:

$$\rho_x = \frac{\sigma_\tau^2}{\sigma_x^2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_e^2} \quad (2.2.1)$$

Thus reliability is represented as the amount of true score variance relative to the observed score variance. However, it is difficult to estimate the reliability since the true and error scores are unobservable elements.

The most frequently used reliability estimation method for a single measure (variable) is the test-retest method. The same test is administered to the same subjects twice, under the same conditions, within a certain time period, and the Pearson product-moment correlation (PPMC) coefficient between the two sets of scores is taken as an estimate of the test reliability. Since this coefficient is based on measurement at two time points, it is often called a stability coefficient. The idea behind this is that only true scores should be correlated between two time points because the error component is random and not correlated with any other elements. Thus, the correlated part is due to only the true score element. The important assumption that should be satisfied in using PPMC as a reliability estimate is that there is

no change in true scores between the two time points. This assumption has been questioned by many researchers (e.g., Heise, 1969; Marsh & Grayson, 1994a) because there is inevitable temporal instability of measures taken at multiple points in time. Another concern is the possibility of correlated errors between two time points (Wiley & Wiley, 1974). Many statistical models have been suggested to overcome these problems (e.g., Heise, 1969; Werts, Jöreskog & Linn, 1971; Wiley & Wiley, 1970, 1974). More recent techniques such as structural equation modeling provide methods to account for these problems analytically in certain situations.

A more general form of the PPMC is the intraclass correlation coefficient (intraclass r). The intraclass r is used when a single item is measured repeatedly, or several items are measured once, or several items are measured repeatedly (Schutz, 1998). Like all reliability coefficients, it is conceptualized as a ratio of true score variance to observed score variance, and in this case ANOVA is used to estimate various sources of variance (mean squares). Depending on what assumptions a researcher wishes to make about error variances and true score variances, different intraclass r s can be calculated. One of the earliest attempts to use the intraclass r for reliability estimation can be attributed to Hoyt (1941). He derived the equation using a "Persons by Items" ANOVA design, and related it to the classical reliability definition by noting that the mean square due to the persons (MS_p) represents the variance of observed scores, and the mean square residual (MS_{res} : the Persons \times Items interaction effect) represents the variance due to the error. The following two equations are most frequently used. The intraclass r for the mean test score over all trials or observations can be estimated as follows;

$$r = \frac{MS_p - MS_{res.}}{MS_p}, \quad (2.2.2)$$

and that for a single item score as;

$$r_i = \frac{MS_p - MS_{res.}}{MS_p + (k-1)MS_{res.}}, \quad (2.2.3)$$

where

$$E(MS_p) = k\sigma_p^2 + \sigma_e^2, \quad (2.2.4)$$

$$E(MS_{res.}) = E(MS_{\text{person} \times \text{item}}) = \sigma_e^2. \quad (2.2.5)$$

MS_p is the mean square due to persons, $MS_{res.}$ is the mean square due to error (the Persons \times Items interaction effect), k is the number of items or repeated measurements, σ_p^2 is population variance due to

between persons and σ^2_e is population variance due to error (Winer, 1971). The equation 2.2.3 yields identical results to the internal consistency reliability, Cronbach's alpha (Crocker & Algina, 1986; Schutz, 1998).

When several items are used to measure an attribute, the degree of agreement among these items is called internal consistency, and the most frequently used coefficient of internal consistency is Cronbach's alpha (Cronbach, 1951). Cronbach's alpha can be considered as an index of reliability of the composite score that is obtained by summing item scores. However, it is not equivalent to a reliability stability coefficient. A large alpha indicates that there is small item-specific variation. However, although it suggests a strong possibility that all items represent a single factor, it is not sufficient evidence to make such a conclusion. That is, a high alpha is a necessary, but not a sufficient condition for unidimensionality. One should also note that with large number of items, Cronbach's alpha could be very high, overestimating the degree of agreement among items (Cortina, 1993).

Estimation of Longitudinal Reliability

Traditional approaches of reliability estimation fall short when applied to longitudinal data because these approaches were developed with static variables in mind. Much of the rationale behind traditional approaches is based on the assumption of unchanging true scores, with any change in observed scores directly attributable to measurement error (Blok & Saris, 1983; Collins, 1991; Werts, Breland, Grandy, & Rock, 1980). The reliability of a measurement tool may also change over time, for several reasons; a change in the characteristics of the subjects (e.g., age), different measurement administrators, etc. As noted in earlier sections, using PPMC or intraclass r for the estimation of reliability requires the assumption that the true score does not change over time. Thus, in most situations one may not use PPMC or intraclass r directly for longitudinal data in the estimation of reliability (Blok & Saris, 1983; Werts et al., 1980). One simple solution for this is to measure the variable twice or more at each time point, and estimate the reliability. However, this is not a very practical solution.

A few statistical models have been suggested to overcome this problem analytically. One of the earliest works may be the path analytic solution that was suggested by Heise (1969). Based on the works of Wright (1934), Blalock (1963), and Siegel and Hodge (1968), he employed an autoregressive model in separating temporal instability of true scores from the measurement error. This basic idea has been extended and widely used by others (e.g. Jöreskog, 1970; Werts, Jöreskog & Linn, 1971; Wheaton, Muthen, Alwin & Summers, 1977; Wiley & Wiley, 1970). A general form of this autoregressive model is depicted in Figure 2.2.1 which shows a model with a variable (X) that is measured repeatedly at five time points. According to this model, the variable at each time point is explained by the immediately preceding variable (time point). This model is called a quasi-simplex model, and is a special case of an autoregressive model (Jöreskog, 1970). The observed score at time t (X_t) is a composite of two elements,

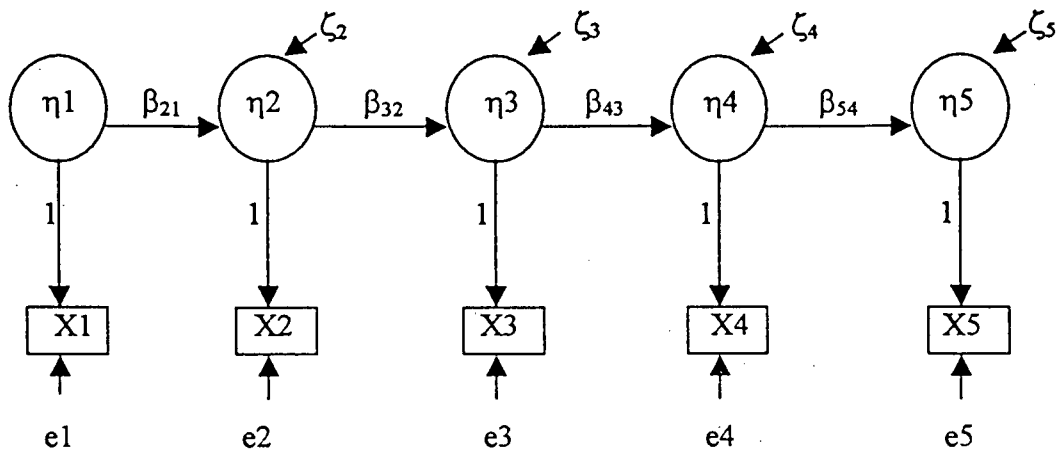


Figure 2.2.1. A quasi-simplex model with five time points

a true score (η_t) and an error score (e_t), as in classical test theory, that is; $X_t = \eta_t + e_t$. The successive η_t are related by the linear equation, $\eta_{t+1} = \beta_t \eta_t + \zeta_{t+1}$, and the reliability of variable X_t at time t is calculated as follows (for $t_s < t_t < t_u$);

$$r_{tt} = \frac{r_{st} r_{tu}}{r_{su}} \quad (2.2.6)$$

where r_{st} is the correlation between X_s and X_t , r_{tu} is the correlation between X_t and X_u , and r_{su} is the correlation between X_s and X_u . Thus, this model takes account of the change in true score over time by the regression coefficient (β , the stability coefficient), and the variance unexplained by this relationship among variables (θ_e) is due to the error.

It is obvious from equation 2.2.6 that the reliability coefficients of the first and the last time points cannot be obtained unless an additional restriction is imposed in the model. Heise (1969), in his three-wave (three time points) model, imposed a restriction of equal reliability over time. He suggested that although the variances of the true and the error scores may vary over time, the ratio between them can remain unchanged. Wiley and Wiley (1970) used a similar approach to reliability estimation, but based on the assumption that error variances are constant over time. Another possible restriction is to assume that the stability coefficient (β) is constant between adjacent waves (Kenny, 1979). Selecting one of these restrictions depends on the theory behind the variable of interest. When there are more than three time points, these restrictions may be relaxed to only parts of the model, involving the parameters of the first two and the last two time points (Jöreskog, 1970, 1974; Jöreskog & Sorbom, 1988). Although there are some limitations, this model has been widely used in the estimation of longitudinal reliability (e.g., Werts, Linn & Jöreskog, 1977, 1978; Morera, Johnson, Freels, Parsons, Crittenden, Flay & Warnecke, 1998).

Recently, McArdle and Epstein (1987), and Tisak and Tisak (1996) suggested another way to estimate reliability with longitudinal data. They employed a Latent Growth Model (LGM) approach, and showed how one can estimate the change parameters and reliability at the same time. The idea behind this approach is that any part of the observed variance that is not explained by the growth (change) parameters is due to error. In the two-factor model presented in Figure 2.2.2, the reliability of time t can be calculated using following equation;

$$r_t = \frac{\lambda_{ti}^2 \psi_i + \lambda_{ts}^2 \psi_s + 2\lambda_{ti}\lambda_{ts}\psi_{is}}{\lambda_{ti}^2 \psi_i + \lambda_{ts}^2 \psi_s + 2\lambda_{ti}\lambda_{ts}\psi_{is} + \theta_t} \quad (2.2.7)$$

where λ represents a factor loading, ψ represents a factor variance, θ represents error variance, t stands

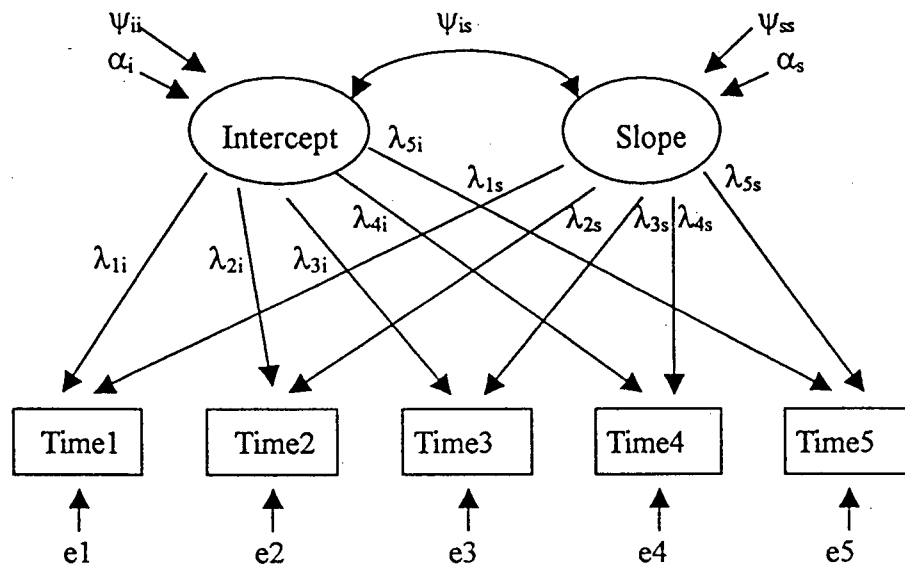


Figure 2.2.2. Two-factor LGM

for time t , i stands for an intercept factor, and s stands for a slope factor. Basically this equation has the same form as the definition of the reliability (equation 2.2.1). That is, the numerator represents the true score variance and the denominator represents the observed score variance. A notable aspect of this equation is that both true score variance and error variance may change over time.

This model has several merits in estimating longitudinal reliability. First, multiple measurements are not needed at each time point, as is required in the test-retest method. Second, measurements are decomposed into separate sets of parameters that represent reliability and the function of change. Third, parameters for both change and reliability are estimated at the same time. Fourth, the model permits reliability to change as a function of time. Fifth, this model is a generalization of test-retest reliability. Sixth, this model requires less strict statistical assumptions than classical methods (Tisak & Tisak, 1996). The first three merits are shared by quasi-simplex model, but last three are unique aspects of LGM approach. However, as noted by Tisak and Tisak (1996), one has to first determine an appropriate longitudinal model before interpreting estimated coefficients in the application of this approach. Because it is relatively new, the LGM approach has seldom been used in practice for the estimation of longitudinal reliability.

There have been some other suggestions regarding reliability estimation models in the situation where several items are measured over time to represent an attribute at several occasions (e.g., Blalock, 1970; Marsh & Grayson, 1994a; Raffalovich & Bohrnstedt, 1987; Wheaton et al., 1977; Wiley & Wiley, 1974). The models that Blalock (1970), Wheaton et al. (1977) and Wiley and Wiley (1974) employed were multivariate extensions of a quasi-simplex model. A very similar approach is using a confirmatory factor analysis (CFA) model. The CFA model has been widely used for reliability estimation of individual items within a scale. However, it has rarely been used for reliability estimation of multi-item-multi-occasion situations. Basically, these models examine how much variance among the observed item variance is due to the underlying latent trait (true score) at each time point. One notable merit of these models is that one can take account of possible correlated errors between repeatedly measured variables in the model (Blalock, 1970; Wheaton et al., 1977; Wiley & Wiley, 1974). The CFA model has been extended so that the model takes into account the sources of systematic variance due to specific items as well as specific times (Marsh & Grayson, 1994a; Raffalovich & Bohrnstedt, 1987). The form of this model is same as that of a multi-trait multi-method (MTMM) model. This model can be regarded as a variance decomposition model, decomposing the total variance (observed variance) into time-specific, item-specific and residual (error) variance (Marsh & Grayson, 1994a). Although Raffalovich and Bornstedt (1987) and others used a second-order factor model for extracting another component of the variance, the common factor variance, the existence of the second-order factor does not affect the reliability estimation of individual items.

Comparing the Latent Growth Model and Quasi-simplex Model

In the case of estimating longitudinal reliability of a single variable, both longitudinal models, a quasi-simplex model and LGM, may be used. Selecting one model over another on purely statistical criteria is not feasible because empirically, these two models are hard to distinguish (Rogosa & Willett, 1985a). The two models differ in that a quasi-simplex model defines changes over time to be independent of prior changes, while a LGM defines changes over time to be dependent upon prior changes (McArdle & Epstein, 1987). In general, there are considerable discrepancies in reliability estimates between these two models. Which of these two provide more accurate reliability estimates is not known.

There have been some studies in which these two models were compared. However, none of these studies focused on the accuracy of reliability estimation. Rogosa and Willett (1985a) showed that a quasi-simplex model fits well to data that were generated based on a growth model. However, they argued that automatic usage of a quasi-simplex model is not desirable because very different types of individual growth curves may yield indistinguishable covariance or correlation structures. They also found that the reliability was overestimated by the quasi-simplex model, and noted that if the partial correlation between any two time points after controlling for any intervening time point is not zero, as in a growth model, the reliability will be systematically overestimated.

On the contrary, Kenny and Campbell (1989) argued that a simplex model is superior to LGM in examining the stability of personality, for several reasons. First, a simplex model treats the random component as a lasting part of the true score while a LGM treats it as an unreliable part. Second, a LGM typically assumes that all scores of a person either steadily increase or steadily decrease over time, but the true score of a person rarely exhibits this pattern of change. Third, a LGM requires equivalent metrics at all time points while a simplex model does not. They also noted that one of the weaknesses with a quasi-simplex model has been the exclusion of means in the model, but this problem can easily be improved by including means in the model following Roskam's (1976) suggestion. However, their view was based mainly on the application of these statistical models to the examination of the stability of personality where the individual growth (i.e., directional change) is not a main interest. Bast and Reitsma (1997) supported this view in favour of a quasi-simplex model.

Mandys, Dolan and Molenaar (1994) made a more detailed comparison between a quasi-simplex model and LGM, and showed several differences between the two models in analyzing longitudinal data. Contrary to Rogosa and Willett's (1985a) findings, they showed that a quasi-simplex model does not fit data that are based on a growth curve when there are eight or more time points. They also showed that decreasing the variance of the errors and increasing the variance of the individual growth rates resulted in deterioration of the fit of the quasi-simplex model to the data. They concluded that one has to be careful in rejecting a quasi-simplex model in favour of LGM on the basis of a single analysis.

Although some discussions of reliability estimates have been made (e.g., Mandys et al., 1994;

Rogosa & Willett, 1985a), these studies focused more on the rationales, strengths and weaknesses of applying two models in the analysis of longitudinal data rather than the accuracy of reliability estimation. The capability of these two longitudinal models to accurately estimate longitudinal reliability needs to be examined.

STUDY 2-CHAPTER III. METHODOLOGY

The purpose of study 2 was to compare the latent growth model (LGM) and the simplex model in estimating longitudinal reliability under various conditions. Several longitudinal data sets representing various conditions were generated and analyzed by a LGM and a simplex model. The data generation was necessary because the true reliability of the data should be known to examine the accuracy of reliability that is estimated by the two models. The results were compared in terms of the accuracy of reliability estimation.

Data and Conditions

Several longitudinal data sets were generated with known parameters. As in practice, it was assumed that each individual subject has his/her own initial status and rate of change. This also means that there is considerable between-person variation in both initial status and change. However, for simplicity, it was assumed that each individual subject changes linearly over time. Thus the difference in true scores between any two adjacent time points was constant within a subject. The number of repeated measurements was fixed at five in all generated data sets, and the sample size was fixed at 5000 for all conditions. This relatively large sample size was used so that each generated data set would yield more accurate parameters (i.e., closely approximate true parameter values).

The means and the variances of both the initial status and change were based on the analysis results of the Jump-and-Reach (JAR) variable from the Michigan data (data set 1). The true mean and variance of the initial status were 9.426 and 2.057, respectively. The true mean and the variance of the linear change were .994 and .082, respectively. These values were used in all generated data sets. Other parameters, the magnitude of the correlation between the initial status and change and the magnitude of error variances at each time point, were varied depending on the conditions that are explained below.

The conditions of the data sets were varied based on three factors that may affect the estimation of the reliability. These three factors were; (a) the magnitude of correlations between the intercept (initial status) and change (growth), (b) the magnitude of true reliability, and (c) the magnitude of correlated errors between repeated measurements. Population data sets rather than samples were used in all analyses to isolate the effect of each condition on the estimation of the reliability from the sampling variation.

Condition A: The Magnitude of the Correlation Between the Intercept and Change (r_{ic})

Three different magnitudes of correlation between the initial status and the change were used in the generation of the data. These three correlations represent no relationship ($r_{ic} = 0$, condition A1), medium relationship ($r_{ic} = -.3$, condition A2) and relatively large relationship ($r_{ic} = -.6$, condition A3) between initial status and change. The reliability coefficients at each time point were fixed

at .65, .75, .75, .75, and .75 at time 1, 2, 3, 4 and 5, respectively. These magnitudes of reliability reflect the reliability of a physical performance field test, such as the JAR. The reliability of the measure at the first time point was fixed at a lower value (i.e., .65) than those of other time points to reflect a changing reliability in longitudinal measurements. Following the assumptions of classical test theory, the correlations between the true scores (initial status and change) and errors, and the correlations of errors between different time points, were fixed at zero. The procedures of the data generation are presented in a following section.

Condition B: The magnitude of reliability

Three different sets of magnitudes of reliability were used. These are relatively small (condition B1: .40, .50, .50, .50, and .50 at time 1, 2, 3, 4, and 5, respectively), medium (condition B2: .65, .75, .75, .75, and .75) and relatively large (condition B3: .90, .95, .95, .95, and .95) reliabilities. In general, these magnitudes reflect the reliability of a questionnaire, a physical performance field test and a laboratory test, respectively. The correlation between the initial status and change (r_{ic}) was fixed at -.3 in all three conditions to isolate the effect of the magnitude of reliability from the magnitude of correlation between the initial status and change. Other parameters regarding assumptions of classical test theory were same as in condition A2. Thus, condition A2 and condition B2 are identical (i.e., the same data set was used for these two conditions).

Condition C: The Correlation Between Errors (r_{ee})

Five different conditions were examined regarding correlated errors. These are no correlated errors (condition C1), relatively small correlated errors ($r_{ee} = .1$) between all time points (condition C2), relatively small correlated errors ($r_{ee} = .1$) between the last two time points only (condition C3), relatively large correlated errors ($r_{ee} = .3$) between all time points (condition C4), and relatively large correlated errors ($r_{ee} = .3$) between the last two time points only (condition C5). Thus in condition C2, correlations of errors between all time points were set at .1 while in condition C3, correlation of errors between only last two time points was set at .1. The purpose of employing conditions C3 and C5 (correlated errors between last two time points only) was to examine if the correlated errors between specific time points affect the estimation of reliability at other time points. Parameters regarding the initial status and change (means, variances and correlation between two) were the same as condition A2. As well, the true reliabilities at each time point were set at .65, .75, .75, .75 and .75, as in condition A2. Thus, condition C1 is identical to condition A2 (and B2). The conditions of the generated data are summarized in Table 2.3.1.

Data Generation Procedure

The data generation involved the following five steps; (a) generating initial status, linear change

Table 2.3.1

Summary of conditions of generated data

Condition	Initial status (I)		Linear change (C)		Correlation between I and C	Correlation between errors	Time1	Time2	Time3	Time4	Time5
	Mean	Variance	Mean	Variance							
Condition A: The magnitude of correlation between the intercept and change (r_{IC})											
A1	9.426	2.057	.994	.082	0	0	.65	.75	.75	.75	.75
A2	9.426	2.057	.994	.082	-.3	0	.65	.75	.75	.75	.75
A3	9.426	2.057	.994	.082	-.6	0	.65	.75	.75	.75	.75
Condition B: The magnitude of reliability											
B1	9.426	2.057	.994	.082	-.3	0	.40	.50	.50	.50	.50
B2	9.426	2.057	.994	.082	-.3	0	.65	.75	.75	.75	.75
B3	9.426	2.057	.994	.082	-.3	0	.90	.95	.95	.95	.95
Condition C: The correlation between errors ($r_{\epsilon\epsilon}$)											
C1	9.426	2.057	.994	.082	-.3	0	.65	.75	.75	.75	.75
C2	9.426	2.057	.994	.082	-.3	.1 (all)	.65	.75	.75	.75	.75
C3	9.426	2.057	.994	.082	-.3	.1 (last two)	.65	.75	.75	.75	.75
C4	9.426	2.057	.994	.082	-.3	.3 (all)	.65	.75	.75	.75	.75
C5	9.426	2.057	.994	.082	-.3	.3 (last two)	.65	.75	.75	.75	.75

Note. Bolded numbers are varied conditions.

and errors at each time point, (b) computing true scores at each time point, (c) changing variance of errors, (d) computing observed scores. The size of the data (sample size) was 5000. Although this size may not reflect what is used in most research projects, this magnitude of sample size was required to satisfy restrictions (conditions) on each data set.

Generating Initial Status, Linear Change and Errors

First, seven normally distributed variables with a mean of zero and a variance of 1.0 were generated using PRELIS (Jöreskog & Sorbom, 1999: Version 2.30) program. The correlations between these seven variables were varied by conditions. These seven variables are initial status, linear change and error1 to error5. Error1 to error5 denote the errors at time1 to time5, respectively. The initial status and linear change variables were then transformed so as to have the specified means and variances. The descriptive statistics of initial status, change and errors of condition A1 are presented in Table 2.3.2 as an example. Variances of errors are the values that were obtained after step (c). Although all the correlations between variables and means of errors were fixed at zero, the generated data showed values that are slightly different from zero. However, these correlations as well as means were very close to the specified values. The descriptive statistics of the data for other conditions are presented in Appendix D.

Computing True Scores at Each Time Point

In the next step, for each subject true scores at each time point were computed. This required a simple linear transformation of initial status and linear change. True scores at each time point were calculated using following equations. True1 to True5 denote the true score at time1 to time5, respectively.

$$\text{True1} = \text{initial status} + (0) \text{ linear change}$$

$$\text{True2} = \text{initial status} + (1) \text{ linear change}$$

$$\text{True3} = \text{initial status} + (2) \text{ linear change}$$

$$\text{True4} = \text{initial status} + (3) \text{ linear change}$$

$$\text{True5} = \text{initial status} + (4) \text{ linear change}$$

Changing the Variance of Errors

To obtain the specified magnitudes of reliability at each time point, the variances of error1 to error5 were transformed. To accomplish this, the variances of true scores at each time point were calculated first, and the variance of the errors were transformed accordingly. Note that this kind of linear transformation does not affect the correlations between variables.

Computing Observed Scores

Finally, the observed score at each time point was calculated by adding the two components,

Table 2.3.2

Descriptive statistics of true and error scores for condition A1 (An example)

	Initial status	Change	Error 1	Error 2	Error 3	Error 4	Error 5
Change	.016						
Error 1	.020	.009					
Error 2	.012	.010	- .017				
Error 3	.007	.010	- .009	.005			
Error 4	- .004	- .003	.001	- .011	.008		
Error 5	- .021	- .019	- .017	- .002	.006	- .001	
Mean	9.426	.994	-3.0E-06	-3.8E-05	-1.4E-17	2.8E-06	2.1E-18
SD	1.434	.286	1.052	.847	.897	.972	1.068
Variance	2.057	.082	1.107	.717	.804	.945	1.140

true score and error, at each time point. Thus, it is calculated as follows:

$$\text{Time1} = \text{true1} + \text{error1}$$

$$\text{Time2} = \text{true2} + \text{error2}$$

$$\text{Time3} = \text{true3} + \text{error3}$$

$$\text{Time4} = \text{true4} + \text{error4}$$

$$\text{Time5} = \text{true5} + \text{error5}$$

Model Fitting and Evaluation

Two longitudinal models, a linear LGM (Figure 1.2.1) and a simplex model (Figure 2.2.1), were fitted to each generated data set. The results were compared in terms of goodness-of-model fit, parameter estimates, and the accuracy of reliability estimates.

STUDY 2-CHAPTER IV. RESULTS

In the following sections the results of using a latent growth model and two simplex models to estimate longitudinal reliability are presented and compared. As explained in Study 2-Chapter III, computer simulated data sets were used for this component of the dissertation.

The Effect of Correlation Between Initial Status and Linear Change

Goodness-of-fit indices and estimated reliability coefficients of three longitudinal models under the various magnitudes of correlations between initial status (scores at time 1) and linear change are presented in Table 2.4.1. The term "Linear" indicates the two-factor linear latent growth model (Figure 2.1), "Simplex 1" indicates a quasi-simplex model with equal error variances for all five time points, and "Simplex 2" indicates a quasi-simplex model with equal error variances between the first two time points and between the last two time points (Figure 2.6).

The Linear model fit the data very well in all three conditions, while the Simplex models showed some conflicting goodness-of-fit results. In terms of χ^2 statistics, all Simplex models should be rejected, however, SRMR and NNFI indicated that the Simplex models fit the data very well. The large χ^2 value was due to the large sample size of the analyzed data ($N = 5000$). For example, if the sample size were 200, the χ^2 for the Simplex 1 model in condition A1 in Table 2.4.1 ($r_{ic} = 0$) would be 3.43, and all other fit indices would show a better fit. All RMSEA values indicated a good ($< .06$) or an acceptable ($< .08$) model fit, except for the Simplex 2 model of condition A3 where the correlation between the initial status and linear change is $-.6$ (see Table 2.3.1). Overall, the model fit of the Linear model was much better than that of the Simplex models in all conditions.

Reliability coefficients estimated by the Linear model were very accurate in all three conditions. The average discrepancies were .0014 (.2%), .0028 (.4%) and .0032 (.5%) for conditions of $r_{ic} = 0$ (condition A1), $r_{ic} = -.3$ (condition A2), and $r_{ic} = -.6$ (condition A3), respectively. The largest discrepancy was .009 (1.2%) and most of the estimates showed discrepancies smaller than .003 (.4%). Contrary to the Linear model, Simplex models overestimated the reliability at all time points in all three conditions. The largest overestimation was associated with the first time point where the true reliability is .65. This was due to the model constraints that force the error variances to be equal between time points (for the purpose of identification). The magnitude of overestimation ranged from .013 (1.7%) to .241 (37.1%) for the Simplex 1 model and from .026 (3.5%) to .210 (32.3%) for the Simplex 2 model.

The parameter estimates of the Linear model for condition A1 are presented in Table 2.4.2. In general, parameter estimates of the Linear model were very accurate, and there was no tendency of overestimation or underestimation where there is a discrepancy between the true and estimated parameter. There was no discrepancy between the estimated and true factor means up to three decimal

Table 2.4.1

Fit indices and estimated reliability coefficients of models with various correlations between initial status and linear change (r_{ic})

Condition						Estimated Reliability				
Model	$\chi^2(df)$	p-value	RMSEA	SRMR	NNFI	T1	T2	T3	T4	T5
True reliability →						.650	.750	.750	.750	.750
Condition A1: $r_{ic} = 0$										
Linear	2.39(10)	.992	< .001	.005	1.00	.651	.751	.752	.748	.749
Simplex 1	86.26(5)	< .001	.057	.011	.99	.785	.763	.787	.814	.844
Simplex 2	40.66(3)	< .001	.050	.008	.99	.816	.795	.797	.776	.811
Condition A2: $r_{ic} = -.3$										
Linear	3.30(10)	.974	< .001	.005	1.00	.648	.759	.750	.751	.752
Simplex 1	83.58(5)	< .001	.055	.013	.99	.845	.803	.806	.821	.843
Simplex 2	74.45(3)	< .001	.068	.012	.99	.855	.816	.814	.802	.826
Condition A3: $r_{ic} = -.6$										
Linear	4.38(10)	.929	< .001	.005	1.00	.652	.746	.748	.755	.753
Simplex 1	139.6(5)	< .001	.072	.017	.98	.891	.842	.815	.803	.818
Simplex 2	109.8(3)	< .001	.083	.016	.99	.860	.799	.824	.825	.839

Note. df = degrees of freedom; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; NNFI = non-normed fit index; T1 to T5 = Time1 to Time5.

places for both the intercept and the linear factor. The variance of the intercept factor was slightly overestimated (.051, 2.5%), and the variance of slope factor was slightly underestimated (.004, 4.9%). Error variances of time 1, 2 and 4 were overestimated and the error variances of time 3 and 5 were underestimated, but the magnitudes of overestimation or underestimation were small. The average discrepancy between the true and estimated error variances was .011 (1.1%). The parameter estimates of the Linear model for conditions A2 and A3 were also relatively accurate. The parameter estimates for conditions A2 and A3 are presented in Appendix D, Table D.3 and Table D.6.

The parameter estimates of the Simplex models showed similar results in all conditions. The parameter estimates of only the Simplex 2 model for condition A1 are presented in Figure 2.4.1 as an example. Because the data were generated based on the growth of a certain attribute over time, the true parameters (i.e., path coefficients, factor mean and factor variances) for the Simplex models are not available except for error variances. The true error variances of observed variances are available, and presented in bolded numbers. The error variances were underestimated at all time points, and the magnitude of underestimation was relatively large. The average underestimation was .236 (23.6%), and it is largest at the first time point (.512, 46.3%). This implies that the true score variances were overestimated by the model, and resulted in the overestimation for the reliability coefficients. The standardized path coefficients (β) were relatively high, indicating that there was a year-to-year stability of relative positions of subjects (cases) in their true scores. The path coefficient predicting time 2 factor from time 1 factor was smaller than other path coefficients due to the low reliability (.65) of the first time point. The mean of the time 1 factor was identical to the mean of the observed variable at the first time point. The factor mean of each time point is calculated as follows;

$$\text{Time 1} = 9.43$$

$$\text{Time 2} = (9.43 \times .80) + 2.91 = 10.45$$

$$\text{Time 3} = (\text{mean of time 2} \times .99) + 1.09 = 11.44$$

$$\text{Time 4} = (\text{mean of time 3} \times 1.01) + .86 = 12.41$$

$$\text{Time 5} = (\text{mean of time 4} \times 1.04) + .54 = 13.45.$$

These means are slightly different from the means of X1 to X5 (Appendix D, Table D.1) due to the estimation error. The parameter estimates of the Simplex models for other conditions are presented in the Appendix D, Table D.4 to Table D.7.

The Effect of the Magnitude of Reliability

The goodness-of-fit indices and estimated reliability coefficients of three models under the various magnitudes of reliability coefficients are presented in Table 2.4.3. Although the medium-level

Table 2.4.2

The true and estimated parameters (standard errors) of the Linear model for condition A1

	Intercept	Linear	Error variances				
	Factor	Factor	Time 1	Time 2	Time 3	Time 4	Time 5
<u>Mean</u>	9.426	.994	1.107	.717	.804	.945	1.140
	9.426	.994	1.130	.724	.795	.948	1.126
	(.023)	(.006)	(.033)	(.020)	(.020)	(.025)	(.036)
<u>Variance</u>	2.057	.082					
	2.108	.078					
	(.055)	(.004)					
<u>Covariance</u>	0						
	-.001						
	(.011)						

Note. Bolded numbers are true values. All parameter estimates were significant at $p < .001$ except for the covariance between two factors that was not significant ($p = .899$).

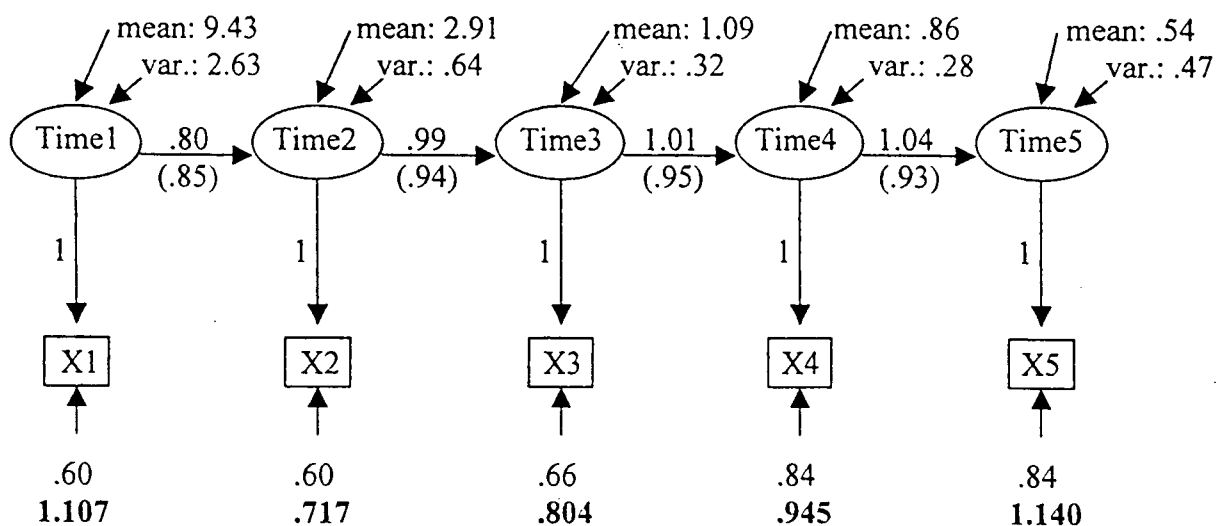


Figure 2.4.1. Parameter estimates of the Simplex 2 model for condition A1

Note. The numbers in brackets are standardized path coefficients. Bolded numbers are true error variances. All parameter estimates were significant at $\alpha = .05$.

Table 2.4.3

Fit indices and estimated reliability coefficients of models with various magnitudes of reliability

Condition						Estimated Reliability				
Model	$\chi^2(df)$	p-value	RMSEA	SRMR	NNFI	T1	T2	T3	T4	T5
Condition B1: Rel.=.40~.50			True reliability →			.400	.500	.500	.500	.500
Linear	3.81(10)	.955	< .001	.007	1.00	.408	.511	.499	.499	.504
Simplex 1	16.52(5)	.006	.021	.009	1.00	.666	.542	.547	.578	.633
Simplex 2	9.861(3)	.020	.021	.007	1.00	.683	.561	.556	.550	.604
Condition B2: Rel.=.65~.75			True reliability →			.650	.750	.750	.750	.750
Linear	3.30(10)	.974	< .001	.005	1.00	.648	.759	.750	.751	.752
Simplex 1	83.58(5)	< .001	.055	.013	.99	.845	.803	.806	.821	.843
Simplex 2	74.45(3)	< .001	.068	.012	.99	.855	.816	.814	.802	.826
Condition B3: Rel.=.90~.95			True reliability →			.900	.950	.950	.950	.950
Linear	4.15(10)	.940	< .001	.002	1.00	.897	.953	.950	.951	.950
Simplex 1	1103(5)	< .001	.199	.028	.95	.995	.994	.994	.994	.995
Simplex 2	1091(3)	< .001	.259	.028	.91	.997	.997	.997	.988	.989

Note. df = degrees of freedom; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; NNFI = non-normed fit index; T1 to T5 = Time1 to Time5; Rel. = reliability.

reliability condition (condition B2) is identical to condition A2 in Table 2.4.1, the results are presented again for comparison purposes.

The Linear model fit the data very well under all three conditions, B1, B2, and B3. The χ^2 statistics were lower than the degrees of freedoms, and all other fit indices indicated that the Linear model fit the data very well under all conditions. Simplex models showed interesting results regarding model fit. Under condition B1 where the reliability is relatively low (.40 to .50), the fit indices indicated that the Simplex models fit the data very well (e.g., RMSEA = .021). In condition B2 where the magnitude of reliability is medium (.65 to .75), the model fit was worse than condition B1, but within an acceptable range. However, in condition B3 where the reliability is relatively high (.90 to .95), the Simplex models did not fit the data well. Although SRMR and NNFI were within an acceptable range, χ^2 and RMSEA indicated that the Simplex models should be rejected in condition B3. Certainly the Simplex models fit the data well when the reliability is low, but as reliability becomes larger the model fit of Simplex models becomes worse. Overall, the Linear model showed much better model fit compared to the Simplex models in all conditions.

The reliability coefficients estimated by the Linear model were very accurate in all conditions. The largest discrepancy between the estimated and the true value was .011 (2.2%). However, Simplex models, regardless of the magnitude of the true reliability, overestimated reliability. The overestimation ranged from .042 (8.4%) to .266 (66.5%), and the largest overestimation within a model was associated with the first time point where the true reliability is lowest among time points. In general, other parameter estimates of the Linear model were relatively accurate. The parameter estimates for the Simplex models showed similar results with those of the Simplex 2 model in condition A1 (Figure 2.4.1). These parameter estimates for the Linear and the Simplex models are presented in Appendix D, Table D.9 to D.13.

The Effect of Correlated Errors (r_{ee})

The goodness-of fit indices and estimated reliability coefficients of three models under various magnitudes of correlations among errors are shown in Table 2.4.4. Although condition C1 in which $r_{ee} = 0$, is identical to condition A2 of Table 2.4.1, it is represented for comparison purposes.

The Linear model fit the data very well in all conditions except for condition C5 where the errors of only the last two time points are correlated with a magnitude of .3. The χ^2 statistic of this model was obviously much larger as compared to those of the Linear model in other conditions. However, in terms of other fit indices, this model also fit the data very well. Simplex models showed similar patterns in the model fit with conditions A and B. The χ^2 statistic was not satisfactory, but other indices indicated that these models could be considered acceptable, except for the Simplex 2 model in condition C4 (RMSEA > .08) where the magnitude of correlations among errors are .3 between all time

Table 2.4.4

Fit indices and estimated reliability coefficients of models with various magnitudes of correlated errors (r_{ee})

Condition						Estimated Reliability				
Model	χ^2	p-value	RMSEA	SRMR	NNFI	T1	T2	T3	T4	T5
True reliability \rightarrow						.650	.750	.750	.750	.750
Condition C1: $r_{ee}=0$										
Linear	3.30(10)	.974	< .001	.005	1.00	.648	.759	.750	.751	.752
Simplex 1	83.58(5)	< .001	.055	.013	.99	.845	.803	.806	.821	.843
Simplex 2	74.45(3)	< .001	.068	.012	.99	.855	.816	.814	.802	.826
Condition C2: $r_{ee}=.1$ between all time points										
Linear	7.39(10)	.688	< .001	.010	1.00	.675	.771	.775	.772	.776
Simplex 1	113.3(5)	< .001	.064	.013	.99	.856	.816	.821	.832	.855
Simplex 2	100.0(3)	< .001	.080	.013	.98	.854	.813	.840	.813	.838
Condition C3: $r_{ee}=.1$ between last two time points										
Linear	5.34(10)	.868	< .001	.005	1.00	.660	.753	.749	.773	.777
Simplex 1	75.46(5)	< .001	.053	.012	.99	.850	.814	.815	.830	.854
Simplex 2	75.39(3)	< .001	.069	.012	.99	.851	.816	.814	.830	.854
Condition C4: $r_{ee}=.3$ between all time points										
Linear	10.64(10)	.386	.004	.011	1.00	.738	.825	.828	.823	.831
Simplex 1	131.7(5)	< .001	.069	.012	.99	.897	.871	.870	.879	.895
Simplex 2	118.1(3)	< .001	.087	.012	.98	.894	.868	.885	.865	.882
Condition C5: $r_{ee}=.3$ between last two time points										
Linear	78.26(10)	< .001	.038	.021	1.00	.656	.746	.734	.802	.832
Simplex 1	123.5(5)	< .001	.068	.014	.99	.864	.823	.823	.838	.860
Simplex 2	81.62(3)	< .001	.072	.012	.99	.841	.794	.805	.875	.893

Note. df = degrees of freedom; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; NNFI = non-normed fit index; T1 to T5 = Time1 to Time5.

points.

The Linear model overestimated reliability coefficients when correlated errors were present (i.e., $r_{ee} > 0$). In condition C2 where errors between all time points are correlated with a magnitude of .1, the reliability coefficients were overestimated in all time points. The average overestimation was .024 (3.26%). In condition C3 where only the errors between the last two time points are correlated with a magnitude of .1, the reliability coefficients of only the last two time points were overestimated. The average overestimation of these two reliability coefficients was .025 (3.3%). In conditions C4 and C5 where the magnitude of correlated errors is .3, the magnitude of overestimations for the reliability coefficients was larger than that of conditions C2 and C3. The average overestimations were .079 (10.9%) and .067 (8.9%) in conditions C4 and C5, respectively. In condition C5 where the errors of between last two time points are correlated with the magnitude of .3, the reliability coefficients for these last two time points were overestimated. Reliability coefficients for other time points (other than last two time points) in conditions C3 and C5 were either slightly overestimated or underestimated. The largest discrepancy was .016 (2.1%) at time 3 in condition C5.

The Simplex models overestimated reliability coefficients at all time points under all conditions. The overestimations were not limited in the last two time points in conditions C3 and C5 where errors between only the last two time points are correlated. The magnitude of overestimation was larger than that of the Linear model. The average overestimation was .117 (16.5%) across all conditions and two Simplex models. The largest overestimation was associated with the first time point where the true reliability is .65. The average overestimation of the reliability for the first time point across all conditions and across two Simplex models was .213 (32.8%), while the average overestimation of all other time points was .093 (12.4%). In addition, the overestimation was larger in condition C4 than that of other conditions. The average overestimation across all time points and across two Simplex models for condition C4 was .151 (21.1%), while the average overestimation for other conditions was .106 (15.0%). Overall, the Simplex models overestimated reliability at all time points regardless of the condition, and the magnitude of the overestimation was much larger than that of the Linear model.

The parameter estimates of the Linear model are presented in Table 2.4.5. Because the magnitude of reliability coefficients is a function of the factor variances and error variances, parameter estimates of variances only are presented. As shown in Table 2.4.5, the overestimation of reliability coefficients for the Linear model was due to both the underestimation of the error variances and the overestimation of the factor variances. For condition C2 and C4 where the errors between all time points were correlated, the error variances at all time points were underestimated, and the magnitude of the underestimation was larger in condition C4 ($r_{ee} = .3$ between all time points) than in condition C2 ($r_{ee} = .1$ between all time points). The average underestimations were .071 (9.4%) and .222 (29.4%) for condition C2 and C4, respectively. This indicated that a larger correlation between errors resulted in a larger underestimation of error variances. This was also evident in conditions C3 and C5 where the

Table 2.4.5

The true and estimated variances of Linear model for condition Cs

Condition	Factor variance		Error variance				
	Intercept	Linear	Time 1	Time 2	Time 3	Time 4	Time 5
C1: $r_{ee} = 0$	2.057	.082	1.107	.717	.804	.945	1.140
	2.108	.078	1.130	.724	.795	.948	1.126
C2: $r_{ee} = .1$ between all time points	2.057	.082	1.107	.627	.624	.675	.781
	2.096	.082	1.010	.574	.561	.617	.697
C3: $r_{ee} = .1$ between last two time points	2.057	.082	1.107	.630	.630	.684	.793
	2.077	.093	1.070	.626	.644	.623	.720
C4: $r_{ee} = .3$ between all time points	2.057	.082	1.107	.626	.622	.672	.777
	2.287	.082	.812	.448	.443	.481	.519
C5: $r_{ee} = .3$ between last two time points	2.057	.082	1.107	.629	.626	.679	.786
	2.070	.108	1.084	.637	.682	.522	.516

Note. Bolded numbers are true values. Standard errors are omitted. All parameter estimates were significant at $p < .001$.

errors of only the last two time points were correlated. For condition C3 and C5, the error variances at the last two time points were underestimated, and the magnitude of the underestimation was also larger in condition C5 ($r_{ee} = .3$ between last two time points) than in condition C3 ($r_{ee} = .1$ between last two time points). The average underestimations of last two time points were .067 (9.1%) and .214 (28.8%) for condition C3 and C5, respectively. Some of the factor variances were overestimated (this also resulted in an overestimated reliability). The variance of the intercept factor was overestimated in all conditions, but the magnitude of overestimation was relatively small except for condition C4 (.23, 11.2%). The variance of the linear factor was overestimated in condition C3 (.011, 13.4%) and C5 (.026, 31.7%), and the magnitude of the overestimation was relatively large. The estimated variance of the linear factor in condition C2 and C4 were accurate (no discrepancy was found up to 3 decimal places). Thus, especially under conditions C3 and C5, where the errors of only the last two time points were correlated, the variance of the linear factor was overestimated. The parameter estimates of the Simplex models showed similar results with those of Simplex 2 model in condition A1 (Figure 2.4.1). These results are presented in Appendix D, Table D.16 to Table D.25.

STUDY 2-CHAPTER V. DISCUSSION

The LGM and the Simplex models were compared in the estimation of longitudinal reliability. Longitudinal data sets with known parameters and reliabilities were generated, through a computer simulation, based on several stipulated conditions, and used in the examination of two models. Conditions were varied by the magnitude of correlation between the initial status and the rate of change, the magnitude of reliability, and the magnitude of correlated errors.

The goodness-of-fit indices indicated that the LGM fit the data very well in all conditions while Simplex models showed a questionable model fit. In general, the goodness-of-fit of the Simplex models was worse than that of LGM, and the χ^2 statistics of the Simplex models were very large in most conditions. The RMSEA, SRMR, ECVI and NNFI also indicated a worse fit for the Simplex models than the LGM. This was expected, as the data sets were generated based on the linear growth of individuals over time. These results partially agree with the conclusion by Mandys, Dolan and Molenaar (1994). They found poor model fits with the Simplex models on growth data with eight or more time points.

Although the model fit of the Simplex models was worse than that of LGM, the model still showed a fairly good fit to the data. The RMSEAs were within an acceptable range in many conditions, and the SRMR and NNFI indicated an excellent model fit for all conditions. Thus, in practice, one may conclude that the Simplex model fit the growth data well. These results agree with the findings by Rogosa and Willet (1985). They argued that this is a problem because the data from a growth model violate the assumption of a Simplex model that the change between any two time points is not affected by the change between previous time points. As they concluded, a caution is needed when employing a Simplex model for the analysis of longitudinal data, especially where a change over time is expected. In terms of the χ^2 and RMSEA statistics, however, the results supported Mandys et al.'s (1994) findings. As noted above, they found that the model fit of the Simplex models on growth data start to deteriorate when there are eight or more time points. In the present study, the χ^2 statistics and RMSEA showed that the deterioration was partially evident with five time points as well.

The reliability coefficients estimated by LGM (Linear model) were very accurate except in conditions where there existed correlated errors between time points. The largest discrepancy between the estimated and true reliability was 2.2%, excluding the conditions with correlated errors. Thus, when the errors were not correlated, LGM accurately decomposed the observed variance into the two components that are due to error and true change. However, the Simplex models overestimated reliability in all the conditions. The magnitude of the overestimation ranged from 1.7% to 66.5%, depending on the time point and conditions. The overestimation of reliability by a Simplex model was observed and discussed by Rogosa and Willet (1985). They argued that in growth data, the partial

correlation between any two time points after controlling for any intervening time point is not zero, thus the reliability estimation by a Simplex model is overestimated. This implies that growth data violate the assumption that is required in a Simplex model. Thus, when one expects growth in a performance variable over time, LGM provides a more accurate reliability estimation. On the other hand, Kenny and Campbell (1989) contend that a simplex model treats the random component of a measure as a lasting part of the true score while a LGM treats it as an unreliable part. However, Kenney and Campbell's view was under the circumstances where one is interested more in the stability of a measure over time rather than in the change.

The requirement of constraints that should be imposed for the purpose of identification is one of major weaknesses of Simplex models in the estimation of reliability. The constraints that were imposed to the Simplex model in the present study were the equality of error variances between the first two and between the last two time points, or across all five time points. This resulted in a larger overestimation of the reliability at the first time point, where the true reliability was lower than other time points. Because of the equality constraints, the magnitude of estimated error variance at the first time point was forced to be equal to that of other time points, although the true error variance is larger than that of other time points. This means that using Simplex models, one may not adequately take into account the nature of longitudinal data, in which the true and error variances (and hence the reliability) may change over time. The constraints of equal error variance over time that were suggested and used by Joreskog (1970) and Wiley and Wiley (1970), are difficult to justify in many longitudinal studies. Other types of constraints have been also used in the literatures such as equal reliability over time (Heise, 1969) and equal stability over time (Kenny, 1979). However, as with the constraints of equal error variances over time, these types of equality constraints of Simplex models are rarely justified in most of longitudinal studies.

The magnitude of the correlation between the initial status and the rate of linear change (slope) did not affect the estimation of longitudinal reliability. The estimated reliability coefficients by LGM were accurate under all conditions. The reliability coefficients were overestimated by the Simplex models, but the magnitude of the overestimation was not systematically affected by the magnitude of the correlation between the initial status and the rate of the linear change. However, it is not conclusive that these results can be generalized to correlation between change factors (e.g., between the linear and quadratic factors in a Quadratic model). The present study employed only a linear change, thus examined only the effect of correlation between the initial status and the rate of linear change. In a quadratic or higher order models, the correlation between the change factors may affect the estimation of change parameters and hence, the reliability.

The magnitude of the true reliability did not show any systematic effect on the estimation of reliability. The LGM accurately estimated reliability and the Simplex models overestimated the reliability under all the conditions with various magnitudes of reliability. However, the magnitude of

reliability showed an effect on the goodness-of-fit of the Simplex models. When the magnitude of the true reliability was relatively low (.40 to .50), the Simplex models fit the data very well. As the magnitude of true reliability became larger the goodness-of-fit of the Simplex model became worse. Conditions with relatively high reliability produced an unacceptable model fit for the Simplex models. On the contrary, LGM fit the data well regardless of the magnitude of the true reliability. These results imply that as the magnitude of true reliability becomes smaller, there is a higher chance of accepting a Simplex model as a good fitting model in the analysis of longitudinal data. This eventually may lead one to make an erroneous conclusion regarding the reliability of a measure because the Simplex models provide overestimated reliability for the growth data. As Rogosa and Willett (1985) noted, selecting one model over another between a LGM and a Simplex model is not feasible because, empirically, these two models are difficult to distinguish. The results of the present study supported this view, especially where the true reliability is relatively low. Because the magnitudes of reliability coefficients that were employed in this study are common in psychological measures, one should be cautious when a low reliability is expected.

As expected, the LGM overestimated reliability in the presence of correlated errors. When the errors were correlated between the last two time points only, the reliability estimation of other time points were not affected. The magnitude of overestimation was dependent on the magnitude of the correlation between errors. In the conditions where the magnitude of correlation between errors was .10, the average magnitude of overestimation was 3.3%, and in the condition where the magnitude of correlation between errors was .30 the average overestimation was 10.3%. Thus, when there exist correlated errors and one fails to take it into account in the model, a LGM provides overestimated reliability coefficients. In addition, the magnitude of overestimation was dependent on the magnitude of correlation between errors, resulting in larger overestimation with larger correlation between errors. Further analyses revealed that both the overestimation in the factor variance (true score variance) and the underestimation in the error variance resulted in the overestimation of the reliability. The model treats the component of correlated errors as a lasting true score component. Thus, the variances of the change factors were overestimated, and hence the reliability was overestimated.

These results agreed with the notes by Werts, Breland, Grandy and Rock (1980), and Wiley and Wiley (1974). Although Wiley and Wiley (1974) explained this in the situation of obtaining the true correlation between variables, they showed that the magnitude of overestimation is directly proportional to the magnitude of correlation between errors. There have been other studies in which correlated errors were used in a longitudinal model (e.g., Blalock, 1970; Marsh & Grayson, 1994; Wheaton, Muthen, Alwin & Summers, 1977), but most of these studies used a multivariate longitudinal model or did not focus on the reliability estimation. When one anticipates that errors are correlated between time points, one should include the correlated errors in the model to obtain accurate parameter estimates. However, including correlations between all possible pairs of time points in a univariate LGM is not possible

because of the identification problem. Thus, one has to limit the number of correlations between errors in a model, depending on the available degrees of freedom of the model. In many cases, it is difficult to justify the inclusion of correlated errors between specific time points. This should be done only when there is a strong theoretical or empirical background that supports the inclusion of correlated errors. The Simplex models overestimated reliability regardless of the correlated errors.

CHAPTER VI. SUMMARY AND CONCLUSIONS

Summary

The present study is presented as two components. In study 1, (a) the latent growth model (LGM) was introduced, (b) the merits and the problems of using LGM were examined, and (c) the development of children's physical performances was examined. These phases of the investigation were accomplished by analyzing a longitudinal data set which includes seven physical performance variables that were measured at five time points, and five predictor variables. In study 2, the validity of the two widely used longitudinal factor analysis models, the LGM and the quasi-simplex model, were compared in estimating longitudinal reliability. For this purpose, data sets with known parameters (e.g., reliability) under various conditions were computer simulated and analyzed. The conditions of the data sets were varied in terms of the magnitude of correlations between initial status and change, the magnitude of reliability, and the magnitude of correlated errors between time points.

In study 1, the univariate LGM analyses revealed that the children's individual development over a 5-year period was adequately explained by variable specific trends. Specifically, the Linear growth model provided a good fit for the jump-and-reach and sit-and-reach, Quadratic for flexed-arm hang, Cubic for standing long jump, and Unspecified Curve models for agility shuttle run, endurance shuttle run and 30-yard dash. The children improved in their physical performances between ages 8 and 12 except for flexibility, in which children's performance declined over time. Among the predictor variables, test practice (the number of previous testing sessions) and age in months showed positive effects on the children's performance at the initial time point. A negative test practice effect on development in physical performances was also found. The effect of other predictor variables varied for different performance variables.

The multivariate analyses showed that the factor structure of three hypothesized factors, "Run", "Power" and "Motor Ability", holds at all five time points. However, only the change in the "Run" factor was adequately explained by any of the latent growth models, with the Unspecified Curve model providing the best fit. There were significant test practice, age, measurement season and measurement year effects on the intercept factor, and significant test practice and measurement year effects on the curve factor. The cross-validation procedure generally supported these findings.

In study 2, the results showed that the simplex model overestimated the reliability in all conditions, while the LGM provided relatively accurate reliability estimates in almost all conditions. The magnitude of correlation between the initial status and change, and the magnitude of reliability did not affect the reliability estimation, while the correlated errors lead to an overestimation of reliability for both models. On the other hand, the magnitude of reliability showed a negative effect on the goodness-of-fit of the simplex model.

Conclusions

Some conclusive statements can be made on the basis of study 1, as follows;

1. Latent growth modelling is a very useful and informative statistical procedure for the analysis of longitudinal physical performance data. Specifically, following conclusions are made with respect to the merits of LGM.

- (a) The capability of modelling change at the individual level is one of the most notable merits of LGM. This further enables one to include the predictors of change in a model, and to estimate the relationship between the initial status and change.
- (b) LGM takes into account the error component of variables in the analysis, and thus it represents the true developmental change in an attribute. In addition, LGM allows one to examine a hypothesis regarding error variances (e.g., equality of error variances over time).
- (c) LGM is a useful statistical model for the analysis of change in a multivariate latent factor.

2. The application of LGM to physical performance longitudinal data produced several unique findings regarding the children's development in physical performances that were not available in previous studies. The conclusions that were drawn from these findings are;

- (a) Individual children show approximately quadratic developmental patterns in upper arm and shoulder girdle muscular strength and endurance, leg muscular endurance, running speed, and agility.
- (b) There are considerable inter-individual variations in the linear, quadratic and cubic components of children's developmental change in physical performances. For some physical performance variables (e.g., the flexed-arm hang and standing long jump in the present study) the positive and negative quadratic and cubic components of individual children's development cancel each other out and produce an approximately linear group level of development (as also indicated by ANOVA results for flexed-arm hang), while the true developmental pattern of individual children is quadratic or cubic. The conclusions regarding children's developmental patterns in physical performances from previous studies in which traditional methods and group statistics were used need to be reexamined.
- (c) The relationship between the level of physical performance at the initial time of testing and the rate of development is not always negative, but depends on a specific performance as well as a selected time interval.
- (d) The "Run" factor which is characterized by a particular type of movement, was the only valid multivariate factor in representing the longitudinal development of latent physical performance.

Other conclusions include;

- (e) Test practice and age in months have positive effects on the physical performances.
- (f) Such construct as "general motor ability" does not exist even for young children. Latent physical performance variables are specific to a particular type of movement or a particular muscle group.

3. The practical problems of using LGM in the analysis of longitudinal physical performances need to be attended. Specifically;

- (a) Choosing the best fitting latent growth model based solely on statistical criteria is not always straight forward (e.g., comparing between the Unspecified Curve model and the Quadratic or Cubic model). In such a case, researchers should make decision based on a conceptual and a theoretical basis of physical performance development.
- (b) The complex relationship between performance variables, and between time points may result in the case where none of the multivariate LGMs (e.g., Linear, Quadratic, Cubic or Unspecified Curve models) fits the data in the curve-of-factors model, while all indicator variables in the model separately fit one of the LGMs well.

The main conclusion from study 2 is;

- 1. The LGM accurately estimates reliability, while the quasi-simplex model overestimates the reliability of longitudinal developmental variables. The availability of this valid statistical model for the estimation of longitudinal reliability is beneficial especially in Human Kinetics research, since the measurement of physical performance variables is often costly.

Some other conclusive statements from study 2 as well can be made, as follows;

- 2. The reliability estimations by the LGM and quasi-simplex models were not affected by the correlation between the initial status and the rate of change, or by magnitude of reliability.
- 3. The correlated errors result in an overestimation of reliability, and the overestimation is isolated at correlated time points.
- 4. The magnitude of reliability of developmental variables has a negative effect on the goodness of model fit of the quasi-simplex model.

References

- AAHPERD. (1988). Physical best test manual. Reston, VA: American Alliance for Health, Physical Education, Recreation and Dance.
- AAHPER. (1976). AAHPER youth fitness test manual. Washington, DC: American Alliance for Health, Physical Education, Recreation and Dance.
- American Health and Fitness Foundation. (1986). Fit youth today test program manual. Austin, TX: Author.
- Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. In S. Leinhardt (Ed.), Sociological methodology (pp.61-98). San Francisco, CA: Jossey-Bass.
- Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), Longitudinal research in the study of behavior and development (pp.1-39). New York, NY: Academic Press.
- Barlow, D. A. (1970). Relation between power and selected variables in the vertical jump. In J. M. Cooper (Ed.), Selected topics on biomechanics (pp. 233-241). Chicago, IL: Athletic Institute.
- Barry, J., & Cureton, T. K. (1961). Factorial analysis of physique and performance in prepubescent boys. Research Quarterly, *32*, 283-300.
- Bast, J., & Reitsma, P. (1997). Matthew effects in reading: A comparison of latent growth curve models and simplex models with structured means. Multivariate Behavioral Research, *32*, 135-167.
- Baumgartner, T. A., East, W. B., Frye, P. A., Hensley, L. D., Knox, D. F., & Norton, C. J. (1984). Equipment improvements and additional norms for the modified pull-up test. Research Quarterly for Exercise and Sport, *55*, 64-68.
- Baumgartner, T. A., & Jackson, A. S. (1970). Measurement schedules for tests of motor performance. Research Quarterly, *41*, 10-14.
- Baumgartner, T. A., & Jackson, A. S. (1999). Measurement for evaluation in physical education and exercise science. Boston, MA: WCB/McGraw-Hill.
- Baumgartner, T. A., & Zuidema, M. A. (1972). Factor analysis of physical fitness tests. Research Quarterly, *43*, 443-450.
- Bischoff, J. A., & Lewis, K. A. (1987). A cross-sectional study of fitness levels in a movement education program. Research Quarterly for Exercise and Sport, *58*, 348-353.
- Blalock, H. M., Jr. (1963). Making causal inferences for unmeasured variables from correlations among indicators. American Journal of Sociology, *69*, 53-62.
- Blalock, H. M., Jr. (1970). Estimating measurement error using multiple indicators and several points in time. American Sociological Review, *35*, 101-111.
- Blok, H., & Saris, W. E. (1983). Using longitudinal data to estimate reliability. Applied Psychological Measurement, *7*, 295-301.
- Bock, R. D., & Tissen, D. (1976). Fitting multi-component models for growth in stature.

Proceedings of the 9th International Biometric Conference, 1, 431-442.

Bock, R. D., & Tissen, D. (1980). Statistical problems of fitting individual growth curves. In F. E. Johnston, A. F. Roche, & C. Susanne (Eds.), Human physical growth and nutrition: Methodologies and factors (pp. 265-290). New York, NY: Plenum.

Browne, M. W., & Cudeck, R. (1993). 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.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101, 147-158.

Burr, J. A., & Nesselroade, J. R. (1990). Change measurement. In A. von Eye (Ed.), Statistical methods in longitudinal research (Vol. I, pp. 3-34). San Diego, CA: Academic Press.

Buxton, C. (1938). The application of multiple factorial methods to the study of motor abilities. Psychometrika, 3, 85-93.

Canada Fitness Survey (1984). Physical fitness of Canadian youth. Ottawa, Ontario: Author.

Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment. Beverly Hills, CA: Sage.

Caskey, S. R. (1968). Effect of motivation on standing broad jump performance of children. Research Quarterly, 39, 54-59.

Cearley, J. E. (1957). Linearity of contributions of ages, heights, and weights to prediction of track and field performances. Research Quarterly, 28, 218-222.

Chrysler Fund-Amateur Athletic Union. (1987). Physical fitness program. Bloomington, IN: Author.

Clarke, H. H., & Wickens, J. S. (1962). Maturity, structural, strength, and motor ability growth curves of boys 9 to 15 years of age. Research Quarterly, 33, 26-39.

Coleman, J. W. (1937). The differential measurement of the speed factor in large muscle activities. Research Quarterly, 8, 123-130.

Collins, L. M. (1991). Measurement in longitudinal research. In L. M. Collins & T. L. Horn (Eds.), Best methods for the analysis of change: Recent advances, unanswered questions, future directions. Washington DC: American Psychological Association.

Collins, L. M., Cliff, N. (1985). Axiomatic foundations of a three-set Guttman simplex model with applicability to longitudinal data. Psychometrika, 50, 147-158.

Considine, W. J. (1970). A validity analysis of selected leg power tests utilizing a force platform. In J. M. Cooper, (Ed.), Selected topics on biomechanics (pp. 243-250). Chicago, IL: Athletic Institute.

Corbin, C. B., & Pangrazi, R. P. (1992). Are American children and youth fit?. Research Quarterly for Exercise and Sport, 63, 96-106.

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78, 98-104.

- Costill, D. L., Miller, S. J., Myers, W. C., Kehoe, F. M., & Hoffman, W. M. (1968). Relationship among selected tests of explosive leg strength and power. Research Quarterly, 39, 785-787.
- Cotton, D. J., & Marwitz, B. (1971). Relationship between two flexed arm hangs and pull-ups for college women. Research Quarterly, 40, 415-416.
- Cousins, G. F. (1955). A factor analysis of selected wartime fitness tests. Research Quarterly, 26, 277-288.
- Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.
- Cromwell, J. B., Labys, W. C., & Terraza, M. (1994). Univariate tests for time series models. Thousand Oaks, CA: Sage.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.
- Cronbach, L. J., & Furby, L. (1970). How should we measure "change" - or should we?. Psychological Bulletin, 74, 68-80.
- Crosbie, J. (1995). Interrupted Time-series analysis with short series: Why it is problematic; How it can be improved. In J. M. Gottman (Ed), The analysis of change (pp. 361-395). Mahwah, NJ: Lawrence Erlbaum.
- Cumbee, F. Z. (1954). A factorial analysis of motor coordination. Research Quarterly, 23, 412-428.
- Cuttance, P. (1987). Issues and problems in the application of structural equation models. In P. Cuttance & R. Ecob (Eds.), Structural modeling by example: Applications in educational, Sociological, and behavioral research (pp. 241-279). New York, NY: Cambridge University Press.
- DuBois, P. H. (1957). Multivariate correlational analysis. New York, NY: Harper.
- Duncan, S. C., & Duncan, T. E. (1996). A multivariate latent growth curve analysis of adolescent substance use. Structural Equation Modeling, 3, 323-347.
- Duncan, T. E., & Duncan, S. C. (1991). A latent growth curve approach to investigating developmental dynamics and correlates of change in children's perceptions of physical competence. Research Quarterly for Exercise and Sport, 62, 390-398.
- Duncan, T. E., & Duncan, S. C. (1994). Modeling developmental processes using latent growth structural equation methodology. Applied Psychological Measurement, 18, 343-354.
- Duncan, T. E., & Duncan, S. C. (1995). Modeling the processes of development via latent variable growth curve methodology. Structural Equation Modeling, 2, 187-213.
- Duncan, T. E., Duncan, S. C., Li, F. (1998). A comparison of model- and multiple imputation-based approaches to longitudinal analysis with partial missingness. Structural Equation Modeling, 5, 1-21.
- Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F. & Alpert A. (1999). An introduction to

latent variable growth curve modeling. Mahwah, NJ: Lawrence Erlbaum.

Duncan, T. E., & Stoolmiller, M. (1993). Modeling social and psychological determinants of exercise behaviors via structural equation systems. Research Quarterly for Exercise and Sport, 64, 1-16.

Dusenberry, L. (1952). A study of the effects of training in ball throwing by children ages three to seven. Research Quarterly, 23, 9-14.

Erbaugh, S. J. (1984). The relationship of stability performance and the physical growth characteristics of preschool children. Research Quarterly for Exercise and Sport, 55, 8-16.

Espenschade, A. (1947). Development of motor coordination in boys and girls. Research Quarterly, 18, 30-44.

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

Frederiksen, C. H. & Rotondo, J. A. (1979). Time-series models and the study of longitudinal change. In J. R. Nesselroade & P. B. Baltes (Eds.), Longitudinal research in the study of behavior and development (pp. 111-153). New York, NY: Academic Press.

Gallahue, D. L. (1982). Understanding motor development in children. New York: John Wiley & Sons.

Goldberger, A. S. (1964). Econometric theory. New York: Wiley.

Goodman, L. A. (1972). A general model for the analysis of surveys. American Journal of Sociology, 78, 1135-1191.

Goodman, L. A. (1978). Analysing qualitative/categorical variables: Loglinear models and latent structure analysis. Cambridge: Abt.

Gottman, J. M. (Ed.). (1995). The analysis of change. Mahwah, NJ: Lawrence Erlbaum.

Guttman, L. A. (1954). A new approach to factor analysis: The radix. In P. F. Lazarsfeld (Ed.), Mathematical thinking in the social sciences (pp. 258-348). New York: Columbia University Press.

Halverson, H. M. (1931). An experimental study of prehension in infants by means of systematic cinema records. Genetic Psychology Monographs, 10, 107-286.

Halverson, L. E., Robertson, M. A., Safrit, M. J., & Roberts, T. W. (1977). Effect of guided practice on overhand-throw ball velocities of kindergarten children. Research Quarterly, 48, 311-318.

Halverson, L., & Williams, K. (1985). Developmental sequences for hopping over distance: A prelongitudinal screening. Research Quarterly for Exercise and Sport, 56, 37-44.

Harman, H. H. (1976). Modern factor analysis. Chicago: University of Chicago Press.

Harris, M. (1969). A factor analytic study of flexibility. Research Quarterly, 40, 62-70.

Hastad, D. N., & Lacy, A. C. (1994). Measurement and evaluation in physical education and exercise science (2nd ed.). Scottsdale, AZ: Gorsuch Scarisbrick.

Hastard, D., Marett, J., & Plowman, S. A. (1983). Evaluation of the health related physical fitness status of youth in the state of Illinois. Dekalb, Illinois: Northern Illinois University.

- Haubenstricker, J., & Seefeldt, V. (1986). Acquisition of motor skills during childhood. In V. Seefeldt (Ed.), Physical activity and well-being (pp. 41-102). Reston, VA: AAHPERD.
- Haywood, K. M. (1993). Life span motor development. Champaign, IL: Human Kinetics.
- Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. American Sociological Review, 34, 93-101.
- Heyward, V. H. (1984). Designs for fitness. Minneapolis, MN: Burgess.
- Hilsendager, D. R., Karnes, E., & Spiritoso, T. (1969). Some dimensions of physical performance. Perceptual and Motor Skills, 28, 479-487.
- Hilsendager, D. R., Stow, M. H., & Ackerman, K. J. (1969). Comparison of speed, strength, and agility exercises in the development of agility. Research Quarterly, 37, 71-75.
- Hoyt, C. J. (1941). Test reliability estimated by analysis of variance. Psychometrika, 6, 153-160.
- Hu, L. -T., & Bentler, P. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1-55.
- Hummel-Rossi, B., & Weinberg, S. (1975). Practical guidelines in applying current theories to the measurement of change. Part 1, Part 2. JSAS Catalog of Selected Documents in Psychology, 5, 226 (ms#916).
- Institute for Aerobics Research. (1987). FITNESSGRAM user's manual. Dallas, TX: Author.
- Ismail, A. H., & Cowell, C. C. (1961). Factor analysis of motor aptitude of pre-adolescent males. Research Quarterly, 32, 507-513.
- Jackson, A. S. (1971). Factor analysis of selected muscular strength and motor performance test. Research Quarterly, 42, 164-172.
- Jackson, A. S., & Baumgartner, T. A. (1969). Measurement schedules of sprint running. Research Quarterly, 40, 708-711.
- Jackson, A. W., & Baker, A. A. (1986). The relationship of the sit and reach test to criterion measures of hamstring and back flexibility in young females. Research Quarterly for Exercise and Sport, 57, 183-186.
- Jones, H. E. (1946). Skeletal maturity as related to strength. Child Development, 17, 173.
- Jöreskog, K.G. (1970). Estimation and testing of simplex models. British Journal of Mathematical and Statistical Psychology, 23, 121-145.
- Jöreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, R. D. Luce & P. Suppes (Eds.), Contemporary developments in mathematical psychology (Vol. 2, pp. 1-56). San Francisco, CA: W. H. Freeman.
- Jöreskog, K. G., & Sorbom, D. (1988). LISREL VII: Analysis of linear structural relations by the method of maximum likelihood. Chicago: SPSS.
- Jöreskog, K. G., & Sorbom, D. (1999). LISREL 8 computer program. Chicago, IL: Scientific Software International.

- Jöreskog, K. G., & Sorbom, D. (1999). PRELIS 2 computer program. Chicago, IL: Scientific Software International.
- Kaplan, D. (2000). Structural equation modeling: Foundations and extensions. Thousand Oaks, CA: Sage.
- Kenney, D. A. (1979). Correlation and causality. New York: Wiley-Interscience.
- Kenny, D. A., & Campbell, D. T. (1989). On the measurement of stability in over-time data. Journal of Personality, *57*, 445-481.
- Klesius, S. E. (1968). Reliability of the AAHPER Youth Fitness items and relative efficiency of the performance measures. Research Quarterly, *39*, 801-811.
- Labouvie, E. W. (1982). Concepts of change and regression toward the mean. Psychological Bulletin, *92*, 251-257.
- Larson, L. A. (1941). A factor analysis of motor ability variables and tests, with tests for college men. Research Quarterly, *12*, 499-517.
- Lawrence, F. R., & Hancock, G. R. (1998). Method, plainly speaking Assessing change over time using latent growth modeling. Measurement and Evaluation in Counseling and Development, *30*, 211-224.
- Liba, M. (1967). Factor analysis of strength variables. Research Quarterly, *38*, 649-663.
- Lord, F. M. (1956). The measurement of growth. Educational Psychological Measurement, *16*, 421-437.
- Lord, F. M., & Novick, M. N. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.
- Malina, R. M., & Bouchard, C. (1991). Growth, Maturation, and physical activity. Champaign, IL: Human Kinetics.
- Mandys, F., Dolan, C. V., & Molenaar, P. C. M. (1994). Two aspects of the simplex model: Goodness of fit to linear growth curve structures and the analysis of mean trends. Journal of Educational and Behavioral Statistics, *19*, 201-215.
- Manitoba Department of Education. (1977). Manitoba physical fitness performance test manual and fitness objectives. Manitoba, Canada: Author.
- Manning, W. H., & DuBois, P. H. (1962). Correlational methods in research on human learning. Perceptual and Motor Skills, *15*, 287-321.
- Marmis, C., Montoye, H. J., Cunningham, D. A., & Kozar, A. J. (1969). Reliability of the multi-trial items of the AAHPER youth fitness test. Research Quarterly, *40*, 240-245.
- Marsh, H. W. (1993). The multidimensional structure of physical fitness: Invariance over gender and age. Research Quarterly for Exercise and Sport, *64*, 256-273.
- Marsh, H. W. (1996). Physical self description questionnaire: Stability and discriminant validity. Research Quarterly for Exercise and Sport, *67*, 249-264.

Marsh, H. W., & Grayson, D. (1994a). Longitudinal confirmatory factor analysis: Common, time-specific, item-specific, and residual-error components of variance. Structural Equation Modeling, 1, 116-145.

Marsh, H. W., & Grayson, D. (1994b). Longitudinal stability of latent means and individual differences: A unified approach. Structural Equation Modeling, 1, 317-359.

Marsh, H. W., & Hau, K. -T. (1996). Assessing goodness of fit: Is parsimony always desirable?. The Journal of Experimental Education, 64, 364-390.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In R. B. Cattell & J. Nesselroade (Eds.), Handbook of multivariate experimental psychology (2nd ed., pp. 561-614). New York, NY: Plenum.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. Child Development, 58, 110-133.

McArdle, J. J., & Hamagami, F. (1991). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. In L. M. Collins & J. C. Horn (Eds.), Best methods for the analysis of change (pp. 276-304). Washington, DC: American Psychological Association.

McCloy, C. H. (1935). The influence of chronological age on motor performance. Research Quarterly, 6, 61-64.

McCloy, C. H., & Young, N. D. (1954). Test and measurements in health and physical education. New York: Appleton-Century-Crofts.

Meredith, W., & Tisak, J. (1984). "Tuckerizing" curves. Paper presented at the Psychometric Society Annual Meetings, Santa Barbara, CA.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. Psychometrika, 55, 107-122.

Metheny, E. (1938). Studies of the Johnson test as a test of motor educability. Research Quarterly, 9, 105-114.

Milne, C., Seefeldt, V., & Reuschlein, P. (1976). Relationship between grade, sex, race, and motor performance in young children. Research Quarterly, 47, 726-730.

Mirwald, R. L., & Bailey, D. A. (1986). Maximal aerobic power: A longitudinal analysis. London, Ontario: Sports Dynamics.

Montoye, H. J. (1984). Age and cardiovascular response to submaximal treadmill exercise in males. Research Quarterly for Exercise and Sport, 55, 85-88.

Montoye, H. J., & Lamphiear, D. E. (1977). Grip and arm strength in males and females, age 10 to 69. Research Quarterly, 48, 109-120.

Morera, O. F., Johnson, T. P., Freels, S., Parsons, J., Crittenden, K. S., Flay, B. R., & Warnecke, R. B. (1998). The measure of stage of readiness to change: Some psychometric considerations. Psychological Assessment, 10, 182-186.

Morris, A. M., Williams, J. M., Atwater, A. E., & Wilmore, J. H. (1982). Age and sex

differences in motor performance of three through 6-year-old children. Research Quarterly for Exercise and Sport, 53, 214-221.

Morrow, J. R., Jackson, A. S., & Bell, J. A. (1978). The function of age, sex, and body mass on distance running. Research Quarterly, 49, 491-497.

Mulaik, S. A. (1972). The foundation of factor analysis. New York: McGraw-Hill.

Muthen, B. (1994). Multilevel covariance structure analysis. Sociological Methods & Research, 22, 376-398.

Muthen, B. (1996). Growth modeling with binary responses. In A. von Eye & C. Clogg (Eds.), Categorical variables in developmental research: Methods of analysis (pp. 37-54). San Diego, CA: Academic Press.

Muthen, B. (1997). Latent variable modeling of longitudinal and multilevel data. In A. E. Raftery (Ed.), Sociological Methodology (pp. 453-480). Washington, DC: Blackwell.

Muthen, B., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. Psychological Methods, 2, 371-402.

Muthen, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of nonnormal Likert variables. British Journal of Mathematical and Statistical Psychology, 38, 171-189.

Muthen, B., & Muthen, L. (1998). Mplus computer program. Los Angeles, CA: Author.

Nelson, K. R., Thomas, J. R., & Nelson, J. K. (1991). Longitudinal change in throwing performance: Gender differences. Research Quarterly for Exercise and Sport, 62, 105-108.

Nesselroade, J. R., & Baltes, P. B. (Eds.) (1979). Longitudinal research in the study of behavior and development. New York, NY: Academic Press.

Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. Psychological Bulletin, 88, 622-637.

Pangrazi, R. P., & Corbin, C. B. (1990). Age as a factor relating to physical fitness test performance. Research Quarterly for Exercise and Sport, 61, 410-414.

Pate, R. R., Burgess, M. L., Woods, J. A., Ross, J. G., & Baumgartner, T. (1993). Validity of field tests of upper body muscular strength. Research Quarterly for Exercise and Sport, 64, 17-24.

Phillips, M. (1949). Study of a series of physical education tests by factor analysis. Research Quarterly, 20, 60-71.

Ponthieux, N. A., & Barker, D. G. (1963). An analysis of the AAHPER youth fitness test. Research Quarterly, 34, 525-526.

President's Council on Physical Fitness and Sports. (1987). The president's physical fitness award program. Washington, DC: Author.

Raffalovich, L. E., & Bohrnstedt, G. W. (1987). Common, specific, and error variance components of factor models: Estimation with longitudinal data. Sociological Methods and Research, 15,

385-405.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. Biometrics, 14, 1 - 17.

Rarick, G. L. (1937). An analysis of the speed factor in simple athletic activities. Research Quarterly, 8, 89-105.

Rarick, G. L. (1980). Cognitive-motor relationships in the growing years. Research Quarterly, 51, 174-189.

Rarick, G. L., & Dobbins, D. A. (1975). Basic components in the motor performance of children six to nine years of age. Medicine and Science in Sports, 7, 105-110.

Reiff, G. G., Dixon, W. R., Jacoby, D., Ye, G. X., Spain, C. G., & Hunsicker, P. A. (1986). The president's council on physical fitness and sports 1985: National school population fitness survey. HHS-Office of the Assistant Secretary for Health, Research Project 282-82-0086, University of Michigan.

Richards, J. M. (1975). A simulation study of the use of change measures to compare educational programs. American Educational Research Journal, 12, 299-311.

Rogosa, D. (1995). Myths and methods: "Myths about longitudinal research" plus supplemental questions. In J. M. Gottman (Ed.), The analysis of change (pp. 3-66). Mahwah, NJ: Lawrence Erlbaum.

Rogosa, D., Brandt, D., & Zimowski (1982). A growth curve approach to the measurement of change. Psychological Bulletin, 92, 726-748.

Rogosa, D., & Willet, J. B. (1985a). Satisfying a simplex structure is simpler than it should be. Journal of educational statistics, 10, 99-107.

Rogosa, D., & Willet, J. B. (1985b). Understanding correlates of change by modeling individual differences in growth. Psychometrika, 50, 203-228.

Roskam, E. E. (1976). Multivariate analysis of change and growth: Critical review and perspectives. In D. N. M. de Gruijter & L. J. T. van der Kamp (Eds.), Advances in psychological and educational measurement (pp. 111-133). New York: Wiley.

Ross, J. G., & Gilbert, G. C. (1985). The national children and youth fitness study: A summary of findings. Journal of Physical Education and Recreation, 58, 51-56.

Ross, J. G., Pate, R. R., Delpy, L. A., Gold, R. S., & Svilar, M. (1987). New health-related fitness norms. Journal of Physical Education, Recreation and Dance, 58, 66-78.

Rowe, F. A. (1933). Growth comparison of athletes and non-athletes. Research Quarterly, 4, 108 -116.

Safrit, M. J., & Wood, T. M. (1995). Introduction to measurement in physical education and exercise science (3rd ed.). St. Louis, Missouri: Mosby.

Sargent, D. A. (1921). The physical test of man. American Physical Education Review, 26, 188-194.

Schutz, R. W. (1970). Stochastic processes: Their nature and use in the study of sport and physical activity. Research Quarterly, 41, 205-212.

Schutz, R. W. (1989). Analyzing change. In J. Safrit, & T. Wood (Eds.), Measurement concepts in physical education and exercise science (pp. 206-228). Champaign, Illinois: Human Kinetics.

Schutz, R. W. (1995). The stability of individual performance in baseball: An examination of four 5-year periods, 1928-32, 1948-52, 1968-72, and 1988-92. The Proceedings of the Annual Meeting of the American Statistical Association.

Schutz, R. W. (1998). Assessing the stability of Psychological Traits and measures. In J. L. Duda (Ed.), Advances in Sport and Exercise Psychology Measurement (pp. 393-408). Morgantown, WV: Fitness Information Technology, Inc.

Schutz, R. W., & Gessaroli, M. E. (1987). The analysis of repeated measures designs involving multiple dependent variables. Research Quarterly for Exercise and Sport, 58, 132-149.

Schutz, R. W., & Park, I. (in press). Some methodological considerations in developmental sport and exercise psychology. In M. R. Weiss (Ed.), Developmental sport and exercise psychology: A lifespan perspective.

Seashore, H. G. (1942). Some relationships of fine and gross motor abilities. Research Quarterly, 13, 259-274.

Sellis, L. G. (1951). The relationship between measures of physical growth and gross motor performance of primary-grade school children. Research Quarterly, 22, 244-260.

Shuleva, K. M., Hunter, G. R., Hester, D. J., & Dunway, D. L. (1990). Exercise oxygen uptake in 3- through 6-year-old children. Pediatric Exercise Science, 2, 130-139.

Siegel, P. M., & Hodge, R. W. (1968). A causal approach to the study of measurement error. In H. M. Blalock, Jr., & A. B. Blalock (Eds.), Methodology in social research (pp. 28-59), New York, NY: McGraw Hill.

Smoll, F. L., & Schutz, R. W. (1990). Quantifying gender differences in physical performance: A developmental perspective. Developmental Psychology, 26, 360-369.

Solley, W. H. (1960). Relationship of selected factors in growth derivable from age-height-weight measurements. Research Quarterly, 31, 92-100.

SPSS Inc. (1997). Statistical package for the social science. Chicago, IL: SPSS.

Start, K. B., Gray, R. K., Glencross, D. J., & Walsh, A. (1966). A factorial investigation of power, speed, isometric strength, and anthropometric measures in the lower limb. Research Quarterly, 37, 553-559

Stevens, J. (1996). Applied multivariate statistics for the social sciences. Mahwah, NJ: L. Erlbaum.

Stoolmiller, M. (1995). Using latent growth curve models to study developmental processes. In J. M. Gottman (Ed.), The analysis of change (pp. 103-138). Mahwah, NJ: Lawrence Erlbaum.

Teeple, J. & Massey, B. (1976). Force-time parameters and physical growth of boys ages 6 to 12 years. Research Quarterly, 47, 464-471.

Thorndike, E. I. (1924). The influence of chance imperfections of measures upon the relationship of initial score to gain or loss. Journal of Experimental Psychology, 7, 225-232.

Tisak, J., & Meredith, W. (1990). Descriptive and Associative developmental models. In A. von Eye (Ed.), Statistical methods in longitudinal research (Vol. II, pp. 387-406). San Diego, CA: Academic Press.

Tisak, J., & Meredith, W. (1986). "Tuckerizing" curves for latent variables. Paper presented at the annual meeting of the Psychometric Society, Toronto, Canada.

Tisak, J., & Tisak, M. S. (1996). Longitudinal models of reliability and validity: A latent curve approach. Applied Psychological Measurement, 20, 275-288.

Tissen, D., & Bock, R. D. (1990). Linear and nonlinear curve fitting. In A. von Eye (Ed.), Statistical methods in longitudinal research (Vol. II, pp. 289-318). San Diego, CA: Academic Press.

Thomas, J. R., & French, K. E. (1985). Gender differences across age in motor performance: A meta-analysis. Psychological Bulletin, 98, 260-282.

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. Psychometrika, 38, 1-10.

Tucker, L. R. (1966). Learning theory and multivariate experiment: Illustration of generalization learning curves. In R. B. Cattell (Ed.), Handbook of multivariate experimental psychology (pp. 476-501). Chicago: Rand McNally.

Tucker, L. R., Damarin, F., & Messick, S. (1966). A base-free measure of change. Psychometrika, 31, 457-473

Vilchkovsky, E. S. (1972). Motor development in pre-school and school age children. Theory and Practice of Physical Culture, 6, 29-33.

von Eye, A. (Ed.). (1990). Statistical methods in longitudinal research, Vol. I, II. San Diego, CA: Academic Press.

Werner, P. (1974). Education of selected movement patterns of preschool children. Perceptual and Motor Skills, 39, 795-798.

Werts, C. E., Breland, H. M., Grandy, J., & Rock, D. R. (1980). Using longitudinal data to estimate reliability in the presence of correlated measurement errors. Educational and Psychological Measurement, 40, 19-29.

Werts, C. E., Jöreskog, K. G., & Linn, R. L. (1971). Comment on the estimation of measurement error in panel data. American Sociological Review, 36, 110-113.

Werts, C. E., Linn, R. L., & Jöreskog, K. G. (1977). A simplex model for analyzing academic growth. Educational and Psychological Measurement, 37, 745-756.

Werts, C. E., Linn, R. L., & Jöreskog, K. G. (1978). Reliability of college grades from

longitudinal data. Educational and Psychological Measurement, 38, 89-95.

Wheaton, B., Muthé, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed), Sociological methodology (pp. 84-136), San Francisco, CA: Jossey-Bass.

Wilder, J. (1957). The law of initial value in neurology and psychiatry. Journal of Nervous and Mental Disease, 125, 73-86.

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. Psychological Bulletin, 116, 363-381.

Willett, J. B., & Sayer, A. G. (1996). Cross-domain analyses of change over time: Combining growth modeling and covariance structure analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.) Advanced structural equation modeling: Issues and techniques (pp. 125-157). Mahwah, NJ: Lawrence Erlbaum.

Wiley, D. E., & Wiley, J. A. (1970). The estimation of measurement error in panel data, American Sociological Review, 35, 112-117.

Wiley, J. A., & Wiley, M. G. (1974). A note on correlated errors in repeated measurements. Sociological Methods and Research, 3, 172-188.

Winer, B. J. (1971). Statistical principles in experimental design (2nd ed.) New York, NY: McGraw-Hill.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). Statistical principles in experimental design (3rd ed.) New York, NY: McGraw-Hill.

Wright, S. (1934). The method of path coefficients. Annals of Mathematical Statistics, 5, 161-215.

Zieve, L. (1940). Note on the correlation of initial scores with gain. Journal of Educational Psychology, 31, 391-394.

APPENDICES

Appendix A: Example Data Records for Five Selected Subjects (Michigan Data Set 1)

Subject#	Predictor Variables					Flexed-Arm Hang (seconds)				
	Practice	Age	Grade	Season	Year	Age 8	Age 9	Age 10	Age 11	Age 12
1	5	97	0	1	1970	22	17	22	27	29
2	1	94	1	0	1975	8	7	18	16	22
3	7	99	0	1	1978	6	8	8	12	9
4	6	96	1	0	1982	23	21	16	13	13
5	6	95	1	0	1987	22	17	17	22	28

Subject#	Jump-and-Reach (inches)					Sit-and-reach (inches)				
	Age 8	Age 9	Age 10	Age 11	Age 12	Age 8	Age 9	Age 10	Age 11	Age 12
1	8.0	10.5	12.0	13.0	14.5	8.5	7.5	8.5	9.0	7.0
2	10.5	11.0	12.0	11.0	13.0	8.0	7.5	6.0	4.0	7.0
3	10.5	13.5	13.5	15.0	17.0	8.5	8.0	7.0	5.0	5.0
4	7.0	10.5	10.5	11.0	11.5	8.0	7.0	8.0	9.0	8.5
5	8.5	11.5	14.0	13.0	16.0	9.0	10.5	8.5	9.0	7.0

Subject#	Agility Shuttle Run (seconds)					Endurance Shuttle Run (seconds)				
	Age 8	Age 9	Age 10	Age 11	Age 12	Age 8	Age 9	Age 10	Age 11	Age 12
1	12.3	12.2	11.7	11.3	10.8	44.8	43.8	43.0	40.8	40.9
2	12.0	12.9	11.4	11.1	12.0	40.2	43.0	45.6	40.4	42.5
3	13.2	13.6	11.6	11.0	10.0	45.6	45.4	41.6	39.4	36.7
4	11.7	11.0	10.9	11.4	10.9	46.0	42.9	38.9	40.9	39.5
5	11.2	11.3	10.8	10.4	10.0	39.9	41.2	39.8	38.2	36.3

Subject#	30-yard Dash (seconds)					Standing Long Jump (inches)				
	Age 8	Age 9	Age 10	Age 11	Age 12	Age 8	Age 9	Age 10	Age 11	Age 12
1	5.4	5.2	4.9	4.6	4.6	58.0	62.0	67.5	66.0	67.5
2	5.8	5.0	4.8	5.3	4.5	55.0	65.0	62.0	64.5	73.0
3	5.2	4.9	4.3	4.4	4.3	65.0	70.0	75.0	77.0	81.0
4	5.6	5.2	4.7	4.9	4.9	47.0	52.0	55.0	62.0	66.0
5	5.0	5.3	4.8	4.8	4.7	58.5	62.0	68.5	72.0	68.5

Note. Practice = the number of measurement taken before age 8, Age = age in months at the first time point (age 8), Grade = grade at age 8 (0 = grade 2, 1 = grade 3), Season = measured season (0 = summer, 1 = winter), Year = measurement year at age 8.

Appendix B: Program Commands for Latent Growth Models

LISREL Commands for Univariate Models

Linear Model (flexed-arm-hang)

```

1      DA NI=63 NO=210
2      RA FI=C:\THESIS\DATA\MOTOR\MOTOR1.DAT FO
3      (41F8.2)
4      LABEL
5      ID, PR_ME_NO, AGE, GRADE, ME_SESN, ME_YR,
6      FAH8, JAR8, ASR8, SLJ8, DASH8, SAR8, ESR8,
7      FAH9, JAR9, ASR9, SLJ9, DASH9, SAR9, ESR9,
8      FAH10, JAR10, ASR10, SLJ10, DASH10, SAR10, ESR10,
9      FAH11, JAR11, ASR11, SLJ11, DASH11, SAR11, ESR11,
10     FAH12, JAR12, ASR12, SLJ12, DASH12, SAR12, ESR12
11
12     SE
13     7 14 21 28 35 /
14
15     MO NY=5 TY=ZE NE=2 TE=SY,FI AL=FR BE=ZE PS=SY,FR
16
17     LE
18     INTERCEPT SLOPE
19
20     MA LY
21     1 0
22     1 1
23     1 2
24     1 3
25     1 4
26
27     FR TE 1 1 TE 2 2 TE 3 3 TE 4 4 TE 5 5
28
29     OU RS SE SC TV ND=3 IT=1000 AD=OFF

```

Note. Line numbers are added for a presentation purpose. ID = subjects' ID; PR_ME_NO = the number of measurement before age 8 (initial time point); GRADE = grade at age 8; AGE = age in months at age 8; ME_SESN = measurement season; ME_YR = measurement year at age 8; FAH8 = flexed-arm-hang at age 8; JAR = jump-and-reach; ASR = agility shuttle run, SLJ = standing long jump, DASH = 30-yard dash, SAR = sit-and-reach, ESR = endurance shuttle run.

For an equal error variance model, add following commands between lines 27 and 29.

```
EQ TE 1 1 TE 2 2 TE 3 3 TE 4 4 TE 5 5
```

Quadratic Model (flexed-arm-hang)

Replace lines from 15 to 25 with following commands.

MO NY=5 TY=ZE NE=3 TE=SY, FI AL=FR BE=ZE PS=SY, FR

LE

INTERCEPT SLOPE QUADRATC

MA LY

1 0 0

1 1 1

1 2 4

1 3 9

1 4 16

Cubic Model (flexed-arm-hang)

Replace lines from 15 to 25 with following commands.

MO NY=5 TY=ZE NE=4 TE=SY, FI AL=FR BE=ZE PS=SY, FR

LE

INTERCEPT SLOPE QUADRATC CUBIC

MA LY

1 0 0 0

1 1 1 1

1 2 4 8

1 3 9 27

1 4 16 64

Unspecified Curve Model (flexed-arm-hang)

Add following commands between lines 25 and 27.

FR LY 3 2 LY 4 2 LY 5 2

Linear Model With One Predictor (PR ME NO)

Replace lines from 12 to 15 with following commands.

SE

7 14 21 28 35 2 /

MO NY=5 NX=1 TX=ZE TY=ZE NE=2 NK=1 TD=FI TE=SY, FI KA=FR AL=FR BE=ZE

```

PS=SY,FR

LK
LEARNING

VA 1 LX 1 1

```

**Linear Model With Three Predictors, PR ME NO, AGE and ME_YR (the Effect of
"PR ME NO" on the Slope Factor is Fixed at Zero)**

Replace lines from 13 to 16 with following commands.

```

SE
7 14 21 28 35 2 3 6/

MO NY=5 NX=1 TX=ZE TY=ZE NE=2 NK=3 TD=FI TE=SY,FI KA=FR AL=FR BE=ZE
PS=SY,FR

LK
LEARNING AGE ME_YR

MA LX
1 0 0
0 1 0
0 0 1

FI GA 2 1

```

Program Commands for Multivariate Models (Curve-of-Factors Model)

**LISREL Commands for the 5-factor Measurement Model ("Run" Factor: Equal Factor Loadings
and Correlated Errors Over Time)**

```

DA NI=63 NO=210
RA FI=C:\THESIS\DATA\MOTOR\MOTOR1.DAT FO
(41F8.2)
LABEL
ID, PR_ME_NO, AGE, GRADE, ME_SESN, ME_YR,
FAH8, JAR8, ASR8, SLJ8, DASH8, SAR8, ESR8,
FAH9, JAR9, ASR9, SLJ9, DASH9, SAR9, ESR9,
FAH10, JAR10, ASR10, SLJ10, DASH10, SAR10, ESR10,
FAH11, JAR11, ASR11, SLJ11, DASH11, SAR11, ESR11,
FAH12, JAR12, ASR12, SLJ12, DASH12, SAR12, ESR12

SE
9 11 13 16 18 20 23 25 27 30 32 34 37 39 41 /

MO NX=15 NK=5 PH=SY,FR TD=FU,FI

```

```

LK
RUN1 RUN2 RUN3 RUN4 RUN5

PA LX
3(1 0 0 0 0) 3(0 1 0 0 0) 3(0 0 1 0 0) 3(0 0 0 1 0) 3(0 0 0 0 1)

FIX LX 1 1 LX 4 2 LX 7 3 LX 10 4 LX 13 5
VA 1 LX 1 1 LX 4 2 LX 7 3 LX 10 4 LX 13 5

EQ LX 2 1 LX 5 2 LX 8 3 LX 11 4 LX 14 5
EQ LX 3 1 LX 6 2 LX 9 3 LX 12 4 LX 15 5

FR TD 1 1 TD 2 2 TD 3 3 TD 4 4 TD 5 5 TD 6 6 TD 7 7 TD 8 8
FR TD 9 9 TD 10 10 TD 11 11 TD 12 12 TD 13 13 TD 14 14 TD 15 15

FR TD 1 4 TD 1 7 TD 1 10 TD 1 13 TD 4 7 TD 4 10 TD 4 13 TD 7 10 FR TD
7 13 TD 10 13 TD 2 5 TD 2 8 TD 2 11 TD 2 14 TD 5 8 TD 5 11 FR TD 5 14
TD 8 11 TD 8 14 TD 11 14 TD 3 6 TD 3 9 TD 3 12
FR TD 3 15 TD 6 9 TD 6 12 TD 6 15 TD 9 12 TD 9 15 TD 12 15

OU RS SE SC TV ND=3 IT=1000 AD=OFF

```

LISREL Commands for the Linear Model ("Run" factor)

```

1      DA NI=63 NO=210
2      RA FI=C:\THESIS\DATA\MOTOR\MOTOR1.DAT FO
3      (41F8.2)
4      LABEL
5      ID, PR ME_NO, AGE, GRADE, ME_SESN, ME_YR,
6      FAH8, JAR8, ASR8, SLJ8, DASH8, SAR8, ESR8,
7      FAH9, JAR9, ASR9, SLJ9, DASH9, SAR9, ESR9,
8      FAH10, JAR10, ASR10, SLJ10, DASH10, SAR10, ESR10,
9      FAH11, JAR11, ASR11, SLJ11, DASH11, SAR11, ESR11,
10     FAH12, JAR12, ASR12, SLJ12, DASH12, SAR12, ESR12
11
12     SE
13     9 11 13   16 18 20   23 25 27   30 32 34   37 39 41 /
14
15     MO NY=15 TY=ZE NE=5 NK=2 TE=FU,FI AL=ZE KA=FR BE=ZE
16     PS=DI,FR GA=FU,FI PH=FU,FR
17
18     LE
19     RUN1 RUN2 RUN3 RUN4 RUN5
20
21     LK
22     INTERCEPT SLOPE
23
24     PATTERN LY
25     3(1 0 0 0 0) 3(0 1 0 0 0) 3(0 0 1 0 0) 3(0 0 0 1 0)
26     3(0 0 0 0 1)

```

```

27
28          FI LY 1 1 LY 4 2 LY 7 3 LY 10 4 LY 13 5
29          VA 1 LY 1 1 LY 4 2 LY 7 3 LY 10 4 LY 13 5
30
31          EQ LY 2 1 LY 5 2 LY 8 3 LY 11 4 LY 14 5
32          EQ LY 3 1 LY 6 2 LY 9 3 LY 12 4 LY 15 5
33
34          MA GA
35          1 0
36          1 1
37          1 2
38          1 3
39          1 4
40
41          FR TE 1 1 TE 2 2 TE 3 3 TE 4 4 TE 5 5 TE 6 6 TE 7 7
42          TE 8 8 FR TE 9 9 TE 10 10 TE 11 11 TE 12 12 TE 13 13
43          TE 14 14 TE 15 15
44
45          OU RS SE SC TV ND=3 IT=1000 AD=OFF

```

For a correlated errors model, add following commands between the lines 43 and 45.

```

FR TE 1 4 TE 1 7 TE 1 10 TE 1 13 TE 4 7 TE 4 10 TE 4 13 TE 7 10 FR TE
7 13 TE 10 13 TE 2 5 TE 2 8 TE 2 11 TE 2 14 TE 5 8 TE 5 11 FR TE 5 14
TE 8 11 TE 8 14 TE 11 14 TE 3 6 TE 3 9 TE 3 12
FR TE 3 15 TE 6 9 TE 6 12 TE 6 15 TE 9 12 TE 9 15 TE 12 15

```

LISREL Commands for the Quadratic Model ("Run" factor)

Replace lines from 15 to 39 with the following commands.

```

MO NY=15 TY=ZE NE=5 NK=3 TE=FU,FI AL=ZE KA=FR BE=ZE
PS=DI,FR GA=FU,FI PH=FU,FR

```

```

LE
RUN1 RUN2 RUN3 RUN4 RUN5

```

```

LK
INTERCEPT LINEAR QUADRATIC

```

```

PATTERN LY 3(1 0 0 0 0) 3(0 1 0 0 0) 3(0 0 1 0 0) 3(0 0 0 1 0)
          3(0 0 0 0 1)

```

```

FI LY 1 1 LY 4 2 LY 7 3 LY 10 4 LY 13 5
VA 1 LY 1 1 LY 4 2 LY 7 3 LY 10 4 LY 13 5

```

```

EQ LY 2 1 LY 5 2 LY 8 3 LY 11 4 LY 14 5
EQ LY 3 1 LY 6 2 LY 9 3 LY 12 4 LY 15 5

```

```

MA GA
1 0 0

```



```

1 1 1
1 2 4
1 3 9
1 4 16

```

LISREL Commands for the Cubic Model ("Run" factor)

Replace lines from 15 to 39 with the following commands.

```

MO NY=15 TY=ZE NE=5 NK=4 TE=FU, FI AL=ZE KA=FR BE=ZE
  PS=DI, FR GA=FU, FI PH=FU, FR

```

```

LE
RUN1 RUN2 RUN3 RUN4 RUN5

```

```

LK
INTERCEPT LINEAR QUADRATIC Cubic

```

```

PATTERN LY 3(1 0 0 0 0) 3(0 1 0 0 0) 3(0 0 1 0 0) 3(0 0 0 1 0)
          3(0 0 0 0 1)

```

```

FI LY 1 1 LY 4 2 LY 7 3 LY 10 4 LY 13 5
VA 1 LY 1 1 LY 4 2 LY 7 3 LY 10 4 LY 13 5

```

```

EQ LY 2 1 LY 5 2 LY 8 3 LY 11 4 LY 14 5
EQ LY 3 1 LY 6 2 LY 9 3 LY 12 4 LY 15 5

```

```

MA GA
1 0 0 0
1 1 1 1
1 2 4 8
1 3 9 27
1 4 16 64

```

LISREL Commands for the Unspecified Curve Model ("Run" factor)

Add following commands between lines 39 and 41.

```

FR GA 3 2 GA 4 2 GA 5 2

```

MPLUS Commands for the Unspecified Curve Model With One Predictor, PR ME NO ("Run" factor)

```

TITLE: MPLUS RUN FOR MULTIVARIATE LGM (UNSPECIFIED CURVE MODEL)

```

```

DATA:  FILE IS C:\THESIS\DATA\MOTOR\MOTOR1.DAT;
      FORMAT IS 41F8.2;
VARIABLE: NAMES ARE ID, PR_ME_NO, AGE, GRADE, ME_SESN, ME_YR,
          FAH8, JAR8, ASR8, SLJ8, DASH8, SAR8, ESR8,
          FAH9, JAR9, ASR9, SLJ9, DASH9, SAR9, ESR9,
          FAH10, JAR10, ASR10, SLJ10, DASH10, SAR10, ESR10,
          FAH11, JAR11, ASR11, SLJ11, DASH11, SAR11, ESR11,
          FAH12, JAR12, ASR12, SLJ12, DASH12, SAR12, ESR12;

USEVARIABLES ARE PR_ME_NO, ASR8, DASH8, ESR8, ASR9, DASH9, ESR9,
                ASR10, DASH10, ESR10, ASR11, DASH11, ESR11, ASR12,
                DASH12, ESR12;

ANALYSIS: TYPE IS MEANSTRUCTURE;
          ITERATIONS=1000;

MODEL:  RUN1 BY ASR8; RUN1 BY DASH8*.5(1); RUN1 BY ESR8(2)*3.5;
        RUN2 BY ASR9; RUN2 BY DASH9(1); RUN2 BY ESR9(2);
        RUN3 BY ASR10; RUN3 BY DASH10(1); RUN3 BY ESR10(2);
        RUN4 BY ASR11; RUN4 BY DASH11(1); RUN4 BY ESR11(2);
        RUN5 BY ASR12; RUN5 BY DASH12(1); RUN5 BY ESE12(2);
        I BY RUN1-RUN5 @1;
        C BY RUN1@0 RUN2@1 RUN3*1.8 RUN4*2.3 RUN5*2.9;
        [ASR8-ESR12@0];
        [RUN1-RUN5@0 I*12.5 C];
        DASH8 DASH9 DASH10 DASH11 DASH12(3);
        ASR8 WITH ASR9*0 ASR10*0 ASR11*0 ASR12*0;
        ASR9 WITH ASR10*0 ASR11*0 ASR12*0;
        ASR10 WITH ASR11*0 ASR12*0;
        ASR11 WITH ASR12*0;
        DASH8 WITH DASH9*0 DASH10*0 DASH11*0 DASH12*0;
        DASH9 WITH DASH10*0 DASH11*0 DASH12*0;
        DASH10 WITH DASH11*0 DASH12*0;
        DASH11 WITH DASH12*0;
        ESR8 WITH ESR9*.7 ESR10*.2 ESR11*.3 ESR12*0;
        ESR9 WITH ESR10*0 ESR11*.1 ESR12*0;
        ESR10 WITH ESR11*.3 ESR12*.5;
        ESR11 WITH ESR12*.2;

        I C ON PR_ME_NO

OUTPUT:  SAMPSTAT; STANDARDIZED; RESIDUAL; TECH4;

```

LISREL Commands for the Simplex Model With Mean Structure
(Equal Error Variance Between the First and Last Two Time Points)

```

DA NI=5 NO=200
CMATRIX FI=DAT11.COV
MEANS FI=DAT11.MEA

```

```

LABEL

```

*

Y1, Y2, Y3, Y4, Y5

MO NY=5 NE=5 LY=ID TE=SY,FI BE=FU PS=DI TY=ZE AL=FR

LE

TIME1 TIME2 TIME3 TIME4 TIME5

FR BE 2 1 BE 3 2 BE 4 3 BE 5 4

FR TE 1 1 TE 2 2 TE 3 3 TE 4 4 TE 5 5

EQ TE 1 1 TE 2 2

EQ TE 4 4 TE 5 5

OU RS SE SC TV ND=3 IT=1000 AD=OFF

Appendix C: Descriptive Statistics and Parameter Estimates of Latent Growth Models

Univariate Results

Data Set 1

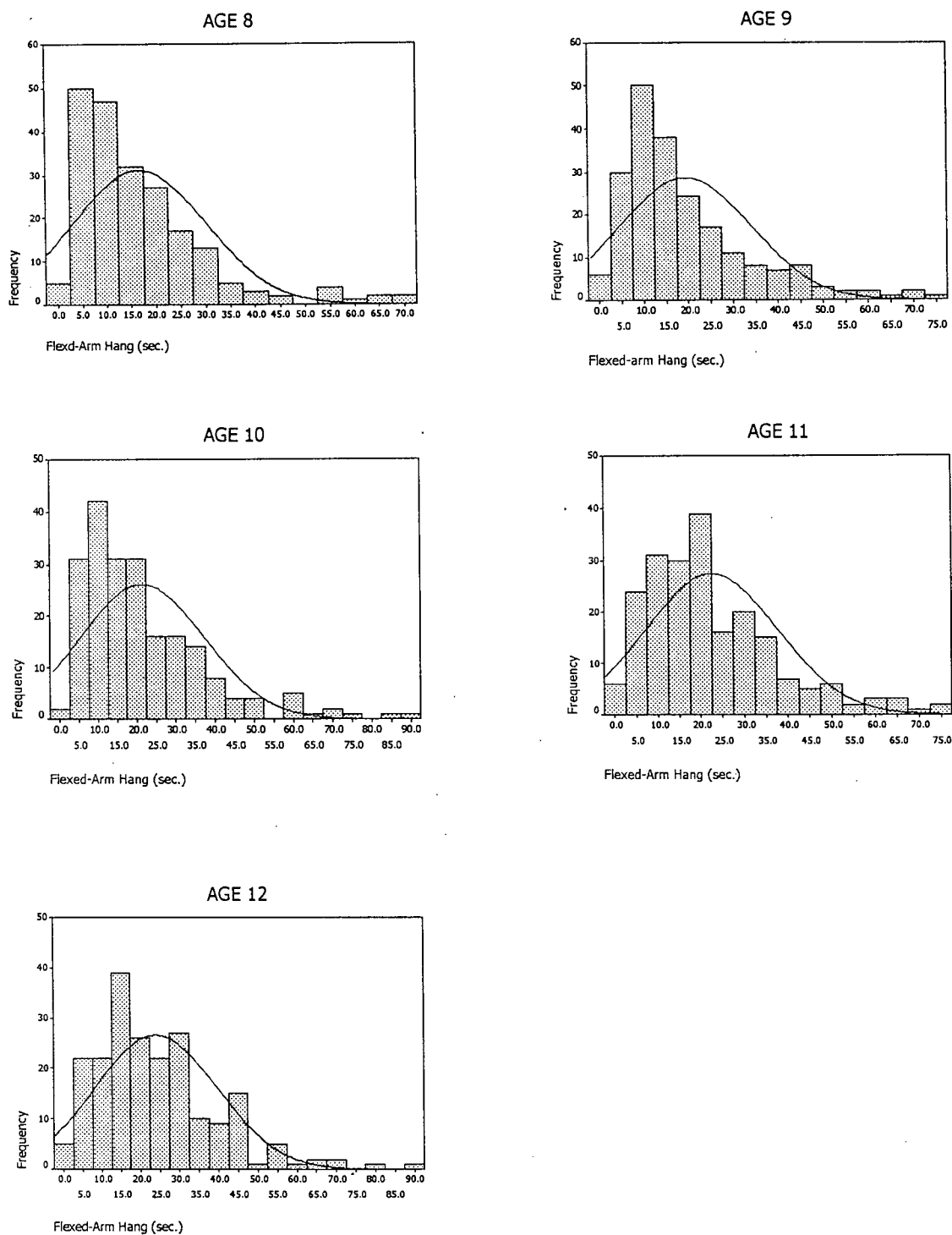


Figure C.1. Histograms for FAH scores at five time points

Table C.1

Correlation coefficients and distributional statistics for predictor variables

	PR_ME_NO	AGE	GRADE	ME_SESN	ME_YR
AGE	.058				
GRADE	-.103	.279			
ME-SESN	.157	-.055	-.657		
ME-YR	.444	-.022	-.140	.004	
Mean	4.84	96.33	.51	.50	1976.5
SD	2.00	1.93	.50	.50	5.50
Minimum	0	93	0	0	1968
Maximum	11	100	1	1	1992

Note. PR_ME_NO = the number of measurement before age 8 (initial time point); GRADE = grade at age 8; AGE = age in months at age 8; ME_SESN = measurement season; ME_YR = measurement year at age 8.

Jump-and-reach.

Table C.2

Correlation coefficients and distributional statistics for jump-and-reach

	Age 8	Age 9	Age 10	Age 11	Age 12
Age 9	.586				
Age 10	.519	.659			
Age 11	.524	.619	.693		
Age 12	.488	.599	.636	.698	
Mean (inch)	9.42	10.34	11.58	12.33	13.40
SD	1.78	1.81	1.89	1.85	2.07
Skewness	-.23	-.01	.11	-.12	.31
Kurtosis	.70	.81	.22	.09	.54

Table C.3

Parameter estimates (standard errors) of the best fitting growth model for jump-and-reach: Linear, equal error variances

	Intercept factor	Linear factor	Error variance	
<u>Mean</u>	9.43** (.113)	.994** (.031)	<u>Age 8</u>	1.22** (.069)
<u>Variance</u>	1.94** (.265)	.082** (.021)	<u>Age 9</u>	1.22** (.069)
<u>Covariance between factors</u>	-.025 (.056)		<u>Age 10</u>	1.22** (.069)
			<u>Age 11</u>	1.22** (.069)
			<u>Age 12</u>	1.22** (.069)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Sit-and-reach.

Table C.4

Correlation coefficients and distributional statistics for sit-and-reach

	Age 8	Age 9	Age 10	Age 11	Age 12
Age 9	.821				
Age 10	.799	.857			
Age 11	.769	.802	.826		
Age 12	.742	.766	.770	.812	
Mean (inch)	7.89	7.70	7.37	7.12	6.89
SD	2.29	2.20	2.22	2.34	2.51
Skewness	-.40	-.29	-.53	-.39	-.11
Kurtosis	.05	.51	.60	.08	-.05

Table C.5

Parameter estimates (standard errors) of the best fitting growth model for sit-and-reach: Linear, unequal error variances

	Intercept factor	Linear factor		Error variance
<u>Mean</u>	7.91** (.150)	-.260** (.028)	<u>Age 8</u>	1.05** (.154)
<u>Variance</u>	4.13** (.462)	.050** (.018)	<u>Age 9</u>	.720** (.099)
<u>Covariance between factors</u>	-.015 (.065)		<u>Age 10</u>	.782** (.101)
			<u>Age 11</u>	.970** (.126)
			<u>Age 12</u>	1.44** (.199)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Agility shuttle run.

Table C.6

Correlation coefficients and distributional statistics for agility shuttle run

	Age 8	Age 9	Age 10	Age 11	Age 12
Age 9	.585				
Age 10	.560	.634			
Age 11	.461	.581	.599		
Age 12	.516	.547	.577	.644	
Mean (sec.)	12.46	11.92	11.39	11.06	10.76
SD	1.05	.90	.82	.74	.75
Skewness	.79	.65	.91	.67	.65
Kurtosis	1.03	.46	1.37	.83	.68

Table C.7

Parameter estimates (standard errors) of the best fitting growth model for agility shuttle run:Unspecified Curve, unequal error variances

	Intercept factor	Curve factor		Factor loading	Error variance
<u>Mean</u>	12.45** (.073)	-.536** (.062)	<u>Age 8</u>	- (fixed)	.481** (.070)
<u>Variance</u>	.637** (.092)	.025* (.010)	<u>Age 9</u>	1.00 (fixed)	.297** (.039)
<u>Covariance between factors</u>	-.084** (.027)		<u>Age 10</u>	1.97** (.175)	.263** (.032)
			<u>Age 11</u>	2.61** (.239)	.208** (.027)
			<u>Age 12</u>	3.15** (.297)	.203** (.031)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Endurance shuttle run.

Table C.8

Correlation coefficients and distributional statistics for endurance shuttle run

	Age 8	Age 9	Age 10	Age 11	Age 12
Age 9	.631				
Age 10	.560	.631			
Age 11	.522	.647	.623		
Age 12	.521	.609	.655	.668	
Mean (sec.)	43.93	42.00	40.55	39.46	38.32
SD	3.64	3.04	3.03	2.68	2.67
Skewness	.89	.59	.86	.85	.66
Kurtosis	1.25	.30	1.42	1.21	.22

Table C.9

Parameter estimates (standard errors) of the best fitting growth model for endurance shuttle run:Unspecified Curve, unequal error variances

	Intercept factor	Curve factor		Factor loading	Error variance
<u>Mean</u>	43.93** (.253)	- 1.92** (.203)	<u>Age 8</u>	- (fixed)	5.40** (.787)
<u>Variance</u>	7.95** (1.10)	.359** (.138)	<u>Age 9</u>	1.00 (fixed)	2.91** (.391)
<u>Covariance between factors</u>	- 1.04** (.328)		<u>Age 10</u>	1.77** (.142)	3.53** (.409)
			<u>Age 11</u>	2.33** (.189)	2.49** (.317)
			<u>Age 12</u>	2.92** (.245)	2.07** (.351)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

30-yard dash.

Table C.10

Correlation coefficients and distributional statistics for 30-yard dash

	Age 8	Age 9	Age 10	Age 11	Age 12
Age 9	.635				
Age 10	.599	.577			
Age 11	.563	.565	.679		
Age 12	.617	.550	.579	.685	
Mean (sec.)	5.21	4.93	4.75	4.65	4.50
SD	.44	.40	.36	.35	.33
Skewness	.59	.82	.40	.52	.47
Kurtosis	.90	1.21	.21	.50	.03

Table C.11

Parameter estimates (standard errors) of the best fitting growth model for 30-yard dash: Unspecified Curve, unequal error variances

	Intercept factor	Curve factor		Factor loading	Error variance
<u>Mean</u>	5.21** (.030)	-.275** (.026)	<u>Age 8</u>	- (fixed)	.063** (.012)
<u>Variance</u>	.125** (.018)	.005* (.003)	<u>Age 9</u>	1.00 (fixed)	.069** (.008)
<u>Covariance between factors</u>	-.017** (.006)		<u>Age 10</u>	1.68** (.131)	.051** (.006)
			<u>Age 11</u>	2.04** (.157)	.042** (.005)
			<u>Age 12</u>	2.59** (.202)	.035** (.006)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Standing long jump.

Table C.12

Correlation coefficients and distributional statistics for standing long jump

	Age 8	Age 9	Age 10	Age 11	Age 12
Age 9	.751				
Age 10	.723	.826			
Age 11	.655	.719	.800		
Age 12	.657	.723	.753	.770	
Mean (inch)	53.38	57.46	61.40	64.48	67.76
SD	7.56	7.61	7.21	6.49	6.91
Skewness	-.76	-.62	-.58	-.39	-.25
Kurtosis	.65	.45	.31	-.21	.19

Table C.13

Parameter estimates (standard errors) of the best fitting growth model for standing long jump: Cubic, equal error variances

	Intercept factor	Linear factor	Quadratic factor	Cubic factor		Error variance
<u>Mean</u>	53.36** (.520)	4.52** (.648)	-.344 (.396)	.028 (.064)	<u>Age 8</u>	7.89** (.772)
<u>Variance</u>	48.79** (5.59)	37.55** (9.91)	12.44** (3.77)	.300** (.099)	<u>Age 9</u>	7.89** (.772)
<u>Covariance between factors</u>					<u>Age 10</u>	7.89** (.772)
Linear factor	-2.49 (5.11)				<u>Age 11</u>	7.89** (.772)
Quadratic factor	-2.76 (3.02)	-20.28** (5.92)			<u>Age 12</u>	7.89** (.772)
Cubic factor	.613 (.483)	2.91** (.920)	-1.90** (.604)			

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Data Set 2

Table C.14

Correlation coefficients and distributional statistics for predictor variables

	PR_ME_NO	AGE	GRADE	ME_SESN	ME_YR
AGE	.065				
GRADE	-.112	.391			
ME-SESN	-.003	-.176	-.433		
ME-YR	.474	-.113	-.208	.029	
Mean	5.81	102.47	3.03	.46	1977.5
SD	2.18	1.95	.43	.50	5.89
Minimum	0	99	2	0	1968
Maximum	11	106	4	1	1992

Note. PR_ME_NO = the number of measurement before age 8.5 (initial time point); GRADE = grade at age 8.5; AGE = age in months at age 8.5; ME_SESN = measurement season; ME_YR = measurement year at age 8.5.

Flexed-arm-hang.

Table C.15

Correlation coefficients and distributional statistics for flexed-arm-hang

	Age 8.5	Age 9.5	Age 10.5	Age 11.5	Age 12.5
Age 9.5	.815				
Age 10.5	.746	.829			
Age 11.5	.721	.790	.856		
Age 12.5	.623	.686	.778	.856	
Mean (sec.)	17.60	20.09	22.46	23.71	24.31
SD	13.85	14.86	15.69	16.42	16.71
Skewness	1.83	1.58	1.21	1.08	1.11
Kurtosis	3.82	2.98	1.25	.76	.95

Table C.16

Parameter estimates (standard errors) of the best fitting growth model for flexed-arm-hang: Quadratic, equal error variances

	Intercept factor	Linear factor	Quadratic factor		Error variance
<u>Mean</u>	17.53** (.963)	3.10** (.558)	-.349** (.130)	<u>Age 8.5</u>	32.74** (2.30)
<u>Variance</u>	159.09** (18.78)	22.58** (6.90)	1.12** (.380)	<u>Age 9.5</u>	32.74** (2.30)
<u>Covariance between factors</u>				<u>Age 10.5</u>	32.74** (2.30)
Linear factor	10.35 (7.93)			<u>Age 11.5</u>	32.74** (2.30)
Quadratic factor	- 3.59* (1.82)	- 4.08** (1.55)		<u>Age 12.5</u>	32.74** (2.30)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Jump-and-reach.

Table C.17

Correlation coefficients and distributional statistics for jump-and-reach

	Age 8.5	Age 9.5	Age 10.5	Age 11.5	Age 12.5
Age 9.5	.636				
Age 10.5	.614	.715			
Age 11.5	.527	.649	.755		
Age 12.5	.518	.587	.631	.691	
Mean (inch)	9.80	10.90	11.98	12.81	14.01
SD	1.90	1.83	1.85	1.88	2.31
Skewness	.02	-.01	-.20	-.23	.10
Kurtosis	.17	-.20	.01	-.12	.02

Table C.18

Parameter estimates (standard errors) of the best fitting growth model for jump-and-reach: Linear, unequal error variances

	Intercept factor	Linear factor	Error variance	
<u>Mean</u>	9.85** (.122)	1.02** (.034)	<u>Age 8.5</u>	1.48** (.214)
<u>Variance</u>	2.23** (.308)	.087** (.026)	<u>Age 9.5</u>	1.04** (.136)
<u>Covariance between factors</u>	-.038 (.068)		<u>Age 10.5</u>	.860** (.114)
			<u>Age 11.5</u>	.886** (.132)
			<u>Age 12.5</u>	2.01** (.269)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Sit-and-reach.

Table C.19

Correlation coefficients and distributional statistics for sit-and-reach

	Age 8.5	Age 9.5	Age 10.5	Age 11.5	Age 12.5
Age 9.5	.797				
Age 10.5	.780	.836			
Age 11.5	.760	.790	.828		
Age 12.5	.704	.764	.775	.844	
Mean (inch)	7.96	7.59	7.38	7.24	7.18
SD	2.17	2.26	2.33	2.48	2.52
Skewness	-.30	-.34	-.43	-.45	-.29
Kurtosis	.10	-.03	.52	-.10	.00

Table C.20

Parameter estimates (standard errors) of the best fitting growth model for sit-and-reach: Linear, equal error variances

	Intercept factor	Linear factor		Error variance
<u>Mean</u>	7.86** (.149)	-.192** (.030)	<u>Age 8.5</u>	.972** (.056)
<u>Variance</u>	3.93** (.449)	.091** (.020)	<u>Age 9.5</u>	.972** (.056)
<u>Covariance between factors</u>	.027 (.067)		<u>Age 10.5</u>	.972** (.056)
			<u>Age 11.5</u>	.972** (.056)
			<u>Age 12.5</u>	.972** (.056)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Agility shuttle run.

Table C.21

Correlation coefficients and distributional statistics for agility shuttle run

	Age 8.5	Age 9.5	Age 10.5	Age 11.5	Age 12.5
Age 9.5	.603				
Age 10.5	.599	.654			
Age 11.5	.650	.615	.701		
Age 12.5	.526	.596	.590	.753	
Mean (sec.)	12.14	11.59	11.19	10.88	10.56
SD	1.00	.80	.86	.76	.72
Skewness	1.07	.41	.66	.80	.47
Kurtosis	1.95	-.07	.78	.78	.06

Table C.22

Parameter estimates (standard errors) of the best fitting growth model for agility shuttle run:Unspecified Curve, unequal error variances

	Intercept factor	Curve factor		Factor loading	Error variance
<u>Mean</u>	12.14** (.069)	-.552** (.056)	<u>Age 8.5</u>	- (fixed)	.368** (.057)
<u>Variance</u>	.601** (.084)	.029** (.011)	<u>Age 9.5</u>	1.00 (fixed)	.249** (.032)
<u>Covariance between factors</u>	-.075 (.025)		<u>Age 10.5</u>	1.70** (.137)	.278** (.032)
			<u>Age 11.5</u>	2.28** (.180)	.131** (.019)
			<u>Age 12.5</u>	2.86** (.236)	.143** (.024)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Endurance shuttle run.

Table C.23

Correlation coefficients and distributional statistics for endurance shuttle run

	Age 8.5	Age 9.5	Age 10.5	Age 11.5	Age 12.5
Age 9.5	.710				
Age 10.5	.685	.727			
Age 11.5	.590	.579	.686		
Age 12.5	.550	.537	.678	.720	
Mean (sec.)	42.86	41.26	39.97	38.94	37.64
SD	3.35	3.22	2.90	2.81	2.54
Skewness	.99	.57	.75	.93	.62
Kurtosis	1.47	.02	.28	1.34	.47

Table C.24

Parameter estimates (standard errors) of the best fitting growth model for endurance shuttle run: Linear, unequal error variances

	Intercept factor	Linear factor		Error variance
<u>Mean</u>	42.67** (.227)	- 1.27** (.050)	<u>Age 8.5</u>	2.97** (.520)
<u>Variance</u>	8.74** (1.05)	.293** (.057)	<u>Age 9.5</u>	3.33** (.422)
<u>Covariance between factors</u>	- 1.03** (.203)		<u>Age 10.5</u>	2.24** (.280)
			<u>Age 11.5</u>	2.61** (.319)
			<u>Age 12.5</u>	1.33** (.319)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

30-yard dash.

Table C.25

Correlation coefficients and distributional statistics for 30-yard dash

	Age 8.5	Age 9.5	Age 10.5	Age 11.5	Age 12.5
Age 9.5	.720				
Age 10.5	.693	.719			
Age 11.5	.649	.664	.668		
Age 12.5	.620	.579	.576	.651	
Mean (sec.)	5.06	4.88	4.73	4.56	4.44
SD	.41	.41	.37	.35	.34
Skewness	.84	.80	.57	.68	.30
Kurtosis	1.13	.74	.82	.47	-.13

Table C.26

Parameter estimates (standard errors) of the best fitting growth model for 30-yard dash: Linear, equal error variances

	Intercept factor	Linear factor		Error variance
<u>Mean</u>	5.05** (.028)	-.156** (.006)	<u>Age 8.5</u>	.045** (.003)
<u>Variance</u>	.132** (.016)	.002** (.001)	<u>Age 9.5</u>	.045** (.003)
<u>Covariance between factors</u>	-.011** (.003)		<u>Age 10.5</u>	.045** (.003)
			<u>Age 11.5</u>	.045** (.003)
			<u>Age 12.5</u>	.045** (.003)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Standing long jump.

Table C.27

Correlation coefficients and distributional statistics for standing long jump

	Age 8.5	Age 9.5	Age 10.5	Age 11.5	Age 12.5
Age 9.5	.833				
Age 10.5	.777	.836			
Age 11.5	.692	.716	.749		
Age 12.5	.634	.684	.681	.694	
Mean (inch)	55.44	59.29	62.47	66.01	69.42
SD	7.10	7.33	7.16	7.30	7.17
Skewness	-.51	-.35	-.13	-.47	-.21
Kurtosis	.27	-.21	-.33	.39	-.42

Table C.28

Parameter estimates (standard errors) of the best fitting growth model for standing long jump: Linear, unequal error variances

	Intercept factor	Linear factor		Error variance
<u>Mean</u>	55.59** (.503)	3.47** (.100)	<u>Age 8.5</u>	8.44** (1.60)
<u>Variance</u>	45.96** (5.14)	.838** (.236)	<u>Age 9.5</u>	8.51** (1.20)
<u>Covariance between factors</u>	-2.66** (.826)		<u>Age 10.5</u>	10.30** (1.30)
			<u>Age 11.5</u>	15.93** (1.95)
			<u>Age 12.5</u>	16.53** (2.47)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Multivariate Results

Data Set 1

Descriptive statistics for "Power" (JAR, SLJ, DASH).

Table C.29

Correlation coefficients and descriptive statistics for the JAR, SLJ and DASH across all time points

	<u>Age 8</u>			<u>Age 9</u>			<u>Age 10</u>			<u>Age 11</u>			<u>Age 12</u>		
	JAR	SLJ	DASH	JAR	SLJ	DASH	JAR	SLJ	DASH	JAR	SLJ	DASH	JAR	SLJ	DASH
<u>Age 8</u>	JAR	1.00													
	SLJ	.63	1.00												
	DASH	-.55	-.67	1.00											
<u>Age 9</u>	JAR	.59	.64	1.00											
	SLJ	.55	.75	.67	1.00										
	DASH	-.34	-.49	-.49	-.51	1.00									
<u>Age 10</u>	JAR	.52	.55	.66	.54	1.00									
	SLJ	.53	.72	.65	.83	.42	1.00								
	DASH	-.43	-.56	-.54	-.54	.58	-.49	1.00							
<u>Age 11</u>	JAR	.52	.54	.62	.55	-.43	.69	.68	1.00						
	SLJ	.51	.66	.66	.72	-.46	.62	.80	.72	1.00					
	DASH	-.41	-.47	-.49	-.48	.57	-.50	-.55	-.58	-.60	1.00				
<u>Age 12</u>	JAR	.49	.53	.60	.52	-.40	.64	.60	.70	.59	-.54	1.00			
	SLJ	.51	.66	.65	.72	-.49	.62	.75	.64	.77	-.59	.71	1.00		
	DASH	-.46	-.49	-.48	-.51	.55	-.51	-.51	-.56	-.54	.69	-.59	-.62	1.00	
	Mean	9.42	53.38	5.21	10.34	57.46	4.93	11.58	61.40	4.75	12.33	64.48	4.65	13.40	67.76
	SD	1.78	7.56	.44	1.81	7.61	.40	1.89	7.21	.36	1.85	6.49	.35	2.08	6.91
															.33

Note. JAR = jump-and-reach; SLJ = standing long jump; DASH = 30-yard dash.

Descriptive statistics for "Motor Ability" (FAH, SLJ, SAR, DASH, ESR).

Table C.30

Correlation coefficients and descriptive statistics for the FAH, SLJ, SAR, DASH and ESR across all time points

	Age 8					Age 9					Age 10					Age 11					Age 12					
	FAH	SLJ	SAR	DASH	ESR	FAH	SLJ	SAR	DASH	ESR	FAH	SLJ	SAR	DASH	ESR	FAH	SLJ	SAR	DASH	ESR	FAH	SLJ	SAR	DASH	ESR	
Age 8	FAH	1.00																								
	SLJ	.27	1.00																							
	SAR	.25	.42	1.00																						
	DASH	-.28	-.67	-.23	1.00																					
Age 8	ESR	-.31	-.60	-.28	.65	1.00																				
	FAH	.78	.29	.26	-.27	-.30	1.00																			
	SLJ	.31	.75	.42	-.59	-.56	.33	1.00																		
	SAR	.22	.34	.82	-.20	-.26	.27	.35	1.00																	
Age 9	DASH	-.28	-.49	-.18	.64	.46	-.33	-.51	-.20	1.00																
	ESR	-.40	-.53	-.29	.54	.63	-.43	-.61	-.25	.59	1.00															
	FAH	.70	.33	.24	-.28	-.32	.84	.33	.23	-.29	-.42	1.00														
	SLJ	.30	.72	.47	-.54	-.55	.33	.83	.38	-.42	-.54	.37	1.00													
Age 10	SAR	.18	.32	.80	-.17	-.21	.19	.33	.86	-.15	-.23	.18	.37	1.00												
	DASH	-.31	-.56	-.23	.60	.47	-.30	-.54	-.22	.58	.50	-.35	-.57	-.19	1.00											
	ESR	-.30	-.52	-.28	.51	.56	-.32	-.51	-.24	.49	.63	-.35	-.56	-.25	.56	1.00										
	FAH	.63	.33	.26	-.25	-.31	.77	.39	.23	-.32	-.46	.79	.40	.21	-.37	-.36	1.00									
Age 11	SLJ	.36	.66	.47	-.53	-.49	.38	.72	.40	-.46	-.52	.39	.80	.39	-.56	-.51	.44	1.00								
	SAR	.21	.30	.77	-.17	-.18	.24	.30	.80	-.19	-.23	.21	.35	.83	-.20	-.23	.27	.42	1.00							
	DASH	-.25	-.47	-.19	.56	.37	-.26	-.48	-.19	.57	.47	-.28	-.55	-.18	.68	.47	-.33	-.60	-.25	1.00						
	ESR	-.32	-.47	-.30	.45	.52	-.35	-.51	-.25	.53	.65	-.35	-.53	-.25	.47	.62	-.40	-.60	-.25	.54	1.00					
Age 12	FAH	.65	.34	.34	-.25	-.28	.72	.41	.29	-.33	-.41	.71	.43	.26	-.35	-.35	.80	.45	.30	-.35	-.44	1.00				
	SLJ	.31	.66	.43	-.57	-.46	.31	.72	.33	-.49	-.57	.32	.75	.32	-.58	-.55	.40	.77	.34	-.59	-.57	.47	1.00			
	SAR	.20	.24	.74	-.15	-.20	.21	.24	.77	-.14	-.22	.17	.26	.77	-.17	-.19	.21	.35	.81	-.18	-.25	.26	.34	1.00		
	DASH	-.30	-.49	-.23	.62	.43	-.29	-.51	-.17	.55	.46	-.27	-.51	-.11	.58	.46	-.33	-.54	-.24	.69	.46	-.40	-.62	-.18	1.00	
Age 12	ESR	-.30	-.51	-.28	.58	.52	-.32	-.56	-.25	.55	.61	-.30	-.55	-.24	.52	.66	-.38	-.58	-.24	.57	.67	-.44	-.66	-.23	.63	1.00
	Mean	16.64	53.38	7.89	5.21	43.93	19.29	57.46	7.70	4.93	42.00	21.35	61.40	7.37	4.75	40.55	22.46	64.48	7.12	4.65	39.46	23.91	67.76	6.89	4.50	38.32
SD	13.40	7.56	2.29	.44	3.64	14.62	7.61	2.20	.40	3.04	15.97	7.21	2.22	.36	3.03	15.23	6.49	2.34	.35	2.68	15.73	6.91	2.51	.33	2.67	

Note. FAH = flexed-arm-hang; SLJ = standing long jump; SAR = sit-and-reach; DASH = 30-yard dash; ESR = endurance shuttle run.

Data Set 2

Descriptive statistics and parameter estimates for "Run" (ASR, ESR, DASH).

Table C.31

Correlation coefficients and descriptive statistics for the JAR, SLJ and DASH across all time points

	Age 8.5			Age 9.5			Age 10.5			Age 11.5			Age 12.5		
	ASR	ESR	DASH	ASR	ESR	DASH	ASR	ESR	DASH	ASR	ESR	DASH	ASR	ESR	DASH
Age 8.5	ASR	1.00													
	ESR	.76	1.00												
	DASH	.65	.66	1.00											
Age 9.5	ASR	.60	.61	.56	1.00										
	ESR	.66	.71	.59	.79	1.00									
	DASH	.52	.54	.72	.71	.66	1.00								
Age 10.5	ASR	.60	.59	.54	.65	.62	.59	1.00							
	ESR	.64	.69	.62	.70	.73	.62	.80	1.00						
	DASH	.50	.58	.69	.56	.55	.92	.68	.65	1.00					
Age 11.5	ASR	.65	.60	.62	.62	.63	.59	.70	.74	.59	1.00				
	ESR	.59	.59	.56	.56	.58	.55	.61	.69	.53	.75	1.00			
	DASH	.48	.49	.65	.56	.50	.66	.56	.57	.67	.62	.59	1.00		
Age 12.5	ASR	.53	.45	.55	.60	.54	.56	.59	.62	.54	.75	.64	.54	1.00	
	ESR	.55	.55	.53	.55	.54	.53	.64	.68	.55	.74	.72	.65	.82	1.00
	DASH	.50	.49	.62	.49	.48	.58	.49	.51	.58	.60	.54	.54	.67	.68
Mean	ASR	12.14	42.86	5.06	11.59	41.26	4.88	11.19	39.97	4.73	10.88	38.94	4.56	10.56	37.64
	ESR	1.00	3.35	.41	.80	3.22	.41	.86	2.90	.37	.76	2.81	.35	.72	2.54
SD															

Note. ASR = agility shuttle run; ESR = endurance shuttle run; DASH = 30-yard dash.

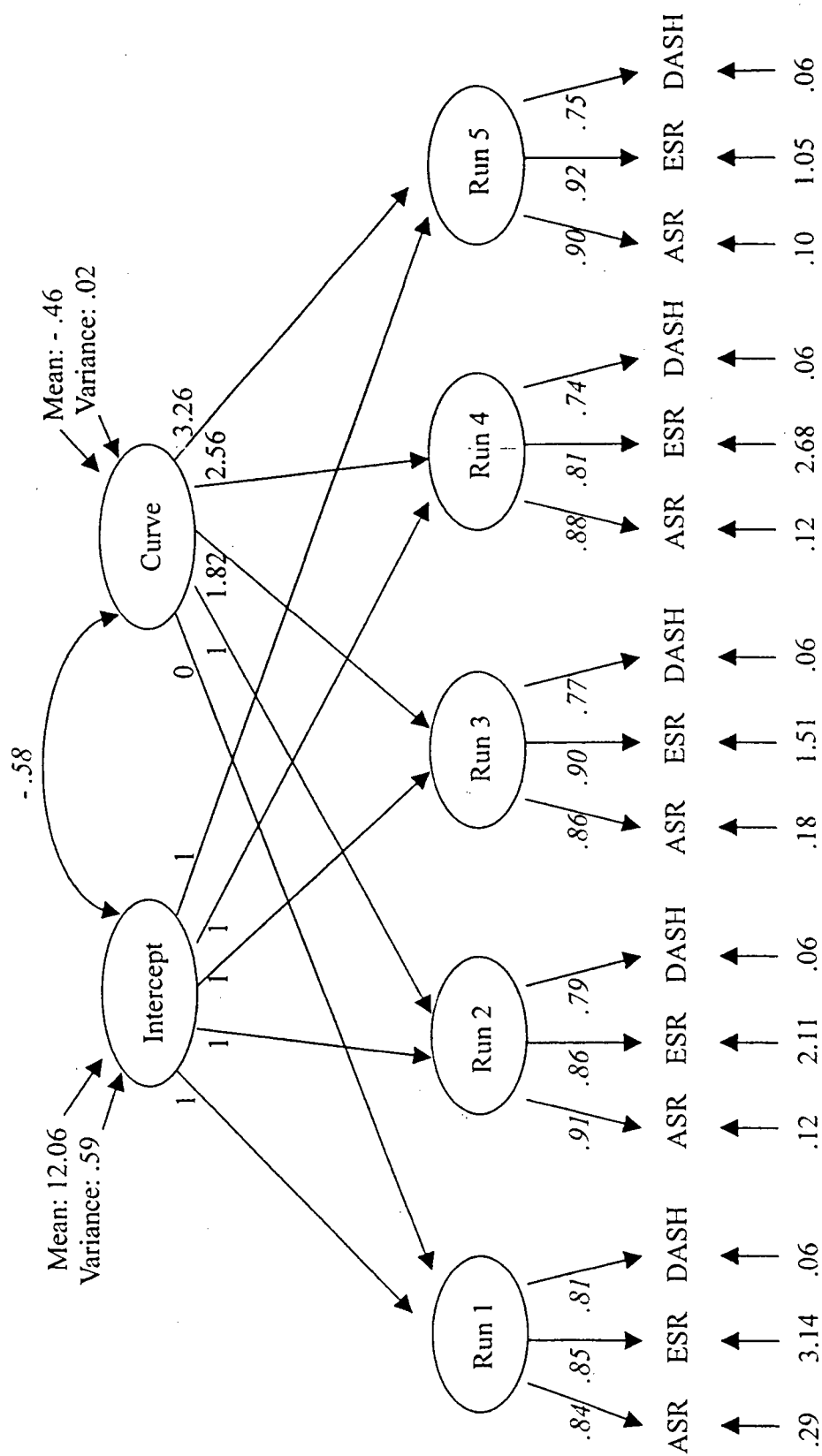


Figure C.2. Curve-of-factors model for "Run"

Note. The numbers in italics are standardized values. Correlated errors are omitted.

Descriptive statistics and parameter estimates for "Power" (JAR, SLJ, DASH).

Table C.32

Correlation coefficients and descriptive statistics for the JAR, SLJ and DASH across all time points

	Age 8.5			Age 9.5			Age 10.5			Age 11.5			Age 12.5		
	JAR	SLJ	DASH	JAR	SLJ	DASH	JAR	SLJ	DASH	JAR	SLJ	DASH	JAR	SLJ	DASH
<u>Age 8.5</u>	JAR	1.00													
	SLJ	.63	1.00												
	DASH	-.53	-.65	1.00											
<u>Age 9.5</u>	JAR	.64	.55	-.52	1.00										
	SLJ	.59	.83	-.64	.67	1.00									
	DASH	-.48	-.58	.72	-.55	-.62	1.00								
<u>Age 10.5</u>	JAR	.61	.60	-.55	.72	.64	-.52	1.00							
	SLJ	.56	.78	-.59	.61	.84	-.58	.67	1.00						
	DASH	-.51	-.63	.69	-.58	-.64	.72	-.65	-.70	1.00					
<u>Age 11.5</u>	JAR	.53	.54	-.54	.65	.61	-.53	.76	.62	-.62	1.00				
	SLJ	.56	.69	-.57	.60	.72	-.59	.66	.75	-.64	.64	1.00			
	DASH	-.42	-.52	.65	-.55	-.60	.66	-.53	-.57	.67	-.61	-.65	1.00		
<u>Age 12.5</u>	JAR	.52	.49	-.45	.58	.54	-.43	.63	.53	-.51	.69	.55	-.50	1.00	
	SLJ	.55	.63	-.50	.55	.68	-.51	.62	.68	-.57	.65	.70	-.55	.77	1.00
	DASH	-.47	-.54	.62	-.48	-.57	.58	-.53	-.55	.58	-.62	-.56	.65	-.68	-.69
															1.00
Mean	9.80	55.44	5.06	10.90	59.29	4.88	11.98	62.47	4.73	12.81	66.01	4.56	14.01	69.42	4.44
SD	1.90	7.10	.41	1.83	7.33	.41	1.85	7.16	.37	1.88	7.30	.35	2.31	7.17	.34

Note. JAR = jump-and-reach; SLJ = standing long jump; DASH = 30-yard dash.

Table C.33

Parameter estimates of the 5-factor model with correlated errors and the equality of factor loadings over time for "Power"

Standardized factor loading			Correlations of factors between time points				
Time	Variables	Loading	Age 8.5	Age 9.5	Age 10.5	Age 11.5	Age 12.5
Age 8.5	JAR	.77					
	SLJ	.84	1.00	.92	.90	.84	.77
	DASH	-.68					
Age 9.5	JAR	.81					
	SLJ	.84		1.00	.93	.93	.78
	DASH	-.68					
Age 10.5	JAR	.82					
	SLJ	.84			1.00	.94	.80
	DASH	-.77					
Age 11.5	JAR	.80					
	SLJ	.82				1.00	.85
	DASH	-.75					
Age 12.5	JAR	.78					
	SLJ	.90					1.00
	DASH	-.81					

Note. Correlated errors are omitted. All estimates were significant at an alpha level of .01.

Descriptive statistics and parameter estimates for "Motor Ability" (FAH, SLJ, SAR, DASH, ESR).

Table C.34

Correlation coefficients and descriptive statistics for the FAH, SLJ, SAR, DASH and ESR across all time points

	Age 8.5					Age 9.5					Age 10.5					Age 11.5					Age 12.5					
	FAH	SLJ	SAR	DASH	ESR	FAH	SLJ	SAR	DASH	ESR	FAH	SLJ	SAR	DASH	ESR	FAH	SLJ	SAR	DASH	ESR	FAH	SLJ	SAR	DASH	ESR	
Age 8.5	FAH	1.00																								
	SLJ	.33	1.00																							
	SAR	.25	.42	1.00																						
	DASH	-.38	-.65	-.30	1.00																					
	ESR	-.36	-.66	-.32	.66	1.00																				
Age 9.5	FAH	.82	.34	.37	-.40	-.41	1.00																			
	SLJ	.31	.83	.45	-.64	-.59	.39	1.00																		
	SAR	.21	.36	.80	-.26	-.28	.32	.42	1.00																	
	DASH	-.30	-.58	-.30	.72	.54	-.38	-.62	-.27	1.00																
	ESR	-.30	-.63	-.35	.59	.71	-.38	-.62	-.28	.66	1.00															
Age 10.5	FAH	.75	.37	.30	-.40	-.38	.83	.42	.27	-.39	-.37	1.00														
	SLJ	.30	.78	.47	-.59	-.57	.34	.84	.40	-.58	-.60	.42	1.00													
	SAR	.22	.35	.78	-.29	-.28	.29	.42	.84	-.28	-.28	.27	.47	1.00												
	DASH	-.31	-.63	-.31	.69	.58	-.35	-.64	-.29	.72	.55	-.42	-.70	-.32	1.00											
	ESR	-.33	-.60	-.35	.62	.69	-.38	-.61	-.30	.62	.73	-.41	-.66	-.31	.65	1.00										
Age 11.5	FAH	.72	.36	.28	-.36	-.35	.79	.39	.25	-.32	-.35	.86	.39	.26	-.36	-.41	1.00									
	SLJ	.34	.69	.37	-.57	-.55	.40	.72	.31	-.59	-.56	.44	.75	.35	-.64	-.58	.46	1.00								
	SAR	.21	.27	.76	-.27	-.23	.28	.33	.79	-.24	-.24	.22	.35	.83	-.27	-.25	.23	.31	1.00							
	DASH	-.33	-.52	-.27	.65	.49	-.38	-.60	-.30	.66	.50	-.40	-.57	-.31	.67	.57	-.41	-.65	-.28	1.00						
	ESR	-.26	-.55	-.29	.56	.59	-.32	-.56	-.26	.55	.58	-.36	-.55	-.29	.53	.69	-.39	-.56	-.25	.59	1.00					
Age 12.5	FAH	.62	.35	.31	-.37	-.34	.69	.40	.29	-.36	-.35	.78	.40	.28	-.40	-.40	.86	.49	.30	-.46	-.42	1.00				
	SLJ	.27	.63	.36	-.50	-.45	.34	.68	.33	-.51	-.48	.37	.68	.33	-.57	-.50	.39	.69	.33	-.55	-.49	.47	1.00			
	SAR	.23	.24	.70	-.27	-.20	.30	.29	.76	-.25	-.21	.29	.32	.78	-.29	-.27	.27	.32	.84	-.28	-.27	.34	.38	1.00		
	DASH	-.21	-.54	-.29	.62	.49	-.28	-.57	-.25	.58	.48	-.31	-.55	-.27	.58	.51	-.36	-.56	-.29	.65	.54	-.44	-.69	-.31	1.00	
	ESR	-.31	-.53	-.30	.53	.55	-.35	-.55	-.24	.53	.54	-.42	-.51	-.20	.55	.68	-.43	-.56	-.22	.54	.72	-.51	-.65	-.28	.68	1.00
Mean	17.60	55.44	7.96	5.06	42.86		20.09	59.29	7.59	4.88	41.26	22.46	62.47	7.38	4.73	39.97	23.71	66.01	7.24	4.56	38.94	24.31	69.42	7.18	4.44	37.64
SD	13.85	7.10	2.17	.41	3.35		14.86	7.33	2.26	.41	3.22	15.69	7.16	2.33	.37	2.90	16.42	7.30	2.48	.35	2.81	16.71	7.17	2.52	.34	2.54

Note. FAH = flexed-arm-hang; SLJ = standing long jump; SAR = sit-and-reach; DASH = 30-yard dash; ESR = endurance shuttle run.

Table C.35

Parameter estimates of the 5-factor model with correlated errors and the equality of factor loadings over time for "Motor Ability"

Standardized factor loading			Correlations of factors between time points				
Time	Variables	Loading	Age 8.5	Age 9.5	Age 10.5	Age 11.5	Age 12.5
Age 8.5	FAH	.47					
	SLJ	.86					
	SAR	.43	1.00	.94	.90	.87	.76
	DASH	-.77					
	ESR	-.73					
Age 9.5	FAH	.46					
	SLJ	.85					
	SAR	.41		1.00	.93	.92	.80
	DASH	-.76					
	ESR	-.74					
Age 10.5	FAH	.44					
	SLJ	.86					
	SAR	.39			1.00	.92	.79
	DASH	-.82					
	ESR	-.79					
Age 11.5	FAH	.39					
	SLJ	.80					
	SAR	.34				1.00	.88
	DASH	-.78					
	ESR	-.74					
Age 12.5	FAH	.40					
	SLJ	.82					
	SAR	.34					1.00
	DASH	-.84					
	ESR	-.83					

Note. Correlated errors are omitted. All estimates were significant at an alpha level of .01.

Appendix D: Descriptive Statistics and Parameter Estimates for Generated Data Sets

Condition A1: $r_{ig} = 0$, true reliability = .65 ~ .75

Table D.1

Correlation coefficients and distributional statistics for the data set of condition A1

	Time 1	Time 2	Time 3	Time 4	Time 5
Time 2	.686				
Time 3	.651	.741			
Time 4	.606	.707	.744		
Time 5	.547	.668	.719	.740	
Mean	9.426	10.420	11.414	12.408	13.402
SD	1.796	1.704	1.800	1.940	2.110

Condition A2: $r_{ig} = -.30$, true reliability = .65 ~ .75

Table D.2

Correlation coefficients and distributional statistics for the data set of condition A2

	Time 1	Time 2	Time 3	Time 4	Time 5
Time 2	.686				
Time 3	.638	.737			
Time 4	.564	.691	.739		
Time 5	.485	.623	.693	.738	
Mean	9.426	10.420	11.414	12.408	13.402
SD	1.787	1.583	1.593	1.662	1.775

Table D.3

Parameter estimates (standard errors) of the Linear model for the data set of condition A2

	Intercept factor	Linear factor		Error variance
<u>Mean</u>	9.426** (.023)	.994** (.006)	<u>Age 8</u>	1.128** (.033)
<u>Variance</u>	2.081** (.053)	.085** (.004)	<u>Age 9</u>	.605** (.018)
<u>Covariance between factors</u>	-.132** (12.158)		<u>Age 10</u>	.631** (.016)
			<u>Age 11</u>	.683** (.019)
			<u>Age 12</u>	.789** (.027)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Table D.4

Parameter estimates (standard errors) of the Simplex 2 model for the data set of condition A2

Parameter	Time 1	Time 2	Time 3	Time 4	Time 5
β		.710** (.011)	.916** (.014)	.954** (.013)	.984** (.014)
Standardized β		.821	.911	.921	.907
Factor mean	9.426** (.025)	3.725** (.109)	1.866** (.147)	1.523** (.152)	1.198** (.171)
Error variance of the factor	2.732** (.068)	.665** (.033)	.350** (.023)	.336** (.023)	.459** (.039)
Error variance of the observed variable	.462** (.023)	.462** (.023)	.473** (.019)	.548** (.022)	.548** (.022)

Note. All parameter estimates were significant at an alpha level of .05. β = regression coefficient predicting time 2 factor from time 1 factor, predicting time 3 factor from time 2 factor and so on.

Condition A3: $r_{ig} = -.60$, true reliability = .65 ~ .75

Table D.5

Correlation coefficients and distributional statistics for the data set of condition A3

	Time 1	Time 2	Time 3	Time 4	Time 5
Time 2	.690				
Time 3	.644	.726			
Time 4	.553	.669	.727		
Time 5	.447	.570	.664	.734	
Mean	9.426	10.420	11.414	12.408	13.402
SD	1.779	1.484	1.366	1.323	1.379

Table D.6

Parameter estimates (standard errors) of the Linear model for the data set of condition A3

	Intercept factor	Linear factor	Error variance	
<u>Mean</u>	9.426** (.022)	.994** (.005)	<u>Age 8</u>	1.095** (.031)
<u>Variance</u>	2.048** (.051)	.083** (.003)	<u>Age 9</u>	.558** (.016)
<u>Covariance</u> <u>between factors</u>	-.245** (.011)		<u>Age 10</u>	.473** (.012)
			<u>Age 11</u>	.431** (.012)
			<u>Age 12</u>	.466** (.018)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Table D.7

Parameter estimates (standard errors) of the Simplex 2 model for the data set of condition A3

Parameter	Time 1	Time 2	Time 3	Time 4	Time 5
β	.669** (.010)	.843** (.013)	.862** (.013)	.927** (.014)	
Standardized β	.832	.902	.889	.882	
Factor mean	9.426** (.025)	4.112** (.100)	2.635** (.137)	2.570** (.144)	1.901** (.169)
Error variance of the factor	2.722** (.067)	.541** (.028)	.287** (.018)	.303** (.017)	.353** (.026)
Error variance of the observed variable	.443** (.021)	.443** (.021)	.329** (.015)	.307** (.015)	.307** (.015)

Note. All parameter estimates were significant at an alpha level of .05. β = regression coefficient predicting time 2 factor from time 1 factor, predicting time 3 factor from time 2 factor and so on.

Condition B1: $r_{ij} = 0$, true reliability = .40 ~ .50

Table D.8

Correlation coefficients and distributional statistics for the data set of condition B1

	Time 1	Time 2	Time 3	Time 4	Time 5
Time 2	.443				
Time 3	.409	.495			
Time 4	.353	.462	.498		
Time 5	.302	.406	.454	.491	
Mean	9.426	10.420	11.414	12.408	13.402
SD	2.278	1.937	1.952	2.036	2.171

Table D.9

Parameter estimates (standard errors) of the Linear model for the data set of condition B1

	Intercept factor	Linear factor		Error variance
<u>Mean</u>	9.426** (.027)	.994** (.008)	<u>Age 8</u>	3.095** (.083)
<u>Variance</u>	2.129** (.077)	.093** (.008)	<u>Age 9</u>	1.833** (.049)
<u>Covariance</u> <u>between factors</u>	-.153** (.020)		<u>Age 10</u>	1.896** (.047)
			<u>Age 11</u>	2.057** (.052)
			<u>Age 12</u>	2.363** (.072)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Table D.10

Parameter estimates (standard errors) of the Simplex 2 model for the data set of condition B1

Parameter	Time 1	Time 2	Time 3	Time 4	Time 5
β		.552** (.016)	.901** (.029)	.942** (.025)	.952** (.026)
Standardized β		.717	.898	.908	.852
Factor mean	9.426** (.032)	5.213** (.157)	2.023** (.301)	1.653** (.292)	1.584** (.320)
Error variance of the factor	3.543** (.122)	1.023** (.069)	.409** (.055)	.399** (.057)	.779** (.100)
Error variance of the observed variable	1.648** (.064)	1.648** (.064)	1.691** (.055)	1.865** (.064)	1.865** (.064)

Note. All parameter estimates were significant at an alpha level of .05. β = regression coefficient predicting time 2 factor from time 1 factor, predicting time 3 factor from time 2 factor and so on.

Note. The statistics for Condition B2 are not presented because it is identical to those of Condition A2.

Condition B3: $r_{ig} = 0$, true reliability = .90 ~ .95

Table D.11

Correlation coefficients and distributional statistics for the data set of condition B3

	Time 1	Time 2	Time 3	Time 4	Time 5
Time 2	.906				
Time 3	.846	.931			
Time 4	.755	.874	.932		
Time 5	.650	.794	.883	.935	
Mean	9.426	10.420	11.414	12.408	13.402
SD	1.52	1.41	1.42	1.48	1.58

Table D.12

Parameter estimates (standard errors) of the Linear model for the data set of condition B3

	Intercept factor	Linear factor	Error variance	
<u>Mean</u>	9.426** (.021)	.994** (.004)	<u>Age 8</u>	.236** (.008)
<u>Variance</u>	2.061** (.043)	.082** (.002)	<u>Age 9</u>	.094** (.003)
<u>Covariance between factors</u>	-.123** (.007)		<u>Age 10</u>	.100** (.003)
			<u>Age 11</u>	.107** (.003)
			<u>Age 12</u>	.127** (.006)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Table D.13

Parameter estimates (standard errors) of the Simplex 2 model for the data set of condition B3

Parameter	Time 1	Time 2	Time 3	Time 4	Time 5
β		.844** (.006)	.937** (.006)	.975** (.006)	1.016** (.006)
Standardized β		.908	.933	.939	.946
Factor mean	9.426** (.021)	2.462** (.056)	1.647** (.060)	1.278** (.066)	.799** (.072)
Error variance of the factor	2.292** (.046)	.346** (.011)	.257** (.008)	.254** (.008)	.260** (.010)
Error variance of the observed variable	.006 (.005)	.006 (.005)	.005 (.004)	.027** (.004)	.027** (.004)

Note. All parameter estimates were significant at an alpha level of .05. β = regression coefficient predicting time 2 factor from time 1 factor, predicting time 3 factor from time 2 factor and so on.

Note. The statistics for Condition C1 are not presented because it is identical to those of Condition A2.

Condition C2: $r_{ig} = 0$, $r_{ee'} = .10$ between all time points, true reliability = .65 ~ .75

Table D.14

Correlation coefficients and distributional statistics for the data set of condition C2

	Time 1	Time 2	Time 3	Time 4	Time 5
Time 2	.708				
Time 3	.671	.754			
Time 4	.600	.706	.759		
Time 5	.520	.640	.726	.761	
Mean	9.426	10.420	11.414	12.408	13.402
SD	1.773	1.567	1.588	1.643	1.767

Table D.15

Parameter estimates (standard errors) of the Linear model for the data set of condition C2

	Intercept factor	Linear factor		Error variance
<u>Mean</u>	9.426** (.023)	.994** (.006)	<u>Age 8</u>	1.010** (.030)
<u>Variance</u>	2.096** (.052)	.082** (.003)	<u>Age 9</u>	.574** (.017)
<u>Covariance between factors</u>	-.124** (.010)		<u>Age 10</u>	.561** (.015)
			<u>Age 11</u>	.617** (.017)
			<u>Age 12</u>	.697** (.025)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Table D.16

Parameter estimates (standard errors) of the Simplex 2 model for the data set of condition C2

Parameter	Time 1	Time 2	Time 3	Time 4	Time 5
β		.733** (.011)	.945** (.013)	.942** (.012)	1.008** (.013)
Standardized β		.850	.918	.925	.922
Factor mean	9.426** (.025)	3.512** (.105)	1.570** (.140)	1.659** (.141)	.900** (.162)
Error variance of the factor	2.686** (.066)	.556** (.030)	.335** (.022)	.315** (.021)	.391** (.035)
Error variance of the observed variable	.458** (.021)	.458** (.021)	.404** (.018)	.504** (.020)	.504** (.020)

Note. All parameter estimates were significant at an alpha level of .05. β = regression coefficient predicting time 2 factor from time 1 factor, predicting time 3 factor from time 2 factor and so on.

Condition C3: $r_{fg} = 0$, $r_{ee'} = .10$ between last two time points, true reliability = .65 ~ .75

Table D.17

Correlation coefficients and distributional statistics for the data set of condition C3

	Time 1	Time 2	Time 3	Time 4	Time 5
Time 2	.690				
Time 3	.635	.735			
Time 4	.575	.691	.745		
Time 5	.488	.624	.700	.765	
Mean	9.426	10.420	11.414	12.408	13.402
SD	1.772	1.593	1.595	1.667	1.795

Table D.18

Parameter estimates (standard errors) of the Linear model for the data set of condition C3

	Intercept factor	Linear factor		Error variance
<u>Mean</u>	9.426** (.023)	.994** (.006)	<u>Age 8</u>	1.070** (.032)
<u>Variance</u>	2.077** (.052)	.093** (.004)	<u>Age 9</u>	.626** (.018)
<u>Covariance</u> <u>between factors</u>	-.132** (.011)		<u>Age 10</u>	.644** (.016)
			<u>Age 11</u>	.623** (.018)
			<u>Age 12</u>	.720** (.026)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Table D.19

Parameter estimates (standard errors) of the Simplex 2 model for the data set of condition C3

Parameter	Time 1	Time 2	Time 3	Time 4	Time 5
β		.729** (.012)	.906** (.014)	.963** (.013)	.992** (.013)
Standardized β		.828	.906	.913	.908
Factor mean	9.426** (.025)	3.552** (.112)	1.972** (.146)	1.416** (.152)	1.095** (.162)
Error variance of the factor	2.674** (.067)	.653** (.034)	.369** (.024)	.385** (.024)	.480** (.038)
Error variance of the observed variable	.466** (.023)	.466** (.023)	.474** (.019)	.472** (.021)	.472** (.021)

Note. All parameter estimates were significant at an alpha level of .05. β = regression coefficient predicting time 2 factor from time 1 factor, predicting time 3 factor from time 2 factor and so on.

Condition C4: $r_{ig} = 0$, $r_{ee'} = .30$ between all time points, true reliability = .65 ~ .75

Table D.20

Correlation coefficients and distributional statistics for the data set of condition C4

	Time 1	Time 2	Time 3	Time 4	Time 5
Time 2	.770				
Time 3	.729	.809			
Time 4	.657	.757	.806		
Time 5	.583	.698	.774	.818	
Mean	9.426	10.420	11.414	12.408	13.402
SD	1.776	1.592	1.579	1.645	1.764

Table D.21

Parameter estimates (standard errors) of the Linear model for the data set of condition C4

	Intercept factor	Linear factor		Error variance
<u>Mean</u>	9.426** (.023)	.994** (.005)	<u>Age 8</u>	.812** (.024)
<u>Variance</u>	2.287** (.054)	.082** (.003)	<u>Age 9</u>	.448** (.013)
<u>Covariance</u> <u>between factors</u>	-.131** (.010)		<u>Age 10</u>	.433** (.011)
			<u>Age 11</u>	.481** (.013)
			<u>Age 12</u>	.519** (.020)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Table D.22

Parameter estimates (standard errors) of the Simplex 2 model for the data set of condition C4

Parameter	Time 1	Time 2	Time 3	Time 4	Time 5
β		.772** (.010)	.928** (.011)	.954** (.010)	1.015** (.011)
Standardized β		.874	.927	.927	.937
Factor mean	9.426** (.025)	3.147** (.092)	1.743** (.113)	1.518** (.121)	.813** (.135)
Error variance of the factor	2.822** (.065)	.519** (.026)	.312** (.018)	.330** (.018)	.338** (.028)
Error variance of the observed variable	.333** (.017)	.333** (.017)	.287** (.014)	.366** (.015)	.366** (.015)

Note. All parameter estimates were significant at an alpha level of .05. β = regression coefficient

predicting time 2 factor from time 1 factor, predicting time 3 factor from time 2 factor and so on.

Condition C5: $r_{lg} = 0$, $r_{ee} = .30$ between last two time points, true reliability = .65 ~ .75

Table D.23

Correlation coefficients and distributional statistics for the data set of condition C5

	Time 1	Time 2	Time 3	Time 4	Time 5
Time 2	.684				
Time 3	.640	.729			
Time 4	.572	.681	.729		
Time 5	.496	.612	.693	.812	
Mean	9.426	10.420	11.414	12.408	13.402
SD	1.789	1.574	1.572	1.635	1.762

Table D.24

Parameter estimates (standard errors) of the Linear model for the data set of condition C5

	Intercept factor	Linear factor		Error variance
<u>Mean</u>	9.426** (.023)	.994** (.006)	<u>Age 8</u>	1.084** (.032)
<u>Variance</u>	2.070** (.052)	.108** (.004)	<u>Age 9</u>	.637** (.018)
<u>Covariance between factors</u>	-.155** (.011)		<u>Age 10</u>	.682** (.017)
			<u>Age 11</u>	.522** (.015)
			<u>Age 12</u>	.516** (.022)

Note. *significant at alpha level of .05; **significant at alpha level of .01.

Table D.25

Parameter estimates (standard errors) of the Simplex 2 model for the data set of condition C5

Parameter	Time 1	Time 2	Time 3	Time 4	Time 5
β	.716** (.012)	.922** (.014)	.948** (.014)	1.000** (.012)	
Standardized β	.837	.916	.875	.918	
Factor mean	9.426** (.025)	3.674** (.110)	1.807** (.150)	1.583** (.156)	.994** (.149)
Error variance of the factor	2.691** (.068)	.588** (.032)	.319** (.023)	.548** (.025)	.434** (.036)
Error variance of the observed variable	.510** (.023)	.510** (.023)	.481** (.020)	.334** (.019)	.334** (.019)

Note. All parameter estimates were significant at an alpha level of .05. β = regression coefficient

predicting time 2 factor from time 1 factor, predicting time 3 factor from time 2 factor and so on.