MULTILEVEL STRUCTURAL ANALYSIS OF THE ATTITUDES TOWARD PHYSICAL EDUCATION: METHODOLOGICAL CONSIDERATIONS

A Dissertation Submitted to the Faculty of Physical Education and Sport

Charles University in Prague

written by

Mgr. Ondřej Pecha

supervised by

Prof. PhDr. Petr Blahuš, DrSc.

Prague, Czech Republic

January 2009

DEDICATION

I would like to dedicate this thesis to my parents, Emílie and Petr, as a token of my gratitude for their support and help throughout my life.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Professor Petr Blahuš, for his guidance as well as his mental support. Thank you for your mentorship and many useful methodological lectures and discussions over the last five years. Nobody has had a larger impact on my scientific training. It was an honor for me to be your doctoral student, Petr.

I would be remiss if I didn't recognize Professor Knut A. Hagtvet. Many thanks to him for his excellent teaching in generalizability theory and structural equation modeling classes, his willingness to consult, and for advising me during my three stays in Norway. His comments and questions were very beneficial in forming of my scientific way of thinking.

I am also grateful to Professor Roderick P. McDonald for consulting my work. I have been impressed that a person of his stature was willing to read my thesis and to provide me such useful suggestions. I have appreciated all his essential comments and remarks.

There are a number of people at the FTVS UK that deserve my thanks as well. Namely, I would like to thank Dr. Karel Kovář, and Dr. Jan Štochl for their professional discussions, opinions, and presumptions. Moreover, I would like express my appreciation to the other PhD students who are interested in methodology, especially Lenka, Andrea, Libor, Tomáš, Michal, and Eva.

Last but certainly not least, I must acknowledge my girlfriend Zdeňka for her support, tolerance, and patience.

DECLARATION

I declare that this thesis is my own original work, conducted under the supervision of Prof. PhDr. Petr Blahuš, DrSc. It is submitted for the degree of Doctor of Philosophy at the Faculty of Physical Education and Sport at the Charles University in Prague. No part of this thesis has been submitted in the past or is being submitted for a degree at any other University. To the best of my knowledge, it contains no material or work performed by others, published or unpublished without due reference being made within the text.

SIGNED: _____ DATE: 30. 1. 2009

AGREEMENT

I agree that the copy of my thesis deposited in the library shall be available for consultation under conditions laid down by the library.

The borrowers should be carefully registered in a table given below.

First and last name	Number of identity card	Date	Note

CONTENTS

LIST OF TABLES....ix
LIST OF FIGURES....x
ABSTRACT....xi
INTRODUCTION....1
CONTRIBUTION OF THE PRESENT THESIS TO THE KINANTHROPOLOGY AS A BRANCH OF BEHAVIORAL SCIENCE....1
NESTING STRUCTURES AND MULTILEVEL ANALYSIS IN BEHAVIORAL AND ESPECIALLY IN KINANTHROPOLOGICAL RESEARCHES....3
Nesting structures in behavioral research....3
The theoretical reasons for multilevel modeling....5
The statistical reasons for multilevel modeling....7
Nesting structures and multilevel analysis in kinanthropology....8
MEASURING AND ANALYZING THEORETICAL CONCEPTS IN KINATHROPOLOGY....10
General rules, definitions, and theories....10
Multilevel nature of theoretical concepts....13
Measurement and identification of theoretical concepts....14
THEORY OF ATTITUDES....19
Definitions, origins, and structure of attitudes....19
Measurement of attitudes....21
Attitudes in Kinanthropology: Overview and Theories....25
HYPOTHESES AND RESEARCH QUESTION....28
Problem statement....28
Research question....28
Hypotheses....28
ASSUMPTIONS AND LIMITATIONS....30

METHOD....32
PARTICIPANTS....32
Sampling procedure....32
Inventory return and missing data....32
Sample descriptive statistics....33
INSTRUMENT....35
Inventory development....35
Structure of the inventory....35
PROCEDURE....37
Inventory administration....37
Data screening....37
ANALYSIS METHODS....38
Introduction to structural equation modeling and confirmatory factor analysis....38
Brief historical overview of SEM and CFA development....41
The process of SEM and CFA....43
Multilevel confirmatory factor analysis....54
Multilevel confirmatory factor analysis for ordinal variables....63
RESULTS....67
CONTINUOUS VARIABLES RESULTS....67
Descriptive statistics....67
Reliability analysis....69
Single-level CFA results....72
Estimation of the between-group variation....79
Within-group CFA results....79
Between-group structure....85
Simultaneous estimation of the within- and between-group structures....90

ORDINAL VARIABLES RESULTS 92
Descriptive statistics 92
Single-level CFA results for ordinal variables 94
Multilevel CFA results for ordinal variables 96
DISCUSSION 101
CONCLUSION 107
REFERENCES 108
APPENDICES 119
INVENTORY ITEMS 119
LISREL SYNTAX FILES 120
Single-level CFA 120
Within-group CFA 123
Between-group CFA 124
MPLUS SYNTAX FILES 125
POLYCHORIC CORRELATION 126
EXTENDING METHODOLOGICAL STUDY 130
Introduction 130
Brief overview of the Generalizability Theory 131
Research questions 133
Methods 134
Results 139
Discussion 146
FOLLOW-UP STUDY 148
Descriptive statistics 148
Multilevel analysis of alcohol drinking 153

LIST OF TABLES

1. Examples of units at the within-group level and respective between-group level.....4
2. Examples of theoretical concepts and some of their empirical indicators.....11
3. Examples of the statements about physical education and Likert-type scoring.....22
4. Gender and age: Absolute and relative frequencies.....33
5. Basic characteristics at the school-level.....34
6. Three aspects of all items: The structural hypotheses of the inventory.....36
7. Basic univariate statistical characteristics of the items.....68
8. Reliability-like coefficients of four attitude-related domains.....71
9. Number of items needed for desirable levels of reliability.....72
10. Total sample covariance matrix $\boldsymbol{S}_T$ and respective correlations.....73-74
11. Fitting history of the hypothesized models based on $\boldsymbol{S}_T$.....75
12. Estimated standardized factor loadings and uniqueness of model 4*b* based on $\boldsymbol{S}_T$.....78
13. Intraclass Correlation coefficients.....79
14. Pooled within-group covariance matrix $\boldsymbol{S}_{PW}$ and respective correlations.....81-82
15. Fitting history of the hypothesized models based on $\boldsymbol{S}_{PW}$.....82
16. Estimated standardized factor loadings and uniqueness of model 4*b* based on $\boldsymbol{S}_{PW}$.....83
17. Estimation of the between group covariance matrix $\boldsymbol{\Sigma}_B$ and respective correlations....86-87
18. Fitting history of the hypothesized models based on $\hat{\boldsymbol{\Sigma}}_B$.....87
19. Estimated standardized factor loadings and uniqueness of the three-factor model based on $\hat{\boldsymbol{\Sigma}}_B$.....89
20. Relative frequencies.....92
21. Fitting history of the hypothesized models for ordinal variables.....94
22. Estimated factor loadings and uniqueness of model 4*b* for ordinal variables.....95
23. Structural part of model 4*b*: Estimated correlations among factors.....96
24. Model 2*a*: Estimated factor loadings using four different approaches.....98
25. Covariances between factors in model 2*a* using four different approaches.....99
26. Standardized uniqueness of model 2*a* using all four approaches.....100

LIST OF FIGURES

1. Hypothetical five-level structure in fictive educational research....3
2. An illustration of the ecological fallacy....6
3. No relationship between variables X and Y when the nesting structure is ignored....7
4. An example of the path-diagram....12
5. The school level structure of theoretical concepts....13
6. Latent variable modeling of theoretical concepts....15
7. Empirical indicators and theoretical concepts....16
8. An illustrative example of non-monotonous function....18
9. An illustrative example of the semantic differential....23
10. Wear's theory of attitudes toward physical education....25
11. Kenyon's theory of attitudes toward physical activity....26
12. Top 10 types of physical activity....33
13. An illustrative SEM model in LISREL "all-y" notation....39
14. One of the hypothesized models: Second-order confirmatory factor analysis....47
15. Two-level factor model: illustration of relationship between levels....62
16. Scatterplot and regression line between item 3 and item 8....68
17. Path diagram of the single-factor model: Health and fitness domain....69
18. Path diagram of the single-factor model: Social domain....70
19. Path diagram of the single-factor model: Emotional domain....70
20. Path diagram of the single-factor model: General domain....71
21. Structural part of model 1*b*....73
22. Structural part of model 2*b*....75
23. Structural part of model 3*b*....76
24. Structural part of model 4*b*....77
25. Structural part of model 2*a*, within-group level....83
26. Structural part of model 3*a*, within-group level....84
27. Modification of model 4*b*, within-group level....85
28. Scree plot of the first 15 factors....88
29. Structural part of the two-level model....91

ABSTRACT

This study introduces multilevel confirmatory factor analysis (MCFA) as an appropriate methodology to use when data are hierarchically structured and an identification of theoretical concepts of a given inventory is desired. The main purpose of this thesis is to introduce all possible methods and issues which are connected with attitude research. The MCFA methodology is demonstrated on the example of the attitudes toward physical education (ATPE). Data were collected from high school students (n = 1157) nested within PE classes (m = 87). A different factor structure of ATPE was determined at the student level (within-group) and at the PE class level (between-group). At the within-group level, Wear's four-factor theory of ATPE was confirmed. However, the existence of two wording factors (positive and negative) and the school PE-related factor was indicated. At the between-group level, the resulting three-factor model exhibited only a marginal fit. The structure of theoretical concepts and their indicators was indistinct, because the items were originally developed to measure ATPE at the within-group level. The analysis included four cases: (1) single-level continuous, (2) multilevel continuous, (3) single-level ordinal, and (4) multilevel ordinal. All four approaches were compared in terms of parameter estimation and their standard errors. It was found that both ordinal variable approaches produced higher factor loadings but also higher standard errors, and lower unique variances. The MCFA for ordinal variables was computationally demanding and is not recommended for extensive models with a high number of factors at both levels. Finally, this thesis is complemented by two interrelated studies that are situated in the Appendix. The first one is a theoretical work about the so-called hidden facets in the ratings of item relevance. The second study is the multilevel analysis of alcohol consumption among the same 1157 high school students.

Keywords: Kinanthropology, attitudes toward physical education, structural equation modeling, multilevel analysis, adolescents, generalizability, alcohol consumption

INTRODUCTION

CONTRIBUTION OF THE PRESENT THESIS TO KINANTHROPOLOGY AS A BRANCH OF BEHAVIORAL SCIENCE

Science is commonly divided into *hard* and *soft* sciences. The hard sciences include the natural sciences (e.g., physics, biology, and chemistry) and the formal sciences (e.g., mathematics, logic, and methodology). The soft sciences incorporate the social sciences (e.g., sociology, economic, or political science) and the behavioral sciences (psychology, social psychology, management science, anthropology, etc.). The main difference between the soft and hard sciences is that the results within soft sciences are less *cumulative*. The scientific results are cumulative if empirical laws and theoretical structures are built on one another so that later development extends and unifies earlier work (Hedges, 1987). In other words, research results in the soft sciences are often conceded to be less replicable than research results in the hard sciences. Theories and rules in the soft sciences are not deterministic and need to be often supported, confirmed, and verified by empirical results.

Kinanthropology is quite a young science which belongs to the family of behavioral sciences. It includes all possible problems dealing with the subject matter of human motoricity. There are several well established sciences dealing with the same problem abroad like human movement science, human kinetics, anthropokinetics, kinesiology, etc. The political and social changes enabled to constitute kinanthropology as a fully recognized science with a larger extent of study in the Czech Republic after 1989 (Blahuš, 2004). In 1991, kinanthropology was accredited as a dissertation discipline thanks to the effort of a group of professors coming from the Faculty of Physical Education and Sport, namely: Blahuš, Dobrý, Hohler, Hošek, Svatoň, and Svoboda, among others (Blahuš, 1993; Hodaň, 2004). Kinanthropology has to be understood as a trans-disciplinary, interdisciplinary, and methodologically synthetic scientific discipline. For this reason, the present research of attitudes toward physical education is also an object of investigation in kinanthropology, even though the research of attitudes is generally central to social psychology.

A multilevel approach to structural equation modeling is one of the most recent and most comprehensive formal statistical methods which are being used in the behavioral sciences.

Generally speaking, the role of statistical methods in the social and behavioral sciences is to contribute to our knowledge of scientific laws (Blahuš, 2008). The role can be either exploratory or confirmatory. The exploratory role is in discovering new regularities. In the present thesis, however, the statistical methods discussed later play a confirmatory role. They help us to verify the theoretically derived hypotheses about lawfulness (Blahuš, 2008). The presented methodology of multilevel structural equation modeling has become a part of the network of systemizing relationships and links between our observable data and the scientific theory of kinanthropology.

From a practical point of view, the present thesis could serve as a guideline for understanding when and how to apply the discussed methodology to observable data. This work should assist professional investigators and Ph.D. students of kinanthropology in conducting their own research and provides them a necessary theoretical background.

NESTING STRUCTURES AND MULTILEVEL ANALYSIS IN BEHAVIORAL AND ESPECIALLY IN KINANTHROPOLOGICAL RESEARCHES

Nesting structures in behavioral research

Behavioral research often involves problems that investigate relationship between individuals and their contexts. The main idea is that individual persons are influenced by the social groups or contexts to which they belong. Contextual analysis is a development in behavioral sciences which has focused on the effects of the social context on individual behavior. For instance, a pupil's educational achievement is substantially affected by teacher, classroom, school, and/or district characteristics. There are several hierarchically structured units of observation in this example. Pupils are nested within classrooms, classrooms are nested within schools, schools are nested within cities, and cities can be further nested within countries. For better illustration, this hierarchical structure is portrayed on figure 1. Five *levels* of hierarchy are considered in this case.

It depends on research purpose of an investigator, how many levels will be included in a study; however, studies with two or three levels of hierarchy are most common (Snijders & Bosker, 1999). The term *level* is central to studies with nesting or hierarchical structures. Since multiple levels are analyzed simultaneously in such studies, the term *multilevel analysis* has become most widely known nowadays (Luke, 2004). Alternative names for multilevel modeling are hierarchical modeling (Raudenbush & Bryk, 2002), and/or random-coefficient modeling (Longford, 1993).

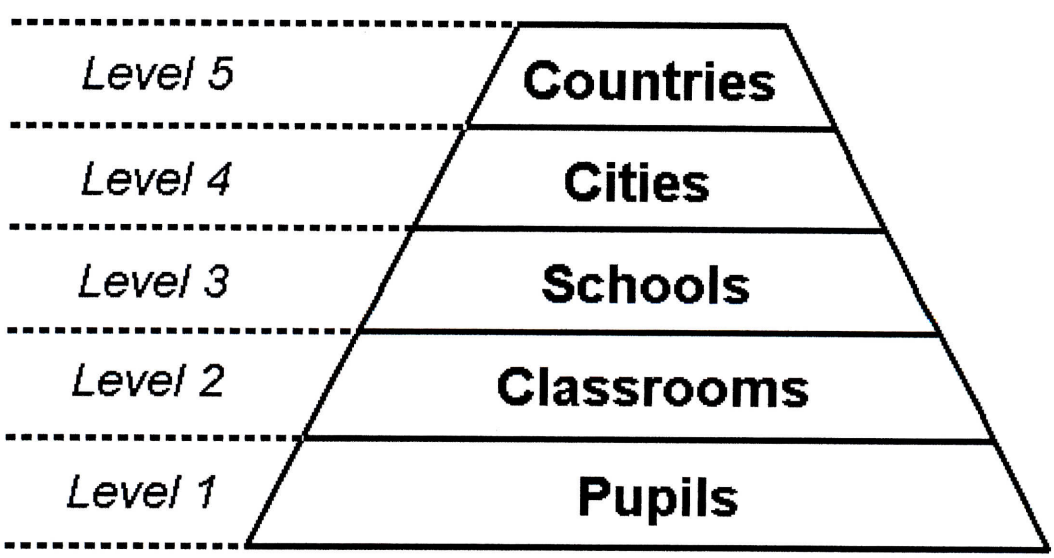


Figure 1. Hypothetical five-level structure in fictive educational research

Now, we will restrict ourselves to the simplest two-level case. There are two types of units or objects of observation. From here on we will refer to level-1 units as the *within-group units* (pupils, employees, children, athletes etc.) and to the level-2 units as the *between-group units* (classrooms, schools, clubs, companies etc.). Synonyms for the within-group units are *micro-level units, micro-units, secondary units or elementary units*, and the between-group units are also called as *macro-level units, clusters, macro-units or primary units* (Snijders & Bosker, 1999).

The notion of two-level structures can be extended to situations beyond where individuals are being influenced by their contexts. Special applications of multilevel analysis are longitudinal research where a series of distinct repeated measures are nested within individuals and meta-analysis where the subjects are nested within different studies (Hox, 2002). The multilevel approach has several important advantages when applied to longitudinal data compared to traditional analytic methods like repeated measures ANOVA or MANOVA. First advantage is that we have growth curves of the same type (line, parabolic curve etc.) that is unique for each person. Second, the number of repeated measures and their spacing might differ across individuals. Other advantages can be found in Hox (2002) or in specialized textbooks aimed at longitudinal modeling (e.g., Singer & Willett, 2003). Typical examples of between-group units and the within-group units nested within them are presented in Table 1.

Table 1. Examples of units at the within-group level and respective between-group level

Example No.	Within-group	Between-group
1	pupils	classrooms, schools
2	employees	bosses, companies
3	repeated measures	individuals
4	subjects	studies
5	patients	doctors, hospitals
6	athletes	coaches, teams
7	children	families

The most fundamental question arises, if one ore more levels are ignored: What would happen, if single-level techniques are applied to data with nesting structures? This can lead to serious problems in interpretation which are discussed in more detail in the following section.

The theoretical reasons for multilevel modeling

The main reasons for using multilevel models are *theoretical* and *statistical* (Luke, 2004). Theoretical reasons were already insinuated in previous section. If the phenomenon what we study is multilevel in nature, we should use analytic techniques that are also multilevel. However, in spite of multilevel techniques are roughly for twenty years at disposal, many investigators still using the single-level techniques in multilevel case (Raudenbush & Bryk, 2002). What is wrong on this approach? There are two ways how to analyze such a data at a single level.

First way how to do this is *aggregation*. Data are collected from individuals and then aggregated (most often as means) into groups to which those individuals belong. An investigator can make incorrect inference, if the relationships observed in groups are assumed to hold also for individuals. Such an error in interpretation is well known as the *ecological fallacy*. This term comes from Robinson's study (Robinson, 1950). He investigated the relationship between literacy rate and the proportion of the population born outside USA in 48 states (it means $N = 48$). These two aggregated between-group variables correlated 0.53, which means that the grater the proportion of immigrants in a state, the higher its average literacy. However, the correlation between these two non-aggregated within-group variables was -0.11, which was expected and means that immigrants were slightly less literate than native citizens. The explanation is straightforward: the correlation at the between-group level was so high and positive, because immigrants tended to settle in states where the native population was more literate.

For better understanding of the ecological fallacy, an illustrative example is presented in Figure 2. Let us assume that total of 25 within-group units (individuals) are nested within 5 between-group units (groups, 5 in each). We would like to investigate the relationship between X and Y variables. Each group is denoted by different symbol and all individuals belonging to the same group have the same symbols in the scatter plot. Five dotted nearly parallel regression lines within each group indicate that there is a strong positive association between X and Y variables at the individual level. The full squares constitute means derived from all five individuals on both X and Y variables within each group. The full (not dotted) regression line fitted to this set of means indicate also very strong but negative relationship between these two variables at the between-group level.

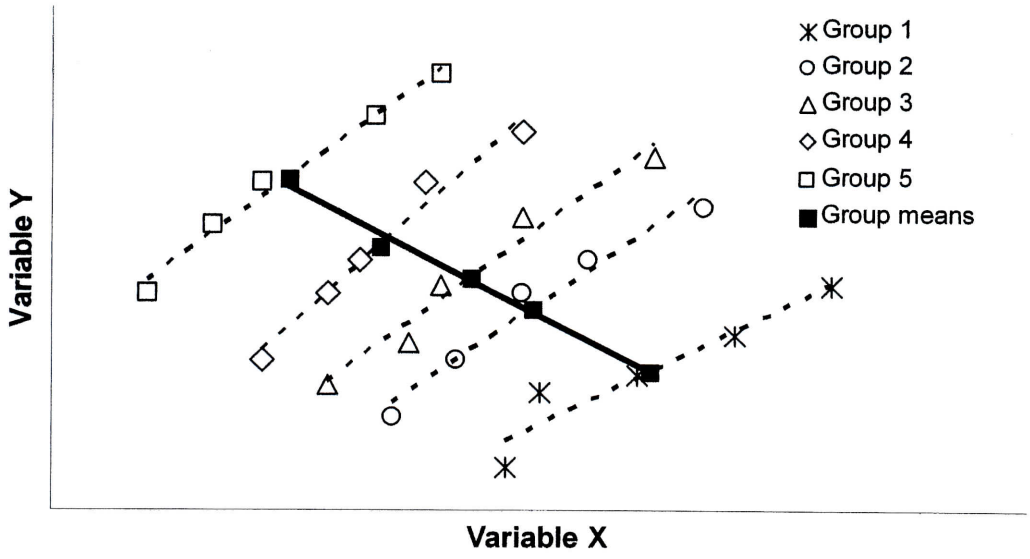


Figure 2. An illustration of the ecological fallacy

Second possibility how to analyze a two-level data at a single level is to *disaggregate* the between-group variables to the individual level. Also here, the interpretation has to be done very carefully to avoid the so called *atomistic fallacy*. This fallacy is the opposite of the ecological fallacy. The inferences about groups are incorrectly drawn from the within-group level information in this case (Hox, 2002). Another dangerous approach is to investigate a relationship between two variables regardless of the nesting structure. Data are analyzed mixed together without any information about group characteristics. Such an instance is illustrated in Figure 3, where the same artificial data set has been used as in previous example. The full regression line fits all 25 points and indicates that there is no relationship between X and Y variables. This is a special case of disaggregated variables, where only within-group variables are included in a study. The inference about association between variables is also biased in this case and thus such an approach mustn't be used. However, the single-level approach to multilevel data might be relevant if the clustering effect is very weak which is usually expressed by the intraclass correlation coefficient (ICC). If the ICC (a number between 0 and 1) is negligible, say lower than 0.05, the hierarchical structure need not to be taken into account (Hox, 2002). The ICC is mathematically formulated in the method section of this thesis.

The statistical reasons for multilevel modeling

The single-level statistical techniques are based on assumption that all units are sampled independently from each other. However, *simple random sampling* is not very cost-efficient strategy in behavioral sciences. More often, the *multi-stage random sampling* is employed, where the highest-level units are randomly sampled in the first stage and then within already chosen highest-level units are randomly sampled units at lower levels. In a random two-stage sample, a random sample of schools, for instance, is taken in the first stage and then pupils are sampled at random from the selected schools in the second stage. In all cases, the multi-stage (or just only two-stage) sampling design leads to dependent observations (Snijders & Bosker, 1999).

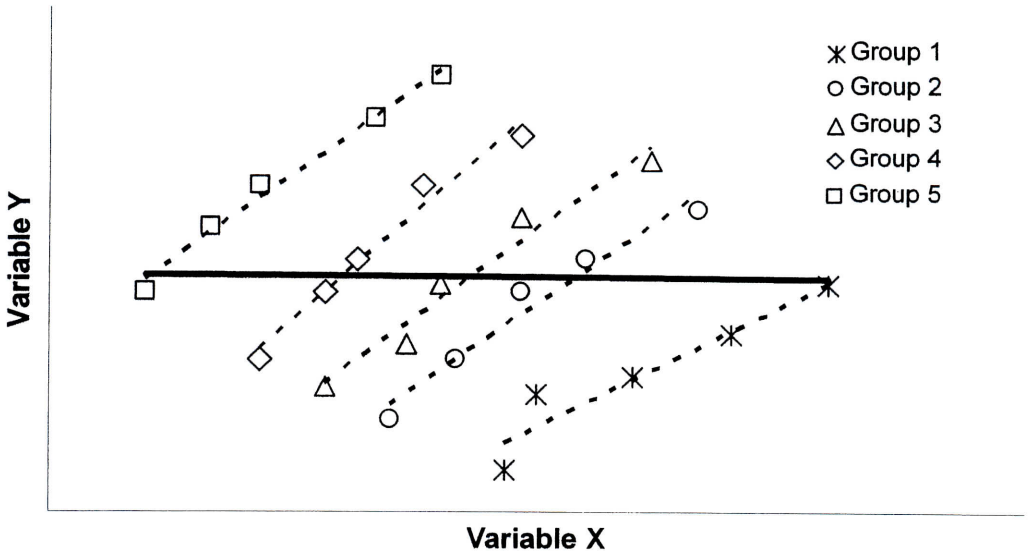


Figure 3. No relationship between variables *X* and *Y* when the nesting structure is ignored

Another important statistical issue is related to disaggregation. As noted by Snijders and Bosker (1999), disaggregation might lead to *miraculous multiplication* of the number of units. For example, if we sample 10 schools in the first stage and 100 pupils by each school in the second stage, we may end up with $10 \times 100 = 1000$ schools. This means that disaggregating the data as if they are independent may result in a dramatic growth of the sample size, which may lead to statistical type I error (rejecting a null hypothesis when it is true). Disaggregated variables are also the cause of problems in multiple regression model (Luke, 2004). Namely, the individuals belonging to the same group will presumably have correlated errors, which violates one of the assumptions of multiple regression.

Statistical background of the multilevel approach to the structural equation modeling is described in more detail in the *method section* of this thesis. An interested reader is referred to Goldstein (1995) to see more technical details for statistical reasons for multilevel approach to other statistical models.

Nesting structures and multilevel analysis in kinanthropology

Besides traditional nesting structures (like pupils nested within schools) that are shared with other branches of behavioral sciences, kinanthropology has several unique nesting structures. Athletes can be nested within coaches, clubs, teams, cities and/or countries, performances are nested within individuals or competitions, judges in artistic gymnastics are nested within apparatuses in all-around competition, sport products of the same type are nested within companies, and many others. Time points nested within individuals are very often used for description of developmental curves of anthropometric measurements or physical tests, attitudes toward physical activity, and/or physical activity participation.

One of the first applications of multilevel analysis in kinanthropology was a paper by Zhu (1997). He reanalyzed ten years old data set from research about effects of school factors associated with health-related fitness (e.g., percentage of classes taught by a PE specialist, and/or minutes in physical education per week) on pupil's 1-mile run-walk performance. He found slightly different results when multilevel analysis was applied compared to single-level ordinary multiple regression.

Roughly at the same time, first application of multilevel analysis in longitudinal study was conducted (Zhu & Erbaugh, 1997). They described the development of several aquatic skills in children over several years. The advantages of multilevel approach to longitudinal data are demonstrated in this study.

To date, a variety of cross-sectional studies have worldwide appeared with application of multilevel analysis in some field of kinanthropology (e.g., Dexter, 1999; Wendel-Vos et al., 2004; Fisher et al., 2004; Ball et al., 2007). Several multilevel studies with longitudinal data have been also accomplished in kinanthropology (e.g., Martin et al., 2004; Elferink-Gemser et al., 2006; Barnett, et al., 2007).

In our country, the application of multilevel analysis is very rare in kinanthropological studies. The inappropriate single-level techniques are still used in typical multilevel cases. It

is necessary to promote the multilevel analysis in our environment. But yet, few studies have been done on this topic. Pecha (2004) carried out a study, where pupils were nested within physical educational classes. The effect of individual-level and class-level variables on a long jump performance was investigated. The study was an application of so called intercepts-and-slopes-as-outcomes model (Raudenbush & Bryk, 2002). Application of the multilevel approach to the factor analysis (Pecha, 2006) is described in more detail in the *method section*, because is more relevant to the purpose of the present thesis.

Let us now make a brief remark on nesting structures. We have to draw attention to two types of nesting structure:

1) *Theory-guided nesting*, where the nesting structure is given a priori. Units are belonging to their natural clusters or groups here. The multilevel analysis is such a case, because we know exactly, within which classrooms are the students nested. Other examples are testlets and/or parcels of items. More general approach is known from the generalizability theory (discussed later on), where facets of measurement or differentiation can be nested within other facets. Such type of nesting is usually not verified by data analysis.

2) *Data-driven nesting*, where clusters are created a posteriori by some mathematical method. Typical example is the exploratory factor analysis, where a pool of indicators is divided into several groups measuring the same factor. A data-driven analogy to multilevel analysis is the *cluster analysis*, where objects (e.g. students) are classified into different groups that share some common trait (often proximity according to some defined distance measure).

There are also some other methods that join together both approaches. From our point of view, however, the confirmatory factor analysis (CFA) is the most important. Therefore, a separate part on CFA is reserved in *method section*.

MEASURING AND ANALYZING THEORETICAL CONCEPTS IN KINATHROPOLOGY

General rules, definitions, and theories

In the previous section, the attention has been paid to objects of observation and their hierarchical structures. The main goal of this part is to target the object's attributes or characteristics. Age, gender, body weight, body high, 100m dash performance, body fatness, intelligence, temperament, and/or endurance ability are the examples of the object's attributes. However, some attributes are not directly observable. Basically, we can distinguish between two levels[1] of attributes: *empirical* and *theoretical*.

Empirical attributes are observable directly, like a measured performance in different physical tests, anthropometric measures, and/or expressing a certain level of agreement with a statement in an attitude inventory. They are of concrete, partial and specific nature and they are denoted as *empirical research variables* (Blahuš, 2008).

Theoretical attributes are unobservable directly, like intelligence, endurance ability, temperament, well-being, quality of life, body fatness, and/or attitudes. They are known under the term *theoretical concepts* that are of abstract, synthetic, and generic nature. Alternative names for theoretical concepts are traits, constructs, domains, categories, dimensions etc. I the present thesis, the terms *concept* and *construct* are used most often and are considered interchangeably, even though they are of different origin (Blahuš, 2008).

The procedure of identifying or constructing theoretical concept is known as *concept formation,* which is based upon the rules of empirical-theoretical correspondence (Blahuš, 1999, 2000). In behavioral sciences, the rules of correspondence are represented by associative measurement, which treats the concept formation procedure as *latent variable modeling*. The standardized directly observable research variables are called *empirical indicators* of the theoretical concept. The relationship between indicators and concepts can be viewed in two ways. First is the classical *representative way of thinking*, where empirical indicators are considered as "effects" of the concepts. On the other hand, indicators may play a role of determinants or "causes" of the underlying concepts, which is used especially in

[1] The term "level" is now related entirely to object's attributes. Hence, it doesn't play any role on which level of hierarchy is the object itself situated.

natural sciences (Grace & Bollen, 2008). Examples of the theoretical concepts and their empirical indicators taken from kinanthropological researches are given in Table 2.

Table 2. Examples of theoretical concepts and some of their empirical indicators

Concept	Respective Indicators
strength ability	bench-press, full-ball-throw, push-ups
endurance ability	12-min run, 10-min stationary bicycle, endurance shuttle run
body fatness	skinfold measures, bioelectrical impedance, DEXA
body coordination [1)]	one-knee balance, stork stand, cross-leg squat
somatic anxiety [2)]	"My body feels tense.", "My heart is racing.", "I feel jittery." (*Inventory items*)
social health [3)]	"Are you able to participate in social activities?" "Are you comfortable with your personal relationships?" "Do you move around your community?" (*Items in questionnaire*)
coach's motivation competence [4)]	"How competent is your coach to motivate his athletes?" "How competent is your coach to build team cohesion?" "How competent is your coach to build team confidence?" (*Inventory items*)
aquatic skills [5)]	head ducking, breathing into the water, feet-first dive

Sources: [1)] Čepička, 2004 ; [2)] Lundqvist & Hassmén, 2005; [3)] Tannenbaum, Ahmed & Mayo, 2007; [4)] Myers, 2005; [5)] Štochl, Čechovská & Pavelková, 2005

The terms *causes* and *effects* have been written in quotation marks because it may indicate a causal-effect. However, it doesn't have to be true. More appropriate terminology for these terms is *independent variables* instead of causes and *dependent variables* instead of effects. More precisely stated, causal variable is always a special case of independent variable and effect variable is always a special case of dependent variable.

The relationship between dependent variable (denoted as y) and independent variable (denoted as x) is mutually symmetric. Causal relationship can be identified entirely via an experimental study (e.g., Myers, 1972), whereas the relationship between x and y is conceived as *estimation* in our case. The estimation is graphically represented by an arrow, which in ordinary case; the arrow leads from x to y (but also from y to x): $x \rightarrow y$. This notation is a base for so-called path-diagram which graphically represents the structure of relationship

among all variables included in a study. In this thesis, let us restrict only to the case, where theoretical constitute independent variables and empirical indicators are dependent variables. An example of the path-diagram is given in Figure 4.

The picture shows concepts and their indicator taken from first two rows in Table 2.

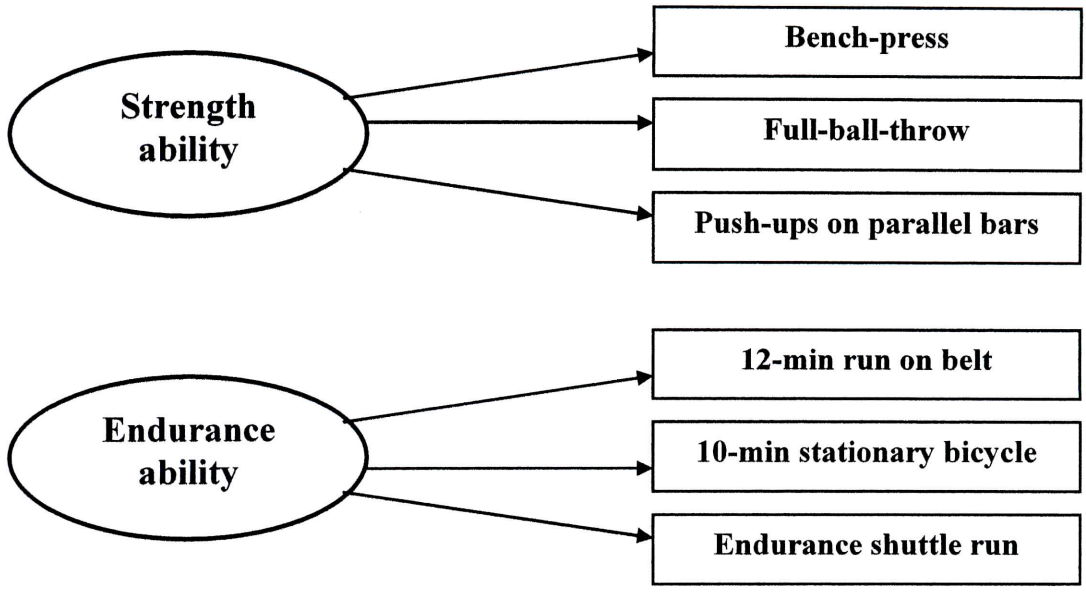


Figure 4. The relationship between theoretical concepts and their indicators: an example of the path-diagram

The theoretical concepts are usually depicted as ellipses (or circles) and empirical indicators are displayed as rectangles (or squares). The organization of concepts can be also hierarchical[2] where one or more *general constructs* are related to several *sub-constructs*, or more precisely spoken, the concepts of higher order are primarily related to construct of lower order. Some authors also use the expression order of hierarchical layer (Blahuš, 2008).

The hierarchical theory of health related fitness might serve as an example of two-level concepts (Bouchard & Shephard, 1994). According to this theory, the general constructs of health related fitness is related to five sub-constructs: morphological, muscular, motor, cardio-respiratory, and metabolic. These sub-constructs are further related to their empirical indicators. For instance, morphological sub-construct is indicated by bone density, body composition, relative body weight, and/or body-fat placement.

[2] The term *hierarchical structure* is now related to concept organization. It mustn't be mistaken for hierarchical structure of observation (pupils, schools, and/or classes) discussed in previous chapter.

Such hierarchical structures are important from the perspective of the present thesis, because some theories of attitudes are also hierarchical and are described in more detail in the next part dedicated to attitudes. An interested reader is referred e.g. to Kline (2005) to se more hierarchical types of concept organization.

Multilevel nature of theoretical concepts

As we noted above, theoretical concepts and empirical indicators are object's unobservable and observable attributes, respectively. In previous section, the hierarchical structure of objects has been discussed. The same rules hold for empirical indicators that are observed at the same level as their corresponding concepts. Serious problems in interpretation may arise, if indicators observed at the lower level are aggregated to the higher level. They don't have to indicate the same theoretical concept at both higher and lower levels. An extreme example of this event was proposed by Pecha (2006). In his study, the objects of observation were students nested within schools. Six physical tests (see Figure 4) indicate different concepts at the school level, if we aggregate them to the school level. This situation is graphically illustrated by a path-diagram in Figure 5.

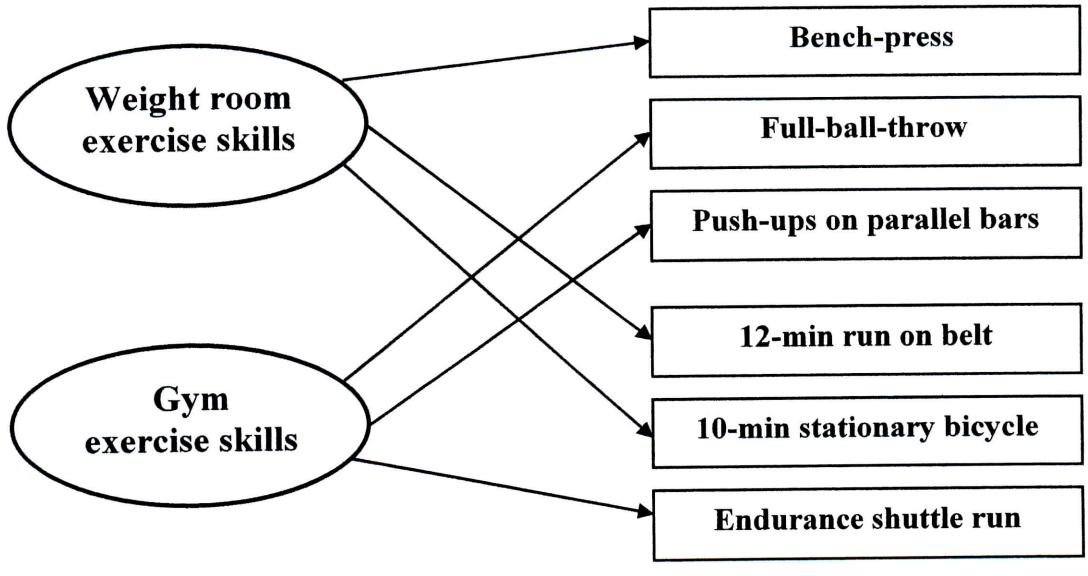


Figure 5. The school level structure of theoretical concepts and aggregated indicators observed at student level

While these six physical tests indicate student's strength and endurance ability concepts at the within-group level, their role is quite different at the between-group level. In some schools, the students are habituated to exercise within a school gym (equipped for running and

furnished with gymnastics apparatuses), whereas students from other schools might be used to exercise within a weight-room (furnished with devices like running belts or stationary bicycles). Therefore, the students coming from schools with gyms are presumably on average (!) better in push-ups on parallel bars, endurance shuttle run and full-ball-throw tests. Analogously, the students coming from schools with weight rooms might be on average better in bench-press, 10-min stationary bicycle, and 12-min run on belt tests.

This phenomenon of possibly different concepts-indicators structures at both levels has been also discussed in scientific literature. Different structure of indicators and concepts at between-group compared to the within-group level has been identified by Zimprich et al. (2005), and/or Farmer (2000). On the other hand, some other authors have considered the same structure at both levels (e.g., Li et al., 1997; Kuhlemeier et al., 2002; Myers, 2005). This topic is further discussed in more detail in the method section.

Measurement and identification of theoretical concepts

Up until now, we have considered theoretical concept and their empirical indicators as unformalized research variables. If we want to model these variables, we have to formalize them. Formalized mirror of theoretical concepts are *latent variables* and formalized counterparts of empirical indicators are *manifest variables*. Manifest observed variables are formalized attributes, whose categories are coded by formal symbols. Latent variables are mathematically constructed variables presumed to model theoretical concepts. Manifest and latent variables are interconnected by a formal *association function*, which is representation of the correspondence rules between concepts and indicators (Blahuš, 2008). In behavioral sciences, the association function is represented by a regression function in the model of associative measurement.

To make this distinction more transparent, the relationships between formalized and unformalized level on one side and between observed and unobserved level on the other side is portrayed in Figure 6, which is inspired by Blahuš (1996a). The unobservable level is on the left hand side, whereas observable level is situated on the right hand side. The formalized level is at the bottom and unformalized level is displayed at the top.

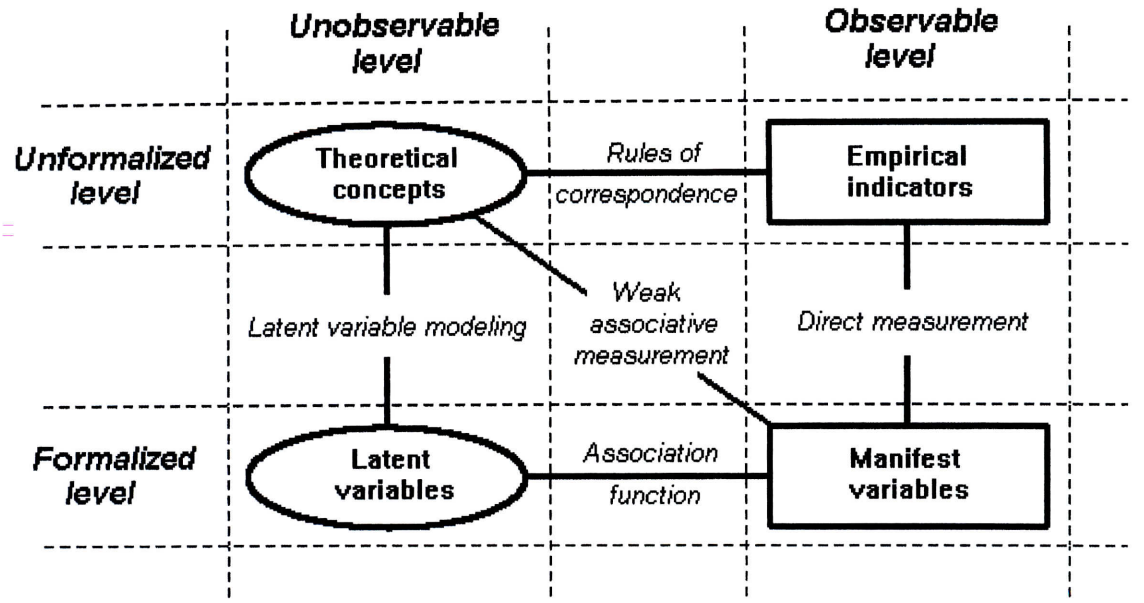


Figure 6. Latent variable modeling of theoretical concepts

From our point of view, the latent variable modeling is the most important component of this picture. Two basic conditions enable to use models with latent variables (Blahuš, 2008): (1) concept formation by the weak associative measurement and (2) completeness of explanation by axiom of local independence.

Several classifications of the models with latent variables have emerged up to recently. For example, Bartholomew and Knott (1999) classified models from the perspective of scale-type of both latent and manifests variables: (1) *factor analysis* for all continuous variables, (2) *latent class analysis* for all categorical variables, (3) *latent profile analysis* for continuous manifest and categorical latent variables, and (4) latent trait analysis for categorical manifest and continuous latent variables. Another typology of latent variable models has been provided by Blahuš (1985). According to his terminology, the general latent variable model consists of 17 sub-models. Another consideration about latent variable models has been carried out by Muthén (2002). He has started from general statistical model not-necessarily with latent variables, which incorporates 7 main types of analysis: (1) *structural equation modeling*, (2) *growth curve modeling*, (3) *multilevel modeling*, (4) *missing data modeling*, (5) *finite mixture modeling*, (6) *latent class modeling*, and (7) *survival modeling*.

For our purpose, the most important is the structural equation modeling (SEM). This term contains a family of models with latent variables like factor analysis, latent variable growth curve modeling, multiple indicator multiple causes (MIMIC), path analysis, and/or

MANOVA. Especially factor analysis and its multilevel extension is explained in more detail in the method section of the present thesis.

Another question is linked to the problem of standardization of diagnostic methods, which is an attitude inventory in our case. The main and minimum standardization requirements of diagnostic quality contain: (1) *construct validity* toward the theoretical concepts, (2) *dimensionality* of the diagnosed domain, (3) *reliability* of diagnostic methods, and (4) *practical validity* to an external criterion (Blahuš, 2004). Very interesting graphical visualization of these issues was proposed by Hagtvet (1995). In Figure 7, the three cones symbolize theoretical concepts and their respective empirical indicators.

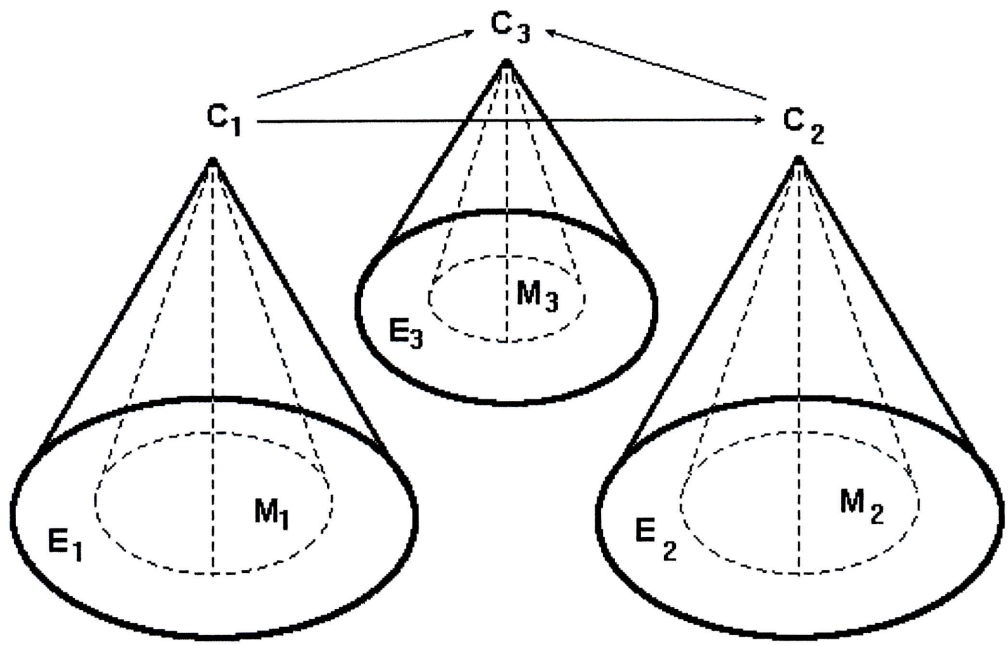


Figure 7. Empirical indicators and theoretical concepts: C – theoretical concepts (constructs), E – empirical domains (universe), M – actual measurements (indicators at hand)

It is assumed here that all three concepts are irreducible (i.e. don't have any sub-constructs). Dimensionality of the diagnosed domain corresponds to the number of depicted cones. In our case, the three-dimensional diagnostic method is considered.

M → E imaginary paths relate to *generalizability* or so-called *generic reliability*. The concept C is represented by infinitely large (at least theoretically) empirical domain or universe E of all admissible, relevant indicators. However, we usually can not cover the whole universe, but we have only smaller sample of indicators M at disposal to be used in the measurement procedure. The universe (scores) can be further only estimated from

accomplished measurement. This estimation is usually expressed via Cronbach's *alpha* (Cronbach, 1951) or by McDonald's *omega* (McDonald, 1978) and is sometimes known under the terms *internal consistency* or *domain validity*.

E – C relations refer to *content validity*. Two main questions are taken into account in content validation procedure and Messick (1995) pointed out two main dangers in this process: (1) *concept irrelevance* and (2) *concept under-representation*. Is a particular item indicator relevant or admissible to measure the respective concept? If not, the item is then concept irrelevant. Nevertheless, all items together in E can be admissible but if they don't cover the entire breadth or all aspects of the concept C, some items may under-represent the concept. Also content validity is discussed in detail in the extended study of this thesis.

M – C connections constitute *construct validity* of the actual measure. This type of validity is concerned in how well person's level of concept C (e.g. attitude) can be estimated by person's scores in available indicators M (e.g. attitude inventory). In contrast to content validity, the construct validity may be formalized and expressed by numerical indices. Construct validity toward the theoretical concept is often modeled by factor validity as factor loadings in the latent common factor model or as parameters of discrimination in IRT (item response theory). We have to emphasize that construct validity is an inference based on scores from the actual measurement device often supported by factor loadings and/or expected relationships to other sub-constructs. (Hagtvet, personal communication) For this reason, the paths between the C's are included in Figure 7.

Finally, let us discuss the problems related to types of *association function* between latent and manifest variables. This is important namely from attitude research viewpoint. As we mentioned above, the associative function represents formalization of the correspondence. This formalization can be carried out by different methods like formal logic, fuzzy logic, conditional probabilities, and/or by mathematical function (Blahuš, 2008). Most frequently, the associating function is represented by linear regression function.

Furthermore, the linking functions are usually assumed to be monotonous, which means that with increasing level of latent variable (e.g. ability) increases probability of positive response on item or accomplishment of a physical task. Nevertheless, attitudes need not be the case, even though the theory is ambiguous in this sense. However, some techniques for measuring attitudes are based on assumption of non-monotonous variables. For instance, we might not agree with a statement that is far from our *attitude ideal*. The statement's distance

from the attitude ideal can be understood in either negative or negative sense. This means, that we could disagree with a certain statement, because it is too strong or too weak. Blahuš (1996b) proposed a graphical visualization of this event that is depicted in Figure 8.

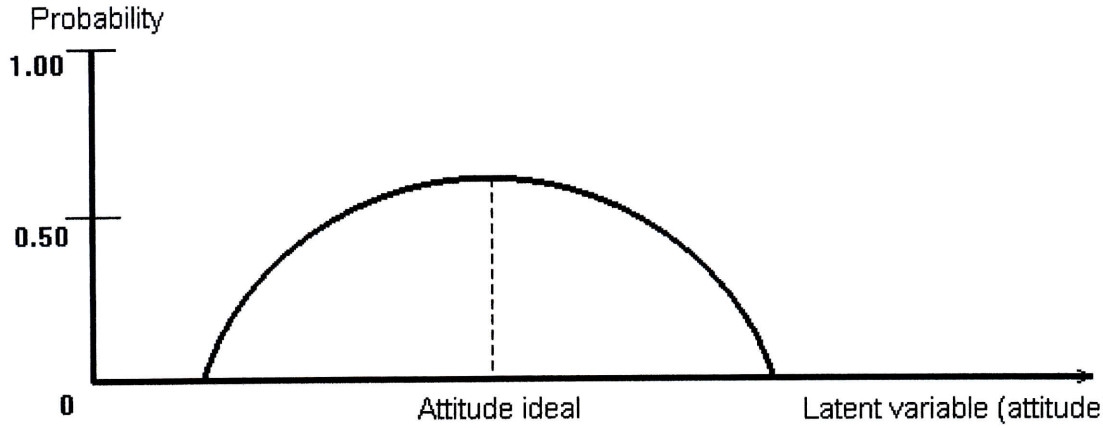


Figure 8. An illustrative example of non-monotonous function (attitude statement)

Note that cumulative models are a special case of so-called *unfolding models* with extreme statements. Unfolding models became widely used in IRT modeling most recently. These models may be useful in some sense, because resulting inferences about attitudes are adjusted for the response style (like in Figure 8) and should be less biased (Javaras & Ripley, 2007). Nevertheless, their incorporating is beyond the scope of this thesis, which is focused predominantly on classical approaches toward attitudes. Following section deals with theory of attitudes, measurement, and their applications in kinanthropology.

THEORY OF ATTITUDES

Definitions, Origins, and Structure of Attitudes

The *attitude* concept (construct) is defined in many ways. Core to most definitions is that attitudes reflect evaluations of objects (e.g., person, event, situation, etc.) on a dimension (or a scale) ranging from positive to negative (Fabrigar, MacDonald & Wegener, 2005). Although attitudes can be defined as simple object-evaluation association, attitudes may be part of larger sets of knowledge structures (Petty & Krosnick, 1995). This knowledge structure is stored in memory or is created only temporarily at the time of judgment.

From historical perspective, attitudes have been central to social psychology since its inception. Several researchers even defined social psychology as the scientific study of attitudes (Krosnick, Judd & Wittenbrink, 2005). First attitudes-related studies have emerged in first half of the 20th century (e.g., Thomas & Znaniecki, 1918; Thurstone, 1928; Folsom, 1931; Bogardus, 1931; Likert, 1932; Allport, 1935). To date, thousands of studies have been published on attitudes theory or its application in behavioral sciences.

Probably, the most comprehensive recently published book on attitudes theory is the co-edited monograph "*The Handbook of Attitudes*" (Albarracín, Johnson & Zanna, 2005). This book consists of 841 pages and 39 top experts in social psychology (e.g., Ajzen and/or Fishbein) have contributed to this monograph. It integrates the existing knowledge of methods, theories, and data in attitudes. The book is divided into two parts, where first part is more theoretical and descriptive, whereas second part is focused on attitude change, which is useful namely in experimental studies.

Researchers have investigated several different concepts related to attitudes. The evaluative responses could be classified into the categories of *behavior, affect,* and *cognition* (Rosenberg & Hovland, 1960). The fourth concept related to attitudes and investigated by many scholars are *beliefs*.

The concept of *behavior* was known earlier than the attitude concept in social psychology. The relationship between these two concepts was firstly summarized by Allport (1935). He stated that: "*An attitude is a mental disposition of the human individual to act for or against a definite object.*" Another similar definition was later formulated by Fishbein and Ajzen (1975). They stated that: "*An attitude is a learned predisposition to respond to an object in a*

consistently favorable or unfavorable way". However, the correlation between attitude and behavior need not to be positive. For example, one can have positive attitude toward physical activity but doesn't participate in physical activities at all due to his or her innate physical defect. On the other hand, behavior may also influence attitudes. This corresponds to the well-known fact that correlation doesn't necessary means causality. For instance, students are asked to evaluate a number of sports like soccer or basketball. Consequently, they are given a choice between two sports to be played within a physical educational class. For some students, the decision can be difficult because they rated the alternatives as very similar in attractiveness. For another student, the decision can be easier because one alternative was rated as much more attractive than the other. After making their decision, students will evaluate all of the sports again. Students who make a difficult choice will presumably increase their favorability toward the chosen alternative (toward rejected will be decreased), whereas other students will show a negligible change in their ratings of the alternatives. This procedure is known in psychology as the *free-choice paradigm* (Brehm, 1956). Olson and Stone (2005) have discussed three other theoretical perspectives on how behavior can influence attitudes: *biased scanning, dissonance motivation, and self-perception inferences.*

Affect is a prevalent concept in psychology which is defined as a momentary pleasant or unpleasant state (Schimmack & Crites, 2005). Authors also refer to a taxonomy of affective experiences which is characterized by type (emotions × moods), quality (pleasant × unpleasant), and aspects (duration, intensity, and frequency). In principle, *emotion* is temporally constrained object-focused state, *mood* is temporally constrained state without object constraint, *attitude* is the object-focused temporally unconstrained tendency (trait), and *temperament* is temporally unconstrained tendency (trait) without object constraint. Hence, attitudes and emotions are similar, because they are dedicated to specific objects.

Beliefs are defined as estimates of the likelihood that an event or state of affairs will occur (Fishbein & Ajzen, 1975). For example, we are often uncertain about whether the food we are eating at the same restaurant tastes as good as it did the last time; or whether a muscular athlete we are watching TV didn't abuse doping. These uncertainties also constitute beliefs (Wyer & Albarracín, 2005). Attitudes can be then viewed as a subcategory of beliefs. An interested reader is referred to Kruglanski and Stroebe (2005) for more details about association between beliefs and attitudes.

Measurement of Attitudes

As we noted in a previous section, an attitude is a theoretical concept, which is certainly not a simple sum of scores that our measurement procedure produces. Nevertheless, we are daily acquainted with the results of a variety of surveys claiming that they have measured attitudes. For instance, the attitudes toward politicians, new laws, and/or certain phenomenon in society are often expressed by a single numerical index reflecting the position of an attitude object on an evaluative continuum. However, such an approach is totally unacceptable from the scientific point of view. Already Thurstone (1928) noted that: "*Attitudes are multifaceted and attempting to describe them with a single numerical index is analogous to attempting to describe an object like a kitchen table with a single numerical index*". The main purpose of this section is to review the development of the main methods measuring attitudes from historical perspective and to summarize knowledge, ideas, and inventions used in measurement procedures nowadays.

Before we introduce procedures available for measuring attitudes, it is necessary to describe the process, how the evaluation is made by a respondent. The *evaluation* is a central term in attitude research. Krossnick, Judd, and Wittenbrink (2005) have distinguished between three stages of the evaluation process: a) an initial spontaneous activation of memory contents, b) a deliberation phase, c) a response phase. In the first stage, the memory activation occurs fast (few hundreds milliseconds) after encountering the attitude object (like brainstorming). Automatic activation should occur more likely for strong (central) attitudes which are more accessible. A deliberation phase is characteristic by controlled search for relevant information in memory. The order and the wording of questions in a questionnaire may impact this stage. The deliberation phase requires motivation and opportunity, because otherwise the input from the first stage will have a direct effect on the final stage - person's evaluative response.

In many researches, self-reports are the most common strategy. Self-reports involve asking participants explicitly do describe their own attitudes. However, also other approaches to measure participants´ attitude should be used, because self-reports need not to be reliable (Krossnick, Judd, & Wittenbrink, 2005). For instance, a particular student may express negative attitude toward smoking, but simultaneously we obtain information from his or her schoolmates or siblings that this particular student have very positive attitude toward smoking. As well, we can meet this particular student by accident and observe how he or she

enjoys a cigarette. This problem is related to reliability where different persons are sources of measurement error and to discriminant and convergent validity discussed in previous section.

One of the eldest classic self-report methods is the *Thurstone's equal-appearing interval method* (Thurstone, 1928). This preparation-intensive method consists of seven steps. First, up to 150 statements of favorable or unfavorable evaluations of an object are generated. This number is subsequently reduced to around 80. In the third stage, at least 200 judges place each statement into one of 11 ordered clusters, where the clusters are intended to be equally spaced points on a continuum ranging from extremely negative to extremely positive. Next, each statement is assigned a numeric value ranged from 1 to 11. The statements with large variances are then dropped. In the following step, two or three statements with means close to the points along the continuum are selected. Finally, the battery is ready and participants indicate those statements with which they agree. The resulting attitude score is an average over values of the statements that were endorsed.

Probably, the most widely known classic self-report method is the *Likert's method of summated ratings* (Likert, 1932). First, about 100 statements are developed which are oriented either positively or negatively toward object. The main difference between Likert's method and Thurstone's method is that statements expressing neutrality are not included in Likert's case. Subsequently, participants are given a set of five (or any other number) response options in a pilot study. The response options ranged from strongly disagree to strongly agree and the responses are coded from 1 to 5 for positively oriented statements and from 5 to 1 for negatively oriented statements. An illustrative example of evaluation process of attitudes toward physical education is presented in Table 3.

Table 3. Examples of the statements about physical education* and Likert-type scoring

Statements	Response options
Positive statement: *"Physical education is one of the most important subjects in the school program."*	1. Strongly disagree 2. Disagree 3. Undecided 4. Agree 5. Strongly agree
Negative statement: *"If for any reason a few subjects have to be dropped from the school program, physical education should be one of the subjects dropped".*	5. Strongly disagree 4. Disagree 3. Undecided 2. Agree 1. Strongly agree

Note: The statements have been taken from Wear (1955).

Participants are next assigned a total score by summing their scores over all statements. For each single item, the persons' scores are correlated with their overall sum scores. The

statements with low item score × sum score correlation are dropped. The final battery usually consists of about 20 items (Krosnick, Judd & Wittenbrink, 2005).

The most recent classic self-report method is the *semantic differential* (Osgood et al., 1957). Semantic differential was developed to measure connotative meaning of attitudes. Respondents mark a spot on a scale between two bipolar adjectives (good – bad, diligent – lazy, strong – weak, pleasant – unpleasant, and others). The response scale consists of several horizontal lines anchored by two antonyms intersect by six vertical lines. Hence, we obtain seven sections (scale points) scored from 1 (for the most negative response) to 7 (for the most positive response). The participants' attitude scores are averages of the scores they receive on each item in the battery. An illustrative example of the semantic differential of attitudes toward an athlete is depicted in Figure 9.

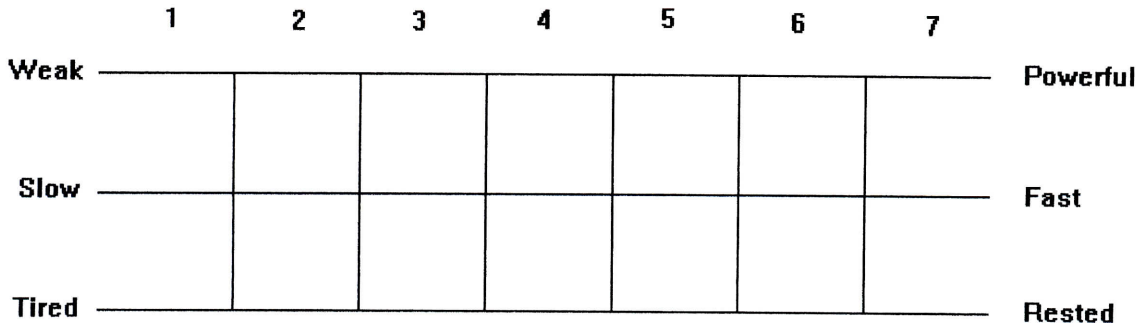


Figure 9. Attitudes toward an athlete: An illustrative example of the semantic differential

These three classic methods have several disadvantages but also advantages. They involve a large set of items to measure a single attitude. This is demanding and time-wasting for participants and investigators themselves. On the other hand, using so many items may lead to high reliability of the total sum score (according to the Spearman-Brown prophecy formula).

Another issue in attitude measurement is whether the open-ended or closed-ended questions and/or statements have to be used as an alternative self-report method. The greatest advantage of the closed-ended questions is that they enable to reach very large sample size with several hundreds or even thousands respondents. If the closed-ended questions are chosen, the decision about number of points on rating scale must be made. In most cases, 2-, 5-, and 7-point scales have been used whereas the scales included 3, 4, 6, 9, 10, and/or 11 scales are less common (Krosnick, Judd & Wittenbrink, 2005). In the present thesis, the Likert-type scales with 7 points have been used as the response format. The main advantage of the open-ended questions is that they enable to reveal special unique problems or aspects of

the construct which are not covered by closed-ended questions. However, the open-ended questions are demanding on encoding and analysis.

Till now, only the direct self-report methods have been considered. However, these methods are based on assumption that respondents are willing to describe themselves accurately. Therefore, several alternatives to self-report attitude measurement method have been developed to overcome the participants' misrepresentation. Namely, the *behavioral observation,* and/or *physiological measures* belong (among others) to this alternatives.

The classical examples of the behavioral observation techniques are the *lost-letter technique* (Milgram et al., 1965) and the *seating task* (Westie, 1953). In the first case, a letter with a certain address like "*to the Czech Olympic Committee*" is placed by an investigator on the ground nearby a mail box. Persons with positive attitudes toward the addressee will presumably pick up the envelope and put it in the mail box. The speed and form of return of this letter is considered as indicators of the attitudes. In the second case, the participants are asked to take a seat in a waiting room where the target person (attitude object) is already sitting. A person with negative attitude toward the object will presumably choose to sit farther from this target.

Physiological attitude measures express an association between physiological reaction and evaluative response (Krosnick, Judd & Wittenbrink, 2005). The main advantage of these measures is that participants have no access to their response and its connection with a specific evaluation. The *galvanic skin conductance* and *pupillary responses* belong among the eldest physiological methods measuring attitudes. (Cacioppo & Sandman, 1981). The first mentioned method is sensitive to arousal but can't distinguish whether this arousal was evoked by positive, negative or novel stimulus. The distinction between negative and positive evaluations was able in pupillary responses research where dilation of the pupil is associated with positive attitude (and constriction with negative). However, pupillary responses are influenced by a new stimulus (Petty & Cacioppo, 1983). Currently, another more sophisticated physiological attitude measures like *facial EMG activity*, *brain imaging techniques*, and/or *functional magnetic resonance imagery* (Krosnick, Judd & Wittenbrink, 2005) are used, nevertheless, the explanation of their principles is beyond the scope of this thesis.

Attitudes in Kinanthropology: Overview and Theories

An attitude research becomes a kinanthropological research as far as the attitude object is related to kinanthropology. In most studies, the *attitudes toward physical education* and *attitudes toward physical activity* have been investigated (Pecha, 2005). From the historical perspective, two main theories predominated in researches of attitudes toward either physical education or physical activity. First was the *Wear's theory* of attitudes toward physical education (Wear, 1951; Wear, 1955), and *Kenyon's theory* of attitudes toward physical activity (Kenyon, 1968).

Wear (1951) developed an inventory consists of 120 likert-type statements measuring attitudes toward physical education as an activity course. The main purpose of his first study was to reduce the number of items in this attitude inventory. As a result, he reduces 120 statements down to 40 items, which was called as a short form. In his second study (Wear, 1955), he constructed two equivalent forms of an attitude scale with 30 items within each form. The items were desired to measure 4 sub-constructs (or *categories* in Wear's original terminology) of attitudes: *physiological-physical*, *mental-emotional*, *social*, and *general*. These 4 sub-constructs were then indicators of the super-construct of *overall* attitude toward physical education. For better illustration, the Wear's theory of attitudes toward physical education is portrayed in Figure 10.

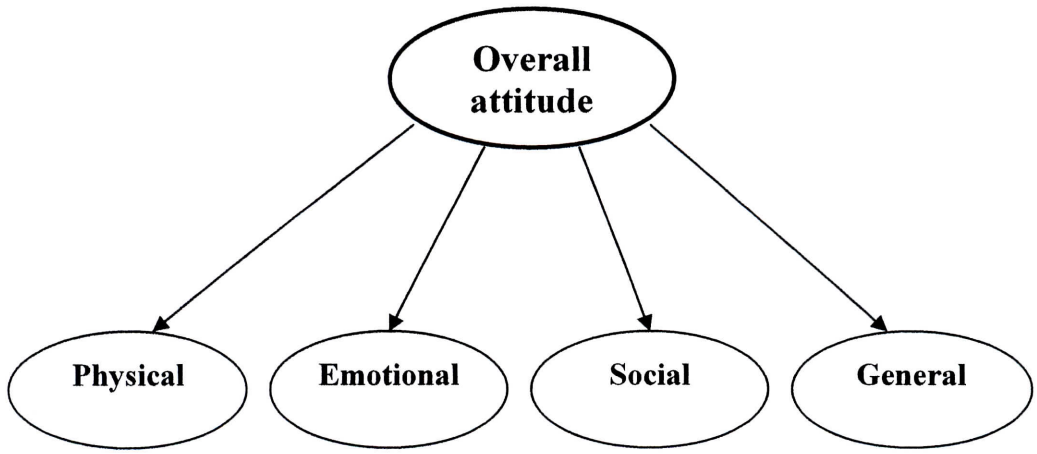


Figure 10. Wear's theory of attitudes toward physical education (empirical indicators omitted)

The sense of the arrows leading from the overall area of attitudes has been explained in previous chapter in section denoted to hierarchy of theoretical constructs. Both forms of the Wear's inventory have been further modified and used in various researches in

kinanthropology (Wessel & Nelson, 1964; Young, 1970). In our country, the Wear's inventory has been translated and revalidated by Blahuš (1984).

Wear's theory was criticized by some scholars because it was restricted only to school physical education which was perceived only as a part of much broader concept of physical activity. One of the best known was the *Kenyon's theory* of attitudes toward physical activity (Kenyon, 1968). He used an inventory consisting of 59 seven-alternative Likert-type attitude statements. He proposed that the overall attitude toward physical activity is composed of 6 sub-constructs (i.e. *subdomains* in Kenyon's terminology): *health and fitness*, *social experience*, *the pursuit of vertigo*, *aesthetic experience*, *catharsis*, and *ascetic experience*. Also Kenyon's theory is depicted for better notion as a path diagram in Figure 11.

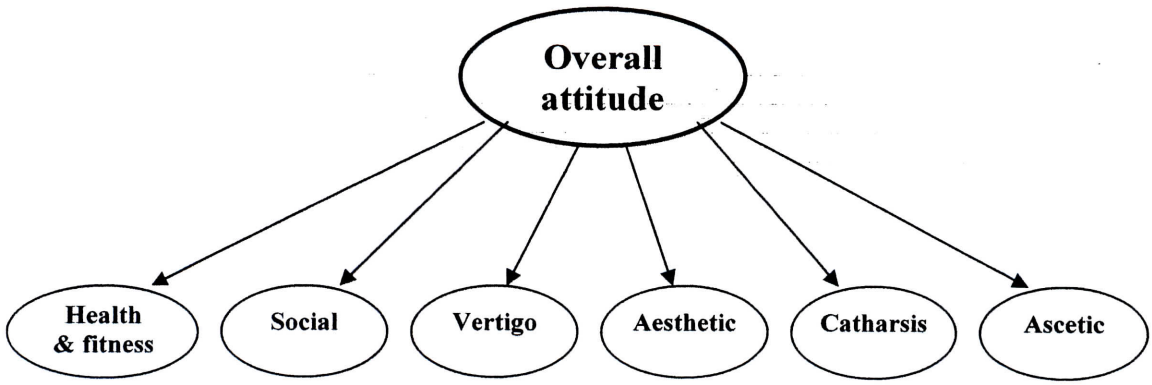


Figure 11. Kenyon's theory of attitudes toward physical activity (empirical indicators omitted)

If we compare these two introduced theories, it is obvious that namely ascetic experience and pursuit of vertigo sub-constructs should not be the aspects of physical education classes at schools. On the other hand, health and fitness and social experience sub-constructs are common for both theories.

Kenyon's theory served as a base for several subsequent researches. For example, Simon and Smoll (1974) modified Kenyon's attitude inventory to be appropriate also for small children. They have developed the CATPA inventory (Children's Attitude toward Physical Activity). This instrument was also used in several other studies (Smoll, Schutz & Keeney, 1976) and longitudinal study by Smoll and Schutz (1980). Sidney and Shephard (1976) conducted an experimental study of the effect of physical training program on attitudes toward physical activity in the elderly. Among other studies coming out of Kenyon's theory belong also studies carried out by Schutz et al. (1985) or Rice (1988). In our country, the

Kenyon's inventory was translated and modified by Svoboda (1971) who entitled this modification as DIPO (dimensions of attitudes).

Another theory of attitudes toward physical education was proposed by Mowatt et al. (1988). According to this theory, there are only 3 subdomains of attitudes toward physical education: *general*, *school curriculum*, and *scientific based domain.* Most recently, variety of studies on attitudes toward physical education or activity have been accomplished (e.g. Chung & Phillips, 2002; Hicks et al., 2001; Koca & Demirhan, 2004).

Till now, we have discussed physical education and physical activity as the only objects of the attitude researches in kinanthropology. However, also other objects have been investigated by some authors. For example, mainly negative attitudes toward obese people have been found among young athletes in USA (Chambliss et al., 2004). Other studies have dealt with public's attitudes toward the handicapped people (e.g. Hutzler, 2003; Miller & Cordova, 2002). Finally, another remarkable study was accomplished by Ryan et al. (2003), where authors examined covariate variables that may relate to student's attitudes toward physical education. The attitudes were substantially affected by teacher (in negative sense), curriculum, and social climate in a classroom (competitive or non-competitive).

In the Czech Republic, several research studies on attitudes have been accomplished. A variety of aspects of the attitudes toward physical education or physical activity have been investigated here. The most widely known works are the studies conducted by Jansa (2002), Slepička and Slepičková (2002), Frömel et al. (2007), and already mentioned paper by Svoboda (1971), among others.

HYPOTHESES AND RESEARCH QUESTION

Problem statement

In the present study, the high school students are nested within PE classes. Modified Wear's inventory of attitudes toward physical education has been used to determine the structure of students' attitudes toward physical education (ATPE). Since an attitude is a theoretical concept, the number of concepts, their kind, and relations among them of ATPE is of the main interest. While previous studies incorrectly modeled these concepts using single-level approaches even in multilevel case, the resulting structure of theoretical concepts and their indicators has been confounded by the between-group variation. ATPE should be therefore modeled at both within-group (students) and between-group (PE classes) levels, regardless whether there is an empirical support for multilevel modeling expressed by sufficiently high intraclass correlation. In addition, some statements of modified Wear's inventory are related to PE as a school subject whereas the others do not. From such reasons, the structure of ATPE concepts and indicators observed at within-group might not necessarily hold at between-group level.

Research question

What kind of theoretical concepts, their number and structure of relationships among them are indicated by the items of the modified Wear's inventory at both within-group and between-group levels?

Hypotheses

Based on described research problems and fundamental research question mentioned above, let us formulate three following hypotheses:

***H_1*:** While Wear's inventory was originally developed to measure attitudes toward physical education among individuals, it is assumed that the structure of attitudes will correspond to the Wear's multidimensional theory at the within-group level.

***H_2*:** With regard to either school or out-of-school related types of inventory items, we assume that the existence of two theoretical concepts, namely attitudes toward PE in school and attitudes toward PE out-of-school, indicated by aggregated Wear's inventory items will be empirically supported at the between-group level.

***H_3*:** Statistical reason for multilevel analysis is commonly expressed by substantive magnitude of the intraclass correlation coefficient. It is therefore assumed that the intraclass correlation coefficients of all manifest variables will be, on average, higher than 0.05.

ASSUMPTIONS AND LIMITATIONS

It might appear that joint application of two powerful methodologies like structural equation modeling (SEM) and multilevel analysis (MA) is very comprehensive and applicable to most situations in behavioral research. Nevertheless, also such studies (including our study) have several following limitations mainly for simplification reasons.

- *Two-level analysis*

In the present study, only two levels of nesting are taken into account, even though three levels of analysis are included here. High school students are nested within PE classes that are further nested within schools. There are at least four reasons, why we have decided to ignore the school level. First, there is insufficient sample size at the school level to carry out SEM (see Method part for more details). Second, ignoring school level has no impact on student level. In multilevel SEM, an investigator is mainly interested in structure of concepts and indicators at the lowest level (McDonald, 2006, personal communication). At the higher levels, multilevel SEM may serve only for explorative purpose. If the main aim is to investigate differences between in factor structure between groups, the multi-group approach is more appropriate. Hence, we have to keep in mind that the relations at the PE class level might be contaminated by the school level variation. Third, most applied studies in behavioral sciences resorted to such a simplification. Finally, most SEM software programs like LISREL or Mplus allow accomplishing structural analysis only at two levels. The three-level regression analysis is introduced in the follow-up study focused on alcohol consumption and smoking.

- *Linearity*

Another limitation of the present study is related to linearity assumption. It is assumed that all relationships between all manifest and latent variables are linear. The nonlinear SEM and especially nonlinear factor analysis has been theoretically elaborated long time ago (McDonald, 1962, Blahuš et al., 1973). Nevertheless, most of applications in behavioral sciences are data driven (Song & Lee, 2002). There is entire lack of theoretical reasons for nonlinear latent variable modeling in kinanthropology, even though some attempts have been

performed in modeling of strength ability (Blahuš et al., 1973). The possibility of nonlinear relation between manifest and latent variable has been discussed in section on measurement of theoretical concepts.

- *Statistical inference*

Some of the statistical methods used in the present thesis involve any type of random sampling. While the students have been sampled from already randomly sampled high schools in Prague, the findings are valid only for a population of students, PE classes, and schools in Prague. However, such a limitation is not in contrariety to our research question and hypotheses. This fact could be an obstruction for an investigator, who is, for example, interested in comparison of ATPE between students coming from cities and students coming villages.

- *The self-report mode*

This limitation was already discussed in the section on measurement of attitudes. Participants provide information about themselves. They describe their own ATPE, but they don't have to respond truly. It may happen that some other persons like parents, PE teacher, and/or classmates evaluate ATPE of a particular person differently. In other words, we might obtain different scores on the same inventory item related to one particular student. Hence, the reliability is subsequently overestimated when we use only the self-report mode. The report mode might be considered as so-called *hidden facet* which is detailed explained in the extended study of this thesis. There were two straightforward reasons for using self-report mode in our study. First, this approach is cost-effective and makes possible to reach large sample sizes. Second, our survey had to be anonymous at the student level with regard to affiliate questions about smoking and drinking.

Other limitations and assumptions are discussed in the extending and follow-up studies separately.

METHOD

PARTICIPANTS

Sampling procedure

The three-stage cluster sampling has been used in our study. In the first stage, a sample of high schools was randomly chosen from a population of high schools in Prague. A total of 11 schools participated in the survey. In the second stage, 87 PE classes were sampled from already chosen schools. The number of PE classes nested within schools ranged from 4 to 15 with median value of 7. In the last stage, the comprehensive sampling of students was accomplished from sampled PE classes. At the lowest level, participants were 1157 students. The number of students within PE classes ranged from 4 to 25 with median value of 13. Finally, the number of students within schools ranged from 38 to 201 with median equaled 95. Let us summarize the three levels of hierarchy:

Level three – 11 high schools

Level two – 87 PE classes

Level one – 1157 students

Inventory return and missing data

Originally, 13 high schools were randomly chosen, but two of them refused to participate in the study. At the PE class level, one sampled class could not be included in the survey from organizational reasons. A total of 1176 students were requested to complete inventories. Two of them rejected to cooperate. Another 17 inventories have been removed from analysis due to a large number of missing values or nonsensical responses, which results in a final number of 1157 students. The surprisingly small rate (1.6%) of missing values at the student level is not common in such surveys. The reason was that the survey was approved and organizationally supported by a school staff. Hence, all students fulfilled inventories during their regular education instead of objects like PE (most often), mathematic, and/or natural sciences. Therefore, also those students, who were not present at school during the day of

survey, are considered as missing values. About three students were usually missing within each PE class.

Sample descriptive statistics

Absolute and relative frequencies sorted by students' gender and age are provided in Table 4.

Table 4. Gender and age: Absolute and relative frequencies

	≤15 years	16 years	17 years	18 years	19 years	20 years ≤	Total
Boys	31	125	146	169	88	31	590 (51%)
Girls	39	126	146	172	67	17	567 (49%)
Total	70 (6.1%)	251 (21.7%)	292 (25.2%)	341 (29.5%)	155 (13.4%)	48 (4.2%)	1157

To have more complex notion about students, we were also interested in their participating in physical activity (PA) beyond the framework of school PE. We had no ambition to rigorously monitor their PA with the aspects like frequency and intensity. We have simplified this problem only to questions whether they are participating in PA or not and in which PA type they are involving. No PA was reported by 34% respondents (39% of all girls and 29% of all boys). Ten most frequent PA types practiced by the rest 763 students are depicted as a horizontal bar graph in Figure 12.

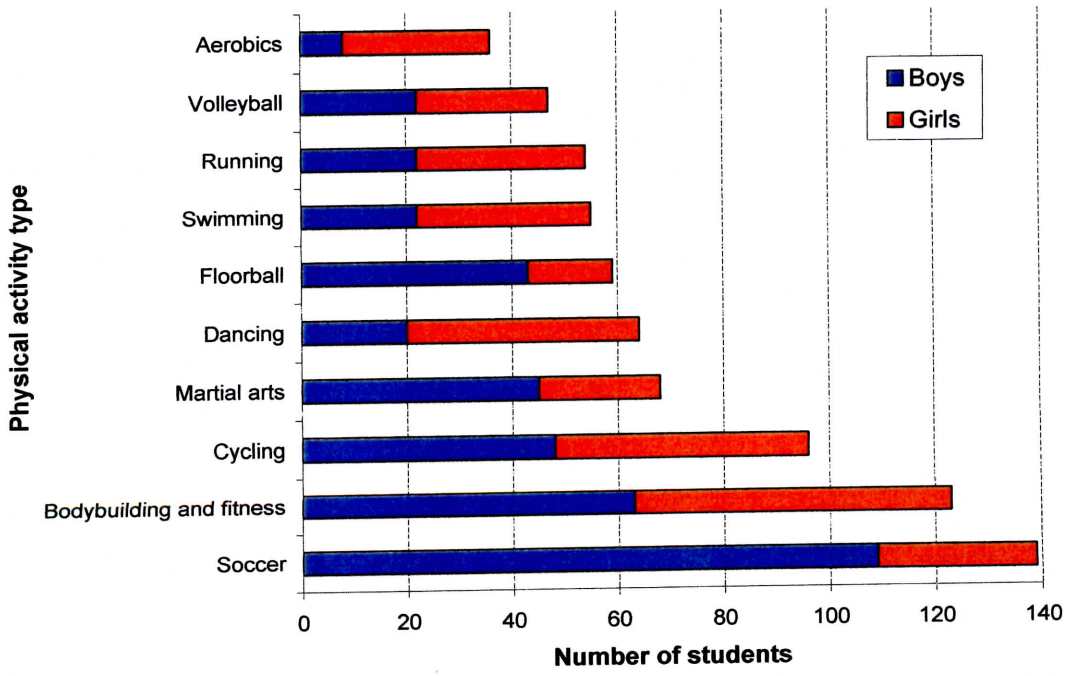


Figure 12. Top 10 PA types, in which respondents (N = 763) were involved

Furthermore, students were also involved in the following PA: tennis, squash, basketball, walking, rock-climbing, break-dance, in-line skating, ice hockey, horseback riding, table tennis, track and field, skiing, yoga, and/or handball. All participants were allowed to specify more than one PA type.

At the school-level, diverse types of schools have been included in the study. Participating schools differ in curriculum, level of providing education as well as number, types and quality of equipment for physical education. A brief description of all schools and their equipment for PE is presented in Table 5. This table is supplemented by information about percentage of graduate students accepted to the university last year. Such information may serve as an indicator of level of education on particular school.

Table 5. Basic characteristics at the school-level

School id	School type	Equipment for PE	Admission to university
1	grammar school	2 large gyms, small gym, outdoor playground, weight room	95%
2	commercial academy	small gym, access to external playground	44%
3	technical school	2 gyms, 2 playgrounds, tennis court, swimming pool (25m)	65%
4	social and health high school	access to external gyms	23%
5	educational establishment	2 small gyms	5%
6	grammar school	small gym	82%
7	vocational school	2 gyms, playground	40%
8	grammar school	large gym, small gym, outdoor playground	92%
9	vocational school	2 gyms, playground	68%
10	educational establishment	small gym, outdoor basketball court	18%
11	grammar school	2 large gyms, playground	94%

INSTRUMENT

Inventory development

A modified Wear's inventory has been used for data collecting. As we noted in previous section on attitudes in kinanthropology, the origin of this inventory traces back to C. L. Wear (1951), who developed 120 statements measuring ATPE. Subsequently, two equivalent forms consisting of 30 items were created by the same author (Wear, 1955). Later on, these two forms were translated to Czech language by Blahuš (1984), who also re-validated the inventory using confirmatory factor analysis (CFA). In the present thesis, a total of 40 items have been randomly sampled from the pool of all items to be evaluated in the content validation process. Strictly speaking, only the item relevance of the statements has been evaluated by "experts" here. The rating of item relevance is very detailed discussed in the extending study of this thesis. We addressed the question of whether the same items re-validated more than 20 years ago are comprehensible also for contemporary teenagers. Moreover, an evaluation of the content validity is indispensable stage in inventory or questionnaire development (Thomas & Nelson, 2001). After this procedure, 20 items have been accepted without any change, 10 items have been modified, and 10 items have been evaluated as unacceptable and irrelevant to measure ATPE among high school students. The remaining 30 items are listed in Appendix.

Structure of the inventory

All items have got three basic aspects. First aspect is related to the sub-construct of ATPE, which is assumed to be measured by an item (health and fitness, social, emotional, and general). Second viewpoint is whether an item is related to school PE or to exercise beyond the obligatory school subject. The last aspect is linked to orientation of a statement. The statements are either negatively or positively oriented. If only positively oriented statements are included in an inventory, the participants may tend to response stereotypically to all items. Therefore, both orientation of statement have been used in our study. However, the use of positively (negatively) worded items may produce artificial relationships on the inventory (Podsakoff et al., 2003). Hence, we may expect the occurrence of two wording-related factors.

The summary of the aspects of all items is provided in Table 6. All items are listed in the same order as in the Appendix.

Table 6. Three aspects of all items: The structural hypotheses of the inventory

Item number	Sub-construct of ATPE	Related to ATPE as a school subject?	Orientation of the statement
1	general	yes	negative
2	social	no	positive
3	emotional	yes	negative
4	social	no	positive
5	general	no	negative
6	emotional	no	positive
7	health and fitness	no	negative
8	general	yes	negative
9	health and fitness	no	negative
10	general	no	negative
11	health and fitness	yes	positive
12	social	no	positive
13	health and fitness	no	positive
14	general	yes	negative
15	social	no	positive
16	emotional	no	negative
17	health and fitness	no	negative
18	emotional	yes	negative
19	general	yes	positive
20	social	yes	negative
21	general	yes	negative
22	emotional	no	negative
23	health and fitness	no	positive
24	health and fitness	yes	negative
25	social	no	positive
26	general	no	negative
27	emotional	yes	positive
28	social	no	positive
29	health and fitness	no	positive
30	general	yes	negative

Each statement has been rated on a 7–point Likert-type response format (7 – Strongly agree; 6 – Agree; 5 – Slightly agree; 4 – Neither agree nor disagree; 3 – Slightly disagree; 2 – Disagree; 1 – Strongly disagree). We have decided for this number of scale points mainly because of problems related to ordinal origin of the scale, which is discussed in detail it the final part of this Method section.

PROCEDURE

Inventory administration

The inventory survey was accomplished during the 2006-2007 school year. Inventories were completed at least three months after beginning of the school year to ensure that students had enough experience to make informed judgment regarding their ATPE. All necessary permissions were obtained from the faculty staff and the 11 school headmasters prior to data collection. An explanation of the study was presented to each teacher, whose class was used for inventory completion. The teacher was not present during this procedure in the classroom. The inventory of ATPE was expected to take around 15 minutes to complete. In reality, however, the inventories were together with briefing and part about alcohol and smoking completed within 25 minutes. The inventory administration and an introduction for students were provided by the author of this thesis personally. Nevertheless, some briefings had to be provided by two other helpers from organizational reasons. These two helpers were trained in the same way and equipped by written instructions.

The main terms, purpose of the study, and principles how to complete the inventory were explained to the participants within the preliminary briefing. Namely, the term exercise had to be explained to the students. This term was defined in almost same way as was proposed by LaMonte, Ainsworth, and Reis (2006) According to the authors, exercise is planned, structured, and repetitive bodily movement done to improve or maintain one or more components of physical fitness. For example, sex activity or housework could not be considered as exercise. All inventories were completed anonymously, because the extending part contained personal information about participants.

Data screening

Transcription of the data from paper forms to the electronic MS EXCEL spreadsheet was accomplished manually within 3 months. Subsequently, data were screened by EXPLORE procedure in SPSS software to identify missing values and to eliminate keying mistakes. Finally, data were transformed to proper WordPad format to serve as an input files for two SEM programs LISREL (Jöreskog & Sorböm, 2005) and Mplus (Muthén & Muthén, 2005). For reliability and generalizability analysis, the pack of GENOVAs was used.

ANALYSIS METHODS

In the present section, we introduce basic principles, assumptions, mathematical background, and limitations of single-level SEM and especially CFA. In the next step, a multilevel extension of CFA is discussed. Finally, SEM and CFA for ordinal variables are considered here. The ordinal variables paradigm within SEM framework is discussed in both single-level and two-level cases.

Introduction to structural equation modeling and confirmatory factor analysis

In principle, CFA and SEM are statistical techniques that one can use to reduce the number of observed variables into a smaller number of latent variables by examining the covariation among the observed variables (Schreiber et al., 2006).

CFA is confirmatory, theory driven technique. When CFA is conducted, an investigator uses a hypothesized model to estimate a population covariance (or correlation) matrix that is compared with the observed covariance (or correlation) matrix. The technique of CFA analyzes a priori *measurement models* in which the number of factors and their correspondence to the indicators are explicitly specified (Kline, 2005).

SEM can be viewed as a combination of CFA and path analysis. In comparison with CFA, SEM extends the possibility of relationships among the latent variables and encompasses two components:

1. *Measurement model* – essentially the CFA
2. *Structural model* – multiple regression with latent variables

One of the most widely known and used SEM model is the LISREL (LInear Structural RELations) model (Jöreskog, 1970). We would like to introduce this model even before historical overview of SEM development, because the LISREL notation will be henceforth used in the present thesis. The full LISREL model is specified as follows:

Measurement model:
$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{\Lambda_x}\boldsymbol{\xi} + \boldsymbol{\delta} \\ \boldsymbol{y} &= \boldsymbol{\Lambda_y}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \end{aligned} \tag{1}$$

Structural model: $\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$, (2)

where ξ is a random vector of latent exogenous variables, x is a vector of observed exogenous variables, Λ_x is a coefficients matrix of the regression of x on ξ, and δ is a vector of measurement errors in an observed exogenous variable x. Similarly, η is a random vector of latent endogenous variables, y is a vector of observed endogenous variables, Λ_y is a coefficients matrix of the regression of y on η, and ε is a vector of measurement errors in an observed endogenous variable y. In equation 2, B is a coefficients matrix of the η in the structural relationship, Γ is a coefficient matrix for latent exogenous variables ξ, and ζ is a vector of regression residual errors in the structural relationship of η on ξ.

It is assumed here, that all observed variables are centered. Therefore, no intercept terms are included in equation (1) and (2). Exogenous and endogenous variables can be observed or unobserved, depending on the model being tested. Exogenous variables represent those concepts that exert an influence on other concepts under study and are not influenced by any concepts in the system. Those concepts identified as endogenous can be affected by exogenous and other endogenous variables in the model. This specification was used from several reasons mainly related to historical context of the model development. However, the model can be written in the so called "all-y" LISREL notation. The all-y notation is a much simpler system. It does not distinguish between exogenous and endogenous variables. Almost all SEM models can be conducted within the all-y notation. Hence, this notation is used also for the purpose of the present thesis. An illustrative example of the LISREL model in all-y notation is portrayed as a path diagram in Figure 13.

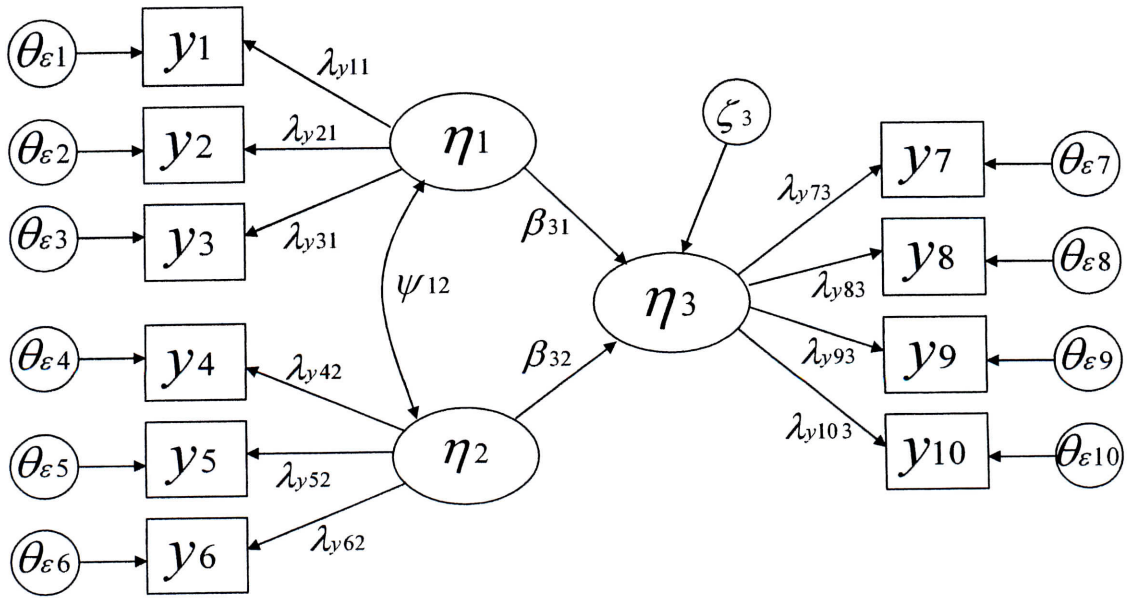


Figure 13. An illustrative SEM model in LISREL "all-y" notation

The path diagram in Figure 13 involves 10 observed variables and 3 latent variables. Third latent variable is hypothesized to be estimated by the first two here. The same hypothetical model from Figure 13 can be expressed using matrix algebra as follows:

Measurement model:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \end{pmatrix} = \begin{pmatrix} \lambda_{y11} & 0 & 0 \\ \lambda_{y21} & 0 & 0 \\ \lambda_{y31} & 0 & 0 \\ 0 & \lambda_{y42} & 0 \\ 0 & \lambda_{y52} & 0 \\ 0 & \lambda_{y62} & 0 \\ 0 & 0 & \lambda_{y73} \\ 0 & 0 & \lambda_{y83} \\ 0 & 0 & \lambda_{y93} \\ 0 & 0 & \lambda_{y103} \end{pmatrix} \times \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \end{pmatrix}$$

Structural model:

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{pmatrix} \times \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{pmatrix}$$

The measurement model of SEM is the CFA. For instance, the illustrative example depicted in Figure 4 in introduction section can be expressed in all-y LISREL notation as follows:

$$\begin{pmatrix} Bench-press(y_1) \\ Full-ball-throw(y_2) \\ Push-ups-on-parallel-bars(y_3) \\ 12-\min-run-on-belt(y_4) \\ 10-\min-stationary-bicycle(y_5) \\ Endurance-shuttle-run(y_6) \end{pmatrix} = \begin{pmatrix} \lambda_{y11} & 0 \\ \lambda_{y21} & 0 \\ \lambda_{y31} & 0 \\ 0 & \lambda_{y42} \\ 0 & \lambda_{y52} \\ 0 & \lambda_{y62} \end{pmatrix} \times \begin{pmatrix} Strength(\eta_1) \\ Endurance(\eta_2) \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}$$

Covariance structure analysis is an alternative name for SEM. This name captures more explicitly the fact that the covariance matrix of observed variables is used as initial information to be analyzed. From the sample data it is possible to construct the sample covariance matrix $\boldsymbol{S}$ with elements s_{ij}. Off-diagonal elements are covariances, and diagonal elements are variances of the observed variables. If the data are standardized, $\boldsymbol{S}$ matrix

contains correlations. The population covariance matrix $\boldsymbol{\Sigma}$ is estimated through the following covariance equation:

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Lambda}}_y \hat{\boldsymbol{\Psi}} \hat{\boldsymbol{\Lambda}}_y' + \hat{\boldsymbol{\Theta}}\boldsymbol{\varepsilon}, \qquad (3)$$

where $\hat{\boldsymbol{\Lambda}}_y$ is an estimate of the population matrix $\boldsymbol{\Lambda}_y$ defined in equation 1. $\hat{\boldsymbol{\Psi}}$ is an estimator of $\boldsymbol{\Psi}$, which is the population covariance matrix of the latent disturbances $\boldsymbol{\zeta}$. In the case of CFA, the covariance matrix of $\boldsymbol{\zeta}$ is identical with covariance matrix of latent variables $\boldsymbol{\eta}$. However, if at least one dependent latent variable is included in the model, $\boldsymbol{\Psi}$ matrix and covariance matrix of $\boldsymbol{\eta}$ are different. Finally, $\hat{\boldsymbol{\Theta}}\boldsymbol{\varepsilon}$ estimates the population matrix $\boldsymbol{\Theta}\boldsymbol{\varepsilon}$ that is covariance matrix of measurement errors $\boldsymbol{\varepsilon}$. The $\boldsymbol{\Theta}\boldsymbol{\varepsilon}$ matrix is usually considered as a diagonal matrix with variances on diagonal known as uniqueness. Estimation involves finding values of $\hat{\boldsymbol{\Lambda}}_y$, $\hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\Theta}}\boldsymbol{\varepsilon}$ that generate an estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ that is as close as possible to the sample covariance matrix $\boldsymbol{S}$ (Long, 1983). It is assumed here that $E(\boldsymbol{\eta}) = \boldsymbol{0}$, $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$, and $\boldsymbol{\varepsilon}$ is uncorrelated with $\boldsymbol{\eta}$.

Brief historical overview of SEM and CFA development

It this section, we briefly introduce the development of models with latent variables from historical point of view. A special attention is paid to the role of factor analysis as a research tool in our country and namely in kinanthropology. The summary of contemporary SEM software programs can be found in the final part of this section.

The historical overview should begin with Francis Galton (1822 – 1911), the forerunner of factor analysis and a cousin of Charles Darwin. A primary objective of Galton was to explain differences among people, which is also the main objective of contemporary psychometrics (Furr & Bacharach, 2007). The best known of Galton's pupils was Karl Pearson (1857 – 1936), one of the greatest statisticians ever. Pearson solved the mathematical problems related to the correlation coefficient that is nowadays known as a Pearson's correlation coefficient. Pearson developed also the χ^2 - test, however, he did not invented factor analysis, because he disregarded measurement error. Dealing with unobserved phenomena (true scores) seemed unthinkable to him (Cudeck et al., 2001). The father of factor analysis was Charles Spearman (1863 – 1945). He laid the foundation of psychometrics as a separate branch of quantitative science by writing two fundamental articles in 1904. He

introduced the concept of reliability and discussed the principle of attenuated correlations. In his second article from 1904, Spearman's one factor model for *general intelligence* was introduced. He correlated grades of six school subjects, where the data was collected from school children in London. His follower within the field of intelligence research, Cyril Burt (1883 – 1971), argued that Spearman's one factor model was too simple and proposed *group factor models*. These models were completed with some factors only accounting for some part of the observed variables. In accordance with this idea was also Raymond Cattell (1905 – 1998), who proposed to take two group factors (fluid and crystallized intelligence) into account. In 1921, Sewell Wright, the geneticist, developed the basics of path analysis. In 1933 Harold Hotelling (1895 – 1973) developed *principal components* as an estimation method. As we already mentioned in the section on attitudes, another important person was Louis Leon Thurstone (1887 – 1955). He was a forerunner of the *confirmatory approach* to factor analysis and he developed such methods like rotation to simple structure, matrix algebra notation, and/or generalization to the oblique factors (Cudeck et al., 2001). Alternative approaches were elaborated by Louis Guttman (1916 – 1987). He was mainly interested in scaling, but also in factor analysis made some important contributions like, for example, communality estimation, radex models, and hierarchical principle (this means hierarchy of factors). Finally, the Keesling-Wiley model (Wiley, 1973) was the first structural equation model, at least from current understanding. Subsequently, the LISREL model arose out of the work of Keesling and Wiley and was developed to its current state by Jöreskog (1973). This model is also known as the JKW model (Bentler, 1980). It was also proven that LISREL can be viewed as a special instance of more general models (McDonald, 1978*b*; Muthén, 2002). Nevertheless, their presentation and description is beyond the scope of this thesis.

Up to date, a large amount of important SEM monographs have been published (e.g., McDonald, 1985; Bollen, 1989; Kaplan, 2000; Kline, 2005). Furthermore, a specialized multidisciplinary journal *Structural Equation Modeling* began published in 1994. SEM has become widely used method by researchers mainly after computer software development and elaboration. Currently, several SEM computer programs exist. In this thesis, the LISREL (Jöreskog & Sörbom, 2005) and the Mplus (Muthén & Muthén, 2005) programs have been used. These two SEM programs are discussed in more detail in result section and syntax files are provided in Appendix. Another commercial program is AMOS (Arbuckle, 1997). It includes a graphic module that enables specifying a model by drawing a path-diagram on screen. Another well-known commercial SEM statistical package is the EQS program

(Bentler, 1985). The CALIS (Covariance Analysis and Linear Structural Equations) is a procedure that runs under SAS software package (Hartmann, 1992). Similarly, RAMONA (Browne, Mels & Cowan, 1994) is currently the module for SEM in SYSTAT. SEPATH (Steiger, 1995) is the SEM module in STATISTICA statistical package. Finally, available freeware includes COSAN (Fraser & McDonald, 1988), GEFA (Blahuš, 1992), and Mx Graph (Neale, 1997).

The process of SEM and CFA

The process of SEM and CFA could be thought of several stages. For example, Ullman (2006) listed four steps of SEM: 1. *model specification* (includes issues as model hypotheses and diagram, model identification, sample size and missing data consideration, and multivariate normality and outliers), 2. *model estimation* (estimation method like maximum likelihood and/or weighted least squares), 3. *model evaluation* (fit indices, and interpreting parameter estimates), 4. *model modification* (statistical tests like Chi-square difference test, and/or Wald test). More general stages of SEM and CFI were proposed by Kline (2005). He identified six iterative steps of SEM:

1. *Specify the model* – hypotheses are expressed in the form of SEM
2. *Determine whether the model is identified* – is it possible to estimate all parameters?
3. *Select measures* – this step includes data collection (!), data screening and preparation
4. *Conduct the analysis* – use SEM program and evaluate model fit, interpret the parameter estimates, and consider equivalent models
5. *Respecify the model* – evaluate the fit of the revised model (if necessary)
6. *Write the report* – follow suggested rules for describing the analysis in written reports

We have inserted the exclamation mark in the point 3 to emphasize that a SEM researcher's work doesn't begin with data collection. As was noted in the first two steps, several important considerations have to precede the data collection stage. However, there is a common and not appropriate practice that investigators first collect the data and try to apply SEM to analyze their data. This approach is purely exploratory and data-driven without any a priori notion about structure of concepts and their indicators.

Some of the steps mentioned above have been already discussed in previous sections of the present thesis (data collection design, and structural hypotheses). In this part, the following topics are briefly discussed: (a) *checking assumptions*, (b) *model identification*, (c) *parameter estimation methods*, (d) *evaluating model fit,* and (e) *model modification.* Finally, we summarize some recommendations for presenting SEM and CFA results in a written report.

(a) Checking assumptions

In previous section, we have described simple data screening methods related to the sample like outliers or missing data. Furthermore, the data must be examined to see whether they fit statistical assumptions of SEM and CFA. The most commonly used estimation methods in SEM assume (Kline, 2005):

- *Multivariate normality* – it is assumed here that all univariate distribution of observed variables are normal, that joint distribution of any pair of the variables is bivariate normal, and that all bivariate scatterplots are linear and homoscedastic. The inspection of all scatterplots is suggested by some authors as a nontechnical evaluative issue of multivariate normality (Schreiber et al., 2006).
- *Univariate normality* – it is assumed here that all observed variables are normally distributed. Skewness and kurtosis are two ways that a distribution can be nonnormal. Positive skewness indicates that most of the scores are below the mean. Positive kurtosis (leptokurtic distribution) indicates heavier tails and higher peak. Standardized indices of skewness (*SI*) and kurtosis (*KI*) that permit the comparison of different distributions to the normal curve are calculated as follows:

$$SI = \frac{S^3}{(S^2)^{\frac{3}{2}}}, \text{ and } KI = \frac{S^4}{(S^2)^2} \qquad (4)$$

where S^2, S^3, S^4 are the second, third, and fourth moments about the mean, respectively. The sign of the *SI* indicates the direction of the skewness, and the value of zero indicates a symmetrical distribution. The value of *KI* in a normal distribution equals 3 and the value greater than 3 indicates positive kurtosis. Note that the same rules hold for negative kurtosis and skewness in opposite sense. A way to deal with univariate nonnormality is with *transformations* of the original scores. There are several types of

transformation like centroid, square root, probit, logarithmic, and/or inverse. However, transformation means that the variable's original metric is lost and the interpretation of results for that variable must be made in the transformed metric.

- *Linearity and homoscedasticity* – these two aspects of multivariate normality can be evaluated through the inspection of bivariate scatterplots. As we noted in limitations section, linearity is one of the assumption of the present thesis. Homoscedasticity is also known as homogeneity of variance around regression line. In a scatterplot of data, homoscedasticity looks like an oval or an ellipse, whereas any other geometrical shapes (for example triangle) indicate violation of this assumption – heteroscedasticity.
- *Lack of multicollinearity* – it is assumed that all covariance matrices are not singular. Multicollinearity occurs when intercorrelations among some variables are too high (e.g., larger than 0.9). It can occur because what appear to be two separate variables measure the same thing (Kline, 2005). Such two variables are essentially redundant. For example, it makes little sense to consider 12 min run and 3 km run as two different indicators of the endurance ability because they will be presumably extremely highly correlated and bring no additional information. Either one or the other variable could be included in the same analysis. It is easy to spot multicollinearity between two variables by inspecting correlation (not covariance) matrix. For detecting multicollinearity among three or more variables, it is suggested to calculate a squared multiple correlation and apply the same rules, namely that the value higher than 0.9 indicates multicollinearity. There are two ways to deal with multicollinearity. We can either eliminate variable or combine redundant ones into a composite variable.

(b) Model identification and sample size

Two main issues must be considered to detect whether the SEM analysis can be mathematically conducted. The requirements of sufficient sample size of objects of measurement (e.g., students) and model identification are discussed in this section.

- *Sample size* is an important issue because it relates to the stability of the parameter estimates. SEM is still a large-sample technique. Several factors affects sample size requirements in SEM: *complexity of the model* (more complex models involve higher sample size than simpler), the type of *estimation algorithm*, and/or type of *variable scales* (discussed later on). Small sample size means that the power of statistical tests may be

very limited. Several recommendations are provided by some authors about minimal sample size. For example, Kline (2005) claims that sample size should exceed 200 cases. Another recommendation is that sample size should be at least 8 times the number of observed variables (Kaplan, 2000). Bentler and Chou (1987) recommend at least 5 cases per estimated parameter. There is no general rule for the number of observations needed, but at least 10 per estimated parameter appears to be the general consensus nowadays (Schreiber et al., 2006). In our case, the model depicted in Figure 14 has 64 parameters, which means that we have an acceptable ratio of 18 students to 1 parameter estimated for this model in a single-level approach. The sample size needed in multilevel SEM and CFA is discussed later.

- *Model identification* – only models that are identified can be estimated. A model is said to be identified if there is a unique numerical solution for each of the parameters in the model (Ullman, 2006). At this point, the data are not students themselves but covariances and variances in the sample covariance matrix ***S***. Identification is a property of the model regardless how many students participated in a study. There are two basic requirements for identification of any kind of model in SEM or CFA (Kline, 2005): (1) there must be at least as many observations (it means variances and covariances) as free model parameters and (2) every latent variables must be assigned a scale. The number of observations is the number of nonredundant sample variance and covariances that can be calculated as follows:

$$\text{Number of observations} = \frac{p(p+1)}{2}, \tag{5}$$

where p equals the number of observed variables.

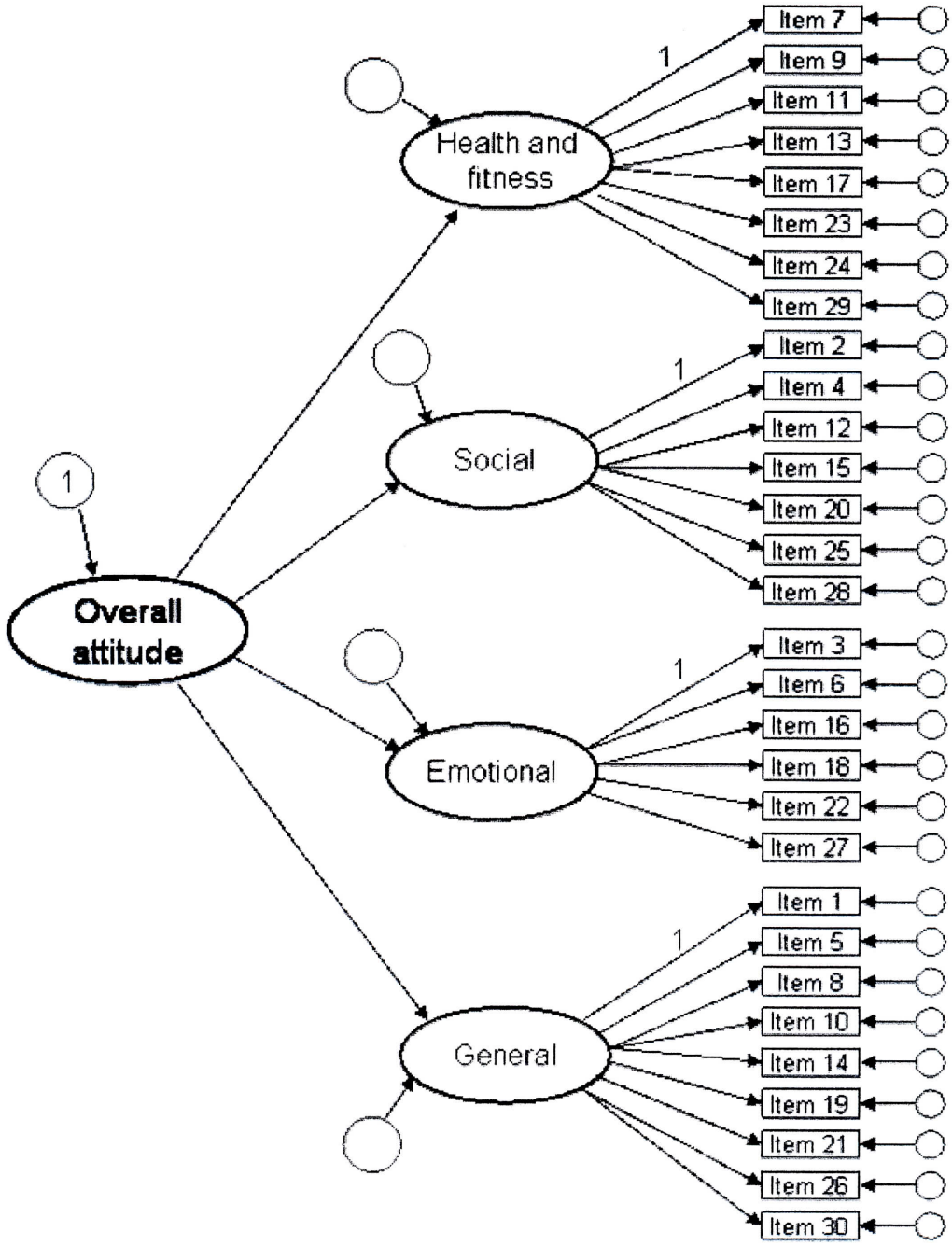


Figure 14. One of the hypothesized models: Second-order confirmatory factor analysis

As was already noted above, the model depicted in Figure 14 has 64 free parameters, namely 26 free parameters (factor loadings) are within the Λ_y matrix, 30 free parameters (uniqueness) are situated within the diagonal $\Theta\varepsilon$ matrix, 4 free parameters (regression coefficients) has the B matrix, and 4 free parameters (residual/error variances) inhere in the diagonal Ψ matrix. Number of nonredundat covariances and variances is 30 × 31 / 2 = 465. Our model has more observations/data than parameters to be estimated. Such models are said to be *over identified.* In other words, this model has positive degrees of

freedom (*df*), where *df* = number of observation – number of free parameters, it means: *df* = 465 – 64 = 401. If there are the same number of observations/data as free parameters (*df* = 0), the model is said to be *just identified.* In this case, the estimated parameters perfectly reproduce the sample covariance matrix ***S***. If there are fewer observations/data than free parameters, the model is said to be *under identified* and parameters cannot be estimated. Therefore, the number of parameters needs to be reduced by fixing and/or constraining. Besides the free parameters, there are also five nonzero fixed parameters in our hypothesized model (see Figure 14): four factor loadings and variance of the "overall attitude" factor are set to 1. The reason is that some metric/scale must be added to all latent variables. Consequently, the four first-order factors have the same metric as the respective indicators with fixed factor loadings and the second-order factor is measured in standardized *z*-scores. The requirement of latent variables assigned with scale (second requirement) is hereby fulfilled.

(c) Parameter estimation methods

As we noted earlier, estimation involves finding parameters that generate an estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ that is as close as possible to the sample covariance matrix ***S***. A function that measures how close a model-generated $\hat{\boldsymbol{\Sigma}}$ is to the sample covariance matrix ***S*** is called a *fitting function.* There are several estimation methods in CFA and an appropriate choice of one of them must be done by an investigator. In making such a decision, robustness questions play a crucial role (Boomsma, 2000). In the present section, we will introduce two of them that assume multivariate normality and relatively large sample size (several hundred): *generalized least squares* (GLS) and *maximum likelihood* (ML).

- The fitting function for GLS is (cf. in textbook by Long, 1983):

$$F_{GLS}(\boldsymbol{S};\hat{\boldsymbol{\Sigma}}) = tr\left[\left(\boldsymbol{S}-\hat{\boldsymbol{\Sigma}}\right)\boldsymbol{S}^{-1}\right]^2, \tag{6}$$

  where "tr" is the trace of the square matrix $\left[\left(\boldsymbol{S}-\hat{\boldsymbol{\Sigma}}\right)\boldsymbol{S}^{-1}\right]^2$ and is defined as the sum of the elements on the main diagonal of this matrix. As $\hat{\boldsymbol{\Sigma}}$ approaches ***S***, the value F_{GLS} becomes smaller. If ***S*** equals $\hat{\boldsymbol{\Sigma}}$, the function necessarily equals zero. This method has the advantage of being scale free.

- The ML estimator minimizes the fitting function defined as follows (e.g. Long, 1983):

$$F(\boldsymbol{S};\hat{\boldsymbol{\Sigma}}) = tr\left(\boldsymbol{S}\hat{\boldsymbol{\Sigma}}^{-1}\right) + \left[\log\left|\hat{\boldsymbol{\Sigma}}\right| - \log\left|\boldsymbol{S}\right|\right] - p, \qquad (7)$$

where $\left|\hat{\boldsymbol{\Sigma}}\right|$ and $\left|\boldsymbol{S}\right|$ are the determinants of $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{S}$ matrices, respectively. If $\boldsymbol{S}$ and $\hat{\boldsymbol{\Sigma}}$ are similar, also their inverses will be similar. Likewise, the term $\boldsymbol{S}\hat{\boldsymbol{\Sigma}}^{-1}$ becomes closer to the $p \times p$ identity matrix. Since the trace of this identity matrix equals p, the first term the equation (7) approaches the value of p as $\boldsymbol{S}$ and $\hat{\boldsymbol{\Sigma}}$ become closer. Subsequently, the second term in equation (7) approaches zero. The last term in equation (7) is the constant p, which serves to cancel the value approached by the first term. Accordingly, if $\boldsymbol{S}$ equals $\hat{\boldsymbol{\Sigma}}$, the fitting function will equal zero.

Both these methods can give valid parameter estimates even when the data are nonnormal, whereas standard errors do not. ML and GLS estimation can give biased standard errors and incorrect test statistics in the presence of excessive skewness and/or kurtosis (McDonald & Ho, 2002). There are many other estimation methods that should be chosen when some of the assumption are violated, however their exact mathematical description is beyond the scope of this thesis. Accordingly, they are already implemented within most of SEM software. For instance, Mplus program (Muthén & Muthén, 2005) contains following 9 parameter estimation methods (in addition to ML and GLS): ULS – unweighted least squares; RML – robust maximum likelihood estimator and other three modifications of RML – MLM, MLMV, and MLF; WLS – weighted least square estimator that use a full weight matrix and two modifications – WLSM, and WLSMV; MUML – Muthén's maximum likelihood estimator that is used in multilevel SEM. The MLR and WLS method are discussed in more detail in the section of CFA for ordinal variables section. The MUML estimator will be discussed in the part aimed at multilevel CFA.

(d) Model evaluation

Evaluating of model fit includes non-technical and technical considerations. The parameters examination, residual inspection, and *fit indexes* are discussed in this section. The core of the non-technical analysis should be an examination of the estimated coefficients of hypothesized relationships and should indicate whether the hypothesized model was a good fit to the observed data (Schreiber et al., 2006). Namely, the estimation method can produce some unexpected or even impossible values of the coefficients like negative variance estimates and/or correlation higher than 1. In other words, the estimation method produces a

close fit at the price of some computational error. Such a problem can be solved by fixing the parameter and reestimate the model. For example, negative estimated variances are usually fixed to zero (Kline, 2005). Furthermore, it makes sense to examine the statistical significance of the structural coefficients in some cases. Hence, standard errors of the estimated parameters and/or *t* values should be inspected (Ullman, 2006).

An examination of the residuals should be also conducted as another indicator of model fit. Some authors even prefer the inspection of residuals beyond fit indexes (McDonald, 1999; McDonald & Ho, 2002; Blahuš, 2008). In written reports, authors should always inform readers about residual correlation matrix to see whether residuals are close to zero (Boomsma, 2000). The term "residual" has nothing to do with observational (individual) residual that is usually investigated within the regression analysis as (individual) regression residuals. An interested reader is referred to Bollen and Arminger (1991) to see more details about observational (individual) residuals in SEM and CFA. For our purpose, the term residual is conceived as residual correlation between variables y_1 and y_2 that remains unexplained by the common latent factor *f*. The main role is played by the principle of partial correlation $r_{y1y2\,|\,f}$ that can be expressed by following formula:

$$r_{y1y2/f} = \frac{r_{y1y2} - r_{y1f} r_{y2f}}{\sqrt{(1 - r_{y1f}^2)(1 - r_{y2f}^2)}} \tag{8}$$

Nevertheless, we don't know the values of *f* and, thus, we can not simply calculate the correlations between *f* and y_1 and/or y_2. We would like to find such variable(s) *f* so that their partialing out would bring the partial correlation of the original variables to zero (Blahuš, 2008). This phenomenon is a special case of the well-known *principle of local independence*. In practice, the correlations (or generally regression coefficients) are found and generated via SEM model as the standardized elements of the $\boldsymbol{\Lambda}_y$ matrix (factor loadings). Subsequently, the difference matrix between $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{S}$ is the residual covariance matrix. All residual correlations lower than 0.1 seem to be a general consensus of an acceptable model fit (McDonald, 1999; Blahuš, 2008; Ullman, 2006).

An alternative way, how to evaluate model fit, is to use so-called *fit indices*. In reference to model fit, investigators use numerous goodness-of-fit indicators to assess a model. The use of this technical analysis of model evaluation became popular mainly in recent years, however, they mustn't be used blindly and separately without those inspections mentioned above. A large variety of fit indices is on disposal to evaluate a model under certain various

conditions and assumptions (Kline, 2005); however, we will limit ourselves to a few of them. Namely, the following indicators/indices of model fit are used in the present thesis: Chi-square statistic (χ^2), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Parsimony Comparative Fit Index (PCFI), and Akaike Information Criterion (AIC).

- χ^2 test statistic together with *df* and its *p*-value should always be included in reports of model fit (Boomsma, 2000). This statistic is approximated as follows:

$$\chi^2 = (N-1)F_{ML}, \tag{9}$$

where $(N-1)$ are the overall degrees of freedom in the sample and F_{ML} has been defined in equation (7). An analogous chi-square statistic can be also defined for the F_{GLS} from equation (6). If the distribution is nonnormal, the Satorra-Bentler chi-square correction (Satorra & Bentler, 1994) is usually used. Basically, χ^2 is a badness-of-fit index, because the higher its value, the worse correspondence of the model to the data. Null hypothesis is tested that the model has perfect fit in the population. For example, if the *p*-value equals 0.1, the researcher would not reject the null hypotheses at the 0.05 level. However, this statistic is sensitive to high sample size and/or to complexity of the model and, therefore, also some other indices should be simultaneously reported (Kline, 2005).

- The RMSEA index (Steiger, 1990) approximates a noncentral chi-square distribution, which doesn't require a true null hypothesis. This means that fit of the model in the population is not assumed to be perfect. Let us first define the estimation of so-called noncentrality parameter $\hat{\delta}$:

$$\hat{\delta} = \chi^2 - df, \tag{10}$$

where $\hat{\delta}$ is set to zero, if negative. Consequently, the RMSEA is also a badness-of-fit index, which is estimated by (Kline, 2005):

$$\text{estimated RMSEA} = \sqrt{\frac{\hat{\delta}}{df(N-1)}}, \tag{11}$$

where both *df* and $\hat{\delta}$ are linked with the estimated model. A rule of thumb is that RMSEA ≤ 0.05 indicates close approximate fit, the values between 0.05 and 0.08 suggest still acceptable fit, values between 0.08 and 0.1 indicate mediocre fit, and RMSEA ≥ 0.1

suggests poor fit (Brown & Cudeck, 1993). However, there is no empirical basis for evaluating these conventions (McDonald & Ho, 2002).

- The SRMR index is related to standardized covariance residuals. As we pointed out, residuals are differences between observed and predicted covariance. This index is based on transforming both the sample covariance matrix $\boldsymbol{S}$ and predicted covariance matrix $\hat{\boldsymbol{\Sigma}}$ into correlation matrices $\boldsymbol{R}$ and $\hat{\boldsymbol{R}}$, respectively. The differences between these two matrices are standardized residuals. Perfect model fit is indicated by SRMR = 0, however, values of SRMR less than 0.1 are generally considered favorable (Kline, 2005).

- The CFI index (Bentler, 1990; McDonald & Marsh, 1990) assesses the relative improvement of researcher's tested model compared with the independence model. This independence (also called null and/or baseline) model assumes zero population covariances among the observed variables. As a result, the independence model's chi-square (χ_B^2) is usually quite large compared with the researcher's model chi-square (χ_M^2). The larger difference between these two chi-squares, the higher improvement of the tested model over the independence model. Too, the CFI doesn't assume zero error approximation and the formula is given by:

$$\text{CFI} = 1 - \frac{\hat{\delta}_M}{\hat{\delta}_B} = 1 - \frac{\chi_M^2 - df_M}{\chi_B^2 - df_B}, \tag{12}$$

where $\hat{\delta}_M$ and $\hat{\delta}_B$ are the estimates of the noncentrality parameters for the researcher's model and for the independence model, respectively. A rule of thumb for the CFI is that values greater than 0.90 good fit of the model (Hu & Bentler, 1999). On the other hand, the value of 1 doesn't indicate perfect fit, but only that $(\chi_M^2) < (df_M)$. Finally, the parsimony-adjusted CFI index, which is sensitive to model size, is defined as:

$$\text{PCFI} = \frac{df_M}{df_B}\text{CFI}. \tag{13}$$

- The AIC index (Akaike, 1985) is a predictive fit index that assesses model fit in hypothetical replication samples of the same sample size and randomly drawn from the same population as the researcher's original sample (Kline, 2005). This parsimony-adjusted index is based on an information theory approach to data analysis. The formula for the AIC index is (Kline, 2005):

$$\text{AIC} = \chi^2_M - 2df_M, \qquad (14)$$

where all terms in formula (14) have been already defined in previous paragraph. This index is used in SEM to select among several alternative models estimated with the same data. When comparing nonnested models, the AIC index is a good choice (Schreiber et al., 2006). Consequently, the model with smallest AIC is chosen, because this model has relatively better fit and fewer parameters compared to other competing models. On the other hand, the AIC index was criticized by some authors. McDonald and Marsh (1990) showed that Akaike's principle of selecting a model to minimize the AIC cannot be used in practice in structural modeling, because the procedure necessarily selects models of increasing complexity as sample size increases until the saturated model is chosen.

(e) Model modification

There are at least two main reasons for modifying a SEM model: (a) *to test different hypotheses* (modification guided by theory), and (b) *to improve model fit* (data-driven modification). In the present thesis, we will start from simple models with small number of free parameters and get on with more complex models with higher number of parameters. Our main objective is to decide whether the improvement of fit is substantial with regard to difference between model sizes. The principle of parsimony is applied here. This means that the simplest model with still acceptable fit should be chosen. In this thesis, several models will be fitted in single-level case. They are in accordance with items' features presented in Table 6. The models will be fitted in the following order (from the simplest to the most complex):

1. *Model 1a* is a single-factor model. It is assumed here, that there is only overall ATPE that explains covariances among all 30 items.
2. *Model 1b* is a three-factor model. This model includes *model 1a* and two wording factors (positive and negative). These two factors are assumed to be uncorrelated with overall ATPE factor. If the improvement of fit will be negligible, these two factors will not be included in following models.
3. *Model 2* is a four-factor model. These factors represent dimensions of ATPE (social, health and fitness, emotional, and general). These factors are assumed to be correlated.

4. *Model 3* is actually depicted in Figure 14. It is a five-factor model with second-order overall ATPE factor that explains correlations among four attitude dimensions.
5. *Model 4* is a five-factor model. Basically, a school-related factor (see Table 6) is added to the *model 2*, but it is assumed to be uncorrelated with other four factors.
6. Model 5 is a six-factor model. It is an analogous modification of *model 3*, where the school-related factor is presumably uncorrelated (orthogonal) with other five remaining factors.

As we noted above, one way, how to find the best parsimonious models is to compare their AIC indices. Other basic methods of model modification are the chi-square difference, Lagrange multiplier, and Wald test (Ullman, 2006). All these tests compared nested models. Broadly speaking, while the LM tests which free parameter should be added, the Wald test asks which could be deleted. The software programs (LISREL, for example) allow for calculation of modification indices that indicate, which parameter should be added to improve fit. However, any modification completed should make theoretical sense. SEM is not a world championship in fitting the best model regardless of whether resulting model is theoretically reasonable. A nice commentary to modification indices has been provided by Ullman (2001, p. 750): "...adding post-hoc paths is like eating salted peanuts: One is never enough."

Multilevel confirmatory factor analysis

Multilevel linear models (MLMs) and structural equation models (SEMs) constitute two of the dominant analytic frameworks in contemporary social and behavioral sciences (e.g., Bauer, 2003). Multilevel analysis was originally linked with regression analysis (e.g., Bryk & Raudenbush, 1992), but currently is used together with every statistical method that deal with clustered observations (Luke, 2004). Several investigators have worked to synthesize CFA and MLM (e.g., Goldstein & McDonald, 1988; McDonald & Goldstein, 1989; Muthén, 1994; Muthén & Satorra, 1995). There are two contradictory views of the question whether multilevel analysis extends CFA or otherwise. Multilevel analysis had been used in research studies more recent than CFA (e.g., Hox, 2002), but, for example, Kaplan (2000) argues that multilevel analysis was initially developed as a univariate statistical technique and it was extended to a multivariate case only several years later. In fact, multilevel (regression) analysis (Zhu, 1997) and multilevel confirmatory factor analysis (MCFA) (Li, et al., 1997) were introduced to kinanthropological research approximately simultaneously. The first

application of the two-level factor (and path) analysis in kinanthropology (Li, et al., 1997) was focused on intrinsic motivation toward physical activity of the university students nested within different physical activity classes.

The MCFA model assumes that sampling occurs at two levels: between groups and within groups. Total population covariance matrix (Σ_T)contains both the between-group covariance matrix$(\Sigma_B)$and the within-group covariance matrix(Σ_W). The process of MCFA involves five successive steps (Muthén, 1994; Farmer, 2000):

1. Conventional single-level CFA of the total structure using total sample covariance matrix
2. Estimation of the between-group variation and calculation of the intraclass correlation coefficients (ICC)
3. Estimation of the within-group structure
4. Estimation of the between-group structure
5. Estimation of the MCFA model using both the between-group and within-group covariance matrices

(1) Conventional single-level CFA

At the first step, the researcher should conduct a conventional single-level CFA that is based on analysis of the total sample covariance matrix $\boldsymbol{S}_T$. Obviously, $\boldsymbol{S}_T = \boldsymbol{S}$ that was defined in previous section, but we would like to keep the conventional notation in multilevel case (Muthén, 1994). Although this solution will likely be biased if the data exhibit multilevel structure (Hox & Maas, 2001), it can provide a rough estimation of model fit. For example, we may spot obvious misspecification such as presence of correlated errors (Heck & Thomas, 2000). The process of single-level SEM and CFA has been already exhaustively described in the previous section. Hence this first step has to follow al the rules and suggestions described above.

(2) Estimation of the between-group variation

As a second step, Muthén (1994) suggested estimating the between-group variation for the set of observed variables. This is done to check whether $\Sigma_B = 0$. The degree of between-

group variation in a single variable of interest is expressed by the intraclass correlation coefficient (ICC). Let us decompose the individual's observed vector $\boldsymbol{y}_T$ into a between-class component $\boldsymbol{y}_B$ and a within-class component $\boldsymbol{y}_W$:

$$\boldsymbol{y}_T = \boldsymbol{y}_B + \boldsymbol{y}_W, \tag{15}$$

where $\boldsymbol{y}_B$ is a vector of the group means, while $\boldsymbol{y}_W$ is a vector of individual deviations from the group means. Since it is assumed here that $\boldsymbol{y}_B$ and $\boldsymbol{y}_W$ are uncorrelated, the total variance of each single item $\left(\sigma_T^2\right)$ is decomposed into between-class component $\left(\sigma_B^2\right)$ and within-class component $\left(\sigma_W^2\right)$:

$$\sigma_T^2 = \sigma_B^2 + \sigma_W^2. \tag{16}$$

Subsequently, the ICC can be expressed as the ratio of between-group variance to the total variance:

$$ICC = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}. \tag{17}$$

The ICC indicates, which portion of variance of an observed variable is ascribable to the between-class level. If the ICCs are almost zero (say, less than 0.05), a multilevel approach may be unnecessary. Moreover, it can be proven that the ICC is a special case of so-called generalizability coefficient (defined later), where groups are objects of measurement and individuals are considered as a nested facet of measurement. (Brennan, 2001).

(3) Estimation of the within-group structure

This step involves analysis of a covariance matrix that is observed at the individual level. As we noted above and according to the decomposition in equation (15), we obtain three different population covariance matrices: the total population covariance matrix $\boldsymbol{\Sigma}_T$, which can be now written as a sum of two additive and orthogonal matrices, namely $\boldsymbol{\Sigma}_B$ (for the between part) and $\boldsymbol{\Sigma}_W$ (for the within part):

$$\boldsymbol{\Sigma}_T = \boldsymbol{\Sigma}_B + \boldsymbol{\Sigma}_W. \tag{18}$$

For the within-group population covariance matrix hold the same measurement model (CFA) as for the total covariance matrix defined in equation (3):

$$\Sigma_W = \Lambda_W \Psi_W \Lambda_W' + \Theta_{\varepsilon W}, \tag{19}$$

where Λ_W, Ψ_W, and $\Theta_{\varepsilon W}$ are the population within-group matrices of factor loadings, covariances between factors, and covariances between measurement errors, respectively. However, the population covariance matrices defined in (18) have to be estimated from a given sample. Hox (2002) distinguished two main different approaches to these estimations in MCFA. One of them (introduced later) was proposed by Goldstein (1995); it is based on direct estimation of these matrices via so-called multivariate multilevel model (Bryk & Raudenbush, 1992). Another approach has been introduced by Muthén (1994); it is based on the simultaneous analysis of between and within sample covariance matrices.

Let us consider, for simplicity, a special case of balanced groups, meaning that all M groups (e.g. classes) are of the same size m. It was shown (McDonald & Goldstein, 1989; Muthén, 1994) that a consistent and unbiased estimate of the population within-group covariance matrix Σ_W is the pooled within-group sample covariance matrix S_{PW}, given by:

$$S_{PW} = \frac{\sum_{j=1}^{M} \sum_{i=1}^{m} (Y_{ij} - \bar{Y}_j)(Y_{ij} - \bar{Y}_j)'}{N - M}, \tag{20}$$

where N is the total number of individuals (students), M is the total number of PE classes and m is the common class size (number of students within each PE class equals to m). Strictly speaking, the equation (20) holds for each element of the S_{PW} matrix. It consists of products of the individual deviations from the group means between all pairs of the observed variables in parentheses in the numerator. The subscripts i, and j, are used for students (level-one units), and PE classes (level-two units), respectively. We have N-M in the denominator instead of usual N-1 (that is because we have in each group m-1 and simultaneously $N = mM$ and thus $M(m\text{-}1)$ equals to N-M).

The analysis of the within-group structure is similar to the analysis of the total single-level structure (step 1). Furthermore, all six unchanged models described in previous section on single-level CFA will be analyzed in the same sequence also in this within-group case. Because the S_{PW} matrix is not distorted by the between-level variability, the fit of the models is expected to be better than that found when the total covariance matrix S_T was used (Muthén, 1994). At this point, many investigators could stop the analysis, because they are usually interested only in the individual-level variation.

(4) Estimation of the between-group structure

This step involves analysis of a covariance matrix that is observed at the group (higher) level. The measurement model for the population between-group covariance matrix $\boldsymbol{\Sigma}_B$ defined in equation (18) can be written as:

$$\boldsymbol{\Sigma}_B = \boldsymbol{\Lambda}_B \boldsymbol{\Psi}_B \boldsymbol{\Lambda}_B' + \boldsymbol{\Theta}_{\varepsilon B}, \tag{21}$$

where all matrices in (21) are defined analogously as the within-group counterparts defined in equation (19). Subsequently, by substituting (21) and (19) to (18), we obtain:

$$\boldsymbol{\Sigma}_T = \boldsymbol{\Lambda}_B \boldsymbol{\Psi}_B \boldsymbol{\Lambda}_B' + \boldsymbol{\Theta}_{\varepsilon B} + \boldsymbol{\Lambda}_W \boldsymbol{\Psi}_W \boldsymbol{\Lambda}_W' + \boldsymbol{\Theta}_{\varepsilon W}. \tag{22}$$

The analysis of the between-class structure is a more difficult part of multilevel analysis. As was pointed out in introduction, the between-group components have a different meaning than the within-group components. Generally, the between-group factor structure differs from the within-group factor structure (Muthén & Satorra, 1995). Namely, the $\boldsymbol{\Psi}_W$ and $\boldsymbol{\Psi}_B$ matrices may differ in order and/or the patterns of the $\boldsymbol{\Lambda}_W$ and $\boldsymbol{\Lambda}_B$ can be different (compare, for example, paths from Figures 4 and 5).

Also in this case, the population between-group covariance matrix $\boldsymbol{\Sigma}_B$ must be estimated from a given sample. Analogous to the estimation of the $\boldsymbol{S}_{PW}$ matrix, the so-called scaled between-group sample covariance matrix $\boldsymbol{S}^*_B$ for the disaggregated PE class means (see equation (15)) can be calculated as follows:

$$\boldsymbol{S}^*_B = \frac{\sum_{j=1}^{M} m(\bar{Y}_j - \bar{Y})(\bar{Y}_j - \bar{Y})'}{M-1}. \tag{23}$$

Unfortunately, the scaled between-class covariance matrix $\boldsymbol{S}^*_B$ **is not** an unbiased estimator of the population between-group covariance matrix $\boldsymbol{\Sigma}_B$. As was proven by Muthén (1990), the $\boldsymbol{S}^*_B$ matrix is the maximum likelihood estimator of the composite $\boldsymbol{\Sigma}_W + c\boldsymbol{\Sigma}_B$, where c equals to the common group size m (in the balanced case). In case of unbalanced groups (PE classes), Muthén (1994) suggested computing the scaling number c as a combination of the observed group (PE class) sizes:

$$c = \frac{N^2 - \sum_{j=1}^{M} m^2{}_j}{N(M-1)}. \qquad (24)$$

This approach was called as "pseudobalanced solution" (McDonald, 1994). Hence, this model is also known as the Muthén's pseudobalanced approach (Hox, 2002). The analysis becomes more complicated without this pseudobalanced approach. We must specify a separate between-group model for each distinct group size. This leads to large models, possibly with groups of sample size less than the number of elements in the corresponding covariance matrix, what makes the maximum likelihood (ML) estimation problematic.

As we noted above, an alternative approach how to estimate the Σ_B and Σ_W matrices was proposed by Goldstein (1995). This approach is based on the multivariate multilevel regression model MMRM (Bryk & Raudenbush, 1992). The MMRM is a special case of a three-level multilevel regression model. It is assumed here that two or more dependent variables are included in the model. The hierarchical structure consists of dependent variables – inventory items (level-one units) nested within individuals (level-two units) further nested within groups (level-three units). In our case, the three-level hierarchy is characterized as follows:

Level three – 87 PE classes

Level two –1157 students

Level one – 30 inventory items

Let us assume that we have generally p dependent variables, N individuals and M groups. The multivariate multilevel model at the lowest level (items) is defined by:

$$Y_{hij} = \pi_{1ij} d_{1ij} + \pi_{2ij} d_{2ij} + ... + \pi_{pij} d_{pij}, \qquad (25)$$

where h = 1,..., p; i = 1,..., N ; j = 1,..., M. Y_{hij} is the value of dependent variable h of individual i in group j. The independent dummy variables d_1 to d_p are defined by:

$$\begin{aligned} d_{sij} &= 1 \text{ if } s = h \\ d_{sij} &= 0 \text{ if } s \neq h \end{aligned}, \qquad (26)$$

where s = 1,..., p. From (25) and (26) we may conclude that all dependent variables are uniformly ordered within individuals and the number of dummy variables d_s equals to the number of dependent variables p. The purpose of formula (25) is to specify a multivariate

model using software that is developed only for univariate analyses (Hox, 2002). There are no error terms and intercepts in equation (25). At the individual level (level 2) we have:

$$\pi_{sij} = \beta_{sj} + u_{sij}, \tag{27}$$

where β_{sj} is the mean of the observed variable s (or h) in group j and u_{sij} is the individual deviation of individual i from this mean β_{sj}. The group mean β_{sj} can be further written as a sum of overall mean and group deviation from this overall mean. Hence, at the group level (level 3) we have:

$$\beta_{sj} = \gamma_s + u_{sj}, \tag{28}$$

where γ_s is overall mean of the observed variable s (or h) and u_{sj} is the group deviation from this overall mean. By substitution of (28) and (27) to (25), we obtain the MMRM expressed by:

$$Y_{hij} = \sum_{h=1}^{p} \gamma_h d_{hij} + \sum_{h=1}^{p} u_{hij} d_{hij} + \sum_{h=1}^{p} u_{hj} d_{hij}. \tag{29}$$

The main attention must be paid on the last two terms (random part) of equation (29). Standard multilevel software programs (HLM, for example) estimate variances and covariances of the random part of the model at both individual and group levels. The variances and covariances are the elements of the within-group covariance matrix $\hat{\Sigma}_W$ at level 2 and between-group covariance matrix $\hat{\Sigma}_B$ at level 3, respectively:

$$\begin{aligned} \hat{\Sigma}_W &= \left[\mathrm{cov}\left(u_{hij}\right)\right] \\ \hat{\Sigma}_B &= \left[\mathrm{cov}\left(u_{hj}\right)\right] \end{aligned}. \tag{30}$$

The $\hat{\Sigma}_W$ and $\hat{\Sigma}_B$ matrices are direct estimators of the population matrices Σ_B and Σ_W, respectively. Consequently, they can be inserted to the standard SEM programs as an input data. However, these matrices are not computed, but estimated. This means that the factor models are estimated from estimates and, thus, the error is cumulated here (Hox, 2002). On the other hand, the MMRM generalizes CFA to more than two levels (Goldstein & Browne, 2005), because it produces one covariance matrix for every additional level.

In practical application, there are different strategies for identifying the between-group structure. It is possible to use confirmatory and/or exploratory approach to the between-group

analysis. For example, Kuhlemeier et al. (2002) conducted an exploratory factor analysis of the between-group structure in educational research, whereas Zimprich (2005), for example, used confirmatory approach in revalidization of the self-esteem scale. In the present thesis, three models will be estimated at the between-group level:

1. *Model 1* is a single-factor model. It is assumed here, that students from the same PE class have **on average** positive and/or negative overall ATPE. This hypothesis means that group's position on the seven-point scale is assumed to be similar across all 30 items (with regard to item's orientation).
2. *Model* 2 is a two-factor model, where one factor expresses average attitudes toward school related PE (see Table 6), and second factor is linked with all remaining items. These two factors are allowed to be correlated.
3. *Model* 3 is also a two-factor model. It is a combination of the first two models. All averaged items are presumably indicators of the overall ATPE, but in addition to this overall factor, the occurrence of the school related factor (see Model 2) is expected. These two factors are assumed to be uncorrelated.

(5) Simultaneous analysis of both Σ_B and Σ_W matrices

One way, how to analyze both between and within structures simultaneously, is to consider them as two different groups. At this point, we have to distinguish between multi-group and multi-level analysis. In this case, a two-level CFA becomes a two-group CFA, even though 87 groups are included as the level 2 units. First group constitutes students and the $\hat{\Sigma}_W$ matrix is analyzed here. The sample size of the first group is $(N - M) = (1157 - 87) = 1070$. The PE classes are considered as the second group with sample size $(M - 1) = 86$ and covariance matrix $\hat{\Sigma}_B$. Heck and Thomas (2000) proposed the specification of this model written in syntax of the LISREL program. We will follow their instructions in this thesis. The input matrices will be estimated in PRELIS program (module within LISREL package) via MMRM mentioned above.

Another way, how to analyze both levels simultaneously, is to use the Muthén's approach (Muthén, 1994; Muthén & Satorra, 1995). This estimation procedure is referred to as Muthén's maximum likelihood (MUML). For balanced data, MUML is full information maximum likelihood (FIML). Simulation studies and comparisons of FIML and MUML

estimates suggested that Muthén's approach provides reasonable approximations of FIML in many applications (Hox & Maas, 2001). The MUML fitting function is defined as follows:

$$F_{MUML} = M\left(\ln|\Sigma_W + c\Sigma_B| + tr\frac{S_B}{\Sigma_W + c\Sigma_B} - \ln|S_B| - p \right) + \\ + (N-M)\left(\ln|\Sigma_W| + tr\frac{S_{PW}}{\Sigma_W} - \ln|S_{PW}| - p \right). \tag{31}$$

Here, N is number of level-one units; M is number of level-two units, and p is number of variables. All matrices were defined in previous sections. An illustration of two-level schema is graphically depicted as a path-diagram in Figure 15.

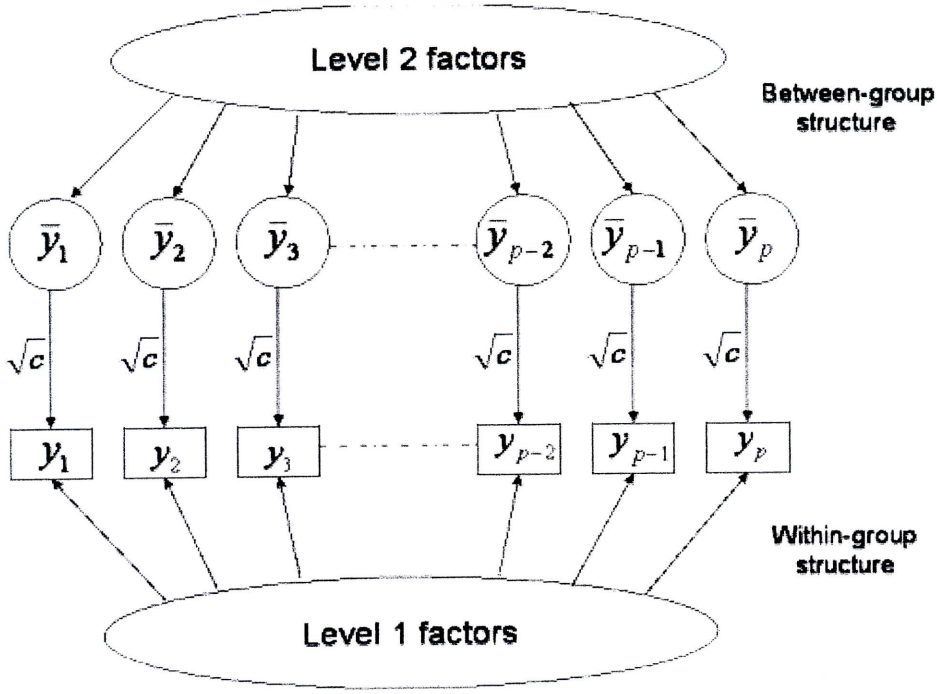


Figure 15. Two-level factor model: illustration of relationship between levels

In figure 15, the factor loadings of the group-level variables are fixed to √c, which is the square root of the scale number from (24). This is done, because the group-level variables in the scaled between-class covariance matrix S^*_B need to be transformed to their proper scale (Muthén, 1994).

At this point, let us return to sample size consideration. Hox and Maas (2001) generated multilevel data to explore the accuracy of Muthén's methodology with pseudobalanced groups and small samples at both levels. They concluded that the within-group part works

well even with small samples at both levels. They also conclude that problems can arise when estimating the between-group part of the model when the number of groups is small (i.e., $M \leq$ 30). Ideally, number of students within each PE class should be ≥ 5 and number of PE classes should be ≥ 50. Typically, researchers have sufficient sample size at the within-group level. However, sometimes could happen that so many level 2 units don't exist. Fortunately, this is not our case. In this thesis, the sample size of PE classes is 87, which meets mentioned requirement.

Multilevel confirmatory factor analysis for ordinal variables

In this section, we present a framework for conducting CFA with ordinal observed variables. We begin first by introducing basic assumptions and rules of this methodology in a single-level case. Subsequently, a two-level approach to CFA for ordinal variables will be introduced.

(1) Single-level CFA for ordinal variables

Most empirical researchers are comfortable conducting commonly used SEM and especially CFA even thought their observed variables are measured on an ordinal scale. Ordinal variables have no origins and units of measurement. As we noted above, however, standard SEM is based on assumption that all observed variables are measured on an interval scale. Thus, consider ordinal variables as being interval continuous variables may not always be suitable. Likert scales are very commonly used with interval procedures, provided the scale item has at least 5 and preferably 7 categories like in our case.

Therefore, the most fundamental question arises: What would happen, if ordinal item indicators are treated as interval continuous? Several problems could occur here, namely:

- Product moment correlations applied to ordinal variables are attenuated (Babakus, Ferguson & Jöreskog, 1987).
- Attenuation increases as the number of response categories decreases (Muthén, 1984).
- Item loadings based on ordinal indicators are also attenuated (Olsson, 1979).
- Likelihood ratio, chi-square and standard errors are incorrect (Muthén & Kaplan, 1985).

For these reasons, one of the most recent guidline of how to conduct SEM for ordinal variables in LISREL was proposed by Jöreskog (2005). The ordinal variables $\boldsymbol{y}$ are taken to be manifestations of some underlying continuous observed variables $\boldsymbol{y}^*$. The connection between single ordinal variable y_i and underlying single variable y_i^* is in our 7-point scale case given by:

$$y_i = s \Leftrightarrow \tau_{s-1}^{(i)} < y_i^* \leq \tau_s^{(i)}, s = 1, 2, ..., 7, \tag{32}$$

where

$$\tau_0^{(i)} = -\infty, \tau_1^{(i)} < \tau_2^{(i)} < \tau_3^{(i)} < \tau_4^{(i)} < \tau_5^{(i)} < \tau_6^{(i)}, \tau_7^{(i)} = +\infty \tag{33}$$

are parameters called thresholds, which are estimated from the univariate marginal distributions. The measurement model is the classical factor analysis model like in equation (1) but now it can be written as (Moustaki, Jöreskog & Mavridis, 2004):

$$\boldsymbol{y}^* = \boldsymbol{\Lambda z} + \boldsymbol{u}, \tag{34}$$

where $\boldsymbol{z}$ is a vector of latent variables and $\boldsymbol{u}$ is a vector of measurement errors. The main difference between equations (1) and (34) is that the left side of equation (1) is directly observed, whereas the left side of (34) is unobserved.

Consequently, the polychoric correlations, ρ, are estimated from the bivariate marginal distributions for given thresholds by minimizing the discrepancies between sample proportions and model-expected proportions. The polychoric correlations are based on assumption of bivariate normality. It may seem that this assumption is too strong, but polychoric correlation is very robust to violations of underlying bivariate normality (Jöreskog, 2005).

The polychoric correlations and their asymptotic covariance matrix (ACM) are then inserted into LISREL program to estimate factor models with weighted least squares (WLS) or robust maximum likelihood (RML) estimation methods. The ACM is usually very large because its order equals to number of non-duplicated elements in matrix of polychoric correlations. In our case, we obtain $p(p\text{-}1)/2 = (30\times29)/2 = 435$ correlations. Thus, the ACM has $(345\times346)/2 = 75\,255$ non-duplicated variances and covariances. Obviously, such a matrix cannot be presented in a paper form. It is calculated in PRELIS program and saved as an external file that is subsequently loaded into LISREL program and used as a weight matrix.

A comparison of LISREL and IRT approaches to single-level latent variable modeling with ordinal variables was performed by Moustaki, Jöreskog and Mavridis (2004). They conclude that IRT models fit the data better and don't require large samples, because the IRT model is non-linear. On the other hand, if models with many factors are fitted, they suggested using LISREL due to the computational burden in the IRT models.

(2) Multilevel CFA for ordinal variables

In one of the first application of the MCFA in physical education and exercise sciences, Li et al. (1997) made a remark to ordinal variables in MCFA. In their first footnote (Li et al., 1997, pp. 225), they stated:

> "The points on the Likert-type scale are frequently assumed to constitute an equal interval scale and scores are treated as constituting a continuous variable, although they are in fact ordered categories in nature. If one treats the scale as ordered categorical response variables, analyses using SEM may require other procedures. Therefore, the interpretation with categorical response variables bears different conceptual meaning depending on the a priori assumptions made by the researcher."

In this section, we present a brief description of that "other procedures" mentioned above. Up to date, there is unfortunately no known application of MCFA for ordinal variables in kinanthropology. This methodology has been elaborated only within theoretical studies (Grilli & Rampichini, 2007; Goldstein & Browne, 2005).

Formally, this approach is a combination of the MCFA approach and single-level CFA for ordinal variables. Let y_{hij} be the *h*th observed ordinal variable (h = 1, 2, ..., 30) for the *i*th student (i = 1, 2, ..., 1157) of the *j*th PE class (j = 1, 2, ..., 87). Also in this case, the variable y_{hij} represents underlying continuous variable y^*_{hij}. Similarly as in expression (32), we obtain:

$$y_{hij} = s_h \Leftrightarrow \tau_{s-1}^{(hij)} < y^*_{hij} \leq \tau_s^{(hij)}, s_h = 1, 2, ..., 7. \qquad (35)$$

Subsequently, the underlying continuous variable y^*_{hij} can be further decomposed as follows:

$$y^*_{hij} = \mu_h + \left(\sum_{m=1}^{M_W} \lambda_{mh}^{(W)} \eta_{mij}^{(W)} + \varepsilon_{hij}^{(W)} \right) + \left(\sum_{m=1}^{M_B} \lambda_{mh}^{(B)} \eta_{mj}^{(B)} + \varepsilon_{hj}^{(B)} \right), \qquad (36)$$

where M_B and M_W are number of common factors η at the between-level and the within-level, respectively. In equation (36), we use the superscript (W) and (B) to denote variables defined at within- and between- levels, respectively. The term μ_h in (36) is the variable mean. This mean has to be included here, because the underlying variable is estimated via the threshold model and, therefore, cannot be in centered or standardized form. All other terms have been defined in equation (1). All other procedures are similar to the MCFA for continuous variables. However, the estimation of model parameters is not so straightforward. Fitting the two-level factor model is computationally intensive. Several estimation methods have been proposed, such as ML with adaptive Gaussian quadrature (Rabe-Hesketh et al., 2005) and/or Bayesian Markov Chain Monte Carlo (MCMC) algorithms (Goldstein & Browne, 2005). The computational burden is heavy even for contemporary efficient computers. Therefore, it is important to base model selection on exploratory analyses to supply algorithms with good starting values. Grilli and Rampichini (2007) suggested following 4 phases of the analysis:

1. *Univariate two-level models*. The ICCs are estimated as a first step. These coefficients will be generally different from their continuous counterparts.
2. *Exploratory single-level factor analysis*. The polychoric correlations are used as the input data here to shed some light on the structure.
3. *Exploratory between and within factor analysis*. The results of this two-stage procedure are expected to be similar to those obtained from the full two-level analysis, as in the continuous case (Grilli & Rampichini, 2007).
4. *Confirmatory two-level factor analysis.* In the present thesis, only the best fitting models as in the continuous case are fitted in this approach for ordinal variables. The main purpose is to compare results of the MCFA for continuous and ordinal variables with respect to estimated parameters, dimensionality, and model fit.

The MCFA with ordinal variables is conducted within Mplus (Muthén & Muthén, 2005) program that allows specifying latent variable models with ordinal variables for clustered observations.

RESULTS

Present section is divided into two main parts. First part is more extensive. All item indicators are considered here to be continuous interval variables. Subsequently, all presented results of analytical procedures are based on this a priori assumption. In the second part, all observed variables are assumed to be ordinal. The results obtained from both parts are compared and further discussed in the *discussion section* of this thesis.

CONTINUOUS VARIABLES RESULTS

Descriptive statistics

In the first instance, basic univariate descriptive statistics were computed. Minimum, maximum, mean, standard deviation, skewness, and kurtosis of all 30 items are presented in Table 7. The minimum value of all items is 1, and the maximum value of all items is 7. This means that all scale points have been used[3]. The grand mean of all 30 items equals 4.79, which is slightly above 4 - the neutral point of the scales. The smallest mean value (3.66) has the item number 4, whereas the maximum mean value (5.70) has the item number 29. Standard deviations ranged from 1.24 (item 25) to 2.08 (item 8). Most of the items were negatively skewed, which means that more scores are observed above rather than below the respective item's mean. The skewness ranged from -1.46 (item 23) to 0.14 (item 10). For completeness' sake, the kurtosis of all items is reported in Table 7. The kurtosis ranged from -1.39 (item 8) to 2.71 (item 23).

In spite of slight violations of assumed univariate normality (especially items 23 and 29), one could proceed to check the bivariate properties, namely homoscedasticity and linearity, by inspecting scatterplots. Obviously, not all (30×29) = 435 scatterplots can be presented here. Furthermore, the scatterplots are not appropriate for Likert-type data. If there is no empty combination of categories between two variables, a scatterplot consists of 7 by 7 overlapping points. As an example, the scatterplot between variable 3 and 8 is depicted in Figure 16.

[3] Note that negatively oriented items were rescaled conversely (7 to 1, 6 to 2, etc.)

Table 7. Basic univariate statistical characteristics of the items

Item number	Mean	St. deviation	Skewness	Kurtosis	Min.	Max.
1	5.35	1.75	-1.05	0.16	1	7
2	4.91	1.31	-0.67	0.17	1	7
3	4.89	1.48	-0.72	-0.15	1	7
4	3.66	1.49	0.08	-0.67	1	7
5	4.32	1.78	-0.34	-0.94	1	7
6	4.94	1.71	-0.70	-0.50	1	7
7	4.17	1.61	-0.17	-0.97	1	7
8	3.98	2.08	0.01	-1.39	1	7
9	5.06	1.56	-0.80	-0.14	1	7
10	3.80	1.79	0.14	-1.11	1	7
11	5.31	1.28	-1.05	1.20	1	7
12	4.11	1.52	-0.13	-0.64	1	7
13	5.48	1.38	-1.24	1.48	1	7
14	5.53	1.62	-0.51	-0.50	1	7
15	5.31	1.40	-1.01	0.75	1	7
16	5.65	1.29	-1.30	1.87	1	7
17	5.31	1.52	-1.08	0.64	1	7
18	5.18	1.48	-0.95	0.43	1	7
19	3.71	1.44	-0.03	-0.63	1	7
20	4.97	1.47	-0.68	0.04	1	7
21	3.92	2.06	0.01	-1.33	1	7
22	4.84	1.54	-0.70	-0.16	1	7
23	5.64	1.25	-1.46	2.71	1	7
24	3.85	1.68	0.03	-1.04	1	7
25	5.09	1.24	-0.86	0.85	1	7
26	5.03	1.44	-0.86	0.32	1	7
27	4.90	1.60	-0.73	-0.18	1	7
28	4.36	1.35	-0.35	-0.25	1	7
29	5.70	1.31	-1.43	2.36	1	7
30	4.76	1.86	-0.69	-0.69	1	7

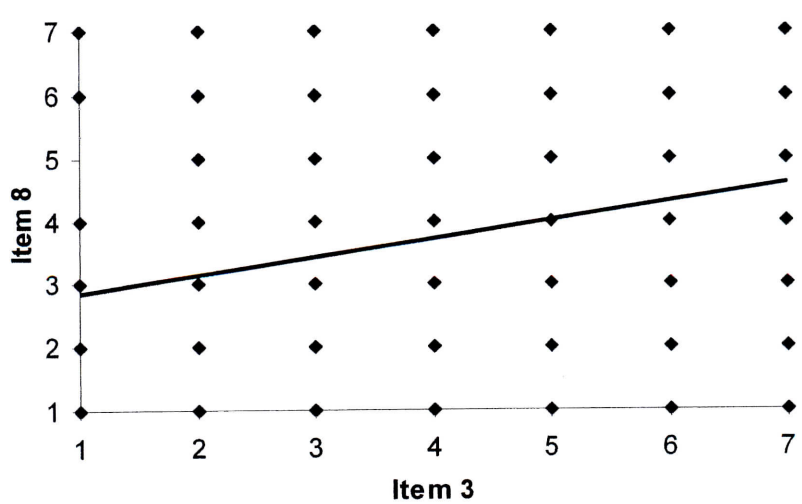


Figure 16. Scatterplot and regression line between item 3 and item 8

As one can see, it makes a little sense to check homoscedasticity and linearity by inspecting the scatterplots in this case. More appropriate approach is to present the bivariate distributions for ordinal variables as 7 by 7 frequency tables. Some of these tables are presented in the result section for ordinal variables.

Reliability analysis

To get an estimate of reliability expressed by Cronbach's alpha or McDonald's omega, we must first fit a single-factor model within each domain (sub-construct, subscale). The path diagrams are depicted in Figures 17, 18, 19, and 20. If the fit is adequate, the sub-construct is said to be unidimensional and homogenous and the estimated parameters (factor loadings and unique variances) may be used to calculate mentioned reliability-like coefficients.

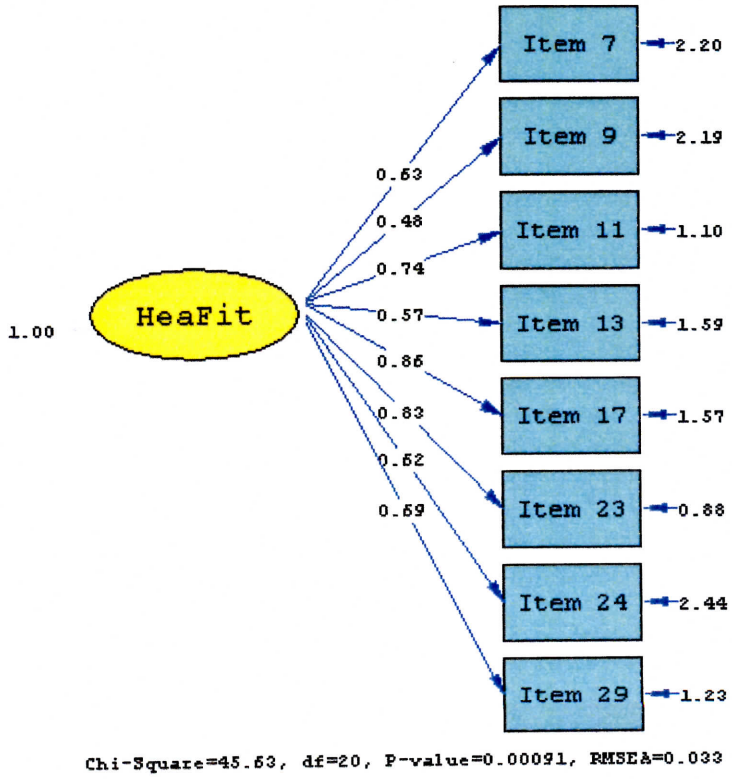


Figure 17. Path diagram of the single-factor model: Health and fitness domain

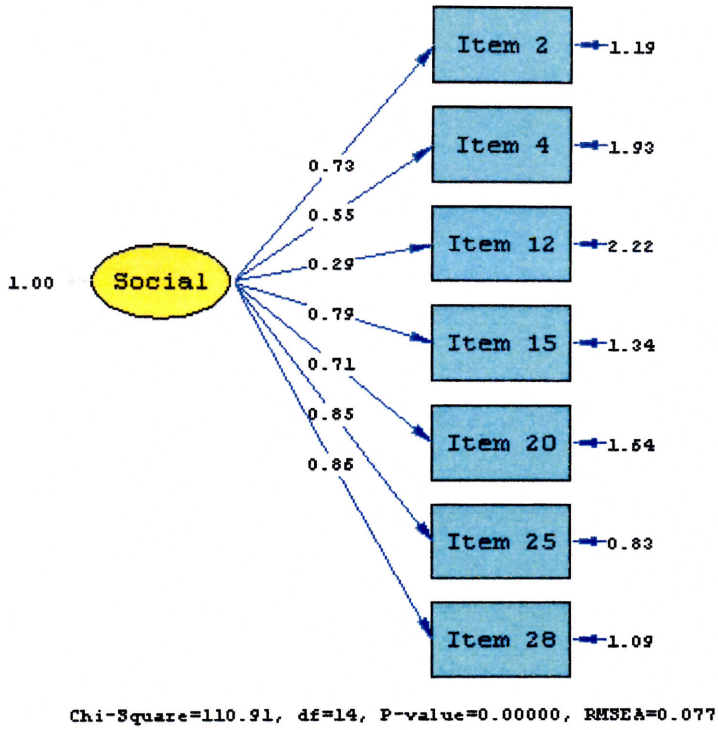


Figure 18. Path diagram of the single-factor model: Social domain

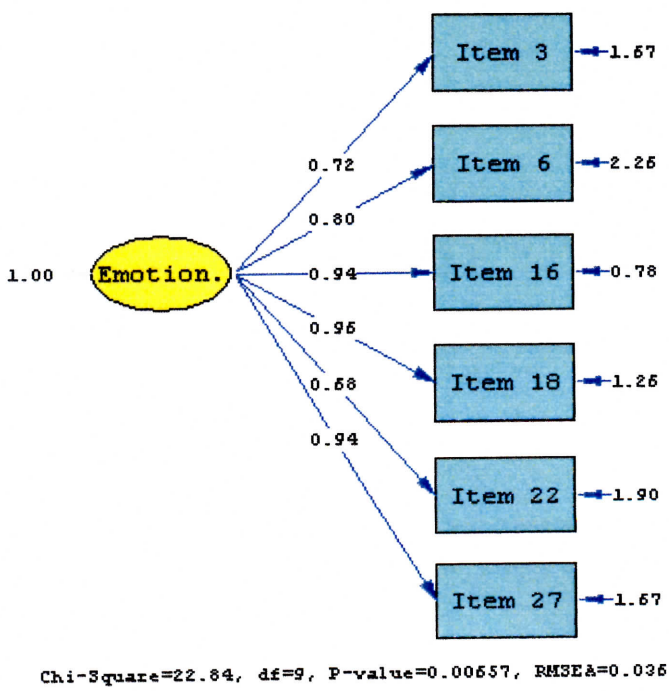


Figure 19. Path diagram of the single-factor model: Emotional domain

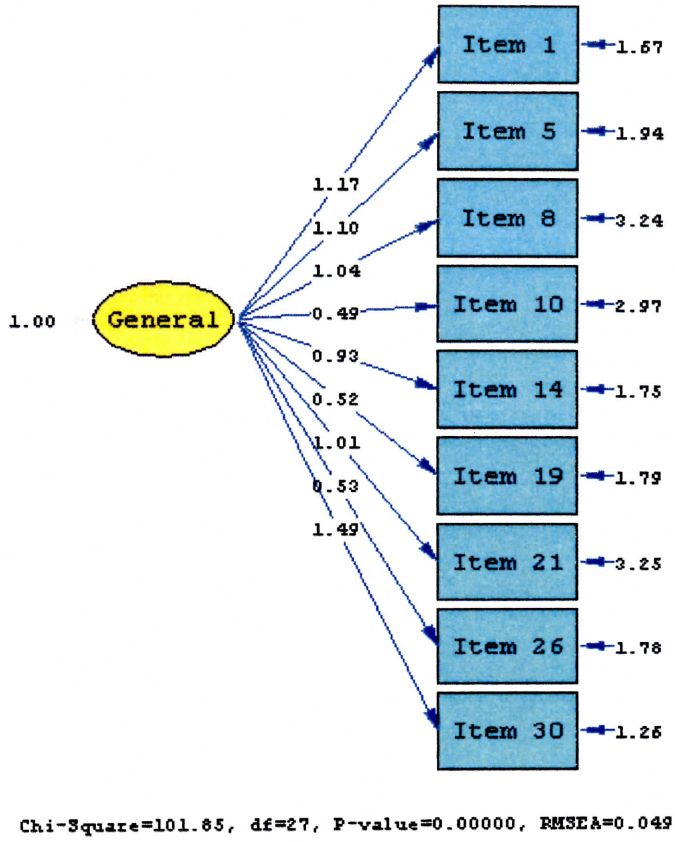


Figure 20. Path diagram of the single-factor model: General domain

All four single-factor models exhibit acceptable fit. Only social domain of ATPE (see Figure 18) fits the data marginally (RMSEA = 0.077). Hence, we have some evidence about unidimensionality of each domain of ATPE and reliability-like coefficient can be therefore calculated. Cronbach's alpha and McDonald's omega for all four attitude-related domains are presented in Table 8.

Table 8. Reliability-like coefficients of four attitude-related domains

	Health and fitness	Social	Emotional	General
Cronbach's alpha	0.69	0.68	0.72	0.77
McDonald's omega	0.69	0.69	0.72	0.78
Number of items	8	7	6	9

As we can see in Table 8, estimated reliabilities of the subscales are not very high. According to some suggestions, the reliability should be at least 0.70. By other psychometricians, the relliability should be higher than 0.80 (Furr & Bacharach, 2007). But how many items within

each subscale are needed to achieve desirable acceptable level of reliability? The answer to this question can be found in Table 9. The emotional domain of ATPE seems to be the most reliable domain. Therefore, only 6 or 10 items are needed to achieve reliability of 0.70 or 0.80, respectively. On the other hand, health and fitness is the domain with lowest reliability. If we sum all items, at least 31 items must be included in the inventory, if we settle for reliability of 0.70. If all domains have to be measured with reliability of 0.80, the inventory should consist of 51 items. The reliability and generalizability issues are discussed in more detail in Appendix.

Table 9. Number of items needed for desirable levels of reliability

	Number of items				
Reliability	Health and fitness	Social	Emotional	General	Total
0.70	9	8	6	8	31
0.80	15	14	10	14	53

Single-level CFA results

As the first step, the proposed ten models were imposed on $\boldsymbol{S}_T$. This total sample covariance matrix is presented in Table 10. The covariances are situated in the left bottom triangle, the variances are on the diagonal, and the respective correlations are in the upper right triangle of $\boldsymbol{S}_T$. All analyses were conducted in the LISREL program (Jöreskog & Sörbom, 2005) and all LISREL syntax files are provided in Appendix.

The fitting history is presented in Table 11. The inspection of all parameters estimation did not reveal any impossible values (e.g., negative variances).

As we noted above, the models are denoted with the letters *b* or *a* accordingly if they include or not two wording factors, respectively (see method section). These two factors essentially improve model fit. The indices of fit are much better for all *b*-models. Therefore, the presence of positive and negative wording factors is strongly empirically supported and these two factors should be included in all models.

The unidimensional model (model 1*a*) exhibited only marginally acceptable fit (see Table 11). Furthermore, the residual covariance matrix includes too large values (0.47 was the highest value). This makes this model unacceptable. On the other hand, the model 1*b*

improved fit substantially. Moreover, the highest residual (covariance) value decreases to 0.37. The structural part of model 1*b* is depicted in Figure 21.

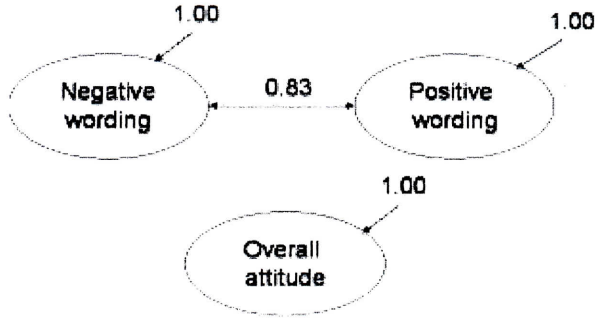


Figure 21. Structural part of model 1*b*

Table 10. Total sample covariance matrix S_T and respective correlations

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11
Item 1	**3.05**	.25	.32	.15	.43	.18	.18	.35	.10	.15	.32
Item 2	.57	**1.72**	.19	.22	.25	.14	.15	.18	.14	.13	.25
Item 3	.83	.38	**2.18**	.01	.26	.19	.14	.21	.11	.18	.23
Item 4	.38	.42	.03	**2.33**	.17	.12	.07	.06	.10	.16	.19
Item 5	1.33	.58	.69	.45	**3.15**	.22	.30	.32	.15	.17	.28
Item 6	.53	.32	.47	.30	.66	**2.91**	.21	.14	.12	.15	.23
Item 7	.52	.32	.33	.16	.86	.58	**2.60**	.17	.14	.20	.20
Item 8	1.28	.49	.64	.18	1.17	.49	.56	**4.33**	.03	.16	.20
Item 9	.27	.29	.25	.24	.41	.31	.36	.10	**2.42**	.09	.17
Item 10	.47	.31	.49	.43	.54	.45	.58	.60	.25	**3.21**	.16
Item 11	.71	.41	.44	.36	.63	.50	.41	.54	.35	.38	**1.64**
Item 12	.41	.22	.52	-.04	.56	.14	.24	.27	.17	.19	.06
Item 13	.16	.33	.10	.19	.45	.47	.42	.14	.34	.17	.44
Item 14	.99	.50	.72	.32	1.09	.61	.62	.87	.53	.52	.69
Item 15	.63	.60	.54	.33	.52	.34	.12	.55	.30	.31	.50
Item 16	.77	.50	.65	.17	.90	.79	.64	.68	.46	.48	.64
Item 17	.77	.45	.62	.08	1.01	.72	.70	.75	.32	.55	.61
Item 18	.87	.47	.68	.15	.92	.81	.62	.85	.35	.50	.52
Item 19	.50	.56	.18	.71	.43	.33	.16	.38	.34	.36	.55
Item 20	.63	.57	.59	.39	.66	.34	.36	.42	.44	.40	.49
Item 21	1.29	.31	.66	.29	1.14	.45	.47	1.14	.19	.57	.56
Item 22	.50	.23	.66	-.04	.62	.46	.36	.33	.30	.25	.26
Item 23	.42	.40	.25	.32	.65	.44	.45	.41	.41	.41	.58
Item 24	.99	.60	.65	.46	.88	.29	.43	.99	.40	.56	.57
Item 25	.56	.58	.37	.42	.63	.50	.33	.42	.34	.39	.58
Item 26	.62	.24	.46	.03	.72	.52	.61	.49	.30	.41	.42
Item 27	1.45	.72	.71	.50	1.16	.76	.53	1.07	.48	.50	.85
Item 28	.67	.63	.32	.63	.77	.65	.51	.65	.35	.56	.56
Item 29	.50	.34	.30	.25	.61	.46	.38	.49	.29	.38	.54
Item 30	1.77	.70	.89	.60	1.57	.66	.71	1.57	.38	.67	.89

Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20	Item 21	Item 22	Item 23	Item 24	Item 25	Item 26	Item 27	Item 28	Item 29	Item 30
.15	*.07*	*.35*	*.26*	*.35*	*.29*	*.34*	*.20*	*.25*	*.36*	*.19*	*.19*	*.34*	*.26*	*.25*	*.52*	*.29*	*.22*	*.54*
.11	*.18*	*.24*	*.32*	*.30*	*.23*	*.24*	*.30*	*.30*	*.12*	*.11*	*.24*	*.27*	*.36*	*.13*	*.35*	*.39*	*.20*	*.29*
.23	*.05*	*.30*	*.26*	*.34*	*.28*	*.30*	*.09*	*.27*	*.22*	*.29*	*.14*	*.26*	*.20*	*.22*	*.30*	*.16*	*.15*	*.32*
-.02	*.09*	*.13*	*.16*	*.09*	*.04*	*.07*	*.33*	*.18*	*.09*	*-.02*	*.17*	*.18*	*.23*	*.02*	*.21*	*.31*	*.13*	*.22*
.21	*.18*	*.38*	*.21*	*.40*	*.38*	*.35*	*.17*	*.25*	*.31*	*.23*	*.29*	*.29*	*.29*	*.28*	*.41*	*.32*	*.26*	*.47*
.05	*.20*	*.22*	*.14*	*.36*	*.28*	*.32*	*.13*	*.13*	*.13*	*.18*	*.21*	*.10*	*.24*	*.21*	*.28*	*.28*	*.21*	*.21*
.10	*.19*	*.24*	*.05*	*.31*	*.29*	*.26*	*.07*	*.15*	*.14*	*.25*	*.23*	*.16*	*.16*	*.26*	*.21*	*.23*	*.18*	*.24*
.08	*.05*	*.26*	*.19*	*.25*	*.24*	*.28*	*.13*	*.14*	*.27*	*.10*	*.16*	*.28*	*.16*	*.16*	*.32*	*.23*	*.18*	*.41*
.07	*.16*	*.21*	*.14*	*.23*	*.14*	*.15*	*.15*	*.19*	*.06*	*.13*	*.21*	*.15*	*.18*	*.14*	*.19*	*.17*	*.14*	*.13*
.07	*.07*	*.18*	*.12*	*.21*	*.20*	*.19*	*.14*	*.15*	*.15*	*.09*	*.18*	*.18*	*.17*	*.16*	*.17*	*.23*	*.16*	*.20*
.03	*.25*	*.33*	*.28*	*.39*	*.31*	*.27*	*.30*	*.26*	*.21*	*.13*	*.36*	*.26*	*.36*	*.23*	*.42*	*.33*	*.32*	*.37*
2.30	*.05*	*.21*	*.13*	*.12*	*.12*	*.18*	*.04*	*.21*	*.06*	*.15*	*.06*	*.18*	*.10*	*.05*	*.16*	*.09*	*.07*	*.19*
.10	**1.91**	*.13*	*.15*	.29	*.20*	*.18*	*.11*	*.14*	*.01*	*.08*	*.29*	*.09*	*.25*	*.21*	*.20*	*.23*	*.22*	*.13*
.52	.29	**2.62**	*.23*	*.31*	*.34*	*.32*	*.31*	*.28*	*.25*	*.18*	*.20*	*.46*	*.26*	*.22*	*.40*	*.32*	*.21*	*.46*
.28	.30	.53	**1.97**	*.31*	*.20*	*.23*	*.24*	*.37*	*.13*	*.12*	*.22*	*.23*	*.39*	*.15*	*.34*	*.31*	*.21*	*.29*
.23	.52	.65	.56	**1.66**	*.46*	*.47*	*.15*	*.32*	*.24*	*.33*	*.38*	*.25*	*.36*	*.39*	*.43*	*.32*	*.33*	*.42*
.28	.41	.84	.42	.89	**2.31**	*.40*	*.12*	*.22*	*.24*	*.22*	*.40*	*.21*	*.30*	*.36*	*.31*	*.28*	*.28*	*.34*
.39	.36	.77	.47	.89	.90	**2.18**	*.12*	*.30*	*.26*	*.28*	*.22*	*.21*	*.27*	*.36*	*.39*	*.25*	*..31*	*.37*
.08	.21	.72	.49	.27	.25	.24	**2.06**	*.25*	*.17*	*.02*	*.19*	*.38*	*.31*	*.03*	*.34*	*.36*	*.22*	*.33*
.47	.28	.68	.75	.61	.48	.66	.52	**2.15**	*.16*	*.26*	*.21*	*.30*	*.30*	*.24*	*.37*	*.23*	*.22*	*.35*
.19	.03	.85	.37	.64	.76	.79	.50	.49	**4.26**	*.10*	*.07*	*.19*	*.14*	*.16*	*.29*	*.14*	*.19*	*.38*
.35	.17	.44	.25	.65	.52	.64	.03	.59	.32	**2.36**	*.13*	*.15*	*.12*	*.28*	*.22*	*.11*	*.19*	*.22*
.11	.50	.41	.39	.61	.75	.41	.33	.38	.19	.26	**1.56**	*.22*	*.35*	*.21*	*.28*	*.36*	*.37*	*.23*
.47	.21	1.24	.54	.54	.54	.51	.92	.75	.68	.40	.47	**2.83**	*.24*	*.12*	*.42*	*.29*	*.19*	*.48*
.20	.42	.52	.68	.58	.57	.50	.55	.55	.36	.23	.55	.51	**1.55**	*.23*	*.41*	*.48*	*.28*	*.34*
.11	.42	.50	.31	.72	.78	.77	.05	.70	.47	.61	.38	.29	.42	**2.06**	*.27*	*.19*	*.21*	*.29*
.40	.45	1.04	.76	.88	.75	.93	.78	.87	.97	.55	.56	1.14	.81	.62	**2.56**	*.43*	*.36*	*.57*
.19	.42	.69	.58	.56	.57	.49	.70	.47	.38	.22	.61	.66	.81	.36	.92	**1.83**	*.26*	*.33*
.15	.39	.45	.39	.55	.56	.60	.41	.42	.50	.38	.60	.41	.45	.40	.76	.46	**1.71**	*.32*
.54	.34	1.40	.76	1.00	.96	1.00	.89	.95	1.44	.61	.54	1.50	.79	.77	1.70	.82	.77	**3.47**

Table 11. Fitting history of the hypothesized models based on S_T

Model number	χ^2	*df*	RMSEA	SRMR	CFI	PCFI	AIC
1a	3174	405	0.077	0.057	0.936	0.871	3294
1b	1531	374	0.052	0.043	0.968	0.832	1713
2a	2497	399	0.067	0.056	0.947	0.869	2629
2b	1311	368	0.047	0.041	0.973	0.823	1505
3a	2534	401	0.068	0.056	0.947	0.873	2662
3b	1334	370	0.047	0.041	0.973	0.828	1524
4a	1819	387	0.057	0.053	0.960	0.854	1975
4b	1005	356	0.040	0.037	0.980	0.802	1223
5a	1868	389	0.057	0.053	0.960	0.858	2021
5b	1164	358	0.044	0.039	0.976	0.803	1378

Note. χ^2 = Chi-square, *df* = degrees of freedom, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, CFI = Comparative Fit Index, PCFI = Parsimony Comparative Fit Index, AIC = Akaike Information Criterion Independence AIC = 30720, Saturated AIC = 930.

The correlation between positive and negative wording factors (0.83) is very high. This indicates that a person who responded very positively to the positive items, responded simultaneously very negatively to negative items and conversely.

The multidimensional model (model 2*a*) improved fit slightly compared to the unidimensional model 1*a* (see Table 11). Also, the *b*-version of model 2 (i.e., model 2*b*) fits the data better than the unidimensional model 1*b*. The structural part of model 2*b* is depicted in Figure 22.

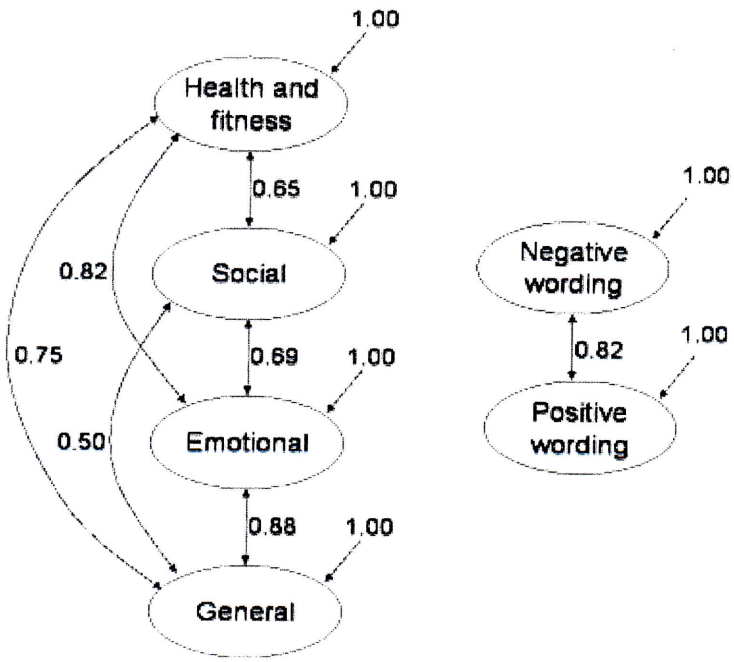


Figure 22. Structural part of model 2*b*

As we can see in Figure 22, all four attitude-related latent variables are quite highly correlated (ranged from 0.50 to 0.88). This finding can serve as an evidence for supporting the existence of the second-order factor "overall attitude" that explains these correlations among these four latent variables. Therefore, it seems reasonable to proceed to estimating model 3. Compared to model 2*a*, model 3*a* is more simple (i.e., has less free parameters to be estimated) and the fit is approximately the same (see Table 11). Similar situation is observed also in the *b*-version of model 3. Estimated parameters of the structural part of model 3*b* are presented in Figure 23. Two wording factors are still highly correlated (0.81). Probably, the most surprising results are negative estimates of the three standardized regression coefficients between second and first order factors (-0.85, -0.70, -0.98). These values are not theoretically impossible, but they are unexpected due to the high correlations among four first-order attitude related factors (see Figure 22). The standardized regression residuals ranged from 0.04 to 0.51.

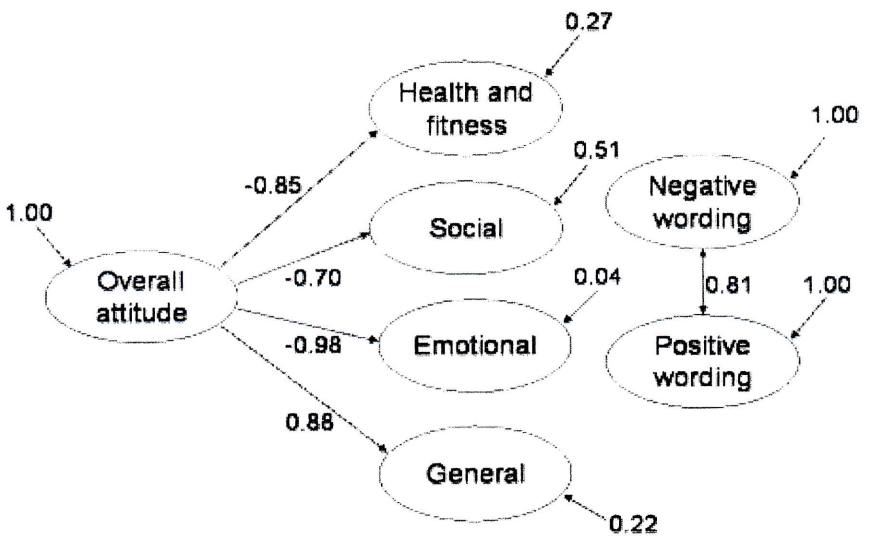


Figure 23. Structural part of model 3*b*

Adding the "school PE" factor to the model resulted in another substantial improvement of model fit. Model 4*a* exhibited better fit compared to model 2*a* (see Table 10). However, the smallest AIC value (1223) and RMSEA value (0.04) is appertained to model 4*b*. The structural part of the model 4*b* with estimated parameters is displayed in Figure 24. The estimated factor loadings within this model are presented in Table 12.

One remarkable difference between model 4*b* and model 2*b* is that correlation among four attitude latent variables become lower in model 4*b* (ranged from 0.32 to 0.82). This can be a reason why the model 5 didn't fit the data well. The second-order "overall attitude" factor did not explained the correlation among all four first-order factors. Furthermore, some negative estimates of variances occurred in model 5*b*. Therefore, we propose to choose the model 4*b* as the best fitting model. For this reason, the estimated parameters in measurement part of model 4*b* (i.e., factor loadings and uniqueness) are presented in Table 12. However, we must still keep in mind that the $\boldsymbol{S}_T$ matrix is confounded by the between-level variation. Hence, it has little sense to make a model modification to improve the model fit (e.g., by inspecting the modification indices). Instead, all models will be imposed on another covariance matrix, $\boldsymbol{S}_{PW}$, that is not contaminated by the between-level variability.

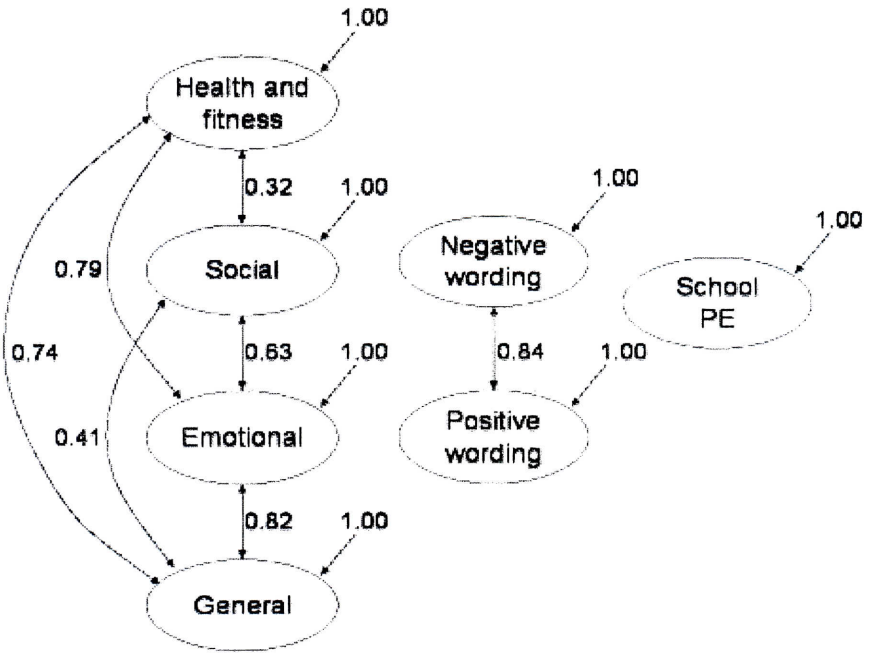


Figure 24. Structural part of model 4*b*

Table 12. Estimated standardized factor loadings and uniqueness of model 4*b* based on S_T

Item	Factor loadings							Uniqueness
	Health & fitness	Social	Emotional	General	Positive wording	Negative wording	School PE	
1	--	--	--	0.39	--	0.37	0.45	0.51
2	--	0.24	--	--	0.49	--	--	0.70
3	--	--	0.38	--	--	0.27	0.19	0.74
4	--	-0.13	--	--	0.46	--	--	0.77
5	--	--	--	0.41	--	0.50	--	0.58
6	--	--	0.34	--	0.31	--	--	0.78
7	0.32	--	--	--	--	0.33	--	0.79
8	--	--	--	0.28	--	0.29	0.29	0.75
9	0.11	--	--	--	--	0.30	--	0.89
10	--	--	--	0.12	--	0.33	--	0.88
11	0.31	--	--	--	0.50	--	0.16	0.63
12	--	0.26	--	--	0.11	--	--	0.92
13	0.26	--	--	--	0.32	--	--	0.83
14	--	--	--	0.16	--	0.55	0.25	0.60
15	--	0.40	--	--	0.43	--	--	0.66
16	--	--	0.59	--	--	0.46	--	0.44
17	0.53	--	--	--	--	0.42	--	0.54
18	--	--	0.54	--	--	0.35	0.08	0.57
19	--	--	--	-0.22	0.59	--	0.23	0.56
20	--	0.41	--	--	--	0.42	0.11	0.64
21	--	--	--	0.28	--	0.25	0.30	0.77
22	--	--	0.42	--	--	0.18	--	0.79
23	0.36	--	--	--	0.48	--	--	0.64
24	-0.01[a]	--	--	--	--	0.56	0.37	0.55
25	--	0.26	--	--	0.62	--	--	0.55
26	--	--	--	0.50	--	0.27	--	0.69
27	--	--	0.35	--	0.55	--	0.44	0.39
28	--	0.08[a]	--	--	0.68	--	--	0.53
29	0.35	--	--	--	0.40	--	--	0.72
30	--	--	--	0.32	--	0.55	0.48	0.36

Note.
[a]non-significant parameter estimates at 0.05 level

The χ^2 value was very likely inflated in all ten models due to the fact that the multilevel structure of the data was ignored.

Estimation of the between-group variation

The intraclass correlation coefficients (ICC) for all 30 items are presented in Table 13. As we noted above, they were estimated via the multivariate multilevel model in PRELIS program. The ICCs ranged from 0.02 to 0.11with mean value of 0.051.

Table 13. Intraclass Correlation coefficients

Item	ICC	Item	ICC
1	0.07	16	0.08
2	0.03	17	0.09
3	0.03	18	0.05
4	0.05	19	0.11
5	0.06	20	0.02
6	0.05	21	0.07
7	0.05	22	0.03
8	0.06	23	0.04
9	0.04	24	0.07
10	0.02	25	0.03
11	0.03	26	0.04
12	0.05	27	0.05
13	0.08	28	0.03
14	0.07	29	0.03
15	0.04	30	0.07

Such a low values of ICCs are a commonplace in educational research where students are nested within classes or schools. For example, Kaplan (2000) reported ICCs ranged from 0.01 to 0.07 in his multilevel CFA introduction. On the other hand, Myers (2005), for example, reported ICCs ranged from 0.08 to 0.22. Nevertheless, some authors suggested that an investigator may apply multilevel approach in all cases regardless of ICC values (e.g. Hox, 2002). In our case, we have also an empirical support for conducting multilevel CFA, because ICCs are not trivial.

Within-group CFA results

As a first step, the pooled within-group sample covariance matrix $\boldsymbol{S}_{PW}$ has been estimated via the MMRM in PRELIS program. This matrix is presented in Table 14. The left lower triangle involves the covariances, the variances are situated on the diagonal, and the italic right upper triangle contains respective correlations. Both $\boldsymbol{S}_{PW}$ and $\boldsymbol{S}_{T}$ matrices are very similar. The covariances, variances and also correlations are slightly lower. Therefore, we can expect better fit of all ten models.

The same ten discussed models have been fitted to this $\boldsymbol{S}_{PW}$ matrix. All analyses have been conducted in the LISREL program as well. The LISREL syntax files are not provided in this case because they are nearly identical with their single-level counterparts. The only differences are lower sample size (1157 – 87 = 1070) and different file name from which is loaded the $\boldsymbol{S}_{PW}$ matrix. As we can see in Table 15, all models exhibit better fit compared to the same model based on $\boldsymbol{S}_{T}$ (see Table 11).

Also in this case, the model 4*b* (see Figure 24) seems to be the best choice from all competing models. For comparison reasons, the estimated parameters of the model 4*b* based on $\boldsymbol{S}_{PW}$ are presented in Table 16. All estimates of factor loading and uniqueness are very similar. The difference between respective standardized estimates of model 4*b* parameters based on analysis of $\boldsymbol{S}_{T}$ and $\boldsymbol{S}_{PW}$ matrices ranged from 0.00 to 0.10. More parameters were slightly higher in $\boldsymbol{S}_{T}$ case, whereas the rest of them were higher in $\boldsymbol{S}_{PW}$ case or equaled to zero.

The occurrence of non-significant estimates was observed also in the within-group case. However, it does not mean that these items (namely items number 24 and 28) do not measure their underlying sub-construct. Obviously, when the negative and positive wording factors and the school PE factor are included in a model, the remaining factor loadings are suppressed. Similar situation already arose in the model 3*b* (see Figure 23). Here, three of four regression coefficients were negative even thought the correlation among all five attitude-related factors were highly positive. In a within-group case, the estimation of the structural part of model 2*a* is depicted in Figure 25.

Table 14. Pooled within-group covariance matrix S_{PW} and respective correlations

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18
Item 1	**2.85**	*.23*	*.30*	*.16*	*.39*	*.18*	*.18*	*.33*	*.11*	*.17*	*.31*	*.12*	*.08*	*.34*	*.25*	*.35*	*.28*	*.34*
Item 2	.50	**1.68**	*.18*	*.22*	*.24*	*.14*	*.16*	*.16*	*.15*	*.13*	*.24*	*.10*	*.20*	*.23*	*.32*	*.31*	*.22*	*.23*
Item 3	.75	.35	**2.13**	*.03*	*.24*	*.17*	*.13*	*.19*	*.12*	*.19*	*.22*	*.23*	*.05*	*.30*	*.26*	*.33*	*.26*	*.38*
Item 4	.39	.41	.07	**2.14**	*.19*	*.14*	*.11*	*.06*	*.10*	*.17*	*.19*	*-.01*	*.10*	*.14*	*.15*	*.12*	*.06*	*.11*
Item 5	1.15	.54	.61	.47	**2.98**	*.22*	*.29*	*.30*	*.16*	*.17*	*.27*	*.18*	*.20*	*.37*	*.20*	*.41*	*.36*	*.35*
Item 6	.50	.29	.42	.33	.63	**2.78**	*.19*	*.11*	*.12*	*.14*	*.21*	*.08*	*.19*	*.21*	*.14*	*.33*	*.25*	*.29*
Item 7	.49	.33	.30	.25	.80	.50	**2.47**	*.16*	*.15*	*.22*	*.20*	*.09*	*.17*	*.23*	*.05*	*.29*	*.27*	*.25*
Item 8	1.11	.43	.56	.18	1.03	.38	.50	**4.07**	*.03*	*.17*	*.19*	*.06*	*.05*	*.22*	*.18*	*.23*	*.20*	*.26*
Item 9	.29	.29	.27	.23	.41	.32	.36	.10	**2.35**	*.07*	*.18*	*.07*	*.14*	*.20*	*.12*	*.24*	*.16*	*.16*
Item 10	.50	.31	.51	.43	.54	.40	.60	.59	.18	**3.16**	*.18*	*.07*	*.06*	*.19*	*.13*	*.23*	*.21*	*.20*
Item 11	.67	.38	.40	.35	.60	.45	.39	.49	.35	.41	**1.60**	*.02*	*.27*	*.33*	*.27*	*.40*	*.31*	*.27*
Item 12	.31	.20	.50	-.03	.46	.19	.21	.19	.15	.20	.04	**2.22**	*.06*	*.20*	*.12*	*.12*	*.11*	*.18*
Item 13	.19	.35	.09	.20	.46	.43	.35	.14	.29	.13	.46	.11	**1.78**	*.14*	*.15*	.26	*.18*	*.16*
Item 14	.89	.47	.69	.32	.99	.54	.57	.71	.49	.52	.66	.47	.29	**2.45**	*.22*	*.31*	*.33*	*.30*
Item 15	.57	.58	.53	.31	.49	.33	.11	.49	.25	.31	.47	.26	.29	.47	**1.92**	*.30*	*.18*	*.21*
Item 16	.74	.50	.60	.21	.87	.69	.57	.58	.45	.50	.63	.22	.44	.60	.52	**1.55**	*.42*	*.44*
Item 17	.69	.41	.56	.13	.91	.61	.61	.56	.25	.55	.57	.24	.34	.74	.35	.75	**2.10**	*.37*
Item 18	.82	.44	.60	.22	.87	.70	.57	.77	.34	.50	.50	.38	.31	.68	.43	.80	.78	**2.08**
Item 19	.44	.50	.19	.59	.39	.38	.25	.36	.30	.35	.52	.09	.23	.64	.45	.35	.28	.27
Item 20	.58	.53	.57	.38	.63	.35	.37	.37	.41	.40	.49	.42	.26	.63	.73	.60	.43	.65
Item 21	1.06	.28	.56	.29	.95	.43	.43	1.01	.24	.62	.52	.08	.07	.70	.36	.60	.71	.75
Item 22	.51	.21	.66	-.02	.65	.43	.37	.28	.28	.23	.25	.35	.18	.44	.23	.63	.48	.65
Item 23	.41	.39	.22	.31	.63	.41	.41	.34	.37	.39	.57	.11	.45	.34	.38	.57	.67	.40
Item 24	.88	.55	.64	.40	.78	.31	.46	.90	.38	.54	.54	.40	.25	1.11	.49	.57	.54	.49
Item 25	.55	.57	.34	.41	.63	.44	.32	.37	.34	.39	.55	.22	.41	.47	.66	.54	.52	.44
Item 26	.63	.22	.43	.08	.67	.41	.52	.42	.27	.41	.41	.10	.36	.43	.26	.62	.65	.69
Item 27	1.31	.68	.64	.48	1.06	.71	.50	.93	.52	.51	.80	.36	.46	.96	.73	.85	.69	.89
Item 28	.63	.61	.30	.65	.71	.59	.45	.57	.35	.56	.57	.20	.39	.63	.56	.53	.48	.44
Item 29	.47	.31	.29	.21	.59	.43	.38	.42	.23	.35	.52	.13	.36	.40	.35	.52	.49	.57
Item 30	1.61	.64	.86	.54	1.43	.64	.72	1.41	.35	.68	.83	.46	.37	1.25	.69	.97	.87	.94

.19	.23	.32	.20	.20	.32	.26	.24	.49	.28	.21	.53	Item 19
.28	.28	*.11*	.10	.25	.26	.35	.12	.35	.33	.19	.28	Item 20
.10	.27	.19	.30	.12	.27	.19	.20	.28	.15	.15	.33	Item 21
.30	.18	.10	-.01	.17	.17	.23	.05	.21	.33	.11	.21	Item 22
.17	.25	.28	.25	.30	.28	.29	.28	.39	.31	.26	.46	Item 23
.17	.14	.13	.17	.20	.11	.21	.17	.27	.26	.20	.21	Item 24
.12	.16	.14	.15	.21	.18	.16	.23	.20	.22	.19	.25	Item 25
.13	.12	.25	.09	.14	.27	.15	.15	.29	.21	.16	.39	Item 26
.14	.18	.08	.12	.20	.15	.18	.13	.21	.17	.12	.13	Item 27
.15	.15	.17	.09	.18	.19	.18	.16	.18	.24	.15	.21	Item 28
.30	.26	.21	.13	.37	.26	.35	.23	.40	.34	.32	.37	Item 29
.05	.19	.03	.15	.06	.16	.12	.05	.15	.10	.07	.17	Item 30
.13	.13	.03	.09	.28	.12	.25	.19	.22	.22	.21	.15	
.30	.28	.23	.19	.18	.44	.24	.19	.39	.30	.20	.44	
.24	.36	.13	.11	.22	.22	.38	.13	.33	.30	.20	.28	
.21	.32	.24	.33	.37	.28	.35	.35	.43	.32	.32	.43	
.14	.20	.25	.22	.38	.23	.29	.32	.30	.25	.26	.33	
.14	.31	.26	.28	.22	.21	.25	.34	.39	.23	.31	.36	
1.85	.23	.16	.03	.20	.34	.32	.05	.33	.37	.20	.31	
.47	**2.15**	.16	.26	.20	.29	.29	.24	.36	.23	.20	.34	
.43	.45	**3.96**	.12	.07	.16	.14	.17	.26	.12	.19	.34	
.07	.58	.36	**2.31**	.15	.17	.12	.28	.23	.10	.19	.24	
.33	.36	.18	.27	**1.51**	.23	.34	.20	.28	.36	.37	.23	
.75	.68	.52	.42	.46	**2.65**	.25	.13	.41	.29	.17	.45	
.53	.54	.35	.22	.52	.49	**1.53**	.21	.40	.48	.27	.33	
.09	.49	.47	.59	.34	.31	.38	**2.00**	.27	.16	.20	.29	
.70	.83	.80	.56	.54	1.05	.79	.59	**2.47**	.42	.37	.56	
.68	.45	.32	.20	.59	.63	.79	.31	.88	**1.79**	.26	.32	
.35	.38	.47	.37	.58	.36	.42	.37	.74	.45	**1.66**	.31	
.76	.91	1.21	.66	.52	1.33	.79	.75	1.58	.77	.72	**3.25**	

Table 15. Fitting history of the hypothesized models based on S_{PW}

Model number	χ^2	*df*	RMSEA	SRMR	CFI	PCFI	AIC
1a	2447	405	0.069	0.053	0.943	0.878	2567
1b	1327	374	0.049	0.042	0.969	0.833	1509
2a	1911	399	0.060	0.051	0.954	0.875	2043
2b	1130	368	0.044	0.039	0.975	0.825	1324
3a	1948	401	0.060	0.051	0.954	0.879	2076
3b	1155	370	0.045	0.040	0.974	0.828	1345
4a	1513	387	0.052	0.049	0.964	0.858	1669
4b	835	356	0.035	0.034	0.983	0.804	1053
5a	1554	389	0.053	0.048	0.964	0.862	1706
5b	850	358	0.036	0.035	0.982	0.808	1064

Note. χ^2 = Chi-square, *df* = degrees of freedom, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, CFI = Comparative Fit Index, PCFI = Parsimony Comparative Fit Index, AIC = Akaike Information Criterion
Independence AIC = 27027, Saturated AIC = 930.

Table 16. Estimated standardized factor loadings and uniqueness of model 4*b* based on S_{PW}

Item	Factor loadings							Uniqueness
	Health & fitness	Social	Emotional	General	Positive wording	Negative wording	School PE	
1	--	--	--	0.39	--	0.38	0.39	0.55
2	--	0.21	--	--	0.50	--	--	0.71
3	--	--	0.36	--	--	0.29	0.17	0.76
4	--	-0.09	--	--	0.44	--	--	0.80
5	--	--	--	0.41	--	0.50	--	0.58
6	--	--	0.28	--	0.32	--	--	0.81
7	0.25	--	--	--	--	0.37	--	0.80
8	--	--	--	0.26	--	0.28	0.28	0.78
9	0.10	--	--	--	--	0.30	--	0.90
10	--	--	--	0.11	--	0.35	--	0.86
11	0.31	--	--	--	0.50	--	0.15	0.63
12	--	0.26	--	--	0.12	--	--	0.92
13	0.22	--	--	--	0.34	--	--	0.84
14	--	--	--	0.14	--	0.52	0.28	0.63
15	--	0.39	--	--	0.44	--	--	0.65
16	--	--	0.51	--	--	0.52	--	0.47
17	0.47	--	--	--	--	0.44	--	0.59
18	--	--	0.51	--	--	0.37	0.06	0.59
19	--	--	--	-0.15	0.55	--	0.22	0.63
20	--	0.45	--	--	--	0.41	0.11	0.61
21	--	--	--	0.31	--	0.24	0.21	0.80
22	--	--	0.44	--	--	0.19	--	0.77
23	0.35	--	--	--	0.49	--	--	0.64
24	0.03[a]	--	--	--	--	0.51	0.39	0.58
25	--	0.19	--	--	0.64	--	--	0.56
26	--	--	--	0.44	--	0.28	--	0.73
27	--	--	0.35	--	0.56	--	0.40	0.41
28	--	-0.02[a]	--	--	0.72	--	--	0.48
29	0.39	--	--	--	0.38	--	--	0.70
30	--	--	--	0.37	--	0.53	0.47	0.37

Note.
[a]non-significant parameter estimates at 0.05 level

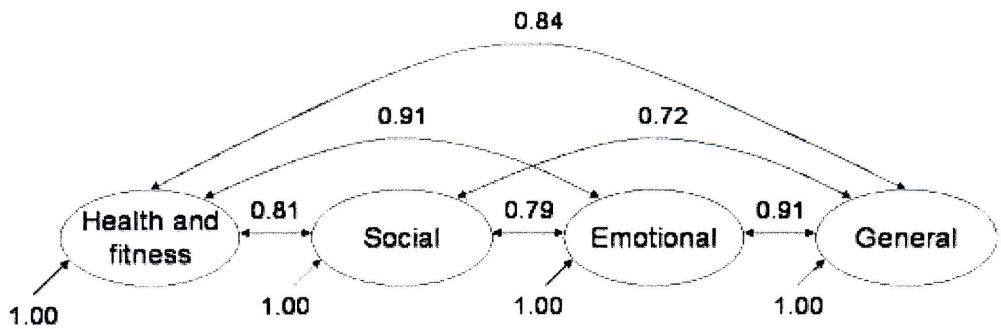


Figure 25. Structural part of model 2*a*, within-group level

Also in this case, all four attitude-related latent variables are again highly correlated. But now, if we add a second-order "overall attitude" factor to explain these correlations, the resulting estimates of factor loadings (the elements of the *beta* matrix) are all positive (see Figure 26). By contrast to Figure 23 and 22, the two wording factors are not included here.

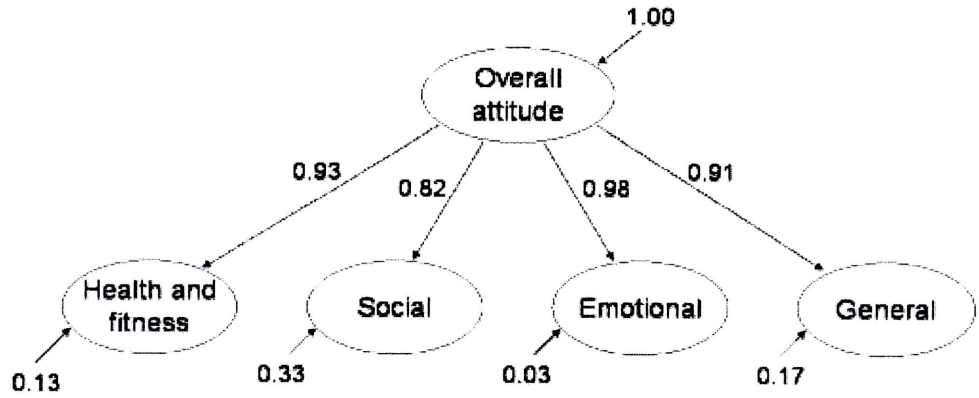


Figure 26. Structural part of model 3*a*, within-group level

In the previous section, we have warned of blind using the modification indices. Still and all, let us make one modification of the best fitting model 4*b*. It has little sense to modify this model by fixing or making free some parameters within the Λ_W and $\Theta_{\varepsilon W}$ matrices. Such a modification would not be theoretically justified. We allow making some changes in the Ψ_W matrix only. Modification indices in LISREL program suggested freeing the correlation between school PE factor and social factor. Subsequently, the model had been re-estimated with this modification. The estimation of the structural part of modified model 4*b* is depicted in Figure 27. The χ^2 decreased to 811, AIC to 1031, and SRMR to 0.032. The RMSEA and CFI indices remain unchanged. The LISREL syntax file of this modification can be found in Appendix.

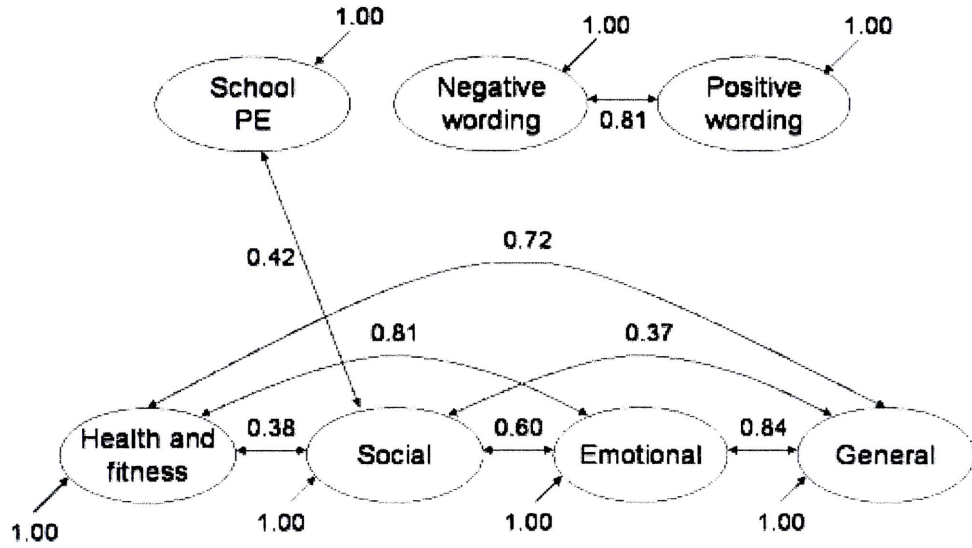


Figure 27. Modification of model 4*b*, within-group level

Between-group structure

The between-group covariance matrix Σ_B has been also estimated via the MMRM in PRELIS program. The estimation of this matrix, $\hat{\Sigma}_B$, is presented in Table 17. Analogous to the within-group case (see Table 14), the correlations are situated in the upper right triangle. All three proposed model for the between-group level were imposed on $\hat{\Sigma}_B$ in the LISREL program.

Let us remind the models that were fitted to this matrix. *Model 1* is the unidimensional model. *Model 2* is a two-factor model, where one factor constitutes school PE related items and second factor is linked with all remaining indicators. *Model 3* involves two orthogonal factors, where first factor is measured by all items and second factor is related only to school PE indicators only. LISREL syntax files of all three models are presented in Appendix.

The fitting history of all models is presented in Table 18. As we can see in this table, no one model fits the data acceptable. Strictly speaking, the models do not fit the data at all. Therefore, we have decided to conduct exploratory factor analysis (EFA) to have some notion about the factor structure and their number.

Table 17. Estimation of the between group covariance matrix Σ_B and respective correlations (in upper right triangle)

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18
Item 1	**.22**	*.56*	*.70*	*-.06*	*.87*	*.15*	*.21*	*.69*	*-.09*	*-.24*	*.44*	*.65*	*-.17*	*.49*	*.44*	*.25*	*.41*	*.30*
Item 2	.06	**.06**	*.43*	*.15*	*.39*	*.20*	*-.15*	*.50*	*.08*	*-.05*	*.46*	*.26*	*-.23*	*.31*	*.35*	*.01*	*.36*	*.28*
Item 3	.08	.03	**.07**	*-.43*	*.63*	*.48*	*.39*	*.60*	*-.17*	*-.24*	*.49*	*.36*	*.08*	*.35*	*.13*	*.54*	*.54*	*-.09*
Item 4	-.01	.01	-.04	**.11**	*-.21*	*-.32*	*-.66*	*-.01*	*.05*	*.05*	*.12*	*-.21*	*-.14*	*-.03*	*.12*	*-.37*	*-.25*	*-.49*
Item 5	.18	.04	.07	-.03	**.19**	*.17*	*.41*	*.64*	*.04*	*-.08*	*.32*	*.69*	*-.05*	*.55*	*.34*	*.23*	*.48*	*.32*
Item 6	.03	.02	.05	-.04	.03	**.15**	*.49*	*.51*	*-.05*	*.18*	*.39*	*-.29*	*.27*	*.37*	*.12*	*.63*	*.59*	*.69*
Item 7	.04	-.01	.04	-.08	.07	.07	**.14**	*.29*	*.00*	*-.15*	*.14*	*.20*	*.42*	*.31*	*.09*	*.55*	*.53*	*.42*
Item 8	.17	.06	.08	.00	.15	.11	.06	**.28**	*.05*	*.00*	*.42*	*.44*	*-.02*	*.70*	*.45*	*.52*	*.75*	*.48*
Item 9	-.01	.01	-.01	.00	.01	-.01	.00	.01	**.10**	*.62*	*-.03*	*.19*	*.38*	*.22*	*.47*	*.08*	*.35*	*.06*
Item 10	-.03	.00	-.02	.01	-.01	.02	-.02	.00	.05	**.08**	*-.30*	*-.10*	*.25*	*.01*	*-.07*	*-.16*	*-.01*	*-.08*
Item 11	.05	.03	.03	-.02	.03	.03	.01	.05	.00	-.02	**.05**	*.13*	*-.23*	*.36*	*.46*	*.23*	*.34*	*.30*
Item 12	.10	.02	.03	-.02	.10	.04	.03	.08	.02	-.01	.01	**.12**	*-.16*	*.34*	*.31*	*.12*	*.27*	*.08*
Item 13	-.03	-.02	.01	.00	-.01	.04	.06	.00	.05	.03	-.02	-.02	**.15**	*-.01*	*.12*	.57	*.40*	*.34*
Item 14	.10	.03	.04	.01	.10	.06	.05	.16	.03	.00	.03	.05	.00	**.19**	*.51*	*.34*	*.52*	*.51*
Item 15	.05	.02	.01	.01	.04	.01	.01	.06	.04	-.01	.03	.03	.01	.06	**.07**	*.35*	*.52*	*.45*
Item 16	.04	.00	.05	-.04	.04	.09	.07	.10	.01	-.02	.02	.01	.08	.05	.03	**.13**	*.78*	*.76*
Item 17	.09	.04	.06	-.04	.10	.11	.09	.18	.05	.00	.04	.04	.07	.10	.06	.13	**.21**	*.68*
Item 18	.05	.02	.05	-.05	.05	.09	.06	.09	.01	-.01	.02	.01	.04	.08	.04	.09	.11	**.12**
Item 19	.06	.06	-.01	.11	.03	-.04	-.08	.02	.03	.01	.03	.00	-.03	.07	.04	-.07	-.02	-.02
Item 20	.05	.04	.02	.01	.04	-.01	.00	.05	.03	.00	.01	.04	.01	.04	.03	.01	.04	.02
Item 21	.22	.04	.09	-.01	.19	.02	.05	.16	-.04	-.05	.05	.11	-.05	.14	.02	.04	.06	.05
Item 22	.00	.02	.00	-.02	-.02	.03	-.01	.04	.01	.02	.00	.00	-.01	.00	.02	.01	.03	.02
Item 23	.01	.00	.02	.00	.03	.04	.04	.07	.03	.01	.01	.00	.05	.06	.01	.04	.07	.02
Item 24	.10	.05	.01	.05	.09	-.02	-.03	.09	.02	.01	.03	.06	-.04	.13	.05	-.04	.01	.02
Item 25	.02	.01	.02	.01	.01	.05	.01	.05	.00	.00	.02	-.02	.02	.05	.02	.04	.05	.05
Item 26	.04	.02	.04	-.04	.05	.09	.08	.08	.02	.00	.02	.00	.05	.06	.04	.08	.12	.08
Item 27	.14	.05	.06	.01	.11	.05	.02	.13	-.03	-.02	.05	.04	-.01	.09	.03	.03	.06	.04
Item 28	.05	.02	.02	-.02	.05	.05	.05	.08	.00	.00	.00	.00	.03	.06	.02	.04	.08	.05
Item 29	.03	.03	.01	.03	.02	.02	-.01	.06	.05	.02	.02	.02	.03	.05	.04	.03	.07	.03
Item 30	.16	.05	.04	.05	.14	.02	-.01	.16	.02	-.01	.05	.08	-.03	.15	.07	.03	.09	.06

.25	.51	.84	.01	.09	.51	.16	.29	.87	.40	.28	.71	Item 19				
.51	.68	.28	.30	.05	.50	.30	.24	.57	.28	.52	.42	Item 20				
-.09	.33	.60	.06	.33	.09	.39	.49	.70	.33	.15	.33	Item 21				
.69	.13	-.03	-.23	.01	.38	.09	-.47	.13	-.23	.37	.30	Item 22				
.16	.43	.78	-.21	.23	.48	.06	.37	.73	.45	.23	.67	Item 23				
-.23	-.12	.11	.29	.35	-.13	.71	.77	.37	.55	.26	.12	Item 24				
-.48	-.01	.22	-.07	.41	-.20	.09	.72	.19	.53	-.06	-.05	Item 25				
.09	.46	.54	.28	.48	.41	.47	.52	.74	.59	.48	.63	Item 26				
.20	.43	-.23	.19	.39	.14	.05	.24	-.23	.02	.62	.11	Item 27				
.06	.04	-.29	.22	.17	.06	-.06	-.02	-.21	-.02	.31	-.08	Item 28				
.25	.22	.37	.05	.16	.30	.53	.34	.58	.01	.41	.44	Item 29				
-.03	.58	.57	.01	.05	.43	-.34	.04	.35	.05	.22	.47	Item 30				
-.14	.15	-.21	-.15	.46	-.25	.20	.47	-.10	.29	.30	-.16	.71				
.35	.47	.57	.03	.50	.68	.58	.49	.57	.56	.46	.71	.74				
.33	.48	.17	.28	.15	.41	.45	.47	.33	.27	.55	.53	.38				
-.41	.12	.18	.17	.44	-.23	.55	.80	.25	.44	.34	.19	.20				
-.11	.41	.23	.26	.58	.04	.51	.84	**.40**	.67	.57	.39	.37				
-.13	.23	.25	.27	.16	.11	.66	.78	.35	.57	.30	.35	.73				
.23	.53	.25	-.20	.02	.78	.25	-.25	.41	.09	.47	.52	.48				
.05	**.04**	.33	.19	.25	.66	.10	.17	.44	.30	.70	.44	.78				
.07	.04	**.31**	-.25	.07	.57	.14	.13	.82	.39	.15	.78	-.19				
-.02	.01	-.03	**.06**	-.18	-.09	.14	.21	-.07	.21	.22	-.19	.18				
.00	.01	.01	-.01	**.07**	.02	.43	.49	.23	.27	.37	.18	.74				
.16	.06	.14	-.01	.00	**.19**	.18	-.11	.59	.25	.47	.74					
.02	.00	.02	.01	.02	.02	**.04**	.61	.43	.37	.43	.38					
-.03	.01	.02	.02	.04	-.01	.04	**.09**	.29	.62	.37	.20					
.07	.03	.16	-.01	.02	.09	.03	.03	**.12**	.49	.33	.73					
.01	.02	.05	.01	.02	.03	.02	.05	.04	**.06**	.27	.37					
.06	.04	.02	.01	.02	.05	.02	.03	.03	.02	**.06**	.48					
.12	.04	.21	-.02	.02	.16	.04	.03	.12	.04	.06	**.23**					

Table 18. Fitting history of the hypothesized models based on $\hat{\Sigma}_B$

Model Number	χ^2	*df*	RMSEA	SRMR	AIC
1	3713	405	0.308	0.212	3833
2	2736	404	0.259	0.232	2858
3	2707	393	0.262	0.205	2851

Note. χ^2 = Chi-square, *df* = degrees of freedom, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, AIC = Akaike Information Criterion, Independence AIC = 5748, Saturated AIC = 930.

EFA is recommended when researches have no hypotheses about the nature of the underlying factor structure of a test. The most common approach to deciding number of factors is to create a scree plot. The factors are on *x*-axis and eigenvalues are on *y*-axis. Eigenvalues represent the variance accounted for by each factor. Our 30-item scale will theoretically have 30 eigenvalues. First 15 factors are arranged in a scree plot in Figure 28.

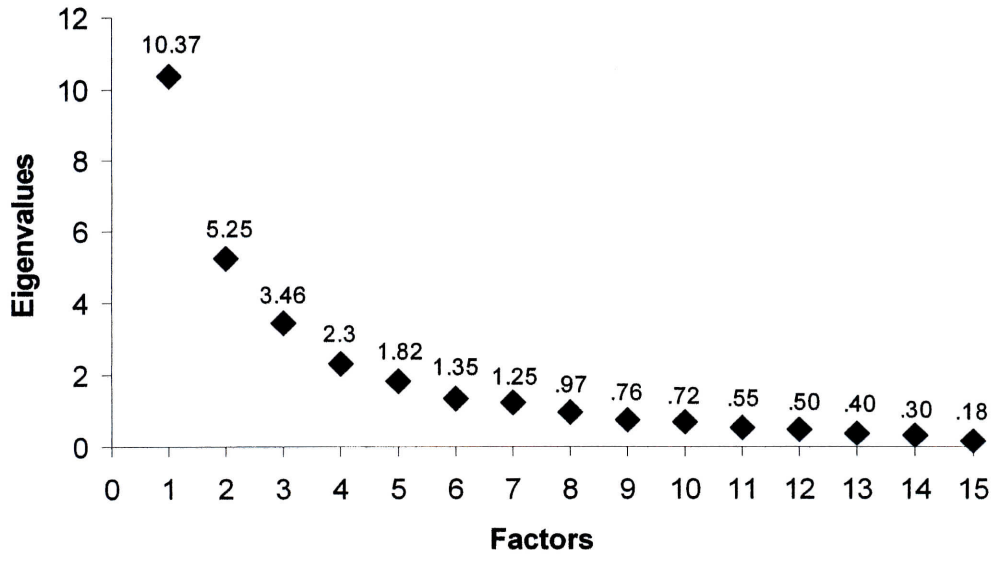


Figure 28. Scree plot of the first 15 factors

We might examine this plot in Figure 28 and decide there are from 3 to 5 underlying factors and remainder of factors is just scree. First 3 factors accounted for 63.6% and first 5 factors accounted for 77.3% of the total variance. This approach to selecting the number of factors involves certain amount of subjective judgment. Another approach is called the Kaiser-Guttman rule and states that the number of latent factors are equal to the number of factors with eigenvalues grater than 1. However, this rule is very rough and seems to produce too many factors. The EFA analysis was conducted in Mplus (Muthén & Muthén, 2005) and the syntax file is presented in Appendix.

First, the three-factor solution was obtained. The loadings were subsequently rotated using *promax* oblique rotation method. An important step is to assign names to the factors. The factor could be called as follows:

- Factor 1 – *Importance* of physical activity. Why is important to incorporate physical activity within curriculum in schools? Why is important to exercise in leisure time?
- Factor 2 – *Health.* Influence of physical activity on physical, social, and mental health of an individual
- Factor 3 – *Utility* of physical activity. Do individuals gain any kind of benefit (e.g., new friends, personality development) when performing physical activity?

Based on EFA results, a three-factor CFA was conducted in LISREL program. The syntax file is presented in Appendix. However, the simple structure did not appear, because some items are indicators of more than one factor. This three-factor model exhibited better fit (χ^2 = 2198, df = 386, RMSEA = 0.234, SRMR = 0.127, AIC = 2356) than the previous three model (see Table 18). However, also this model fits the data poorly. Unfortunately, adding fourth or fifth factor to the model did not resulted in substantial improvement of model fit. For illustrative purposes, therefore, we accept the three-factor model mentioned above.

The correlation among these factors ranged from 0.35 to 0.45. Estimated standardized factor loadings and unique variances are presented in Table 19. All estimates were statistically significant at 0.05 level.

Table 19. Estimated standardized factor loadings and uniqueness of the three-factor model based on $\hat{\Sigma}_B$

Item	Factor loadings			
	Importance	**Health**	**Utility**	**Uniqueness**
1	0.94	--	--	0.11
2	0.33	--	0.45	0.57
3	0.57	0.37	---	0.40
4	--	-0.71	0.67	0.46
5	0.86	--	--	0.26
6	--	0.75	--	0.43
7	--	0.85	-0.44	0.42
8	0.57	0.45	--	0.30
9	-0.48	--	0.74	0.53
10	-0.43	--	0.39	0.81
11	0.49	--	--	0.76
12	0.61	--	--	0.63
13	-0.42	0.63	--	0.61
14	0.34	0.23	0.44	0.37
15	--	0.22	0.58	0.51
16	--	0.85	--	0.27
17	--	0.91	--	0.18
18	--	0.80	--	0.35
19	--	-0.65	0.96	0.21
20	0.20	--	0.65	0.43
21	0.90	--	--	0.18
22	--	0.25	--	0.94
23	--	0.54	--	0.71
24	0.46	-0.53	0.74	0.18
25	--	0.62	--	0.62
26	--	0.95	--	0.10
27	0.91	--	--	0.17
28	--	0.65	--	0.57
29	--	--	0.85	0.28
30	0.60	--	0.41	0.25

Simultaneous estimation of the within- and between-group structures

Up to now, all three covariance matrices, $\boldsymbol{S}_T$, $\boldsymbol{S}_{PW}$, and $\hat{\boldsymbol{\Sigma}}_B$ have been inserted to the SEM programs separately. At this point, the simultaneous analysis of the between- and within-group structures is conducted. The $\boldsymbol{S}_{PW}$ and $\boldsymbol{S}^*_B$ constitute matrices of two groups and can be together inserted into SEM programs that enable multi-group analysis. Nevertheless we have used a raw data to be analyzed in Mplus program. The Mplus syntax file is also in Appendix. The same between- and within-group structures were estimated here. At the within-group level, the modification of model 4*b* discussed above has been estimated. The three-factor structure was assumed to hold at the between-group level. This two-level model is depicted in Figure 29. The scalar *c* defined in equation (24) is 13.26 and its square root equates to 3.64. To this number are fixed factor loadings between indicators at both levels. All estimated parameters are slightly different; because the $\hat{\boldsymbol{\Psi}}_B$ matrix is not positive definite and LISREL and Mplus programs solve this problem variously. Only the structural part of this model is shown in Figure 29, whereas the factor loadings and the uniqueness are omitted for transparency reasons.

All estimated correlations among latent variables differ slightly compared to the separate analyses of the between- and within-group levels. The greatest difference is observed at the between-group level, where all correlations are lower about 0.1 compared to the separate analysis of the $\hat{\boldsymbol{\Sigma}}_B$ matrix. The Mplus program produced following fit indeces of the two-level model:

- $\chi^2 = 1235$
- Degrees of freedom = 741 (df = 355 at the within level) + (df = 386 at the between level)
- CFI = 0.945
- RMSEA = 0.024
- AIC = 119215
- SRMR for the within level = 0.033
- SRMR for the between level = 0.074

The complex two-level model exhibit acceptable fit. The fact that the within-group level accounts for most of the total covariation is taken into account here. If the model fit is

excellent at the within-group level and it accounts for most of variation, the global fit will be presumably also acceptable, even though the between-part of the model fits the data poorly.

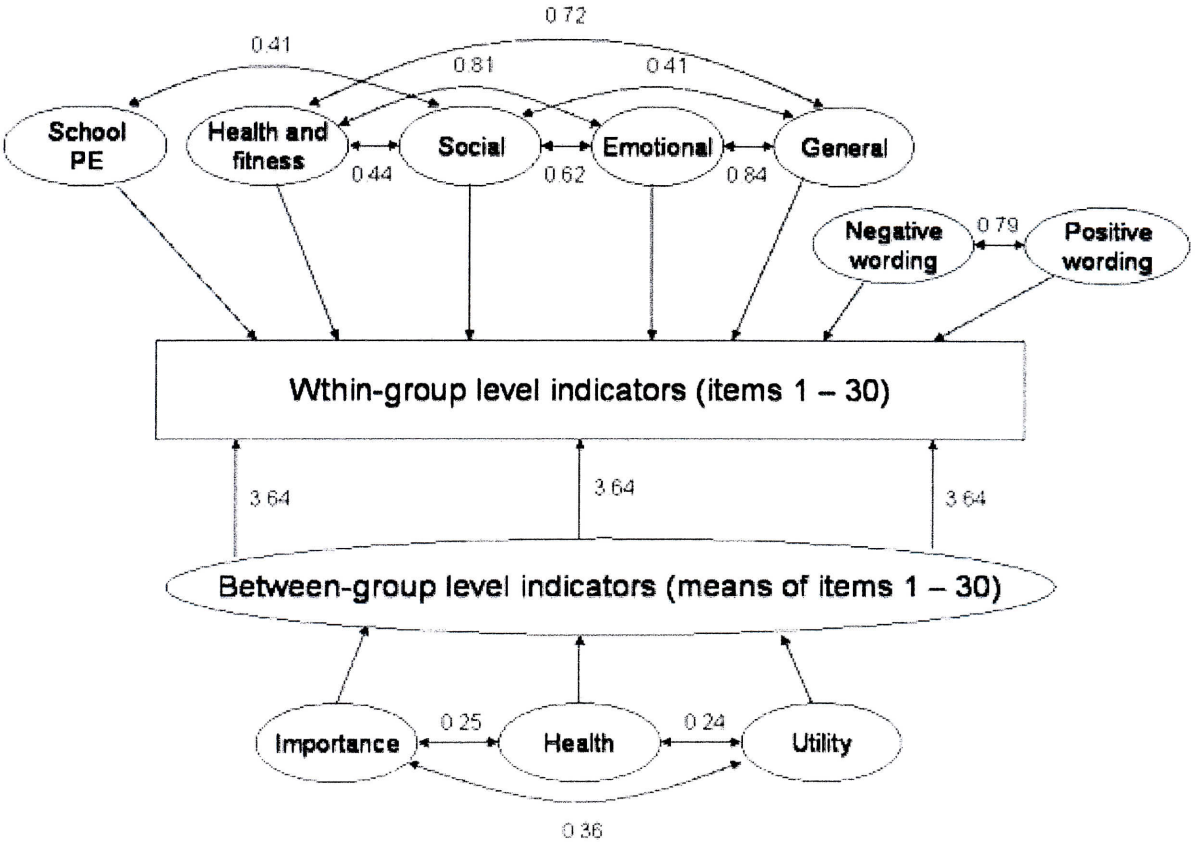


Figure 29. Structural part of the two-level model (all factor variances are fixed to 1)

ORDINAL VARIABLES RESULTS

Descriptive statistics

It is convenient to start with univariate descriptive statistics when dealing with SEM with ordinal variables. PRELIS program provides frequencies, relative frequencies and bar charts for each item. The relative frequencies (%) for all 7 categories are presented in Table 20.

Table 20. Relative frequencies (%)

Item	Scale points						
	1	2	3	4	5	6	7
1	5.4	4.1	7.2	9.3	13.8	28.2	32.1
2	1.4	4.3	7.7	20.5	27.9	30.7	7.4
3	2.6	5.5	9.3	18.9	19.2	35.6	9.0
4	7.4	18.0	18.9	26.8	17.0	9.4	2.4
5	8.4	11.5	11.5	17.9	18.2	23.6	8.9
6	4.2	7.7	10.1	11.7	18.0	30.4	17.9
7	4.0	16.6	13.9	18.5	22.8	18.8	5.4
8	15.8	15.7	13.7	11.7	9.9	19.0	14.2
9	2.5	6.7	7.2	15.2	18.0	34.5	15.8
10	9.5	20.8	16.5	15.3	16.1	14.4	7.3
11	1.5	2.8	4.2	12.9	26.3	38.1	14.2
12	5.0	10.8	17.5	26.9	18.7	16.6	4.5
13	2.2	2.9	4.1	8.8	22.2	37.2	22.6
14	5.7	8.2	9.9	20.8	22.8	23.8	8.7
15	2.1	3.3	4.6	15.1	19.6	37.1	18.3
16	1.6	1.4	2.9	11.6	15.7	40.4	26.3
17	2.9	4.6	5.4	10.5	20.2	35.2	21.2
18	2.9	3.4	7.4	13.8	20.3	35.9	16.3
19	6.3	17.2	17.6	29.0	19.5	8.6	1.9
20	2.8	4.3	6.4	22.4	20.7	29.9	13.5
21	17.4	15.3	10.0	15.5	12.9	14.9	14.0
22	3.9	5.1	9.3	19.0	20.4	31.7	10.5
23	1.8	1.4	3.6	5.5	23.2	40.8	23.6
24	8.4	16.7	20.7	14.8	19.4	15.6	4.5
25	1.2	3.5	4.5	17.1	31.7	32.8	9.1
26	2.7	4.1	7.8	15.4	24.5	34.1	11.5
27	4.2	6.7	6.9	17.3	21.6	29.2	14.2
28	2.5	8.2	12.5	28.9	27.2	17.1	3.5
29	1.9	1.6	2.9	7.5	19.2	37.2	29.6
30	8.8	7.7	9.2	10.6	17.1	30.5	16.1

The most frequently used scale point was the 6th grade (M = 28%), whereas number 1 was the least frequently used scale point (M = 5%). Once the frequencies are known, the thresholds

can be computed. Six thresholds are estimated when the 7-point scale is used. The thresholds for all 30 items computed in PRELIS are listed below.

Variable	Mean	St. Dev.	Thresholds					
IT1	0.000	1.000	-1.611	-1.315	-0.970	-0.645	-0.261	0.465
IT2	0.000	1.000	-2.202	-1.580	-1.107	-0.415	0.302	1.444
IT3	0.000	1.000	-1.944	-1.396	-0.939	-0.352	0.137	1.341
IT4	0.000	1.000	-1.444	-0.661	-0.141	0.559	1.182	1.973
IT5	0.000	1.000	-1.379	-0.845	-0.485	-0.017	0.453	1.346
IT6	0.000	1.000	-1.724	-1.178	-0.770	-0.420	0.043	0.919
IT7	0.000	1.000	-1.753	-0.821	-0.398	0.076	0.702	1.611
IT8	0.000	1.000	-1.001	-0.482	-0.122	0.172	0.434	1.072
IT9	0.000	1.000	-1.958	-1.325	-0.977	-0.477	-0.009	1.001
IT10	0.000	1.000	-1.305	-0.512	-0.076	0.313	0.785	1.456
IT11	0.000	1.000	-2.178	-1.724	-1.374	-0.794	-0.059	1.072
IT12	0.000	1.000	-1.643	-1.001	-0.432	0.259	0.803	1.696
IT13	0.000	1.000	-2.021	-1.635	-1.330	-0.916	-0.248	0.753
IT14	0.000	1.000	-1.580	-1.084	-0.713	-0.135	0.453	1.357
IT15	0.000	1.000	-2.038	-1.611	-1.285	-0.674	-0.135	0.906
IT16	0.000	1.000	-2.134	-1.877	-1.557	-0.932	-0.432	0.634
IT17	0.000	1.000	-1.890	-1.438	-1.132	-0.727	-0.161	0.800
IT18	0.000	1.000	-1.890	-1.529	-1.095	-0.598	-0.054	0.984
IT19	0.000	1.000	-1.529	-0.722	-0.225	0.526	1.255	2.074
IT20	0.000	1.000	-1.916	-1.469	-1.103	-0.361	0.166	1.103
IT21	0.000	1.000	-0.939	-0.448	-0.183	0.207	0.557	1.080
IT22	0.000	1.000	-1.763	-1.341	-0.903	-0.322	0.196	1.255
IT23	0.000	1.000	-2.093	-1.852	-1.488	-1.157	-0.370	0.719
IT24	0.000	1.000	-1.379	-0.672	-0.106	0.268	0.839	1.696
IT25	0.000	1.000	-2.254	-1.669	-1.325	-0.632	0.205	1.336
IT26	0.000	1.000	-1.930	-1.495	-1.057	-0.526	0.111	1.200
IT27	0.000	1.000	-1.733	-1.237	-0.926	-0.384	0.168	1.072
IT28	0.000	1.000	-1.958	-1.241	-0.733	0.052	0.818	1.806
IT29	0.000	1.000	-2.074	-1.806	-1.515	-1.080	-0.434	0.536
IT30	0.000	1.000	-1.352	-0.973	-0.653	-0.352	0.085	0.991

Consequently, the polychoric correlations among all 30 items have been estimated under assumption of bivariate normality. These estimations are presented in Appendix. As we can see, the polychoric correlations are slightly higher compared to the product-moment correlations presented in Table 10. This difference has been expected due to the mentioned fact that the product-moment correlations underestimate the relationship between two ordinal variables. A special attention should be paid to the last two columns. The RMSEA value should not be greater than 0.1 (Jöreskog, 2005). If the RMSEA exceeds this value, the model may not hold and the respective item should be eliminated from the model. Nevertheless, this is not the case of the present data, because all values of RMSEA remain below this cut-point.

Single-level CFA results for ordinal variables

In this part, we present the results of single-level CFA for ordinal variables. Compared to the single-level CFA for continuous variables, the polychoric correlations and their asymptotic covariance matrix are used to be analyzed here. The LISREL syntax file is presented in Appendix. The same ten models as in sinle-level CFA for continuous variables were considered also in this case. All models were estimated using the MLR (robust maximum likelihood) method. The fitting history is presented in Table 21.

Table 21. Fitting history of the hypothesized models for ordinal variables

Model number	χ^2	*df*	RMSEA	SRMR	CFI	PGFI	AIC
1a	99	405	0.000	0.063	1.000	0.703	219
1b	47	374	0.000	0.047	1.000	0.722	229
2a	74	399	0.000	0.062	1.000	0.728	206
2b	39	368	0.000	0.045	1.000	0.722	233
3a	75	401	0.000	0.061	1.000	0.729	203
3b[a)]	41	370	0.000	0.046	1.000	0.723	231
4a	52	387	0.000	0.059	1.000	0.737	208
4b	30	356	0.000	0.041	1.000	0.712	248
5a	53	389	0.000	0.058	1.000	0.738	205
5b	30	358	0.000	0.041	1.000	0.715	244

Note. χ^2 = Satorra-Bentler Scaled Chi-square, *df* = degrees of freedom, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, CFI = Comparative Fit Index, PGFI = Parsimony Goodness of Fit Index, AIC = Akaike Information Criterion.

[a)] The solution has not converged after 950 iterations, MLR method has been used in all cases

As we can see in Table 21, the CFI, and RMSEA fit indices don't work, when Satorra-Bentler chi-square is used. These indices provide no information about model fit because they indicate perfect fit of all models. Four types of the chi-square statistics are presented by the LISREL program. For example, the model 1*a* has following chi-square statistics: Satorra-Bentler $\chi^2 = 99$; minimum fit function $\chi^2 = 3014$; normal theory weighted least squares $\chi^2 = 4127$; χ^2 corrected for non-normality = 235. The Satorra-Bentler χ^2 is appropriate, when MLR estimation method is used. The SRMR fit index seems to be suitable for comparing model fits in this case with fitting history of CFA results based on $\boldsymbol{S}_T$ presented in Table 11. The values of SRMRs are now slightly higher which indicates slightly worse fit. This is caused by higher values of polychoric correlations.

Table 22. Estimated factor loadings and uniqueness of model 4*b* for ordinal variables

Item	Factor loadings							Uniqueness
	Health & fitness	Social	Emotional	General	Positive wording	Negative wording	School PE	
1	--	--	--	0.41	--	0.42	0.46	0.44
2	--	0.23	--	--	0.52	--	--	0.67
3	--	--	0.38	--	--	0.31	0.19	0.73
4	--	-0.13	--	--	0.45	--	--	0.78
5	--	--	--	0.42	--	0.53	--	0.54
6	--	--	0.37	--	0.34	--	--	0.75
7	0.31	--	--	--	--	0.35	--	0.78
8	--	--	--	0.30	--	0.33	0.32	0.71
9	0.13	--	--	--	--	0.35	--	0.86
10	--	--	--	0.12	--	0.34	--	0.87
11	0.33	--	--	--	0.52	--	0.19	0.58
12	--	0.25	--	--	0.13	--	--	0.92
13	0.33	--	--	--	0.36	--	--	0.76
14	--	--	--	0.13	--	0.58	0.27	0.58
15	--	0.45	--	--	0.43	--	--	0.61
16	--	--	0.61	--	--	0.52	--	0.36
17	0.55	--	--	--	--	0.48	--	0.47
18	--	--	0.53	--	--	0.41	0.08	0.55
19	--	--	--	-0.23	0.60	--	0.25	0.52
20	--	0.43	--	--	--	0.46	0.10	0.60
21	--	--	--	0.33	--	0.27	0.34	0.70
22	--	--	0.42	--	--	0.22	--	0.77
23	0.43	--	--	--	0.55	--	--	0.52
24	0.03	--	--	--	--	0.58	0.39	0.51
25	--	0.25	--	--	0.65	--	--	0.50
26	--	--	--	0.46	--	0.32	--	0.69
27	--	--	0.34	--	0.58	--	0.44	0.36
28	--	0.05	--	--	0.71	--	--	0.49
29	0.35	--	--	--	0.45	--	--	0.68
30	--	--	--	0.31	--	0.60	0.48	0.32

So we can compare estimated parameters with the results for continuous variables, the measurement part of model 4*b* (factor loadings and uniqueness) is presented in Table 22. However, the model 4*b* doesn't need to be the most acceptable model from all ten competing models. The PGFI index indicates that the models 4*a* or 5*a* should be preferred, and the AIC index suggests accepting the model 3*a*. Nevertheless, this approach serves us only for analyzing differences between continuous and ordinal variable. Two differences are apparent, when Tables 11 and 22 are compared. First, the estimated standardized factor loadings are higher about 0.05 in ordinal variables case. Second, the uniqueness is lower in ordinal variables case (except items number 4, 12, and 26). The highest difference is observed in item number 23 (0.12). Hence the reliability estimation will be presumably higher, when items are

considered as being measured on ordinal scale. The estimated correlations among factors are presented in Table 23.

Table 23. Structural part of model 4*b*: Estimated correlations among factors, ordinal variables case

	Health & fitness	Social	Emotional	General	Positive wording	Negative wording	School PE
Health & fitness	1.00						
Social	0.24	1.00					
Emotional	0.75	0.54	1.00				
General	0.66	0.32	0.79	1.00			
Positive wording	--	--	--	--	1.00		
Negative wording	--	--	--	--	0.86	1.00	
School PE	--	--	--	--	--	--	1.00

Compared to the continuous variables case (see Figure 24), the correlations among attitude-related factors are now lower about 0.07 (see Table 23). In principle, the correlation between negative and positive wording factors remains unchanged.

Multilevel CFA results for ordinal variables

As we noted in the method section, the multilevel CFA for ordinal variables is very computationally demanding. For this reason, it was impossible to conduct the same analyzes as for continuous variables. The model depicted in Figure 29 with seven factors at the within-level and three factors at the between-level could not be estimated in Mplus program version 3.13. In this program, the estimator for this type of analysis is maximum likelihood with robust standard errors using numerical integration algorithm. Unfortunately, no computer with Mplus license was efficient enough to carry out this analysis. There were not enought memory and hard drive spaces to run this program (1GB of memory and 15GB of hard disk space was not sufficient). The computational time of deteministic numerical integration, Gaussian quadrature, is proportional to the product of the number of quadrature points for all the latent variables used. In our case, total of 5.8×10^{11} integration points and 10 dimensions

were required. In general, models with three or more factors per level may take too much time to be of practical use (Grilli & Rampichini, 2007).

The same authors declare that the Monte Carlo integration may be more convenient in such cases. Therefore, the integration method was changed to Monte Carlo in the Mplus syntax input file. Nevertheless, the program did not terminate during 20 hours. Hence, we made a decision to stop the analysis. It had little sense to wait longer for convergence of the model. To make a comparison with other three approaches possible, the easier model has been estimated. The multilevel model with four factors (model 2*a*) at the within-group level and the one factor (model 1) at the between-level was estimated.

The AIC fit index was the only one provided by the output file, but it brings no information about model fit compared to other approaches. Furthermore, the estimated parameters at the between-group level may not be trustworthy due to a non-positive definite first-order derivative product matrix. For this reason, only the estimations of the within-group structure and their comparison with other approaches are presented here. The time needed for conducting analysis of model 2*a* using four different approaches is presented below:

1. Single-level CFA for continuous variables (LISREL): **0.3 Seconds**
2. Multilevel CFA for continuous variables (Mplus): **5 Seconds**
3. Single-level CFA for ordinal variables (LISREL): **42 Seconds**
4. Multilevel CFA for ordinal variables (Mplus): **53 Minutes, 5 Seconds**

As we can see, the computational burden is really heavy in multilevel CFA with ordinal variables case. The estimated factor loadings with respective standard errors of model 2*a* at the within-group level using all four approaches listed above are presented in Table 24.

The estimated factor loadings are very similar using both single-level and multilevel CFA for continuous variables approaches. Both ordinal variable approaches differ in the magnitude of estimated loadings from subscale to subscale. The factor loadings belonging to the *health and fitness* subscale are higher in both ordinal variable cases. On the other hand, the estimates are smaller within the *general* subscale, when variables are considered as ordinal.

From the standard errors point of view, both continuous variable approaches have low values and all estimated parameters are significantly different from zero. Sligtly higher standard errors are observed when parameters are estimated within the multilevel CFA for ordinal variables framework (see the last column of the Table 24).

Table 24. Model 2*a* : Estimated factor loadings using four different approaches

Sub-consructs	Items	Factor loadings (Standard errors)			
		Single-level continuous	**Multilevel continuous**	**Single-level ordinal**	**Multilevel ordinal**
Health and fitness	Item 7[a]	1.00	1.00	1.00	1.00
	Item 9	0.74 (0.09)	0.75 (0.10)	0.88 (0.38)	0.89 (0.14)
	Item 11	1.19 (0.10)	1.22 (0.11)	1.53 (0.51)	1.75 (0.22)
	Item 13	0.78 (0.09)	0.80 (0.09)	1.06 (0.56)	1.00 (0.15)
	Item 17	1.38 (0.12)	1.37 (0.12)	1.58 (0.48)	1.94 (0.23)
	Item 23	1.07 (0.10)	1.10 (0.10)	1.58 (0.77)	1.68 (0.22)
	Item 24	1.29 (0.12)	1.29 (0.12)	1.19 (0.37)	1.54 (0.20)
	Item 29	1.05 (0.10)	1.07 (0.10)	1.36 (0.75)	1.44 (0.20)
Social	Item 2[a]	1.00	1.00	1.00	1.00
	Item 4	0.71 (0.07)	0.75 (0.07)	0.60 (0.18)	0.56 (0.07)
	Item 12	0.47 (0.07)	0.45 (0.07)	0.39 (0.23)	0.37 (0.07)
	Item 15	1.07 (0.08)	1.06 (0.08)	0.97 (0.27)	0.88 (0.10)
	Item 20	1.07 (0.08)	1.06 (0.08)	0.96 (0.27)	0.95 (0.10)
	Item 25	1.15 (0.07)	1.17 (0.07)	1.22 (0.24)	1.28 (0.12)
	Item 28	1.21 (0.08)	1.21 (0.08)	1.16 (0.22)	1.25 (0.09)
Emotional	Item 3[a]	1.00	1.00	1.00	1.00
	Item 6	1.05 (0.09)	1.07 (0.10)	0.94 (0.39)	0.91 (0.11)
	Item 16	1.28 (0.09)	1.31 (0.09)	1.52 (0.56)	1.96 (0.19)
	Item 18	1.28 (0.09)	1.30 (0.10)	1.30 (0.50)	1.48 (0.13)
	Item 22	0.82 (0.08)	0.86 (0.09)	0.83 (0.44)	0.79 (0.08)
	Item 27	1.66 (0.11)	1.74 (0.12)	1.50 (0.48)	2.12 (0.19)
General	Item 1[a]	1.00	1.00	1.00	1.00
	Item 5	0.99 (0.05)	0.99 (0.06)	0.93 (0.31)	0.83 (0.07)
	Item 8	0.88 (0.06)	0.86 (0.06)	0.75 (0.28)	0.58 (0.06)
	Item 10	0.48 (0.05)	0.53 (0.05)	0.46 (0.21)	0.32 (0.05)
	Item 14	0.84 (0.05)	0.84 (0.05)	0.87 (0.30)	0.73 (0.07)
	Item 19	0.48 (0.04)	0.52 (0.04)	0.58 (0.22)	0.44 (0.06)
	Item 21	0.83 (0.06)	0.79 (0.06)	0.73 (0.30)	0.57 (0.05)
	Item 26	0.54 (0.04)	0.57 (0.04)	0.62 (0.29)	0.49 (0.05)
	Item 30	1.25 (0.06)	1.27 (0.06)	1.14 (0.34)	1.29 (0.11)

Note.
Standard errors are in parentheses.
[a] Factor loadings fixed to 1.00, no standard errors are provided.

Even seven times larger standard errors have arisen in the single-level CFA with ordinal variables instance. In this case, some parameters become no longer significantly different from zero. Concretely, the items number 12, 13, 22, and 29 are no longer statistically significant at the 0.05 level.

The estimated factor variances, covariances and respective correlations are presented in Table 25. Also in this case, the comparison among all four approaches has been accomplished. Much higher diversity in the estimations can be observed in this table.

Table 26. Standardized uniqueness of model 2*a* using all four approaches

Items	Unique variances			
	Single-level continuous	**Multilevel continuous**	**Single-level ordinal**	**Multilevel ordinal**
Item 1	0.57	0.59	0.50	0.41
Item 2	0.70	0.70	0.67	0.68
Item 3	0.78	0.79	0.76	0.73
Item 4	0.88	0.87	0.88	0.87
Item 5	0.60	0.61	0.57	0.51
Item 6	0.82	0.82	0.79	0.76
Item 7	0.84	0.84	0.83	0.82
Item 8	0.77	0.79	0.72	0.67
Item 9	0.91	0.91	0.87	0.85
Item 10	0.91	0.90	0.89	0.87
Item 11	0.65	0.64	0.60	0.60
Item 12	0.95	0.95	0.95	0.94
Item 13	0.87	0.86	0.81	0.82
Item 14	0.65	0.66	0.62	0.57
Item 15	0.70	0.70	0.69	0.73
Item 16	0.53	0.52	0.45	0.41
Item 17	0.66	0.66	0.58	0.55
Item 18	0.64	0.65	0.60	0.55
Item 19	0.85	0.83	0.83	0.78
Item 20	0.72	0.73	0.70	0.71
Item 21	0.79	0.82	0.74	0.69
Item 22	0.86	0.86	0.84	0.81
Item 23	0.70	0.69	0.58	0.62
Item 24	0.76	0.76	0.76	0.66
Item 25	0.55	0.55	0.50	0.57
Item 26	0.82	0.81	0.81	0.75
Item 27	0.48	0.48	0.47	0.38
Item 28	0.59	0.58	0.55	0.58
Item 29	0.74	0.74	0.69	0.69
Item 30	0.41	0.43	0.35	0.30

DISCUSSION

This study provides initial validity and dimensionality evidence for the ATPE, and introduces MCFA as an appropriate methodology to use when data are hierarchically structured. Primary attention is paid to description and discussion of almost all of the potential problems which may arise from conducting such types of research. Generally speaking, the presented approaches and procedures should be used in all quantitative research, where the multidimensional inventories using the Likert-type scales are distributes among measurement objects (e.g., students, pupils, employees, etc.) that are nested within organizations (e.g. classrooms, schools, companies, etc.). The present study discusses two main approaches to Likert-type scales. They can be treated either as continuous interval variables or ordered categorical variables with regard to the a-priori assumptions made by the researcher (Li et al., 1997). This study compares both approaches in both single-level and multilevel cases and describes their influence on validity, reliability, and dimensionality concerns.

The results offer some support for the proposed multilevel multidimensional conceptualization of ATPE at the within-group level (i.e., students within PE classes). On the other hand, the final structure of ATPE at the between-group level (i.e., PE classes) has only marginal empirical support. Results are interpreted to guide future research with the ATPE, to provide recommendations for revision to the instrument, and to assist researchers in physical education and kinanthropology in understanding when and how to apply MCFA to their data.

An analysis of the basic univariate statistical properties of the items indicated that the items distributions were negatively skewed and values of kurtosis fluctuated around zero. The reason for negative skewness of data can be found in the wording of the response categories. Students tended to respond more positively to positive statements and more negatively to negative statements. Subsequently, the scores obtained on negatively oriented items were rescored conversely. Hence, almost all means of the seven-point scales were higher than four. The tendency for preferring higher response categories was later confirmed by conducting CFA, where the existence of the two wording factors was strongly indicated. This finding is typical for attitude surveys using Likert-type scales with both types of item orientations (e.g., Albaum, 1997). All seven response categories were used in all items.

The bivariate distributional properties were examined from a variable type point of view. The continuous variable approach assumes, among other things, homoscedasticity and

linearity. These properties should be examined by inspecting scatterplots (e.g., Kline, 2005). However, it made little sense to draw scatterplots, because it consisted of 7x7 equally spaced points (see Figure 16). The ordinal variable approach revealed that point number 6 (and number 2 for negative statements) was the most favorite scale point. The underlying bivariate normality was confirmed using the threshold model with all values of RMSEA lower than 0.1, which is acceptable from a practical point of view (Jöreskog, 2005).

The reliability estimates were found to be close to the level of acceptability. The unidimensionality of each sub-construct was examined using one-factor models based on total sample covariance matrix $\boldsymbol{S}_T$. The model fit was acceptable within three of four subscales of ATPE (*health and fitness*, *emotional*, and *general*). The One-factor model fit the data only marginally in the remaining *social* sub-construct case. The unidimensionality is a necessary condition for both used reliability-like coefficients, alpha and omega. The values of these coefficients ranged from 0.68 to 0.78. The omega coefficient is always higher than alpha (McDonald, 1978a), nevertheless, the difference was within the 0.01 interval. Such a small difference is surprising. This similarity was probably due to the fact that the uniqueness variances accounted for the majority of the total variance and the absolute differences in factor loading estimates were not distinctive. Since the reliability-like coefficients can be considered as a function of factor loadings, uniqueness variance and number of items, a decision study about the number of items needed to achieve the desired level of reliability was conducted. According to some investigators, the reliability of Likert-type subscales should be at least 0.70 (e.g., Thompson, 2003). Other investigators suggest a value of 0.80 as the lowest acceptable level of reliability (e.g., Furr & Bacharach, 2007). The same number of items within each sub-construct of ATPE would not be necessary to achieve the same level of reliability. The lowest number of items would be needed in the *social* sub-construct case whereas the *health and fitness* domain would involve most of the items. If all sub-constructs had a reliability of 0.70, a total of 31 items would be included. To achieve the level of 0.80, the inventory would consist of 53 items. However, we must still keep in mind that the mentioned levels are only rough suggestions and an investigator himself should justify the desirable level of reliability that is appropriate for his research purpose (Brennan, 2001). From the ordinal variable point of view, the reliability would presumably be higher than in the continuous variable case. This increase in reliability is due to the well-known fact that the estimates of factor loadings are higher and unique variances are generally lower within the

ordinal variable approach (e.g., Olsson, 1979). Obviously, this fact was also confirmed in the present thesis (compare Tables 12 and 22).

There is reason to believe that the presented inventory may measure students' attitudes toward all four dimensions of physical education. The Wear's inventory was originally developed to measure ATPE at the within-group level (Wear, 1951; Wear 1955). Thus, all 10 proposed models at the within-group level (students) were guided by this theory. The four-factor model (model 2*a*) improved model fit substantially compared to the unididmensional "overall attitude" theory (model 1*a*). All four sub-constructs were found to be highly inter-correlated (from 0.72 to 0.90). The highest correlations were observed between *health and fitness* and *emotional* dimensions, and the *emotional* and *general* dimensions. On the contrary, the correlation was lowest between social and general dimensions. Such high rates in all correlations support the existence of the second-order overall attitude factor (model 3*a*). This model fits the data comparably well as the previous one. We may conclude that we have some evidence for accepting the first hypothesis, $\boldsymbol{H_1}$.

It was shown that four or five attitude-related factors were not enough to explain covariances among all items. Therefore, some modification of these models was necessary. The modification of the original model was already discussed by some authors (e.g., Wessel & Nelson, 1964; Blahuš, 1984). We have modified the theory by adding the *school PE* factor (model 4*a*). This was possible due to the fact that some of the items listed in the Appendix were slightly changed compared to Wear's original statements as a result of a content validation study. The improvement of model fit was also not negligible, but the presence of the *school PE* factor caused some problems in the estimation of the second-order model (model 5*a*). The reason for this complication can be explained by the lower values of correlations among four attitude-related factors. Furthermore, the same effect was observed, when two *wording* factors were included (all *b*-models). The presence of these two highly correlated (r = 0.8) factors improved fit substantially. The fit (model 4*b*) became excellent (RMSEA = 0.035, SRMR = 0.034, CFI = 0.98). However, the correlations among the four attitude-related factors decreased. They ranged from 0.32 to 0.82 and, thus, the second-order *overall attitude* factor was no longer empirically justified (model 5*b*). The *school PE* factor was treated as uncorrelated with all remaining factors except the *social* dimension of ATPE. This correlation is theoretically substantiated, because the attitude toward school PE depends to some extent on the social environment within a particular school (Koca, Asci & Demirhan, 2005). These two factors were moderately correlated.

None of the three hypothesized models exhibited acceptable fit at the between-group level (PE classes). The unidimensional model fit the data poorly and the remaining two proposed models improved the fit only slightly. Therefore, the EFA was conducted to find some reasonable factor structure at the between-group level. This strategy has also been used by some other researchers (e.g., Kuhlemeier et al., 2002). Based on the EFA results, a three-factor solution was chosen from all competing models. The factors were subsequently called the *importance*, *health*, and *utility* of physical education. The correlations among these three factors were about 0.3. However, this three-factor model also did not exhibit an acceptable fit, even though the fit improvement was apparent. Hence, it was concluded that no reasonable structure was found at the between-group level. This finding indicates that we have to reject the second hypothesis, $\boldsymbol{H_2}$. At this moment, the most fundamental question arises: *Does it really matter*? The answer to this question is connected with the basic purpose of the MCFA. The purpose of MCFA is to analyze the within-group structure, which is contaminated by the between-group variation. A clear distinction must be made between multilevel CFA and multi-group CFA (Hox, 2002; Muthén & Muthén, 2005; Kaplan, 2000). While the multi-group SEM is primarily aimed at the analysis of differences between groups, the multilevel SEM describes differences between individuals. If the measurement instrument was developed to measure attitudes exclusively at the individual level, the same set of items need not to be indicators of any concepts at the organization level. For this reason, we strongly suggest developing and using such items which would simultaneously indicate certain hypothesized concepts on both levels.

The degree to which all variables are confounded by the between-group variation was partially represented by the intraclass correlation, ICC. They ranged from 0.02 to 0.11 with a mean = 0.051 in the present study. Thus, the hypothesis $\boldsymbol{H_3}$ was closely confirmed. The ICCs are usually discussed in the majority of MCFA studies. For instance, they ranged from 0.01 to 0.07 in Kaplan's book (Kaplan, 2000); from 0.08 to 0.22 in Myers's work (Myers, 2005); from 0.02 to 0.22 in a paper by Cheung and Au (2005); and from 0.07 to 0.11 in a self-esteem study (Zimprich et al., 2005). It should be emphasized that conceptually the two-level SEM refers to the total covariance matrix. The ICCs provide information about the variances of all items and not about the covariances among them. Moreover, the multilevel approach to CFA is always appropriate for multilevel data regardless of the ICCs sizes. Only very low values of ICC make it also possible to use the single-level approach for multilevel data (Hox, 2002).

The present study shows that it is possible to analyze both within- and between-group levels either separately or simultaneously within the continuous variable framework. The first case involves estimating $\hat{\Sigma}_W$ and $\hat{\Sigma}_B$ matrices via the multivariate multilevel model (Raudenbush & Bryk, 2002) and conducting CFA for each matrix separately. This approach has two disadvantages. Namely, the estimates themselves are used for estimating factor models and thus the errors are cumulative. The second disadvantage can be found in the fact that the absolute values of covariances at both levels were not taken into account. This approach implicitly assumes that both levels have the same importance with respect to the magnitude of covariances. However the results showed that the covariances at the within-group level are about ten times higher than the covariances at the between-group level. In such cases, the simultaneous estimation of both levels seems to be more appropriate. This approach was originally proposed by Muthén (1994). The MUML estimation introduced earlier was used here. The results of this simultaneous analysis indicated one remarkable difference compared to the separate analyses of both matrices. The mentioned three-factor model at the between-group level is no longer entirely unacceptable. This model exhibited a marginally acceptable fit (SRMR = 0.07) when Muthén's methodology was used. Furthermore, the global two-level fit was excellent (RMSEA = 0.02, CFI = 0.95) while using this approach.

There are both differences and similarities between the results obtained within the ordinal variable and continuous variable frameworks. In the single-level case, the ordinal variable approach produces higher factor loadings but also higher standard errors, and lower unique variances. This finding was expected due to the well-known fact that correlations and factor loadings are attenuated when the continuous variable framework is used for ordinal variables (Babakus, Ferguson & Jöreskog, 1987; Olsson, 1979). The differences are not prominent, because the attenuation increases as the number of response categories decreases (Muthén, 1984) and seven categories used in our research is quite a high number. The correlations among all factors were similar as well.

The multilevel CFA for ordinal variable was the most problematic part of the present thesis. We were not able to estimate the whole two-level model as in the continuous variable counterpart. We had no big iron (computer) at our disposal that would be able to carry out the analysis using a numerical integration algorithm. Even when the much less computationally demanding Monte Carlo integration was used, the Mplus program did not converge and successfully terminated. Therefore, some simplification was required in terms of factor

number reduction. Hence, the model 2*a* at the within-group level and model 1 at the between-group level were simultaneously estimated. The results showed that unique variances are the lowest and correlations among attitude-related factors are the highest from all four mentioned approaches (i.e., single/multi level, ordinal/continuous variable). The factor loadings were sometimes higher but also sometimes lower than in the other three cases. Finally, the standard errors of factor loadings were higher than the results of both continuous variable approaches but lower than in the case of the single-level CFA for ordinal variables. Currently, theoretically oriented methodologists are working on the elaboration of the multilevel SEM for ordinal variables methodology to become more suitable for practical use (Grilli & Rampichini, 2007).

CONCLUSION

The results of the present thesis provide an improved basis for the further development of instruments designed to measure students' ATPE. However, the students' responses to the inventory items are always influenced by the PE class context. Thus, the relationships among items are distorted by the group-level variation. It seems essential that items designed entirely to measure individual characteristics don't have to refer to the group characteristics.

An introduction to MCFA is provided in this study. MCFA should be considered when individuals are meaningfully nested within groups and evaluation of the factor structure of a set of items is desired. It is not uncommon in kinanthropology to carry out empirical research, where data are collected from individuals (athletes, students, etc.) who are nested within organizations (schools, teams, etc.) and thus be interested in determining the structure of theoretical concepts and their indicators. Such research is usually conducted by PhD students or professional scientists in kinanthropology. We hope that this thesis will assist and serve them as a guideline for understanding when and how to apply MCFA to their data.

Finally, the following concluding remarks are addressed to these investigators:

- If the inventory items are designed to measure attitudes at the within-group level only, the simultaneous analysis of both $\boldsymbol{S}^*_B$ and $\boldsymbol{S}_{PW}$ matrices should be chosen. The amount of covariance at both levels is taken into account here. The fundamental purpose of this approach is to describe the structure of attitudes at the within-group level that is no longer contaminated by the between-level variability.
- If a scale-developer wishes to describe the structure of attitudes at both levels, the items should be designed to be considered as the indicators simultaneously of the within- as well as between-level factor structures. This means that structural hypotheses of theoretical concepts and their respective empirical indicators should be known a-priori. In such a case, both $\hat{\boldsymbol{\Sigma}}_W$ and $\hat{\boldsymbol{\Sigma}}_B$ matrices can be used for separate analysis within some SEM program.
- The MCFA for ordinal variables becomes computationally demanding with an increasing number of factors at each level. We strongly recommend treating the Likert-type scales as single-level ordinal or multilevel continuous if the total number of factors is higher than four. This temporary solution seems to be necessary until faster estimation algorithms are developed.

REFERENCES

Aiken, L.R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement, 40*, 955-959.

Aiken, L.R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement, 45*, 131-142.

Akaike, H. (1985). Prediction and entropy. In A. C. Atkinson & S. E. Feinberg (Eds.), A *celebration of statistics* (pp. 1-24). New York: Springer-Verlag.

Albarracín, D., Johnson, B. T., & Zanna, M. P. (Eds.). (2005). *The handbook of attitudes*. Mahwah, NJ: Lawrence Erlbaum Associates.

Albaum, G. (1997). The Likert scale revisited: An alternate version. *Journal of the Market Research Society, 39*, 331-349.

Alcohol Epidemiology and Monitoring (2008). Retrieved Dec 20, 2008 from http://www.who.int/substance_abuse/activities/gad/en/index.html

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology* (pp. 798-884). Worcester, MA: Clark University Press.

Arbuckle, J. L. (1997). *AMOS user's guide*. Chicago: SmallWaters.

Babakus, E., Ferguson, C. E., & Jöreskog, K. G.(1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional. assumptions. *Journal of Marketing Research, 24*, 222-228.

Ball, K., Timperio, A., Salmon, J., Giles-Corti, B., Roberts, R., & Crawford, D. (2007). Personal, social and environmental determinants of educational inequalities in walking: a multilevel study. *Journal of Epidemiology and Community Health, 61*, 108-114.

Barnett, T. A., Gauvin, L., Craig, C. L., Katzmarzyk, P. T. (2007). Modifying effects of sex, age, and education on 22-year trajectory of leisure-time physical activity. *Journal of Physical Activity and Health, 4*, 153-166.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis*. London: Arnold.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31*, 419-456.

Bentler, P. M. (1985). *Theory and implementation of EQS: A structural equations program*. Los Angeles: BMDP Statistical Software.

Bentler, M. P., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods and Research, 16*(1), 78-117.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246.

Blahuš, P. (1984). K analýze struktury postojů k tělesné výchově s použitím konfirmatorního faktorového modelu. [Analysis of the structure of attitudes toward physical education by confirmatory factor model.] *Teorie a praxe tělesné výchovy, 32*, 350-357.

Blahuš, P. (1985). *Faktorová analýza a její zobecnění*. Praha, SNTL.

Blahuš, P. (1992). A simple model for generalized confirmatory analysis with applications to educational research. *Acta Universitatis Carolinae Gymnica, 28*, 37-53.

Blahuš, P. (1993). Kinantropologie – vědní obor na Fakultě tělesné výchovy a sportu Univerzity Karlovy v Praze. In *Stav a perspektivy kinantropologie*. Praha: FTVS.

Blahuš, P. (1996a). Concept formation via latent variables modeling of motor abilities. *Kinesiology, 28*, 12-21.

Blahuš, P. (1996b). *K systémovému pojetí statistických metod v metodologii empirického výzkumu chování*. Praha: Karolinum.

Blahuš, P. (1999). Concept formation as generalized measurement. *Acta Universitatis Carolinae Kinanthropologica, 35* (2), 5-13.

Blahuš, P (2000). Measuring and modeling motor abilitiesas concept formation. In D. Milanovic (ed.), *Kinesiology for the twentieth century.* (pp. 43-50). Zagreb (Croatia), University of Zagreb Press.

Blahuš, P. (2004). On methodological aspects of building human movement science: Psychomotricity and kinanthropology. *Acta Universitatis Carolinae Kinanthropologica, 40* (2), 5-18.

Blahuš, P. (2008*). Introduction to methodology-based behavioral statistics, test theory, and structural modeling* (Manuscript in preparation).

Blahuš, P., Čelikovský, S., & Kovář, R. (1973). K nelineární faktorové analýze motorických schopností. [On non-linear factor analysis of motor abilities.] *Acta Universitatis Carolinae Gymnica, 9*, 31-44.

Bogardus, E. S. (1931). Attitudes and the Mexican immigrant. In K. Young (Ed.) *Social attitudes* (pp. 291-327). New York: Henry Holt.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., & Arminger, G. (1991). Observational residuals in factor analysis and structural equation models. *Sociological Methodology, 21*, 235-262.

Bonetti, S. (1994). On the use of student questionnaires. *Higher Education Review, 26*, 57-64.

Boomsma, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling, 7*, 461-483.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park (CA): SAGE Publications.

Bouchard C., Shephard R. (1994): Physical activity, fitness, and health: the model and key concepts. In C. Bouchard, R. Shephard, T. Stephens (Eds). *Physical activity, fitness, and health – Consensus statement.* (pp. 11-24). Champaign: Human Kinetic.

Brehm, J. W. (1956). Post-decision changes in desirability of alternatives. *Journal of Abnormal and Social Psychology*, 52, 384-389.

Brennan, R.L. (1992). *Elements of generalizability theory*. Iowa City, IA: American College Testing Program.

Brennan, R.L. (2001a). *Generalizability theory*. New York: Springer-Verlag.

Brennan, R. L. (2001b). *Manual for mGENOVA.* Iowa Testing Programs Occasional Papers Number 47.

Brennan, R.L., & Kane, M.T. (1977). An index of dependability for mastery tests. *Journal of Educational Measurement, 14,* 277-289.

Browne, M. W., Mels, G., & Cowan, M. (1994). *Path analysis: RAMONA: SYSTAT for DOS advanced applications.* Evanston, IL: SYSTAT.

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.

Cacioppo, J. T., & Sandman, C. A. (1981). Psychophysiological functioning, cognitive responding, and attitudes. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion* (pp. 81-103). Hillsdale, NJ: Lawrence Erlbaum Associates.

Čepička, L. (2004) Stepnicka's modification of the brace test: An investigation of dimensionality. *Perceptual and Motor Skills, 98,* 171-178.

Chambliss, H. O., Finley, C. E., & Blair, S. N. (2004). Attitudes toward obese individuals among exercise science students. *Medicine and Science in Sports and Exercise, 36,* 468-474.

Cheung, M. W. L., & Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. *Structural Equation Modeling, 12,* 598-619.

Chung, M., &, Phillips, D. (2002). The relationship between attitude toward physical education and leisure-time exercise in high school students. *The Physical Educator, 59,* 126-139.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

Cohen, J (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70,* 213-220.

Coleman, K.J., Epstein, L.H. (1998). Application of generalizability theory to measurement of activity in males who are not regularly active: a preliminary report. *Research Quarterly for Exercise and Sport, 69,* 58-63.

Crick, J.E., & Brennan, R.L. (1983). Manual for GENOVA: *A generalizability analysis of variance system (American College Testing Technical Bulletin No. 43).* Iowa City, IA: ACT.

Crocker, L., Llabre, M., & Miller, M.D. (1988). The generalizability of content validity ratings. *Journal of Educational Measurement, 23,* 287-299.

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley.

Cudeck, R., Du Toit, S., & Sörbom, D. (2001). *Structural equation models: Present and future. A festschrift in honor of Karl Jöreskog.* Chicago, IL: SSI.

Dexter, T. (1999). Relationships between sport knowledge , sport performance and academic ability: empirical evidence from GCSE physical education. *Journal of Sport Sciences, 17,* 283-295.

Dunn, J.G.H., Bouffard, M., & Rogers, W.T. (1999). Assessing content-relevance in sport psychology scale-construction research: Issues and recommendations. *Measurement in Physical Education and Exercise Science, 3*, 15-36.

Elferink-Gemser, M. T., Visscher, C., van Duijn, M. A., & Lemmink, K. A. (2006). Development of the interval endurance capacity in elite and sub-elite youth field hockey players. *British Journal of Sports Medicine, 40*, 340-345.

Fabrigar, L. R., MacDonald, T., & Wegener, D. T. (2005). The structure of attitudes. In D. Albarracin, B. Johnson, & M. Zanna (Eds.), *The handbook of attitudes*. (pp. 79-124). Mahwah, NJ: Lawrence Erlbaum Associates.

Farmer, G. L. (2000). Use of multilevel covariance structure analysis to evaluate the multilevel nature of theoretical constructs. *Social Work Research, 24*, 180-189.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fisher, K. J., Li, F., Michael, Y., & Cleveland, M. (2004). Neighborhood-level influences on physical activity among older adults: a multilevel analysis. *Journal of Aging and Physical Activity, 12*, 45-63.

Folsom, J. K. (1931). *Social psychology*. New York: Harper.

Fraser, C., & McDonald, R. P. (1988). COSAN: Covariance structure analysis. *Multivariate Behavioral Research, 23*, 263-265.

Frömel, K., Chmelík, F., Řepka, E., Šebrle, Z. et al. (2007). Pohybová aktivita České mládeže: koreláty intenzivní pohybové aktivity. [Physical activity in czech youth: corelates of intensive physical activity.] *Česká kinantropologie, 11* (4), 49-55.

Gill, L. D. (1997). Measurement, statistics, and research design issues in sport and exercise psychology. *Measurement in Physical Education and Exercise Science, 1,* 39-53.

Godbout, P., & Schutz, R.W. (1983). Generalizability of ratings of motor performances with reference to various observational designs. *Research Quarterly for Exercise and Sport, 54,* 20-27.

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

Goldstein, H., & Browne, W. (2005) Multilevel factor analysis models for continuous and discrete data. In A. Mayden-Olivares & McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald.* Mahwah (NJ): Lawrence Erlabaum Associates.

Goodwin, L.D. (2001). Interrater agreement and reliability. *Measurement in Physical Education and Exercise Science, 5*, 13-34.

Grilli, L., & Rampichini, C. (2007). Multilevel factor models for ordinal variables. *Structural Equation Modeling, 14*(1), 1-25.

Hagtvet, K. A. (1995). Using preconceptions in quantitative analysis of empirical data. In G. Imsen, E. Skaalvik & A.Oftedal Telhaug (Eds.) *Frontlinjer i pedagogisk forskning. [Frontlines in educational research]*. (pp.27-56). Trondheim, Norway: Norwegian Center for Child Research/Norwegian Research Council.

Hagtvet, K.A., & Hanin, Y. L. (2007). Consistency of performance-related emotions in elite athletes: Generalizability theory applied to the IZOF model. *Psychology of Sport and Exercise, 8*, 47-72.

Hartmann, W. M. (1992). *The CALIS procedure: Extended user's guide*. Cary, NC: SAS Institute.

Haynes, S.N., Richard, D.C., & Kubany, E.S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*, 238-247.

Hedges, L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist, 42*, 443-455.

Hicks, M. K., Wiggins, M. S., Crist, R. W., & Moode, F. M. (2001). Sex differences in grade three students' attitudes toward physical activity. *Perceptual and Motor Skills, 93*, 97-102.

Hodaň, B. (2004). Možnosti rozvíjení kinantropologie. [Possibilities for the evolution of kinanthropology.] *Česká kinantropologie, 8*, 53-64.

Hox, J. J. (2002). *Multilevel analysis. Techniques and applications.* Mahwah (NJ), Lawrence Erlbaum Associates.

Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling, 8*, 157-174.

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

Hutzler, Y. (2003). Attitudes toward the participation of individuals with disabilities in physical activity: A review. *Quest, 55*, 347-373.

Jackson, D.L., Teal, C.R., Raines, S.J., Nansel, T.R., Force, R.C., Burdsal, C.A. (1999). The dimensions of students' perceptions of teaching effectiveness. *Educational and Psychological Measurement, 59*, 580-596.

Jansa, P. (2002). Názory (postoje) a zájmy adolescentní mládeže o sport, tělesnou výchovu a jiné pohybové aktivity. [Opinions and attitudes of adolescent youth on sport and other movement activities.] *Česká kinantropologie, 6*, 23-39.

Javaras, K. N, & Ripley, B. D. (2007). An "unfolding" latent variable model for likert attitude data: Drawing inferences adjusted for response style. *Journal of the American Statistical Association, 102*, 454-463.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika, 57*, 239-251.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York: Seminar Press.

Jöreskog, K. G. (2005). *Structural equation modeling with ordinal variables using LISREL.* Retrieved October 2008 from the World Wide Web: *http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf*

Jöreskog, K. G., & Sörbom, D. (2005). *LISREL (Version 8.72)*. Lincolnwood, IL: Scientific Software International, Inc.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Newbury Park, CA: SAGE Publications.

Kenyon, G. S. (1968). Six scales for assessing attitude toward physical activity. *Research Quarterly for Exercise and Sport, 39*, 566-574.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd. ed.). New York: Guilford Press.

Koca, C., & Demirhan, G. (2004). An examination of high school students' attitudes toward physical education with regard to sex and sport participation. *Perceptual and Motor Skills, 98*, 754-758.

Koca, C., Asci, F. H., & Demirhan, G. (2005). Attitudes toward physical education and class preferences of Turkish adolescents in terms of school gender composition. *Adolecence, 40*, 365-375.

Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). Attitude measurement. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes*. (pp. 21-76). Mahwah, NJ: Lawrence Erlbaum Associates.

Kruglanski, A..W., & Stroebe, W. (2005). Attitudes, goals and beliefs: Issues of structure, function and dynamics. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes*. (pp. 323-368). Mahwah, NJ: Lawrence Erlbaum Associates.

Kuhlemeier, H., Bergh, H. & Rijlaarsdam, G. (2002). The dimensionality of speaking and writing: A multilevel factor analysis of situational, task and school effects. *British Journal of Educational Psychology. 72*, 467-482.

LaMonte, M. J., Ainsworth, B. E., & Reis, J. P. (2006). Measuring physical activity. In T. M. Wood, & W. Zhu (Eds.), *Measurement theory and practice in kinesiology*. (pp. 237 – 271). Champaign, IL: Human Kinetics.

Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563-575.

Li, F., Duncan, T. E., Duncan, S. C., Harmer, P. & Acock, A. (1997). Latent variable modelling of multilevel intrinsic motivation data. *Measurement in Physical Education and Exercise Science, 1*, 223-244.

Likert, R. (1932). A technique for measurement of attitudes. *Archives of Psychology, 140*, 1-55.

Long, J. S. (1983). *Confirmatory factor analysis: A preface to LISREL.* Beverly Hills, CA: Sage.

Longford, N. (1993). *Random coefficient models*. London: Sage.

Luke, D. A. (2004). *Multilevel modeling*. Sage University Papers Series on Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage.

Lundqvist, C., & Hassmén, P. (2005). *Competitive State Anxiety Inventory -2* (CSAI-2): Evaluating the Swedish version by confirmatory factor analyses. *Journal of Sports Sciences, 23*, 727-736.

Lynn, M.R. (1986). Determination and quantification of content validity. *Nursing Research, 35*, 382-385.

Martin, R. J., Dore, E., Twisk, J., van Praagh, E., Hautier, C. A., & Bedu, M. (2004). Longitudinal changes of maximal short-term peak power in girls and boys during growth. *Medicine and Science in Sports and Exercise, 36*, 498-503.

Mâsse, L.C., Heesch, K.C., Fulton, J.E., & Watson, K.B. (2002). Raters'objectivity in using the compendium of physical activities to codephysical activity diaries. *Measurement in Physical Education and Exercise Science, 6*, 207-224.

McDonald, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika, 27*(4), 397-415.

McDonald, R. P. (1978*a*). Generalizability in factorable domains: "Domain validity and generalizability". *Educational and Psychological Measurement, 38*, 75-79.

McDonald, R. P. (1978*b*). A simple comprehensive model for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 31*, 59-72.

McDonald, R. P. (1985). *Factor Analysis and Related Methods.* Hillsdale, NY: Lawrence Erlbaum.

McDonald, R. P. & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology, 42*, 215-232.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin, 707*(2), 247-255.

McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods & Research, 22*, 399-413.

McDonald, R. P. (1999). *Test theory. A unified treatment.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McDonald, R. P., & Ho, M. R. H. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*, 64-82.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Milgram, S., Mann, L., & Harter, S. (1965). The lost-letter technique: A tool of social research. *Public Opinion Quarterly, 29*, 437-438.

Miller, J. L., & Cordova, J. (2002). Changing attitudes toward people with disabilities. *Palaestra, 18*, 16-21.

Moustaki, I., Jöreskog, K. G., & Mavridis, D. (2004). Factor models for ordinal variables with covariate effects on the manifest and latent variables: A comparison of LISREL and IRT approaches. *Structural Equation Modeling, 11*(4), 487-513.

Mowatt, M., DePauw, K. P., & Hulac, G. M. (1988). Attitude toward physical activity among college students. *The Physical Educator, 45*, 103-108.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika, 49*, 115-132.

Muthén, B. O. (1990). *Means and covariance structure analysis of hierarchical data.* Los Angeles: UCLA Statistics series.

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

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

Muthén, B.O. & Satorra, A. (1995). Complex sample data in structural equation modeling. In P.V. Marsden (Ed.), *Sociological Methodology*, 267-316. Oxford: Blackwell.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29* (1), 81-117.

Muthén, L. K., & Muthén, B. O. (2005). Mplus user's guide. Third edition. Los Angeles, CA: Muthén & Muthén.

Myers, N. D. (2005). *Athletes' evaluations of their head coach's coaching competencies: A multilevel confirmatory factor analysis.* PhD Thesis, Michigan State University, Michigan.

Myers, J. L. (1972). *Fundamentals of experimental design.* Boston: Allyn and Bacon, Inc.

Neale, M. C. (1997). *Mx: Statistical modeling* (4th ed.). Richmond, VA: Author.

Norcini, J., & Grosso, L. (1998). The generalizability of ratings of item relevance. *Applied Measurement in Education, 11*, 301-309.

Olson, J. M., & Stone, J. (2005). The influence of behavior on attitudes. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes.* (pp. 223-271). Mahwah, NJ: Lawrence Erlbaum Associates.

Olsson, U. (1979): Maximum likelihood estimation of the polychoric correlation coefficient, *Psychometrika, 44*, 443–460.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana, IL: University of Illinois Press.

Pecha, O. (2004). Víceúrovňová analýza hierarchicky strukturovaných dat v kinantropologickém výzkumu. [Multilevel analysis of hierarchically structured data in kinanthropological research.] *Česká kinantropologie, 8*, 65-80.

Pecha, O. (2005). K teorii postojů v kinantropologii. [On attitudes theory in kinanthropology.] *Česká kinantropologie, 9*, 61-75.

Pecha, O. (2006). Multilevel extension of the factor analysis: Important aspects for physical educational research. *Acta Universitatis Carolinae Kinanthropologica, 42*(1), 125-139.

Penfield, R.D., & Giacobbi, P.R. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science, 8*, 213-225.

Petty, R. E., & Cacioppo, J. T. (1983). The role of bodily responses in attitude measurement and change. In J. T. Cacioppo & R. E. Petty (Eds.), *Social psychophysiology: A sourcebook* (pp. 51-101). New York: Guilford.

Petty, R. E., & Krosnick, J. A. (Eds.) (1995). *Attitude strength: Antecedents and consequences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.

Polit, D.F., Beck, C.T., & Owen, S.V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*, 459-467.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics, 128*, 301-323.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and data analysis methods* (2nd Edition). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congon, R., & duToit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International, Inc.

Rice, P. L. (1988). Attitude of high school students toward physical education activities, teachers, and personal health. *The Physical Educator, 45*, 94-99.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351-357.

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In M. Rosenberg, C. Hovland, W. McGuire, R. Abelson, & J. Brehm (Eds.), *Attitude organization and change* (pp. 1-14). New Haven, CT: Yale University Press.

Ryan, S., Fleming, D., & Maina, M. (2003). Attitudes of middle school students towards. their physical education teachers and activities. *The Physical Educator, 60* (2), 28-42.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors on covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 399-419). Thousand Oaks, CA: Sage.

Schimmack, U. & Crites, S. L., Jr. (2005). The structure of affect. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes*. (pp. 397-435). Mahwah, NJ: Lawrence Erlbaum Associates.

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99* (6), 323-337.

Schutz, R., Smoll, E, Carre, F., & Mosher, R. (1985). Inventories and norms for children's attitudes toward physical activity. *Research Quarterly for Exercise and Sport, 56*, 256-265.

Shavelson, R.J., & Webb, N.M. (1991). *Generalizability theory. A primer*. London, UK: SAGE.

Shavelson, R.J., Webb, N.M., & Rowley, G.L. (1989). Generalizability theory. *American Psychologist, 6*, 922-932.

Sidney, K. H., & Shephard, R. J. (1976). Attitudes towards health and physical activity in the elderly. Effects of a physical training program. *Medicine and Science in Sports and Exercise, 8*, 246-252.

Simon, J. A., & Smoll, F. L. (1974). An instrument for assessing children's attitudes toward physical activity. *Research Quarterly for Exercise and Sport, 45*, 407-415.

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

Slepička, P., & Slepičková, I. (2002). Sport z pohledu české společnosti. [Sport from a point of view of the czech society.] *Česká kinantropologie, 6*, 7-22.

Smoll, F. L., & Schutz, R. W. (1980). Children's attitudes toward physical activity: A longitudinal analysis. *Journal of Sport Psychology, 2*, 137-147.

Smoll, F., Schutz, R., & Keeney, J. (1976). Relationships among children's attitudes, involvement, and proficiency in physical activities. *Research Quarterly for Exercise and Sport, 47*, 797-803.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis.* London: Sage.

Song, X. Y., & Lee, S. Y. (2002). A bayesian approach for multigroup nonlinear factor analysis. *Structural Equation Modeling, 9*(4), 523-553.

Stegner, A.J., Tobar, D.A, & Kane, M.T. (1999). Generalizability of change scores on the body awareness scale. *Measurement in Physical Education and Exercise Science, 3*, 125-140.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.

Steiger, J. H. (1995). *SEPATH: STATISTICA version 5*. Tulsa: StatSoft, Inc.

Štochl, J., Čechovská, I., & Pavelková, V. (2005). Evaluation of the scale of swimming skills for preschool children. *Acta Universitatis Carolinae Kinanthropologica, 41*, 87-98.

Svoboda, B (1971). Výzkum postojů sportovního trenéra. [Research of sport coaches´ attitudes.] *Acta Universitatis Carolinae Gymnica, 2*, 65-73.

Tannenbaum, C., Ahmed, S, & Mayo, N. (2007). What drives older women's perceptions *of* health-related *quality of life*? *Quality of Life Research, 16*, 593-605.

The European School Survey Project on Alcohol and Other Drugs. (2007). Retrieved Dec 20, 2008 from http://www.espad.org/documents/Espad/Documents/ESPAD_Questionnaire_2007.pdf

Thomas, J. R., & Nelson, J. K. (2001). *Research methods in physical activity* (4th ed.). Champaign, IL: Human Kinetics.

Thomas, W. I., & Znaniecki, R. (1918). *The polish peasant in Europe and America: Monograph of an immigrant group.* Boston: Badger.

Thompson, B. (Ed.) (2003). *Score reliability: Contemporary thinking on reliability issues*. Thousand Oaks, CA: Sage Publications.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*, 529-544.

Tobar, D.A, Stegner, A.J., & Kane, M.T. (1999). The use of generalizability theory in examining the dependability of scores on the profile of mood states. *Measurement in Physical Education and Exercise Science, 3*, 141-156.

Turner, A.L., & Bouffard, M. (1998). Comparison of modified to standard bioelectrical impedance errors using generalizability theory. *Measurement in Physical Education and Exercise Science, 2*, 177-196.

Turner, A.L., Lozano-Nieto, A., & Bouffard, M. (2003). Generalizability of bioimpedance spectroscopy measures. Measurement *in Physical Education and Exercise Science, 7*, 129-148.

Ullman, J. B. (2006). Structural equation modeling: Review the basics and moving forward. *Journal of Personality Assessment, 87*, 35-50.

Wear, C. L. (1951). The evaluation of attitudes toward physical education as an activity course. *Research Quarterly for Exercise and Sport, 22*, 114-126.

Wear, C. L. (1955). Construction of equivalent forms of an attitude scales. *Research Quarterly for Exercise and Sport, 26*, 113-119.

Wendel-Vos, G. C., Schuit, A. J., de Niet, R., Boshuizen, H. C., Saris, W. H., & Kromhout, D. (2004). Factors of the physical environment associated with walking and bicycling. *Medicine and Science in Sports and Exercise, 36*, 725-730.

Wessel, J. A., & Nelson, R. (1964). Relationship between strength and attitudes toward physical education activity among college women. *Research Quarterly for Exercise and Sport, 35*, 562-569.

Westie, F. R. (1953). A technique for the measurement of race attitudes. *American Sociological Review, 18*, 73-78.

Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 69-83). New York: Seminar Press.

Wyer, R. S., & Albarracín, D. (2005). Belief formation, organization and change: Cognitive and motivational influences. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes*. (pp. 273-322). Mahwah, NJ: Lawrence Erlbaum Associates.

Wynd, C.A., Schmidt, B., & Schaefer, M.A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research, 25*, 508-518.

Young, M. L. (1970). Personal – social adjustment, physical fitness, attitude toward physical education of high school students given by socioeconomic level. *Research Quarterly for Exercise and Sport, 41*, 593-603.

Zhu, W. (1997). A multilevel analysis of school factors associated with health-related fitness. *Research Quarterly for Exercise and Sport, 68,* 125-135.

Zhu, W., & Erbaugh, S. (1997). Assesing Change in Swimming Skills Using the Hierarchical Linear Model. *Measurement in Physical Education and Exercise Science, 1*, 179-201.

Zimprich, D., Perren, S. & Hornung, R. (2005). A two-level confirmatory factor analysis of a modified Rosenberg Self-esteem Scale. *Educational and Psychological Measurement, 65*, 565-481.

APPENDICES

INVENTORY ITEMS

1. If for any reason a few subjects have to be dropped from the school program, physical education should be one of the subjects dropped.
2. Associations in physical education activities give people a better understanding of each other.
3. Physical education activities during school physical education class activate negative emotions.
4. Exercise and sport help people to establish and maintain desirable social standards.
5. The time spent in getting ready for and engaging in physical education activities could be more profitably spent in other ways.
6. Vigorous physical activity works off harmful emotional tensions.
7. A person's body usually has all the strength it needs without participation in physical education activities.
8. I would take physical education in our school only if it were required.
9. Participation in physical education activities and exercise makes no contribution to the development of poise.
10. Skill in active games or sports is not necessary for leading the fullest kind of life.
11. Because physical skills loom large in importance is essential that a person be helped to acquire and improve such skills.
12. Participation in sport and exercise tends to make one a more socially desirable person.
13. Calisthenics (setting-up exercises, stretching, yoga, etc.) taken regularly are good for one's general health.
14. Physical education classes in our school provide nothing which will be of value outside of the class.
15. Associating with others in some physical education activity and exercise is fun.
16. A person would be better of emotionally if he/she did not participate in exercise.
17. There is not enough value coming from physical educational activities to justify the leisure time consumed.
18. Physical education in our school does more harm emotionally than it does good.
19. Physical education classes in our school provide situations for the formation of attitudes and taking-over new opinions.
20. In our school, physical education situations are among the poorest for making friends.
21. There should not be over two one-hour periods per week devoted to physical education in schools.
22. Physical education activities tend to upset a person emotionally.
23. Exercise and physical education activities help a person gain and maintain all-round good health.
24. Active participation in physical educational classes in our school makes any apparent effect on person's health.
25. Belonging to a group, for which is opportunity is provided in team activities, is a desirable social experience for a person.
26. People get all the physical exercise they need in just taking care of their daily work.
27. For its contributions to mental and emotional well-being physical education should be included in the program of every school.
28. Engaging in group physical education activities and exercise is desirable for proper personality development.
29. All who are physically able will profit from an hour of physical activity each day.
30. As far as improving physical health is concerned a physical education class in our school is a waste of time.

LISREL SYNTAX FILES

Single-level CFA

Model 1a

```
TITTLE = model 1a
da ni =30 no=1157 ma=cm
cm fi=C:\total.cov
mo ny=30 ne=1 ly=fu,fr ps=sy,fi te=di,fr
va 1 ps 1 1
pd

ou rs se ss sc tv ad=off nd=4
```

Model 1b

```
TITTLE = model 1b
da ni =30 no=1157 ma=cm
cm fi=C:\total.cov
mo ny=30 ne=3 ly=fu,fi ps=sy,fi te=di,fr
fr ly 1 1 ly 2 1 ly 3 1 ly  4 1 ly 5 1 ly 6 1 ly 7 1 ly 8 1 ly 9 1 ly 10 1
fr ly 11 1 ly 12 1 ly 13 1 ly  14 1 ly 15 1 ly 16 1 ly 17 1 ly 18 1 ly 19 1 ly 20 1
fr ly 21 1 ly 22 1 ly 23 1 ly  24 1 ly 25 1 ly 26 1 ly 27 1 ly 28 1 ly 29 1 ly 30 1
fr ly 2 2 ly 4 2 ly 6 2 ly 11 2 ly 12 2 ly 13 2 ly 15 2 ly 19 2 ly 23 2 ly 25 2 ly 27 2 ly 28 2 ly 29 2
fr ly 1 3 ly 3 3 ly 5 3 ly 7 3 ly 8 3 ly 9 3 ly 10 3 ly 14 3 ly 16 3 ly 17 3 ly 18 3 ly 20 3 ly 21 3 ly 22 3 ly 24 3 ly
26 3 ly 30 3
va 1 ps 1 1 ps 2 2 ps 3 3
fr ps 3 2
pd

ou rs se ss sc tv ad=off nd=4
```

Model 2a

```
TITTLE = model 2a
da ni =30 no=1157 ma=cm
cm fi=C:\total.cov
mo ny=30 ne=4 ly=fu,fi ps=sy,fr te=di,fr
va 1 ly 7 1 ly 2 2 ly 3 3 ly 1 4
fr ly 9 1 ly 11 1 ly 13 1 ly 17 1 ly 23 1 ly 24 1 ly 29 1
fr ly 4 2 ly 12 2 ly 15 2 ly 20 2 ly 25 2 ly 28 2
fr ly 6 3 ly 16 3 ly 18 3 ly 22 3 ly 27 3
fr ly 5 4 ly 8 4 ly 10 4 ly 14 4 ly 19 4 ly 21 4  ly 26 4 ly 30 4
pd

ou rs se ss sc tv ad=off nd=4
```

Model 2b

```
TITTLE = model 2b
da ni =30 no=1157 ma=cm
cm fi=C:\total.cov
mo ny=30 ne=6 ly=fu,fi ps=sy,fi te=di,fr
```

fr ly 7 1 ly 9 1 ly 11 1 ly 13 1 ly 17 1 ly 23 1 ly 24 1 ly 29 1
fr ly 2 2 ly 4 2 ly 12 2 ly 15 2 ly 20 2 ly 25 2 ly 28 2
fr ly 3 3 ly 6 3 ly 16 3 ly 18 3 ly 22 3 ly 27 3
fr ly 1 4 ly 5 4 ly 8 4 ly 10 4 ly 14 4 ly 19 4 ly 21 4 ly 26 4 ly 30 4
fr ly 2 5 ly 4 5 ly 6 5 ly 11 5 ly 12 5 ly 13 5 ly 15 5 ly 19 5 ly 23 5 ly 25 5 ly 27 5 ly 28 5 ly 29 5
fr ly 1 6 ly 3 6 ly 5 6 ly 7 6 ly 8 6 ly 9 6 ly 10 6 ly 14 6 ly 16 6 ly 17 6 ly 18 6 ly 20 6 ly 21 6 ly 22 6 ly 24 6 ly 26 6 ly 30 6
va 1 ps 1 1 ps 2 2 ps 3 3 ps 4 4 ps 5 5 ps 6 6
fr ps 2 1 ps 3 1 ps 4 1 ps 3 2 ps 4 2 ps 4 3 ps 5 4
pd

ou rs se ss sc tv ad=off nd=4

Model 3a

TITTLE = model 3a
da ni =30 no=1157 ma=cm
cm fi=C:\total.cov
mo ny=30 ne=5 ly=fu,fi ps=di,fr be=fu,fi te=di,fr
va 1 ly 7 1 ly 2 2 ly 3 3 ly 1 4
fr ly 9 1 ly 11 1 ly 13 1 ly 17 1 ly 23 1 ly 24 1 ly 29 1
fr ly 4 2 ly 12 2 ly 15 2 ly 20 2 ly 25 2 ly 28 2
fr ly 6 3 ly 16 3 ly 18 3 ly 22 3 ly 27 3
fr ly 5 4 ly 8 4 ly 10 4 ly 14 4 ly 19 4 ly 21 4 ly 26 4 ly 30 4
fr be 1 5 be 2 5 be 3 5 be 4 5
pd
ou rs se ss sc tv ad=off nd=4

Model 3b

TITTLE = model 3b
da ni =30 no=1157 ma=cm
cm fi=C:\total.cov
mo ny=30 ne=7 ly=fu,fi ps=sy,fi be=fu,fi te=di,fr
fr ly 7 1 ly 9 1 ly 11 1 ly 13 1 ly 17 1 ly 23 1 ly 24 1 ly 29 1
fr ly 2 2 ly 4 2 ly 12 2 ly 15 2 ly 20 2 ly 25 2 ly 28 2
fr ly 3 3 ly 6 3 ly 16 3 ly 18 3 ly 22 3 ly 27 3
fr ly 1 4 ly 5 4 ly 8 4 ly 10 4 ly 14 4 ly 19 4 ly 21 4 ly 26 4 ly 30 4
fr be 1 5 be 2 5 be 3 5 be 4 5
fr ly 2 7 ly 4 7 ly 6 7 ly 11 7 ly 12 7 ly 13 7 ly 15 7 ly 19 7 ly 23 7 ly 25 7 ly 27 7 ly 28 7 ly 29 7
fr ly 1 6 ly 3 6 ly 5 6 ly 7 6 ly 8 6 ly 9 6 ly 10 6 ly 14 6 ly 16 6 ly 17 6 ly 18 6 ly 20 6 ly 21 6 ly 22 6 ly 24 6 ly 26 6 ly 30 6
va 1 ps 1 1 ps 2 2 ps 3 3 ps 4 4 ps 5 5 ps 6 6 ps 7 7
fr ps 7 6
pd

ou rs se ss sc tv ad=off nd=4

Model 4a

TITTLE = model 4a
da ni =30 no=1157 ma=cm
cm fi=C:\total.cov
mo ny=30 ne=5 ly=fu,fi ps=sy,fi te=di,fr
va 1 ly 7 1 ly 2 2 ly 3 3 ly 1 4
fr ly 7 1 ly 9 1 ly 11 1 ly 13 1 ly 17 1 ly 23 1 ly 24 1 ly 29 1
fr ly 2 2 ly 4 2 ly 12 2 ly 15 2 ly 20 2 ly 25 2 ly 28 2

fr ly 3 3 ly 6 3 ly 16 3 ly 18 3 ly 22 3 ly 27 3
fr ly 1 4 ly 5 4 ly 8 4 ly 10 4 ly 14 4 ly 19 4 ly 21 4 ly 26 4 ly 30 4
fr ly 1 5 ly 3 5 ly 8 5 ly 11 5 ly 14 5 ly 18 5 ly 19 5 ly 20 5 ly 21 5 ly 24 5 ly 27 5 ly 30 5
va 1 ps 1 1 ps 2 2 ps 3 3 ps 4 4 ps 5 5
fr ps 2 1 ps 3 1 ps 4 1 ps 3 2 ps 4 2 ps 4 3
pd
ou rs se ss sc tv ad=off nd=4

Model 4b

TITTLE = model 4b
da ni =30 no=1157 ma=cm
cm fi=C:\total.cov
mo ny=30 ne=7 ly=fu,fi ps=sy,fi te=di,fr
fr ly 7 1 ly 9 1 ly 11 1 ly 13 1 ly 17 1 ly 23 1 ly 24 1 ly 29 1
fr ly 2 2 ly 4 2 ly 12 2 ly 15 2 ly 20 2 ly 25 2 ly 28 2
fr ly 3 3 ly 6 3 ly 16 3 ly 18 3 ly 22 3 ly 27 3
fr ly 1 4 ly 5 4 ly 8 4 ly 10 4 ly 14 4 ly 19 4 ly 21 4 ly 26 4 ly 30 4
fr ly 1 5 ly 3 5 ly 8 5 ly 11 5 ly 14 5 ly 18 5 ly 19 5 ly 20 5 ly 21 5 ly 24 5 ly 27 5 ly 30 5
fr ly 2 6 ly 4 6 ly 6 6 ly 11 6 ly 12 6 ly 13 6 ly 15 6 ly 19 6 ly 23 6 ly 25 6 ly 27 6 ly 28 6 ly 29 6
fr ly 1 7 ly 3 7 ly 5 7 ly 7 7 ly 8 7 ly 9 7 ly 10 7 ly 14 7 ly 16 7 ly 17 7 ly 18 7 ly 20 7 ly 21 7 ly 22 7 ly 24 7 ly 26 7 ly 30 7
va 1 ps 1 1 ps 2 2 ps 3 3 ps 4 4 ps 5 5 ps 6 6 ps 7 7
fr ps 2 1 ps 3 1 ps 4 1 ps 3 2 ps 4 2 ps 4 3 ps 7 6
pd
ou rs se ss sc tv ad=off nd=4

Model 5a

TITTLE = model 5a
da ni =30 no=1157 ma=cm
cm fi=C:\total.cov
mo ny=30 ne=6 ly=fu,fi ps=di,fi be=fu,fi te=di,fr
fr ly 7 1 ly 9 1 ly 11 1 ly 13 1 ly 17 1 ly 23 1 ly 24 1 ly 29 1
fr ly 2 2 ly 4 2 ly 12 2 ly 15 2 ly 20 2 ly 25 2 ly 28 2
fr ly 3 3 ly 6 3 ly 16 3 ly 18 3 ly 22 3 ly 27 3
fr ly 1 4 ly 5 4 ly 8 4 ly 10 4 ly 14 4 ly 19 4 ly 21 4 ly 26 4 ly 30 4
fr ly 1 5 ly 3 5 ly 8 5 ly 11 5 ly 14 5 ly 18 5 ly 19 5 ly 20 5 ly 21 5 ly 24 5 ly 27 5 ly 30 5
!fr ly 2 6 ly 4 6 ly 6 6 ly 11 6 ly 12 6 ly 13 6 ly 15 6 ly 19 6 ly 23 6 ly 25 6 ly 27 6 ly 28 6 ly 29 6
!fr ly 1 7 ly 3 7 ly 5 7 ly 7 7 ly 8 7 ly 9 7 ly 10 7 ly 14 7 ly 16 7 ly 17 7 ly 18 7 ly 20 7 ly 21 7 ly 22 7 ly 24 7 ly 26 7 ly 30 7
fr ps 1 1 ps 2 2 ps 3 3 ps 4 4
va 1 ps 5 5 ps 6 6
fr be 1 6 be 2 6 be 3 6 be 4 6
pd
ou rs se ss sc tv ad=off nd=4

Model 5b

TITTLE = model 5b
da ni =30 no=1157 ma=cm
cm fi=C:\total.cov
mo ny=30 ne=8 ly=fu,fi ps=di,fi be=fu,fi te=di,fr
fr ly 7 1 ly 9 1 ly 11 1 ly 13 1 ly 17 1 ly 23 1 ly 24 1 ly 29 1
fr ly 2 2 ly 4 2 ly 12 2 ly 15 2 ly 20 2 ly 25 2 ly 28 2
fr ly 3 3 ly 6 3 ly 16 3 ly 18 3 ly 22 3 ly 27 3

```
fr ly 1 4 ly 5 4 ly 8 4 ly 10 4 ly 14 4 ly 19 4 ly 21 4  ly 26 4 ly 30 4
fr ly 1 5 ly 3 5 ly 8 5 ly 11 5 ly 14 5 ly 18 5 ly 19 5 ly 20 5 ly 21 5 ly 24 5 ly 27 5 ly 30 5
fr ly 2 6 ly 4 6 ly 6 6 ly 11 6 ly 12 6 ly 13 6 ly 15 6 ly 19 6 ly 23 6 ly 25 6 ly 27 6 ly 28 6 ly 29 6
fr ly 1 7 ly 3 7 ly 5 7 ly 7 7 ly 8 7 ly 9 7 ly 10 7 ly 14 7 ly 16 7 ly 17 7 ly 18 7 ly 20 7 ly 21 7 ly 22 7 ly 24 7 ly
26 7 ly 30 7
fr ps 1 1 ps 2 2 ps 3 3 ps 4 4
va 1 ps 5 5 ps 6 6 ps 7 7 ps 8 8
fr be 1 8 be 2 8 be 3 8 be 4 8
pd
ou rs se ss sc tv ad=off nd=4
```

Model 4b - ordinal variables

```
TITTLE = model 4b ordinal
!da ni =30 no=1157 ma=cm
SY=C:\ordinal.dsf
AC=C:\ordinal.ACM
mo ny=30 ne=7 ly=fu,fi ps=sy,fi be=fu,fi te=di,fr
fr ly 7 1 ly 9 1 ly 11 1 ly 13 1 ly 17 1 ly 23 1 ly 24 1 ly 29 1
fr ly 2 2 ly 4 2 ly 12 2 ly 15 2 ly 20 2 ly 25 2 ly 28 2
fr ly 3 3 ly 6 3 ly 16 3 ly 18 3 ly 22 3 ly 27 3
fr ly 1 4 ly 5 4 ly 8 4 ly 10 4 ly 14 4 ly 19 4 ly 21 4  ly 26 4 ly 30 4
fr ly 1 5 ly 3 5 ly 8 5 ly 11 5 ly 14 5 ly 18 5 ly 19 5 ly 20 5 ly 21 5 ly 24 5 ly 27 5 ly 30 5
fr ly 2 6 ly 4 6 ly 6 6 ly 11 6 ly 12 6 ly 13 6 ly 15 6 ly 19 6 ly 23 6 ly 25 6 ly 27 6 ly 28 6 ly 29 6
fr ly 1 7 ly 3 7 ly 5 7 ly 7 7 ly 8 7 ly 9 7 ly 10 7 ly 14 7 ly 16 7 ly 17 7 ly 18 7 ly 20 7 ly 21 7 ly 22 7 ly 24 7 ly
26 7 ly 30 7
va 1 ps 1 1 ps 2 2 ps 3 3 ps 4 4 ps 5 5 ps 6 6 ps 7 7
fr ps 2 1 ps 3 1 ps 4 1 ps 3 2 ps 4 2 ps 4 3 ps 7 6
pd
ou me=mlr rs se ss sc tv mi ad=off nd=4
```

Within-group CFA

Modification of model 4b

```
TITTLE = modification of model 4b, within-group level
da ni =30 no=1070 ma=cm
cm fi=C:\vysledky_within.cov
mo ny=30 ne=7 ly=fu,fi ps=sy,fi te=di,fr
fr ly 7 1 ly 9 1 ly 11 1 ly 13 1 ly 17 1 ly 23 1 ly 24 1 ly 29 1
fr ly 2 2 ly 4 2 ly 12 2 ly 15 2 ly 20 2 ly 25 2 ly 28 2
fr ly 3 3 ly 6 3 ly 16 3 ly 18 3 ly 22 3 ly 27 3
fr ly 1 4 ly 5 4 ly 8 4 ly 10 4 ly 14 4 ly 19 4 ly 21 4  ly 26 4 ly 30 4
fr ly 1 5 ly 3 5 ly 8 5 ly 11 5 ly 14 5 ly 18 5 ly 19 5 ly 20 5 ly 21 5 ly 24 5 ly 27 5 ly 30 5
fr ly 2 6 ly 4 6 ly 6 6 ly 11 6 ly 12 6 ly 13 6 ly 15 6 ly 19 6 ly 23 6 ly 25 6 ly 27 6 ly 28 6 ly 29 6
fr ly 1 7 ly 3 7 ly 5 7 ly 7 7 ly 8 7 ly 9 7 ly 10 7 ly 14 7 ly 16 7 ly 17 7 ly 18 7 ly 20 7 ly 21 7 ly 22 7 ly 24 7 ly
26 7 ly 30 7
va 1 ps 1 1 ps 2 2 ps 3 3 ps 4 4 ps 5 5 ps 6 6 ps 7 7
fr ps 2 1 ps 3 1 ps 4 1 ps 3 2 ps 4 2 ps 4 3 ps 7 6 ps 5 2
pd
ou rs se ss sc tv mi ad=off nd=4
```

Between-group CFA

Model 1

```
TITTLE = model 1
da ni =30 no=87 ma=cm
cm fi=C:\vysledky_between.cov
mo ny=30 ne=1 ly=fu,fr ps=sy,fi te=di,fr
va 1 ps 1 1
pd
ou rs se ss sc tv ad=off nd=4
```

Model 2

```
TITTLE = model 2
da ni =30 no=87 ma=cm
cm fi=C:\vysledky_between.cov
mo ny=30 ne=2 ly=fu,fi ps=sy,fi te=di,fr
fr ly 2 1 ly 4 1 ly 5 1 ly 6 1 ly 7 1 ly 9 1 ly 10 1 ly 12 1 ly 13 1 ly 15 1 ly 16 1 ly 17 1 ly 22 1 ly 23 1 ly 25 1 ly
26 1 ly 28 1 ly 29 1
fr ly 1 2 ly 3 2 ly 8 2 ly 11 2 ly 14 2 ly 18 2 ly 19 2 ly 20 2 ly 21 2 ly 24 2 ly 27 2 ly 30 2
va 1 ps 1 1 ps 2 2
fr ps 2 1
pd
ou rs ss sc mi ad=off nd=4
```

Model 3

```
TITTLE = model 2
da ni =30 no=87 ma=cm
cm fi=C:\vysledky_between.cov
mo ny=30 ne=2 ly=fu,fi ps=di,fi te=di,fr
fr ly 1 1 ly 2 1 ly 3 1 ly 4 1 ly 5 1 ly 6 1 ly 7 1 ly 9 1 ly 10 1 ly 12 1 ly 13 1 ly 15 1 ly 16 1 ly 17 1 ly 22 1 ly 23 1
ly 25 1 ly 26 1 ly 28 1 ly 29 1 ly 8 1 ly 11 1 ly 14 1 ly 18 1 ly 19 1 ly 20 1 ly 21 1 ly 24 1 ly 27 1 ly 30 1
fr ly 1 2 ly 3 2 ly 8 2 ly 11 2 ly 14 2 ly 18 2 ly 19 2 ly 20 2 ly 21 2 ly 24 2 ly 27 2 ly 30 2
va 1 ps 1 1 ps 2 2
pd
ou rs mi ad=off nd=4
```

Three-factor CFA model

```
TITTLE = model 2
da ni =30 no=87 ma=cm
cm fi=C:\vysledky_between.cov
mo ny=30 ne=3 ly=fu,fi ps=sy,fr te=di,fr
le
Importance Health Utility
fr ly 1 1 ly 2 1 ly 3 1 ly 5 1 ly 8 1 ly 9 1 ly 10 1 ly 11 1 ly 12 1 ly 13 1 ly 14 1 ly 20 1 ly 21 1 ly 24 1 ly 27 1 ly
30 1
fr ly 3 2 ly 4 2 ly 6 2 ly 7 2 ly 8 2 ly 13 2 ly 14 2 ly 15 2 ly 16 2 ly 17 2 ly 18 2 ly 19 2 ly 22 2 ly 23 2 ly 24 2 ly
25 2 ly 26 2 ly 28 2
fr ly 2 3 ly 4 3 ly 7 3 ly 9 3 ly 10 3 ly 14 3 ly 15 3 ly 19 3 ly 20 3 ly 24 3 ly 29 3 ly 30 3
va 1 ps 1 1 ps 2 2 ps 3 3
pd
ou rs ss sc ad=off nd=4
```

MPLUS SYNTAX FILES

EFA of the between-group level

```
TITLE: expoloratory factor analysis of the between-group structure;
DATA:  FILE IS c:\vysledky_between.cor;
       TYPE IS CORRELATION;
       NOBSERVATIONS ARE 87;
VARIABLE:  NAMES ARE y1-y30;
ANALYSIS: TYPE = EFA 3 5;
```

Simultaneous estimation of the between- and within-group structures

```
TITLE:  two-level CFA
DATA:           FILE IS twolevel.dat;
VARIABLE:       NAMES ARE y1-y30 clus;
          CLUSTER IS clus;
ANALYSIS:       TYPE IS TWOLEVEL;
          ESTIMATOR IS MUML;
MODEL:
          %WITHIN%
          fw1 BY y7 y9 y11 y13 y17 y23 y24 y29;
          fw2 BY y2 y4 y12 y15 y20 y25 y28;
          fw3 BY y3 y6 y16 y18 y22 y27;
          fw4 BY y1 y5 y8 y10 y14 y19 y21 y26 y30;
   fw5 BY y2 y4 y6 y11-y13 y15 y19 y23 y25 y27-y29;
   fw6 BY y1 y3 y5 y7-Y10 y14 y16-y18 y20-y22 y24 y26 y30;
   fw7 BY y1 y3 y8 y11 y14 y18-y21 y24 y27 y30;
   fw1 WITH fw5-fw7@0;
   fw2 WITH fw5-fw6@0;
   fw3 WITH fw5-fw7@0;
   fw4 WITH fw5-fw7@0;
   fw7 WITH fw5-fw6@0;
          %BETWEEN%
          fb1 BY y1 y2 y3 y5 y8-y14 y20 y21 y24 y27 y30;
          fb2 BY y3 y4 y6 y7 y8 y13-y19 y22-y26 y28;
   fb3 BY y2 y4 y7 y9 y10 y14 y15 y19 y20 y24 y29 y30;
OUTPUT:
SAMPSTAT STANDARDIZED RESIDUAL;
```

Two-level CFA for ordinal variables

```
TITLE:  two-level CFA with ordinal variables
DATA:           FILE IS twolevel.dat;
VARIABLE:       NAMES ARE y1-y30 clus;
          CATEGORICAL ARE y1-y30;
           CLUSTER IS clus;
ANALYSIS:       TYPE IS TWOLEVEL;
          ESTIMATOR IS MLR;
          ALGORITHM IS INTEGRATION;
          INTEGRATION IS MONTECARLO (100);

MODEL:
          %WITHIN%
```

```
fw1 BY y7 y9 y11 y13 y17 y23 y24 y29;
fw2 BY y2 y4 y12 y15 y20 y25 y28;
fw3 BY y3 y6 y16 y18 y22 y27;
fw4 BY y1 y5 y8 y10 y14 y19 y21 y26 y30;
%BETWEEN%
fb1 BY y1-y30;

OUTPUT:
 STANDARDIZED RESIDUAL;
```

POLYCHORIC CORRELATIONS

(estimated in PRELIS)

			Test of Model			Test of Close Fit	
Var.	vs.	Variable	Correl.	Chi-Squ.	D.F.	P-Value	RMSEA	P-Value
IT29	vs.	IT19	0.246 (PC)	72.680	35	0.000	0.031	1.000
IT29	vs.	IT20	0.239 (PC)	71.073	35	0.000	0.030	1.000
IT29	vs.	IT21	0.215 (PC)	73.338	35	0.000	0.031	1.000
IT29	vs.	IT22	0.218 (PC)	115.200	35	0.000	0.045	1.000
IT29	vs.	IT23	0.426 (PC)	120.799	35	0.000	0.046	1.000
IT29	vs.	IT24	0.218 (PC)	71.914	35	0.000	0.030	1.000
IT29	vs.	IT25	0.318 (PC)	88.514	35	0.000	0.036	1.000
IT29	vs.	IT26	0.239 (PC)	112.278	35	0.000	0.044	1.000
IT29	vs.	IT27	0.406 (PC)	95.829	35	0.000	0.039	1.000
IT29	vs.	IT28	0.318 (PC)	90.080	35	0.000	0.037	1.000
IT30	vs.	IT1	0.595 (PC)	147.241	35	0.000	0.053	1.000
IT30	vs.	IT2	0.328 (PC)	61.188	35	0.004	0.025	1.000
IT30	vs.	IT3	0.343 (PC)	82.838	35	0.000	0.034	1.000
IT30	vs.	IT4	0.237 (PC)	52.128	35	0.031	0.021	1.000
IT30	vs.	IT5	0.510 (PC)	93.106	35	0.000	0.038	1.000
IT30	vs.	IT6	0.232 (PC)	42.152	35	0.189	0.013	1.000
IT30	vs.	IT7	0.255 (PC)	47.122	35	0.083	0.017	1.000
IT30	vs.	IT8	0.453 (PC)	142.464	35	0.000	0.052	1.000
IT30	vs.	IT9	0.163 (PC)	59.486	35	0.006	0.025	1.000
IT30	vs.	IT10	0.232 (PC)	74.326	35	0.000	0.031	1.000
IT30	vs.	IT11	0.414 (PC)	80.417	35	0.000	0.034	1.000
IT30	vs.	IT12	0.205 (PC)	75.552	35	0.000	0.032	1.000
IT30	vs.	IT13	0.161 (PC)	61.773	35	0.003	0.026	1.000
IT30	vs.	IT14	0.503 (PC)	82.102	35	0.000	0.034	1.000
IT30	vs.	IT15	0.327 (PC)	51.858	35	0.033	0.020	1.000
IT30	vs.	IT16	0.467 (PC)	66.670	35	0.001	0.028	1.000
IT30	vs.	IT17	0.372 (PC)	63.606	35	0.002	0.027	1.000
IT30	vs.	IT18	0.406 (PC)	99.717	35	0.000	0.040	1.000
IT30	vs.	IT19	0.369 (PC)	48.518	35	0.064	0.018	1.000
IT30	vs.	IT20	0.387 (PC)	81.013	35	0.000	0.034	1.000
IT30	vs.	IT21	0.432 (PC)	188.380	35	0.000	0.062	1.000
IT30	vs.	IT22	0.245 (PC)	97.207	35	0.000	0.039	1.000
IT30	vs.	IT23	0.289 (PC)	69.351	35	0.000	0.029	1.000
IT30	vs.	IT24	0.527 (PC)	110.390	35	0.000	0.043	1.000
IT30	vs.	IT25	0.381 (PC)	60.681	35	0.005	0.025	1.000
IT30	vs.	IT26	0.315 (PC)	73.768	35	0.000	0.031	1.000
IT30	vs.	IT27	0.608 (PC)	113.582	35	0.000	0.044	1.000
IT30	vs.	IT28	0.372 (PC)	54.231	35	0.020	0.022	1.000
IT30	vs.	IT29	0.353 (PC)	81.520	35	0.000	0.034	1.000

Percentage of Tests Exceeding 0.5% Significance Level: 0.0%

Percentage of Tests Exceeding 1.0% Significance Level: 0.0%

Percentage of Tests Exceeding 5.0% Significance Level: 0.0%

		Test of Model			Test of Close Fit	
Var. vs.	Variable	Correl.	Chi-Squ.	D.F.	P-Value	RMSEA	P-Value
IT2 vs.	IT1	0.288 (PC)	61.158	35	0.004	0.025	1.000
IT3 vs.	IT1	0.354 (PC)	39.068	35	0.292	0.010	1.000
IT3 vs.	IT2	0.210 (PC)	76.018	35	0.000	0.032	1.000
IT4 vs.	IT1	0.173 (PC)	58.523	35	0.008	0.024	1.000
IT4 vs.	IT2	0.233 (PC)	93.362	35	0.000	0.038	1.000
IT4 vs.	IT3	0.004 (PC)	39.619	35	0.271	0.011	1.000
IT5 vs.	IT1	0.482 (PC)	107.206	35	0.000	0.042	1.000
IT5 vs.	IT2	0.274 (PC)	78.729	35	0.000	0.033	1.000
IT5 vs.	IT3	0.288 (PC)	57.921	35	0.009	0.024	1.000
IT5 vs.	IT4	0.183 (PC)	83.957	35	0.000	0.035	1.000
IT6 vs.	IT1	0.201 (PC)	47.160	35	0.082	0.017	1.000
IT6 vs.	IT2	0.152 (PC)	64.888	35	0.002	0.027	1.000
IT6 vs.	IT3	0.207 (PC)	56.825	35	0.011	0.023	1.000
IT6 vs.	IT4	0.131 (PC)	77.762	35	0.000	0.033	1.000
IT6 vs.	IT5	0.239 (PC)	59.449	35	0.006	0.025	1.000
IT7 vs.	IT1	0.205 (PC)	68.715	35	0.001	0.029	1.000
IT7 vs.	IT2	0.152 (PC)	81.665	35	0.000	0.034	1.000
IT7 vs.	IT3	0.154 (PC)	51.701	35	0.034	0.020	1.000
IT7 vs.	IT4	0.069 (PC)	94.273	35	0.000	0.038	1.000
IT7 vs.	IT5	0.319 (PC)	91.063	35	0.000	0.037	1.000
IT7 vs.	IT6	0.240 (PC)	69.719	35	0.000	0.029	1.000
IT8 vs.	IT1	0.414 (PC)	160.454	35	0.000	0.056	1.000
IT8 vs.	IT2	0.195 (PC)	62.895	35	0.003	0.026	1.000
IT8 vs.	IT3	0.245 (PC)	81.147	35	0.000	0.034	1.000
IT8 vs.	IT4	0.062 (PC)	68.358	35	0.001	0.029	1.000
IT8 vs.	IT5	0.350 (PC)	90.263	35	0.000	0.037	1.000
IT8 vs.	IT6	0.160 (PC)	53.268	35	0.025	0.021	1.000
IT8 vs.	IT7	0.185 (PC)	69.291	35	0.000	0.029	1.000
IT9 vs.	IT1	0.124 (PC)	48.659	35	0.062	0.018	1.000
IT9 vs.	IT2	0.179 (PC)	87.304	35	0.000	0.036	1.000
IT9 vs.	IT3	0.146 (PC)	67.073	35	0.001	0.028	1.000
IT9 vs.	IT4	0.115 (PC)	65.697	35	0.001	0.028	1.000
IT9 vs.	IT5	0.189 (PC)	63.858	35	0.002	0.027	1.000
IT9 vs.	IT6	0.144 (PC)	81.162	35	0.000	0.034	1.000
IT9 vs.	IT7	0.179 (PC)	101.946	35	0.000	0.041	1.000
IT9 vs.	IT8	0.057 (PC)	58.083	35	0.008	0.024	1.000
IT10 vs.	IT1	0.182 (PC)	58.500	35	0.008	0.024	1.000
IT10 vs.	IT2	0.132 (PC)	64.331	35	0.002	0.027	1.000
IT10 vs.	IT3	0.193 (PC)	58.301	35	0.008	0.024	1.000
IT10 vs.	IT4	0.173 (PC)	105.917	35	0.000	0.042	1.000
IT10 vs.	IT5	0.185 (PC)	74.335	35	0.000	0.031	1.000
IT10 vs.	IT6	0.153 (PC)	55.586	35	0.015	0.023	1.000
IT10 vs.	IT7	0.218 (PC)	54.402	35	0.019	0.022	1.000
IT10 vs.	IT8	0.192 (PC)	65.800	35	0.001	0.028	1.000
IT10 vs.	IT9	0.108 (PC)	44.927	35	0.121	0.016	1.000
IT11 vs.	IT1	0.356 (PC)	79.754	35	0.000	0.033	1.000
IT11 vs.	IT2	0.271 (PC)	80.868	35	0.000	0.034	1.000
IT11 vs.	IT3	0.265 (PC)	58.306	35	0.008	0.024	1.000
IT11 vs.	IT4	0.197 (PC)	53.519	35	0.023	0.021	1.000
IT11 vs.	IT5	0.303 (PC)	84.664	35	0.000	0.035	1.000
IT11 vs.	IT6	0.259 (PC)	59.951	35	0.005	0.025	1.000
IT11 vs.	IT7	0.224 (PC)	62.834	35	0.003	0.026	1.000
IT11 vs.	IT8	0.233 (PC)	72.432	35	0.000	0.030	1.000
IT11 vs.	IT9	0.223 (PC)	98.965	35	0.000	0.040	1.000
IT11 vs.	IT10	0.175 (PC)	65.353	35	0.001	0.027	1.000
IT12 vs.	IT1	0.164 (PC)	64.606	35	0.002	0.027	1.000
IT12 vs.	IT2	0.110 (PC)	89.692	35	0.000	0.037	1.000
IT12 vs.	IT3	0.241 (PC)	91.870	35	0.000	0.037	1.000
IT12 vs.	IT4	-0.021 (PC)	99.615	35	0.000	0.040	1.000
IT12 vs.	IT5	0.212 (PC)	108.287	35	0.000	0.043	1.000
IT12 vs.	IT6	0.050 (PC)	107.248	35	0.000	0.042	1.000
IT12 vs.	IT7	0.106 (PC)	63.975	35	0.002	0.027	1.000
IT12 vs.	IT8	0.089 (PC)	43.360	35	0.157	0.014	1.000
IT12 vs.	IT9	0.079 (PC)	74.383	35	0.000	0.031	1.000
IT12 vs.	IT10	0.083 (PC)	89.352	35	0.000	0.037	1.000
IT12 vs.	IT11	0.038 (PC)	98.033	35	0.000	0.039	1.000

		Test of Model			Test of Close Fit	
Var. vs.	Variable	Correl.	Chi-Squ.	D.F.	P-Value	RMSEA	P-Value
IT13 vs.	IT1	0.109 (PC)	59.961	35	0.005	0.025	1.000
IT13 vs.	IT2	0.204 (PC)	67.736	35	0.001	0.028	1.000
IT13 vs.	IT3	0.074 (PC)	65.926	35	0.001	0.028	1.000
IT13 vs.	IT4	0.104 (PC)	83.402	35	0.000	0.035	1.000
IT13 vs.	IT5	0.226 (PC)	55.562	35	0.015	0.023	1.000
IT13 vs.	IT6	0.235 (PC)	68.981	35	0.001	0.029	1.000
IT13 vs.	IT7	0.205 (PC)	67.124	35	0.001	0.028	1.000
IT13 vs.	IT8	0.074 (PC)	56.889	35	0.011	0.023	1.000
IT13 vs.	IT9	0.208 (PC)	93.559	35	0.000	0.038	1.000
IT13 vs.	IT10	0.083 (PC)	69.631	35	0.000	0.029	1.000
IT13 vs.	IT11	0.314 (PC)	61.036	35	0.004	0.025	1.000
IT13 vs.	IT12	0.043 (PC)	71.158	35	0.000	0.030	1.000
IT14 vs.	IT1	0.387 (PC)	57.937	35	0.009	0.024	1.000
IT14 vs.	IT2	0.250 (PC)	70.117	35	0.000	0.029	1.000
IT14 vs.	IT3	0.331 (PC)	47.587	35	0.076	0.018	1.000
IT14 vs.	IT4	0.147 (PC)	76.999	35	0.000	0.032	1.000
IT14 vs.	IT5	0.400 (PC)	90.385	35	0.000	0.037	1.000
IT14 vs.	IT6	0.241 (PC)	74.699	35	0.000	0.031	1.000
IT14 vs.	IT7	0.256 (PC)	97.356	35	0.000	0.039	1.000
IT14 vs.	IT8	0.299 (PC)	91.055	35	0.000	0.037	1.000
IT14 vs.	IT9	0.262 (PC)	67.517	35	0.001	0.028	1.000
IT14 vs.	IT10	0.199 (PC)	72.278	35	0.000	0.030	1.000
IT14 vs.	IT11	0.371 (PC)	104.543	35	0.000	0.041	1.000
IT14 vs.	IT12	0.216 (PC)	97.362	35	0.000	0.039	1.000
IT14 vs.	IT13	0.146 (PC)	96.395	35	0.000	0.039	1.000
IT15 vs.	IT1	0.286 (PC)	58.230	35	0.008	0.024	1.000
IT15 vs.	IT2	0.347 (PC)	76.374	35	0.000	0.032	1.000
IT15 vs.	IT3	0.303 (PC)	55.917	35	0.014	0.023	1.000
IT15 vs.	IT4	0.155 (PC)	54.406	35	0.019	0.022	1.000
IT15 vs.	IT5	0.226 (PC)	66.145	35	0.001	0.028	1.000
IT15 vs.	IT6	0.154 (PC)	36.611	35	0.394	0.006	1.000
IT15 vs.	IT7	0.034 (PC)	71.215	35	0.000	0.030	1.000
IT15 vs.	IT8	0.210 (PC)	54.201	35	0.020	0.022	1.000
IT15 vs.	IT9	0.169 (PC)	39.993	35	0.258	0.011	1.000
IT15 vs.	IT10	0.130 (PC)	57.751	35	0.009	0.024	1.000
IT15 vs.	IT11	0.290 (PC)	65.291	35	0.001	0.027	1.000
IT15 vs.	IT12	0.141 (PC)	108.464	35	0.000	0.043	1.000
IT15 vs.	IT13	0.171 (PC)	46.659	35	0.090	0.017	1.000
IT15 vs.	IT14	0.249 (PC)	99.750	35	0.000	0.040	1.000
IT16 vs.	IT1	0.401 (PC)	69.874	35	0.000	0.029	1.000
IT16 vs.	IT2	0.337 (PC)	73.700	35	0.000	0.031	1.000
IT16 vs.	IT3	0.382 (PC)	55.728	35	0.014	0.023	1.000
IT16 vs.	IT4	0.090 (PC)	66.870	35	0.001	0.028	1.000
IT16 vs.	IT5	0.454 (PC)	71.019	35	0.000	0.030	1.000
IT16 vs.	IT6	0.415 (PC)	53.196	35	0.025	0.021	1.000
IT16 vs.	IT7	0.348 (PC)	88.378	35	0.000	0.036	1.000
IT16 vs.	IT8	0.288 (PC)	59.048	35	0.007	0.024	1.000
IT16 vs.	IT9	0.275 (PC)	98.432	35	0.000	0.040	1.000
IT16 vs.	IT10	0.236 (PC)	85.920	35	0.000	0.035	1.000
IT16 vs.	IT11	0.428 (PC)	99.031	35	0.000	0.040	1.000
IT16 vs.	IT12	0.131 (PC)	72.701	35	0.000	0.031	1.000
IT16 vs.	IT13	0.344 (PC)	101.562	35	0.000	0.041	1.000
IT16 vs.	IT14	0.347 (PC)	79.031	35	0.000	0.033	1.000
IT16 vs.	IT15	0.327 (PC)	52.661	35	0.028	0.021	1.000
IT17 vs.	IT1	0.341 (PC)	93.649	35	0.000	0.038	1.000
IT17 vs.	IT2	0.277 (PC)	73.856	35	0.000	0.031	1.000
IT17 vs.	IT3	0.313 (PC)	64.316	35	0.002	0.027	1.000
IT17 vs.	IT4	0.047 (PC)	53.381	35	0.024	0.021	1.000
IT17 vs.	IT5	0.425 (PC)	90.779	35	0.000	0.037	1.000
IT17 vs.	IT6	0.324 (PC)	92.862	35	0.000	0.038	1.000
IT17 vs.	IT7	0.320 (PC)	91.946	35	0.000	0.038	1.000
IT17 vs.	IT8	0.279 (PC)	85.327	35	0.000	0.035	1.000
IT17 vs.	IT9	0.174 (PC)	67.913	35	0.001	0.029	1.000
IT17 vs.	IT10	0.226 (PC)	65.093	35	0.001	0.027	1.000
IT17 vs.	IT11	0.365 (PC)	90.948	35	0.000	0.037	1.000
IT17 vs.	IT12	0.128 (PC)	83.082	35	0.000	0.034	1.000

Var.	vs.	Variable	Correl.	Test of Model Chi-Squ.	D.F.	P-Value	Test of Close Fit RMSEA	P-Value
IT17	vs.	IT13	0.268 (PC)	82.169	35	0.000	0.034	1.000
IT17	vs.	IT14	0.377 (PC)	93.238	35	0.000	0.038	1.000
IT17	vs.	IT15	0.230 (PC)	63.784	35	0.002	0.027	1.000
IT17	vs.	IT16	0.517 (PC)	77.529	35	0.000	0.032	1.000
IT18	vs.	IT1	0.377 (PC)	77.951	35	0.000	0.033	1.000
IT18	vs.	IT2	0.263 (PC)	62.231	35	0.003	0.026	1.000
IT18	vs.	IT3	0.319 (PC)	66.358	35	0.001	0.028	1.000
IT18	vs.	IT4	0.083 (PC)	66.088	35	0.001	0.028	1.000
IT18	vs.	IT5	0.384 (PC)	81.448	35	0.000	0.034	1.000
IT18	vs.	IT6	0.361 (PC)	50.781	35	0.041	0.020	1.000
IT18	vs.	IT7	0.287 (PC)	57.051	35	0.011	0.023	1.000
IT18	vs.	IT8	0.319 (PC)	68.956	35	0.001	0.029	1.000
IT18	vs.	IT9	0.182 (PC)	57.482	35	0.010	0.024	1.000
IT18	vs.	IT10	0.213 (PC)	88.135	35	0.000	0.036	1.000
IT18	vs.	IT11	0.313 (PC)	92.394	35	0.000	0.038	1.000
IT18	vs.	IT12	0.190 (PC)	106.798	35	0.000	0.042	1.000
IT18	vs.	IT13	0.228 (PC)	54.816	35	0.018	0.022	1.000
IT18	vs.	IT14	0.348 (PC)	81.205	35	0.000	0.034	1.000
IT18	vs.	IT15	0.245 (PC)	49.507	35	0.053	0.019	1.000
IT18	vs.	IT16	0.516 (PC)	66.385	35	0.001	0.028	1.000
IT18	vs.	IT17	0.451 (PC)	99.814	35	0.000	0.040	1.000
IT19	vs.	IT1	0.237 (PC)	79.627	35	0.000	0.033	1.000
IT19	vs.	IT2	0.331 (PC)	61.047	35	0.004	0.025	1.000
IT19	vs.	IT3	0.099 (PC)	104.730	35	0.000	0.042	1.000
IT19	vs.	IT4	0.344 (PC)	94.952	35	0.000	0.038	1.000
IT19	vs.	IT5	0.190 (PC)	70.695	35	0.000	0.030	1.000
IT19	vs.	IT6	0.146 (PC)	76.751	35	0.000	0.032	1.000
IT19	vs.	IT7	0.079 (PC)	78.647	35	0.000	0.033	1.000
IT19	vs.	IT8	0.159 (PC)	74.038	35	0.000	0.031	1.000
IT19	vs.	IT9	0.190 (PC)	62.813	35	0.003	0.026	1.000
IT19	vs.	IT10	0.154 (PC)	68.393	35	0.001	0.029	1.000
IT19	vs.	IT11	0.328 (PC)	101.951	35	0.000	0.041	1.000
IT19	vs.	IT12	0.043 (PC)	167.163	35	0.000	0.057	1.000
IT19	vs.	IT13	0.127 (PC)	71.727	35	0.000	0.030	1.000
IT19	vs.	IT14	0.339 (PC)	107.348	35	0.000	0.042	1.000
IT19	vs.	IT15	0.259 (PC)	68.816	35	0.001	0.029	1.000
IT19	vs.	IT16	0.171 (PC)	69.367	35	0.000	0.029	1.000
IT19	vs.	IT17	0.144 (PC)	97.494	35	0.000	0.039	1.000
IT19	vs.	IT18	0.138 (PC)	95.619	35	0.000	0.039	1.000
IT20	vs.	IT1	0.277 (PC)	38.125	35	0.329	0.009	1.000
IT20	vs.	IT2	0.333 (PC)	89.404	35	0.000	0.037	1.000
IT20	vs.	IT3	0.277 (PC)	69.245	35	0.000	0.029	1.000
IT20	vs.	IT4	0.188 (PC)	53.119	35	0.025	0.021	1.000
IT20	vs.	IT5	0.275 (PC)	90.959	35	0.000	0.037	1.000
IT20	vs.	IT6	0.135 (PC)	84.241	35	0.000	0.035	1.000
IT20	vs.	IT7	0.163 (PC)	60.632	35	0.005	0.025	1.000
IT20	vs.	IT8	0.157 (PC)	61.627	35	0.004	0.026	1.000
IT20	vs.	IT9	0.220 (PC)	73.116	35	0.000	0.031	1.000
IT20	vs.	IT10	0.168 (PC)	56.480	35	0.012	0.023	1.000
IT20	vs.	IT11	0.282 (PC)	93.808	35	0.000	0.038	1.000
IT20	vs.	IT12	0.227 (PC)	118.159	35	0.000	0.045	1.000
IT20	vs.	IT13	0.155 (PC)	55.677	35	0.015	0.023	1.000
IT20	vs.	IT14	0.304 (PC)	77.394	35	0.000	0.032	1.000
IT20	vs.	IT15	0.409 (PC)	93.743	35	0.000	0.038	1.000
IT20	vs.	IT16	0.350 (PC)	74.964	35	0.000	0.031	1.000
IT20	vs.	IT17	0.256 (PC)	92.158	35	0.000	0.038	1.000
IT20	vs.	IT18	0.317 (PC)	67.189	35	0.001	0.028	1.000
IT20	vs.	IT19	0.276 (PC)	97.681	35	0.000	0.039	1.000
IT21	vs.	IT1	0.439 (PC)	160.845	35	0.000	0.056	1.000
IT21	vs.	IT2	0.131 (PC)	75.332	35	0.000	0.032	1.000
IT21	vs.	IT3	0.243 (PC)	60.425	35	0.005	0.025	1.000
IT21	vs.	IT4	0.109 (PC)	47.103	35	0.083	0.017	1.000
IT21	vs.	IT5	0.355 (PC)	102.933	35	0.000	0.041	1.000
IT21	vs.	IT6	0.142 (PC)	56.355	35	0.013	0.023	1.000
IT21	vs.	IT7	0.162 (PC)	89.430	35	0.000	0.037	1.000
IT21	vs.	IT8	0.313 (PC)	139.044	35	0.000	0.051	1.000
IT21	vs.	IT9	0.087 (PC)	72.275	35	0.000	0.030	1.000
IT21	vs.	IT10	0.184 (PC)	67.849	35	0.001	0.028	1.000
IT21	vs.	IT11	0.248 (PC)	72.615	35	0.000	0.030	1.000
IT21	vs.	IT12	0.073 (PC)	78.934	35	0.000	0.033	1.000
IT21	vs.	IT13	0.014 (PC)	65.505	35	0.001	0.027	1.000
IT21	vs.	IT14	0.283 (PC)	85.239	35	0.000	0.035	1.000
IT21	vs.	IT15	0.145 (PC)	63.660	35	0.002	0.027	1.000
IT21	vs.	IT16	0.292 (PC)	72.019	35	0.000	0.030	1.000
IT21	vs.	IT17	0.276 (PC)	72.335	35	0.000	0.030	1.000
IT21	vs.	IT18	0.306 (PC)	100.723	35	0.000	0.040	1.000
IT21	vs.	IT19	0.193 (PC)	91.742	35	0.000	0.037	1.000
IT21	vs.	IT20	0.175 (PC)	86.012	35	0.000	0.036	1.000
IT22	vs.	IT1	0.190 (PC)	75.491	35	0.000	0.032	1.000
IT22	vs.	IT2	0.116 (PC)	67.213	35	0.001	0.028	1.000
IT22	vs.	IT3	0.311 (PC)	86.031	35	0.000	0.036	1.000
IT22	vs.	IT4	-0.026 (PC)	108.876	35	0.000	0.043	1.000
IT22	vs.	IT5	0.244 (PC)	57.509	35	0.010	0.024	1.000
IT22	vs.	IT6	0.203 (PC)	71.844	35	0.000	0.030	1.000
IT22	vs.	IT7	0.163 (PC)	86.881	35	0.000	0.036	1.000
IT22	vs.	IT8	0.120 (PC)	53.798	35	0.022	0.022	1.000
IT22	vs.	IT9	0.148 (PC)	47.591	35	0.076	0.018	1.000
IT22	vs.	IT10	0.083 (PC)	87.635	35	0.000	0.036	1.000
IT22	vs.	IT11	0.154 (PC)	99.361	35	0.000	0.040	1.000
IT22	vs.	IT12	0.157 (PC)	115.607	35	0.000	0.045	1.000
IT22	vs.	IT13	0.093 (PC)	62.251	35	0.003	0.026	1.000
IT22	vs.	IT14	0.203 (PC)	98.170	35	0.000	0.040	1.000
IT22	vs.	IT15	0.121 (PC)	73.584	35	0.000	0.031	1.000
IT22	vs.	IT16	0.380 (PC)	116.929	35	0.000	0.045	1.000
IT22	vs.	IT17	0.261 (PC)	98.632	35	0.000	0.040	1.000
IT22	vs.	IT18	0.306 (PC)	94.043	35	0.000	0.038	1.000
IT22	vs.	IT19	0.022 (PC)	78.115	35	0.000	0.033	1.000
IT22	vs.	IT20	0.298 (PC)	87.470	35	0.000	0.036	1.000
IT22	vs.	IT21	0.128 (PC)	67.005	35	0.001	0.028	1.000
IT23	vs.	IT1	0.237 (PC)	55.066	35	0.017	0.022	1.000
IT23	vs.	IT2	0.292 (PC)	74.924	35	0.000	0.031	1.000
IT23	vs.	IT3	0.182 (PC)	60.316	35	0.005	0.025	1.000
IT23	vs.	IT4	0.199 (PC)	63.785	35	0.002	0.027	1.000
IT23	vs.	IT5	0.341 (PC)	46.695	35	0.089	0.017	1.000
IT23	vs.	IT6	0.259 (PC)	55.213	35	0.016	0.022	1.000
IT23	vs.	IT7	0.260 (PC)	92.785	35	0.000	0.038	1.000
IT23	vs.	IT8	0.197 (PC)	42.403	35	0.182	0.014	1.000
IT23	vs.	IT9	0.272 (PC)	124.834	35	0.000	0.047	1.000
IT23	vs.	IT10	0.220 (PC)	83.594	35	0.000	0.035	1.000
IT23	vs.	IT11	0.430 (PC)	62.287	35	0.003	0.026	1.000
IT23	vs.	IT12	0.069 (PC)	81.932	35	0.000	0.034	1.000
IT23	vs.	IT13	0.388 (PC)	95.458	35	0.000	0.039	1.000
IT23	vs.	IT14	0.248 (PC)	85.917	35	0.000	0.035	1.000
IT23	vs.	IT15	0.228 (PC)	54.620	35	0.018	0.022	1.000
IT23	vs.	IT16	0.455 (PC)	109.580	35	0.000	0.043	1.000
IT23	vs.	IT17	0.498 (PC)	132.379	35	0.000	0.049	1.000
IT23	vs.	IT18	0.296 (PC)	110.510	35	0.000	0.043	1.000
IT23	vs.	IT19	0.233 (PC)	74.721	35	0.000	0.031	1.000
IT23	vs.	IT20	0.255 (PC)	68.803	35	0.001	0.029	1.000
IT23	vs.	IT21	0.103 (PC)	71.505	35	0.000	0.030	1.000
IT23	vs.	IT22	0.179 (PC)	87.638	35	0.000	0.036	1.000
IT24	vs.	IT1	0.380 (PC)	85.561	35	0.000	0.035	1.000
IT24	vs.	IT2	0.295 (PC)	59.802	35	0.006	0.025	1.000
IT24	vs.	IT3	0.282 (PC)	56.599	35	0.012	0.023	1.000
IT24	vs.	IT4	0.196 (PC)	66.235	35	0.001	0.028	1.000
IT24	vs.	IT5	0.326 (PC)	79.188	35	0.000	0.033	1.000
IT24	vs.	IT6	0.107 (PC)	64.621	35	0.002	0.027	1.000
IT24	vs.	IT7	0.159 (PC)	99.635	35	0.000	0.040	1.000
IT24	vs.	IT8	0.324 (PC)	76.231	35	0.000	0.032	1.000
IT24	vs.	IT9	0.192 (PC)	69.363	35	0.000	0.029	1.000
IT24	vs.	IT10	0.197 (PC)	75.291	35	0.000	0.032	1.000
IT24	vs.	IT11	0.296 (PC)	100.050	35	0.000	0.040	1.000

Var.	vs.	Variable	Correl.	Test of Model Chi-Squ.	D.F.	P-Value	Test of Close Fit RMSEA	P-Value
IT24	vs.	IT12	0.200 (PC)	126.500	35	0.000	0.048	1.000
IT24	vs.	IT13	0.099 (PC)	84.299	35	0.000	0.035	1.000
IT24	vs.	IT14	0.493 (PC)	123.992	35	0.000	0.047	1.000
IT24	vs.	IT15	0.257 (PC)	64.108	35	0.002	0.027	1.000
IT24	vs.	IT16	0.276 (PC)	81.730	35	0.000	0.034	1.000
IT24	vs.	IT17	0.241 (PC)	109.903	35	0.000	0.043	1.000
IT24	vs.	IT18	0.228 (PC)	112.185	35	0.000	0.044	1.000
IT24	vs.	IT19	0.412 (PC)	102.191	35	0.000	0.041	1.000
IT24	vs.	IT20	0.340 (PC)	90.388	35	0.000	0.037	1.000
IT24	vs.	IT21	0.225 (PC)	83.496	35	0.000	0.035	1.000
IT24	vs.	IT22	0.184 (PC)	106.907	35	0.000	0.042	1.000
IT24	vs.	IT23	0.264 (PC)	83.845	35	0.000	0.035	1.000
IT25	vs.	IT1	0.306 (PC)	43.453	35	0.155	0.014	1.000
IT25	vs.	IT2	0.392 (PC)	105.941	35	0.000	0.042	1.000
IT25	vs.	IT3	0.238 (PC)	56.150	35	0.013	0.023	1.000
IT25	vs.	IT4	0.239 (PC)	95.881	35	0.000	0.039	1.000
IT25	vs.	IT5	0.318 (PC)	76.965	35	0.000	0.032	1.000
IT25	vs.	IT6	0.273 (PC)	53.649	35	0.023	0.021	1.000
IT25	vs.	IT7	0.173 (PC)	73.548	35	0.000	0.031	1.000
IT25	vs.	IT8	0.183 (PC)	66.653	35	0.001	0.028	1.000
IT25	vs.	IT9	0.221 (PC)	94.123	35	0.000	0.038	1.000
IT25	vs.	IT10	0.184 (PC)	61.491	35	0.004	0.026	1.000
IT25	vs.	IT11	0.395 (PC)	82.268	35	0.000	0.034	1.000
IT25	vs.	IT12	0.117 (PC)	89.329	35	0.000	0.037	1.000
IT25	vs.	IT13	0.285 (PC)	87.523	35	0.000	0.036	1.000
IT25	vs.	IT14	0.291 (PC)	73.677	35	0.000	0.031	1.000
IT25	vs.	IT15	0.411 (PC)	79.000	35	0.000	0.033	1.000
IT25	vs.	IT16	0.409 (PC)	76.728	35	0.000	0.032	1.000
IT25	vs.	IT17	0.365 (PC)	72.157	35	0.000	0.030	1.000
IT25	vs.	IT18	0.310 (PC)	70.323	35	0.000	0.030	1.000
IT25	vs.	IT19	0.348 (PC)	96.732	35	0.000	0.039	1.000
IT25	vs.	IT20	0.342 (PC)	85.960	35	0.000	0.035	1.000
IT25	vs.	IT21	0.158 (PC)	59.226	35	0.006	0.024	1.000
IT25	vs.	IT22	0.134 (PC)	89.153	35	0.000	0.037	1.000
IT25	vs.	IT23	0.406 (PC)	90.600	35	0.000	0.037	1.000
IT25	vs.	IT24	0.270 (PC)	64.550	35	0.002	0.027	1.000
IT26	vs.	IT1	0.284 (PC)	50.942	35	0.040	0.020	1.000
IT26	vs.	IT2	0.143 (PC)	43.719	35	0.148	0.015	1.000
IT26	vs.	IT3	0.248 (PC)	62.661	35	0.003	0.026	1.000
IT26	vs.	IT4	0.015 (PC)	60.370	35	0.005	0.025	1.000
IT26	vs.	IT5	0.294 (PC)	67.009	35	0.001	0.028	1.000
IT26	vs.	IT6	0.228 (PC)	89.329	35	0.000	0.037	1.000
IT26	vs.	IT7	0.278 (PC)	80.083	35	0.000	0.033	1.000
IT26	vs.	IT8	0.185 (PC)	81.011	35	0.000	0.034	1.000
IT26	vs.	IT9	0.156 (PC)	63.184	35	0.002	0.026	1.000
IT26	vs.	IT10	0.177 (PC)	58.771	35	0.007	0.024	1.000
IT26	vs.	IT11	0.269 (PC)	80.799	35	0.000	0.034	1.000
IT26	vs.	IT12	0.057 (PC)	94.781	35	0.000	0.038	1.000
IT26	vs.	IT13	0.253 (PC)	50.476	35	0.044	0.020	1.000
IT26	vs.	IT14	0.248 (PC)	91.105	35	0.000	0.037	1.000
IT26	vs.	IT15	0.189 (PC)	47.054	35	0.084	0.017	1.000
IT26	vs.	IT16	0.414 (PC)	55.614	35	0.015	0.023	1.000
IT26	vs.	IT17	0.400 (PC)	75.381	35	0.000	0.032	1.000
IT26	vs.	IT18	0.385 (PC)	69.550	35	0.000	0.029	1.000
IT26	vs.	IT19	0.045 (PC)	56.000	35	0.014	0.023	1.000
IT26	vs.	IT20	0.254 (PC)	49.161	35	0.057	0.019	1.000
IT26	vs.	IT21	0.185 (PC)	77.284	35	0.000	0.032	1.000
IT26	vs.	IT22	0.292 (PC)	79.335	35	0.000	0.033	1.000
IT26	vs.	IT23	0.266 (PC)	114.862	35	0.000	0.044	1.000
IT26	vs.	IT24	0.150 (PC)	85.618	35	0.000	0.035	1.000
IT26	vs.	IT25	0.250 (PC)	49.176	35	0.056	0.019	1.000
IT27	vs.	IT1	0.565 (PC)	125.795	35	0.000	0.047	1.000
IT27	vs.	IT2	0.363 (PC)	46.470	35	0.093	0.017	1.000
IT27	vs.	IT3	0.323 (PC)	47.625	35	0.076	0.018	1.000
IT27	vs.	IT4	0.217 (PC)	55.351	35	0.016	0.022	1.000
IT27	vs.	IT5	0.435 (PC)	72.404	35	0.000	0.030	1.000

Var.	vs.	Variable	Correl.	Test of Model Chi-Squ.	D.F.	P-Value	Test of Close Fit RMSEA	P-Value
IT27	vs.	IT6	0.310 (PC)	76.367	35	0.000	0.032	1.000
IT27	vs.	IT7	0.228 (PC)	90.855	35	0.000	0.037	1.000
IT27	vs.	IT8	0.357 (PC)	127.168	35	0.000	0.048	1.000
IT27	vs.	IT9	0.238 (PC)	54.400	35	0.019	0.022	1.000
IT27	vs.	IT10	0.198 (PC)	83.387	35	0.000	0.035	1.000
IT27	vs.	IT11	0.448 (PC)	131.908	35	0.000	0.049	1.000
IT27	vs.	IT12	0.173 (PC)	109.412	35	0.000	0.043	1.000
IT27	vs.	IT13	0.239 (PC)	64.667	35	0.002	0.027	1.000
IT27	vs.	IT14	0.439 (PC)	84.039	35	0.000	0.035	1.000
IT27	vs.	IT15	0.354 (PC)	94.955	35	0.000	0.038	1.000
IT27	vs.	IT16	0.473 (PC)	67.978	35	0.001	0.029	1.000
IT27	vs.	IT17	0.356 (PC)	58.436	35	0.008	0.024	1.000
IT27	vs.	IT18	0.420 (PC)	72.640	35	0.000	0.031	1.000
IT27	vs.	IT19	0.370 (PC)	73.073	35	0.000	0.031	1.000
IT27	vs.	IT20	0.393 (PC)	76.628	35	0.000	0.032	1.000
IT27	vs.	IT21	0.332 (PC)	119.651	35	0.000	0.046	1.000
IT27	vs.	IT22	0.261 (PC)	135.876	35	0.000	0.050	1.000
IT27	vs.	IT23	0.332 (PC)	93.501	35	0.000	0.038	1.000
IT27	vs.	IT24	0.465 (PC)	104.084	35	0.000	0.041	1.000
IT27	vs.	IT25	0.441 (PC)	74.968	35	0.000	0.031	1.000
IT27	vs.	IT26	0.293 (PC)	81.113	35	0.000	0.034	1.000
IT28	vs.	IT1	0.326 (PC)	65.795	35	0.001	0.028	1.000
IT28	vs.	IT2	0.384 (PC)	96.894	35	0.000	0.039	1.000
IT28	vs.	IT3	0.190 (PC)	55.312	35	0.016	0.022	1.000
IT28	vs.	IT4	0.325 (PC)	103.063	35	0.000	0.041	1.000
IT28	vs.	IT5	0.345 (PC)	54.393	35	0.019	0.022	1.000
IT28	vs.	IT6	0.306 (PC)	55.219	35	0.016	0.022	1.000
IT28	vs.	IT7	0.254 (PC)	88.698	35	0.000	0.036	1.000
IT28	vs.	IT8	0.261 (PC)	92.035	35	0.000	0.038	1.000
IT28	vs.	IT9	0.194 (PC)	82.632	35	0.000	0.034	1.000
IT28	vs.	IT10	0.251 (PC)	95.472	35	0.000	0.039	1.000
IT28	vs.	IT11	0.356 (PC)	56.016	35	0.014	0.023	1.000
IT28	vs.	IT12	0.104 (PC)	127.843	35	0.000	0.048	1.000
IT28	vs.	IT13	0.257 (PC)	62.664	35	0.003	0.026	1.000
IT28	vs.	IT14	0.343 (PC)	74.442	35	0.000	0.031	1.000
IT28	vs.	IT15	0.317 (PC)	80.521	35	0.000	0.034	1.000
IT28	vs.	IT16	0.362 (PC)	54.015	35	0.021	0.022	1.000
IT28	vs.	IT17	0.328 (PC)	87.232	35	0.000	0.036	1.000
IT28	vs.	IT18	0.284 (PC)	79.302	35	0.000	0.033	1.000
IT28	vs.	IT19	0.391 (PC)	121.104	35	0.000	0.046	1.000
IT28	vs.	IT20	0.267 (PC)	83.028	35	0.000	0.034	1.000
IT28	vs.	IT21	0.165 (PC)	91.936	35	0.000	0.038	1.000
IT28	vs.	IT22	0.123 (PC)	96.813	35	0.000	0.039	1.000
IT28	vs.	IT23	0.408 (PC)	87.060	35	0.000	0.036	1.000
IT28	vs.	IT24	0.325 (PC)	96.515	35	0.000	0.039	1.000
IT28	vs.	IT25	0.519 (PC)	107.646	35	0.000	0.042	1.000
IT28	vs.	IT26	0.209 (PC)	92.162	35	0.000	0.038	1.000
IT28	vs.	IT27	0.458 (PC)	102.458	35	0.000	0.041	1.000
IT29	vs.	IT1	0.263 (PC)	57.808	35	0.009	0.024	1.000
IT29	vs.	IT2	0.231 (PC)	44.957	35	0.121	0.016	1.000
IT29	vs.	IT3	0.155 (PC)	49.944	35	0.049	0.019	1.000
IT29	vs.	IT4	0.144 (PC)	39.188	35	0.287	0.010	1.000
IT29	vs.	IT5	0.284 (PC)	56.825	35	0.011	0.023	1.000
IT29	vs.	IT6	0.229 (PC)	48.654	35	0.062	0.018	1.000
IT29	vs.	IT7	0.192 (PC)	75.321	35	0.000	0.032	1.000
IT29	vs.	IT8	0.210 (PC)	55.347	35	0.016	0.022	1.000
IT29	vs.	IT9	0.178 (PC)	72.876	35	0.000	0.031	1.000
IT29	vs.	IT10	0.185 (PC)	76.590	35	0.000	0.032	1.000
IT29	vs.	IT11	0.363 (PC)	96.166	35	0.000	0.039	1.000
IT29	vs.	IT12	0.084 (PC)	79.158	35	0.000	0.033	1.000
IT29	vs.	IT13	0.271 (PC)	88.245	35	0.000	0.036	1.000
IT29	vs.	IT14	0.226 (PC)	53.901	35	0.022	0.022	1.000
IT29	vs.	IT15	0.227 (PC)	61.578	35	0.004	0.026	1.000
IT29	vs.	IT16	0.376 (PC)	107.104	35	0.000	0.042	1.000
IT29	vs.	IT17	0.341 (PC)	88.328	35	0.000	0.036	1.000
IT29	vs.	IT18	0.343 (PC)	78.886	35	0.000	0.033	1.000

EXTENDING METHODOLOGICAL STUDY

Occasions and groups of raters as the hidden facets in ratings of item relevance

Present study tackles a problem of ratings of item relevance. This study is predominantly focused on the reliability of such ratings. It is in accordance with the title of the present thesis, because it belongs among the remarkable methodological considerations. This study constitutes a very rare and even almost unique view on test reliability. But simultaneously, some results have been also used in the attitudes survey. In fact, it is a part of content validity study of the attitudes items. On the other hand, it extends some traditional techniques of the content validation process. Therefore, this study was called as "extending" and is situated in Appendix.

Introduction

Two important methodological interests are included in the studies of reliability of ratings of item relevance. The first one is the joint application of methodology of item relevance or more general concept of content validity, and the second is the inter-rater reliability or inter-rater agreement.

Item relevance is an important indicator's feature for investigators when developing or modifying observable scales used to measure theoretical constructs. Several indices have been developed for this purpose. For example, Aiken's item content-relevance index (Aiken, 1980; Aiken, 1985), Lawshe's content validity ratio (Lawshe, 1975), or the content validity index (Lynn, 1986) can be used for relative and/or absolute decisions about which item should be involved in the ultimate version of the constructed scale. Several studies have been conducted to improve, enhance, or discuss the relevance of the use of these indices in particular research contexts (Dunn, Bouffard, & Rogers, 1999; Penfield, & Giacobbi, 2004; Haynes, Richard, & Kubany, 1995; Wynd, Schmidt, & Schaefer, 2003; Polit, Beck, & Owen, 2006).

Nevertheless, these indices are based on the assumption, that the independent raters in the panel are the only source of measurement error whereas all other possible sources are usually not taken into consideration. However, while some of these sources may be ignored other ones should be taken into account. Our purpose is to examine the extent to which different occasions and different competent groups of raters may become a substantial source of

measurement error. A common practice is to assume that all competent (expert) raters in a panel are randomly sampled from one homogeneous population of competent raters. Further on, the influence of repeated scale evaluations is usually not considered in this context. From a scale developer point of view, it is not meaningful to wait several weeks or even months for repeating the ratings. Nevertheless, no one can be sure that the effect of occasions is really negligible. The current psychological state, motivation, mood, and many other characteristics could play an important role in raters' evaluation. These characteristics cannot be considered to be stable across time. In addition, it cannot be automatically assumed that all raters in a panel have been sampled from one population of all competent raters. This population is very often stratified into several groups of raters that evaluate (on average) the relevance of items differently. Based on particular purpose of the further use of the developed scale, the competent groups can be, for example, students, teachers, parents, investigators, practitioners, language experts or decision makers.

Using standard methods for assessing inter-rater reliability or agreement like Cronbach's alpha or Cohen's kappa (Cohen, 1960; Cohen, 1968) are not appropriate when measurement error arises from multiple sources. Therefore, the generalizability theory (GT) techniques can be applied to assess the influence of occasions and clustered raters on rating of item content relevance. GT is especially recommended to be used in multifaceted designs where inter-rater reliability is under investigation (Goodwin, 2001). In the following section, a brief introduction to the GT is presented.

Brief overview of the Generalizability theory

Generalizability theory (GT) was developed by Cronbach and associates (Cronbach et al., 1972) and further elaborated by Brennan (2001a). In recent years, GT became a well-known methodology to study consistency, reliability and/or dependability of a test scores in physical educational research (Godbout, & Schutz, 1983; Coleman, & Epstein, 1998; Tobar, Stegner, & Kane, 1999) with several applications in sport psychology (Gill, 1997; Stegner, Tobar, & Kane, 1999; Hagtvet, & Hanin, 2007) and studies of overweight reduction and body fat measurement (Coleman, & Epstein, 1998; Turner, & Bouffard, 1998; Mâsse et al., 2002; Turner, Lozano-Nieto, & Bouffard, 2003). GT has been also used in content validity studies (Crocker, Llabre, & Miller, 1988; Norcini, & Grosso, 1998).

GT is showing its greatest advantage when applied to multifaceted measurement designs. The present construct of item relevance is operationalized here by means of three-facet design. In the current study, raters (r) belonged to competent groups of raters (g), rated item relevance on different occasions (o). These three sources of variance are considered as facets of measurement, while items (i) are denoted as object of measurement.

Essential to the framework of GT is the definition of two types of universe. The present *universe of admissible* observation is defined in terms of the three-facet design and how the facets are organized. Facet can be either crossed with (×) or nested within (:) another facet. In this study, objects of measurement are crossed with occasion facet and with raters nested within groups of raters facet, what constitute the $i \times o \times (r : g)$ design. However, only the $i \times r$ design is usually used in the ratings of item relevance, whereas the o and g facets are not included in a measurement design.

Universe of generalization is the universe to which generalizations are made. This universe is defined in GT in terms of random and/or fixed conditions of the facets of measurement. A facet is random, if all conditions within this facet are considered to be randomly sampled from the universe. If we don't wish to generalize over all conditions associated with the facet or if there are no other conditions to which the generalization can be made, then the facet is fixed. In this study, the raters within groups and occasions are considered random, while groups of raters are assumed to be fixed.

Two types of studies are included in the framework of GT. The purpose of *generalizability study* (*G*-study) is to estimate variance and covariance components associated with the different facets in the universe of admissible observations. Based on *G*-study results different types of *decision studies* (*D*-study) can be conducted. The main purpose of the *D*-study is to estimate error variances and reliability-like coefficients for different combinations of conditions associated with each facet. The absolute error variance is the variance of the difference between observed item scores and their universe scores. The relative error variance is the variance of the difference between observed item deviation scores and the their universe deviation scores. Two types of coefficients are usually reported as D-study results. The *generalizability coefficient* (*G*) and the *index of dependability* (*phi*) (Brennan, & Kane, 1977) are the reliability-like coefficients for relative and absolute interpretation of the test scores, respectively. Both indices can be relevant in such studies depending on particular research purposes. If the scale developer wants to choose only the most relevant items, say, scored beyond the middle point of the scale then the absolute interpretation of the scores are of

interest. On the other hand, if he or she wishes to choose, for example, the most relevant 20 from 40 items, the relative interpretation would be necessary.

As noted by Crocker et al. (1988), a content validation is essentially a *D*-study. Our main interest is to examine what would happen, if the number of conditions within occasion facet and groups of raters facet decrease to 1. The coefficients obtained from such conditions are further compared with reliability coefficient calculated from rating of item relevance taken on one occasion by raters sampled from one competent group of raters. From this point of view, occasions and groups of raters are the *hidden facets.* A facet is hidden, when there is only one sampled condition of the facet, which induces bias in statistics such as generalizability coefficient (Brennan, 1992, 2001a). There are infinite numbers (at least theoretically) of a hidden facets in every measurement procedure. Many of them may be negligible, but some of them are substantive contributors to the total variance and their influence on reliability of the test scores should not be ignored. Two hidden facets have been considered in the present study – the occasions and groups of raters. This means that based on the *G*-study results for the $i \times o \times (r : g)$, the *D*-study has been conducted for the $i \times R$ design, where number of occasions = number of groups = 1. These results have been further compared with another *D*-study results for the same $i \times R$ design, where both occasions and groups have been fixed to 1.

Research questions

The present study addressed three research questions:

1. To which extent is the interrater reliability overestimated, if the effect of occasions is ignored?
2. How substantial is the overestimation, if raters are sampled from one group only and all others competent groups of raters are ignored?
3. Which combination of raters within each of fixed groups is optimal to display maximal level of composite generalizability, if the evaluation is accomplished on 1 occasion?

Methods

As we pointed out in the *inventory development* section of the present thesis, the items related to the ATPE are considered here as the objects of evaluation. Obviously, however, the present approach is very general and might be used as a part of any content validation study beyond the framework of inventories development. The following three subsections are presented here: (1) competent groups of raters, (2) measurement design, and (3) data analysis.

(1) Competent groups of raters

As noted by Dunn et al. (1999), the instrument developer should provide some information about characteristics of the judges chosen to assess item relevance. Which groups of raters are competent for rating these attitude items?

First competent group are the *high school students* themselves. A common practice in qualitative stage of scale development is to obtain some sort of feedback from the target group. Nevertheless, it is not quite obvious, whether students are really competent to decide about item relevance, because they don't have clear notion about the construct of attitudes. On the other hand, they provide us with useful information about comprehension of the items. The adequacy of using students as raters has been further discussed in several studies. Some studies suggest to use them (e.g., Jackson et al., 1999), whereas some others do not (e.g., Bonetti, 1994).

Therefore, some adult persons with the knowledge of the same physical educational context are needed to be included in the panel of raters. These adult persons are the high school *teachers of physical education*, who constitute the second group of raters. Teachers can decide whether the items are comprehensible as well as whether items represent the construct or not.

Finally, the members of the third group are the professional investigators with PhD degree sampled from the same faculty of physical education. They have experiences with similar scale development studies and, thus, they can provide us useful information about which items tending to be problematic. They have very good knowledge of the construct but poor knowledge of students and their disposition understand items' meanings.

(2) Measurement design

The data collection design is depicted in Figure A1. A total of 9 raters (*r*) nested within 3 competent groups (*g*) rated all 40 preliminary attitude items (*i*) on 2 occasions (*o*). The time interval was 5 months between first and second evaluation procedure.

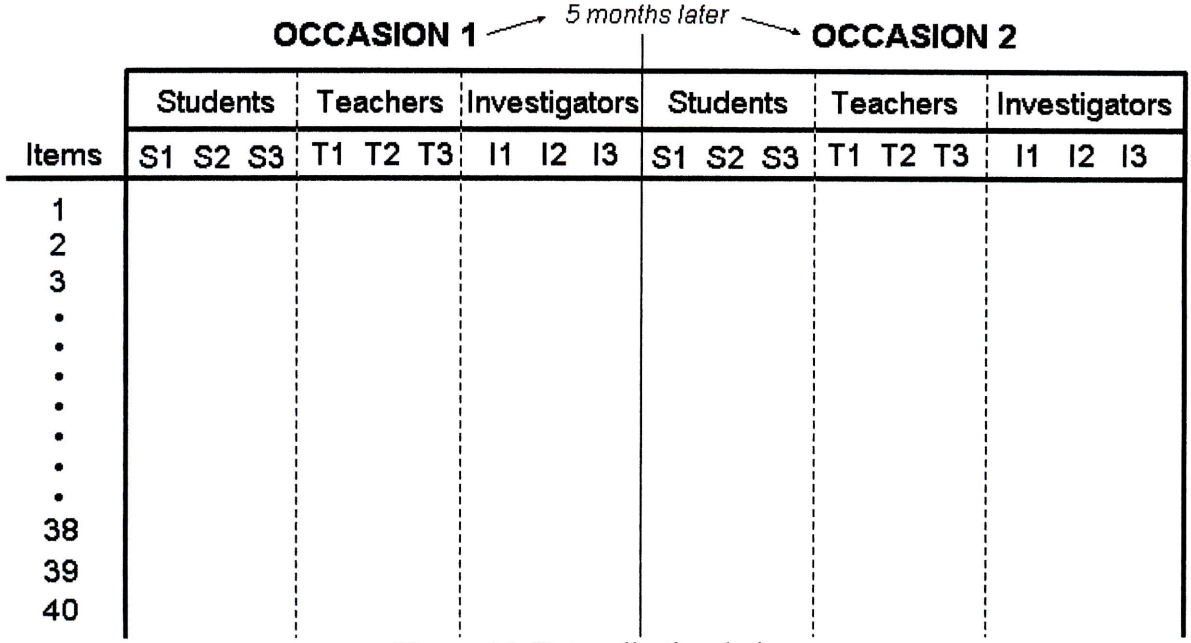


Figure A1. Data collection design
Note. S = student, T = teacher, I = investigator

All items were again administered after 5 months in different item ordering. We supposed that these arrangements were sufficient in that sense, that the raters didn't remember their responses to corresponding items. A generalizability analysis with occasions as a facet assumes that behavior remains constant over occasions (Shavelson, Webb, & Rowley, 1989). In this study, it is reasonable to assume that this assumption was not violated.

(2) Data analysis

The present data can be viewed from univariate and/or multivariate perspectives. The multivariate generalizability analysis is more comprehensive that the univariate one; nevertheless; we suggest to carry out the univariate analysis first to have some preliminary notion about the variance decomposition among effects. The univariate *G*- and *D*- studies results may provide an empirical support for accomplishing the multivariate generalizability analysis. Therefore, the data were analyzed in the following steps:

1. Univariate *G*-study
2. Univariate *D*-study for 1 average group of raters
3. Multivariate *G*-study
4. Univariate *D*-study for each fixed level of the group facet
5. Multivariate *D*-study for 9 raters total and 1 occasion

Univariate *G*-study is performed to estimate the variance components for each source of variability. The total observed score variance $Var(X_{ior:g})$ is decomposed into eleven components – four main effects and seven interactions:

$$Var(X_{ior:g}) = \sigma^2(i) + \sigma^2(o) + \sigma^2(g) + \sigma^2(r:g) + \sigma^2(io) + \\ + \sigma^2(ig) + \sigma^2(ir:g) + \sigma^2(og) + \sigma^2(or:g) + \sigma^2(iog) + \sigma^2(ior:g) \quad (37)$$

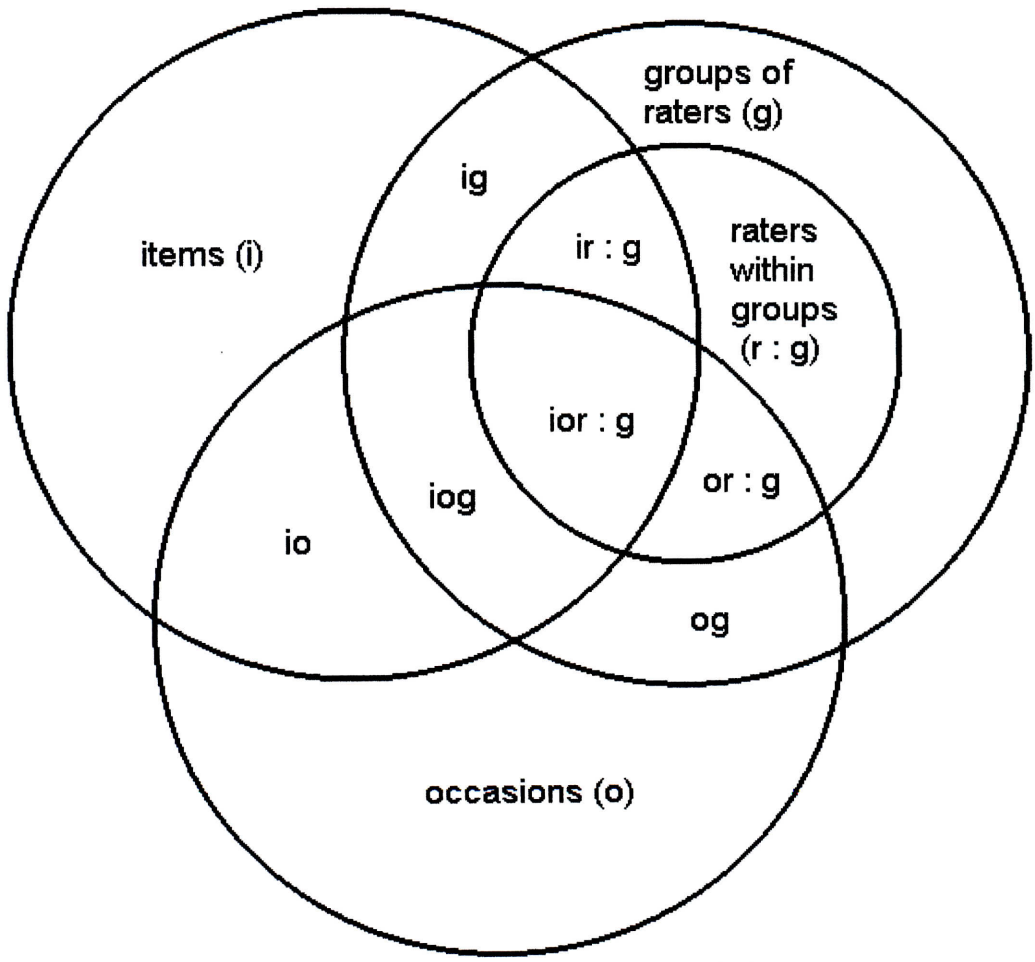


Figure A2. Venn diagram for the *ior*:g design.

$$\hat{\Phi}_1 = \frac{\hat{\sigma}^2(i) + \frac{\hat{\sigma}^2(io)}{1} + \frac{2\hat{\sigma}^2(ig)}{3} + \frac{2\hat{\sigma}^2(iog)}{3}}{\hat{\sigma}^2(i) + \frac{\hat{\sigma}^2(io)}{1} + \frac{2\hat{\sigma}^2(ig)}{3} + \frac{2\hat{\sigma}^2(iog)}{3} + \frac{\hat{\sigma}^2(r:g)}{n'_r} + \frac{\hat{\sigma}^2(i) + \frac{\hat{\sigma}^2(io)}{1} + \frac{2\hat{\sigma}^2(ig)}{3} + \frac{2\hat{\sigma}^2(iog)}{3}}{+ \frac{\hat{\sigma}^2(ir:g)}{n'_r} + \frac{\hat{\sigma}^2(or:g)}{n'_r} + \frac{\hat{\sigma}^2(ior:g)}{n'_r}}} \quad (39)$$

The questions related to sampling from a finite universe and hidden facets are discussed in more details in Brennan (2001a, p. 147-153) or in Brennan (1992, p. 65 and 85). Generalizability coefficient in *io*(*r*:*g*) design for groups (*g*) finite universe size *N'g* = 3 is estimated by:

$$E\hat{\rho}_2^2 = \frac{\hat{\sigma}^2(i)}{\hat{\sigma}^2(i) + \frac{3-n'_g}{3}\cdot\frac{\hat{\sigma}^2(ig)}{n'_g} + \frac{\hat{\sigma}^2(io)}{n'_o} + \frac{\hat{\sigma}^2(ir:g)}{n'_r n'_g} + \frac{3-n'_g}{3}\cdot\frac{\hat{\sigma}^2(iog)}{n'_o n'_g} + \frac{\hat{\sigma}^2(ior:g)}{n'_o n'_g n'_r}} \quad (40)$$

and the corresponding *phi* coefficient is given by:

$$\hat{\Phi}_2 = \frac{\hat{\sigma}^2(i)}{\hat{\sigma}^2(i) + \frac{\hat{\sigma}^2(o)}{n'_o} + \frac{3-n'_g}{3}\cdot\frac{\hat{\sigma}^2(g)}{n'_g} + \frac{\hat{\sigma}^2(r:g)}{n'_r n'_g} + \frac{3-n'_g}{3}\cdot\frac{\hat{\sigma}^2(ig)}{n'_g} + \frac{\hat{\sigma}^2(io)}{n'_o} + \frac{\hat{\sigma}^2(i)}{+ \frac{\hat{\sigma}^2(ir:g)}{n'_r n'_g} + \frac{3-n'_g}{3}\cdot\frac{\hat{\sigma}^2(og)}{n'_o n'_g} + \frac{\hat{\sigma}^2(or:g)}{n'_o n'_r n'_g} + \frac{3-n'_g}{3}\cdot\frac{\hat{\sigma}^2(iog)}{n'_o n'_g} + \frac{\hat{\sigma}^2(ior:g)}{n'_o n'_g n'_r}} \quad (41)$$

Multivariate *G*-study output provides us information about variance components estimated separately for students, teachers, and investigators. Consequently, several *D*-studies are conducted within each group to estimate the error variances, *G*, and *phi* coefficients for different number of occasions and raters. Generalizability coefficient in *ior* design for each fixed group of raters is estimated by following formula:

$$E\hat{\rho}_3^2 = \frac{\hat{\sigma}^2(i)}{\hat{\sigma}^2(i) + \frac{\hat{\sigma}^2(io)}{n'_o} + \frac{\hat{\sigma}^2(ir)}{n'_r} + \frac{\hat{\sigma}^2(ior)}{n'_o n'_r}}, \quad (42)$$

and the corresponding *phi* coefficient is estimated by:

$$\hat{\Phi}_3 = \frac{\hat{\sigma}^2(i)}{\hat{\sigma}^2(i) + \frac{\hat{\sigma}^2(o)}{n'_o} + \frac{\hat{\sigma}^2(r)}{n'_r} + \frac{\hat{\sigma}^2(io)}{n'_o} + \frac{\hat{\sigma}^2(io)}{n'_o} + \frac{\hat{\sigma}^2(or)}{n'_o n'_r} + \frac{\hat{\sigma}^2(ior)}{n'_o n'_r}}. \quad (43)$$

Based on multivariate *G*-study estimates, the multivariate *D*-study is accomplished to optimize the best rate of raters within each group to achieve either minimal composite error variances or maximal composite *G* coefficient. At the same time, the number of occasion is set to 1 and the total number of raters from all groups remains 9. The formula for estimating a composite universe score variance is:

$$\sigma_C^2(\tau) = \sum_v w_v^2 \sigma_v^2(i) + \sum_v \sum_{v'} w_v^2 w_{v'}^2 \sigma_{vv'}(i), \quad (44)$$

where $v \neq v'$ are indices for teachers, students, and investigators; w is the respective nominal weight for each fixed group, and $\sigma_{vv'}(i)$ is the universe score covariance between groups. Subsequently, the composite relative error variance is given by:

$$\hat{\sigma}_C^2(\delta) = \sum_v w_v^2 \hat{\sigma}_v^2(\delta) + \sum_v \sum_{v'} w_v^2 w_{v'}^2 \hat{\sigma}_{vv'}(\delta), \quad (45)$$

where $\hat{\sigma}_{vv'}(\delta) = \hat{\sigma}_v^2(\delta)$ iff $v = v'$ is an element of the Σ_δ matrix defined as:

$$\hat{\sigma}_{vv'}(\delta) = \frac{\hat{\sigma}_{vv'}(io)}{n'_o} + \frac{\hat{\sigma}_{vv'}(ir)}{n'_r} + \frac{\hat{\sigma}_{vv'}(ior)}{n'_o n'_r}. \quad (46)$$

Finally, the multivariate generalizability coefficient is given by:

$$E\hat{\rho}_4^2 = \frac{\hat{\sigma}_C^2(\tau)}{\hat{\sigma}_C^2(\tau) + \hat{\sigma}_C^2(\delta)}. \quad (47)$$

Results

Univariate *G*-study results

The fully random *G*-study results for the *ior:g* design are presented in Table A1. The degrees of freedom, mean squares, estimated variance component, and the percentage of the total variability are shown here for each of 11 sources of variability.

Table A1. Univariate *G*-study results: Mean squares, estimated variance components, and percentage of the total variance for the full-random *ior*:*g* design

Source of Variation	df	Mean Squares	Variance Components	Estimated Variance Components	Percentage of Total Variance
Items (*i*)	39	13.743	$\sigma^2(i)$	.609	18.0
Occasions (*o*)	1	.868	$\sigma^2(o)$	.000	.0
Groups of Raters (*g*)	2	92.668	$\sigma^2(g)$	.256	7.6
Raters within Groups (*r*:*g*)	6	27.730	$\sigma^2(r:g)$	.210	6.2
Items x Occasions (*io*)	39	2.093	$\sigma^2(io)$	.108	3.2
Items x Groups (*ig*)	78	3.730	$\sigma^2(ig)$	.159	4.7
Items x Raters within Groups (*ir*:*g*)	234	1.818	$\sigma^2(ir:g)$	.347	10.3
Occasions x Groups (*og*)	2	12.851	$\sigma^2(og)$	.014	.4
Occasions x Raters within Groups (*or*:*g*)	6	10.219	$\sigma^2(or:g)$	.227	6.7
Items x Occasions x Groups (*iog*)	78	2.085	$\sigma^2(iog)$	.320	9.5
Items x Occasions x Raters within Groups (*ior*:*g*)	234	1.125	$\sigma^2(ior:g)$	1.125	33.2

The items main effect accounted for 18% of the total variance. This variability was quite high due to the systematic differences among attitude items in their "true" relevance. The groups main effect (7.6%) was slightly higher than the raters within groups main effect (6.2%). This indicates a substantive clustering effect. The occasion main effect and occasions-by-groups interactions are negligible. By contrast, the items-by-occasions (3.2%) and the occasions-by-raters-within-groups (6.7%) interaction contributed considerable to the total variance. This is very interesting result which means that the rank order of mean values of the raters within groups varied substantially across occasions. The items-by-raters-within-groups interaction (10.3%) was more than two times higher than the items-by-groups interaction (4.7%). This means that the rank order of items was more unstable across raters within groups than across groups. The last three-way interaction accounted for most of the total variance (33.2%); nevertheless; this component is always confounded by unmeasured and random sources of variation (Shavelson, & Webb, 1991).

Univariate *D*-study results

The estimated sources of variance from the *G*-study are applied in a *D*-study. The *D*-study results are presented in Table A2 and A3. The *G* and *phi* coefficients are provided for occasions and groups fixed to1 in Table A2. The rater facet is considered as the only one

source of measurement error, whereas the occasions and groups are considered as hidden facets here. This means that former $i \times o \times (r : g)$ is now reduced to $i \times r$ design with enhanced universe, because items-by-occasions, items-by-groups, and items-by-occasions-by-groups interactions contributed to the universe-score variance. As a result, occasions and groups are considered as hidden facets in Table A2.

Table A2. Univariate *D*-study results: Generalizability coefficients and indices of dependability for different numbers of raters

Number of Raters	Generalizability Coefficient	Index of Dependability
3	.68	.62
5	.78	.73
7	.83	.79
9	.86	.83
11	.89	.86
13	.90	.88
15	.91	.89

Note. Occasion facet and group facet are hidden (fixed to 1).

Since group is sampled from a finite universe, a *correction factor* (Brennan, 2001a, p. 142) was used here. In this study, the finite universe correction factor for 3 groups was 2/3. The *G* coefficient equals to .86 for the present data collection design with 9 raters included. This value can be obtained by substituting estimated variance components from Table A1 to equation (38):

$$E\hat{\rho}_1^2 = \frac{.609+.108+.106+.213}{.609+.108+.106+.213+\frac{.347}{9}+\frac{1.125}{9}} = \frac{1.036}{1.036+.164} = .86$$

On the other hand, the coefficients presented in Table A3 are substantially lower. The reason is lower universe-score variance, because only items are source of variation in the universe here. The influence of occasions and groups on reliability is taken into account here. The occasions are now considered to be randomly sampled from infinite universe and only one group is sampled from finite universe size equals to 3. Hence, the same correction factor was used for all variance components containing group facet. The *G* coefficient equals to .51 for 9 raters, 1 occasion and 1 average group of raters. This means, that reliability decreases from .86 to .51 only, if we have an access to the information about contribution of occasions and other two competent groups. For comprehensive information, the *phi* coefficients for absolute decisions are also presented in Table A3.

Table A3. Univariate *D*-study Results: Generalizability Coefficients and Indices of Dependability for Different Numbers of Occasions and Raters Sampled from One of Three Fixed Groups.

Number of Occasions (infinite universe)	**Number of Raters** (infinite universe)	**Number of Groups** (finite universe size = 3)	**Generalizability Coefficient**	**Index of Dependability**
1	5	1	.46	.38
1	7	1	.49	.41
1	9	1	.51	.43
1	11	1	.52	.44
1	13	1	.53	.45
2	5	1	.58	.46
2	7	1	.61	.50
2	9	1	.62	.51
2	11	1	.64	.52
2	13	1	.64	.53

For better illustration, the value of .51 (see Table A3) can be obtained when the estimated variance components (see Table A1) are substituted to equation (40), where n'_r=9, n'_o=1, n'_g=1:

$$E\hat{\rho}_2^2 = \frac{.609}{.609 + \frac{2}{3}\cdot\frac{.159}{1} + \frac{.108}{1} + \frac{.347}{9} + \frac{2}{3}\cdot\frac{.309}{1} + \frac{1.125}{9}} = \frac{.609}{1.200} = .51.$$

Multivariate *G*-study results

Till now, the raters within all three groups have been considered to have the same magnitude of all estimated variance components. However, this assumptions need not to be met in practice. The multivariate generalizability analysis was therefore used to estimate variance components within- and covariance components between all three groups of raters. The main purpose of the multivariate analysis is to optimize number of raters within each group to obtain maximal composite generalizability with minimal total number of raters. The estimated variance and covariance components of the $i^{\bullet} \times o^{\bullet} \times r^{\circ}$ design are presented in Table A4. A total of 7 covariance matrices are presented here. First is the universe-score covariance matrix. Estimated variances on diagonal indicate that the relevance of items differs least among high school students (.318), two times more among investigators (.593), and mostly among physical educational teachers (1.232). The highest correlation is observed between investigators and teachers (.960), slightly smaller is the correlation between students and teachers (.872) and least are correlated students and investigators (.691).

Table A4. Multivariate *G*-study results: Variance and covariance components for the item×occasion×rater analysis with groups of raters as a fixed facet

Effect	Teachers	Students	Investigators
Item	**1.232**	*.872*	*.960*
(Σ_i matrix)	.546	**.318**	*.691*
	.821	.300	**.593**
Occasion	**-.088** (.059)		
(Σ_o matrix)	.005	**-.019** (.065)	
	-.005	-.100	**.049**
Rater	**-.059** (.105)		
(Σ_r matrix)		**-.044** (.080)	
			.733
Item × Occasion	**.205**		
(Σ_{io} matrix)	.004	**.382**	
	-.168	.167	**.376**
Item × Rater	**.542**		
(Σ_{ir} matrix)		**.086**	
			.412
Occasion × Rater	**.224**		
(Σ_{or} matrix)		**.168**	
			.290
Item × Occasion × Rater	**1.043**		
(Σ_{ior} matrix)		**1.112**	
			1.222

Note. Bold entries are the variances, italic entries are the correlations, and the entries in parentheses are the standard errors for negative variance estimates.

Too, some negative estimates of variance components occurred here (see Table A4). For all of these cases, the standard errors are presented. If the standard errors are sufficiently large, the negative values are suggested to be fixed to zero (Shavelson, & Webb, 1991). This criterion seems to be met in the present study. Based on the multivariate *G*-study results, we can conclude that the groups of raters differ substantially from each other with respect to the relative and absolute error variances and universe-score variances. For better illustration, the sets of *D*-studies have been separately conducted for each level of the fixed group facet.

Univariate *D*-study results for each level of the fixed facet

D-study results are presented in Figure A3. The *G*-coefficients (on the left-hand three panels) and *phi* indices (on the right-hand three panels) are provided for teachers (the upper two panels), students (the middle two panels), and for investigators (the bottom two panels).

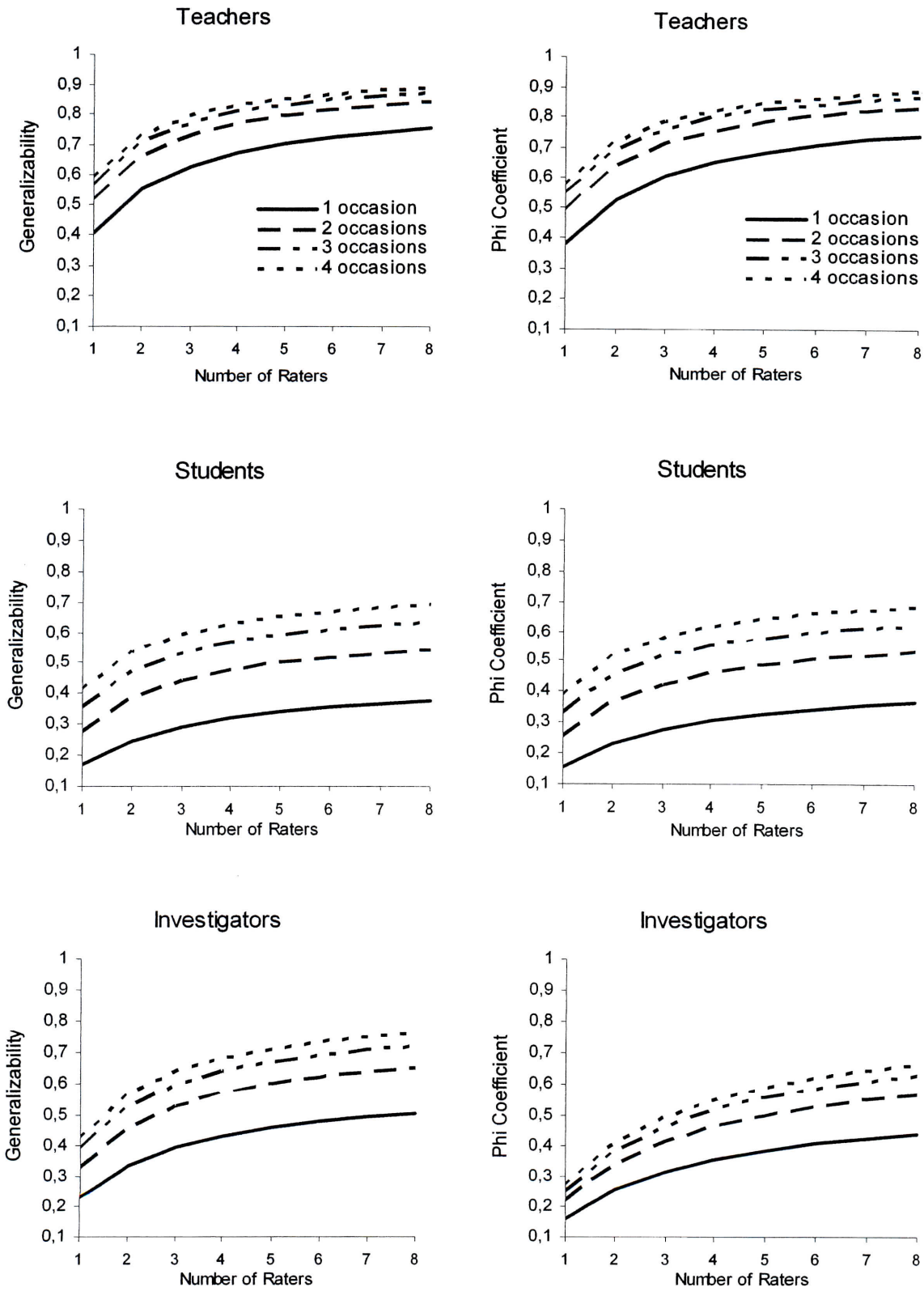


Figure A3. Generalizability and Phi coefficients for each of the three groups of raters for number of raters ranged from 1 to 8 and number of occasions ranged from 1 to 4

The highest coefficients are observed among teachers. The teachers are very consistent across occasions; nevertheless; using more teachers leads to substantial improvement of the coefficients. Thus, the high school teachers are not so consistent compared to other two groups and the coefficients are so high due to the large universe-score variance. The *phi* coefficients are only slightly lover than the *G* coefficients here. Formulas for estimating all coefficients are given in equations (42) and (43).

The curves for students are on average situated much lower than for teachers, what is most likely caused by the four times lover universe-score variance. They are also more flat, what indicates that students are more consistent across raters. The distance among curves is larger than for teachers. This means that students are less consistent over occasions and using large numbers of occasions still increase generalizability. Also here, the *phi* coefficients are only slightly lower than the respective *G* coefficients. The *G* coefficients for investigators are (on average) somewhere in the middle compared to other two groups. On the other hand, the *phi* coefficients are remarkable lower compared to *G* coefficients which indicates that the investigators are less consistent if the absolute interpretation of the test scores is of interest.

Multivariate *D*-study results

As was shown above, the groups of raters differ substantially from each other with respect to their universe-score variances and error variances. Therefore, not necessarily the same number of raters within each group is needed to obtain the highest level of generalizability. The composite relative error variance and the composite *G* coefficient for different combinations of teachers, student, and investigators are presented in Table A5. In all the *D*-studies, the number of random occasion is set to 1 and the sum of all raters is 9 (with at least 2 raters within each group). The composite *G* coefficients ranged from .60 to .75 and for the present data collection design with 3 raters within each group equals to .69. The highest generalizability of .75 is observed either for 4 teachers, 2 students, and 3 investigators or for 5 teachers and 2 students and 2 investigators. We suggest preferring the first combination, due to the lower composite error variance (.2450).

Table A5. Multivariate *D*-study results: Composite relative error variances and multivariate *G*-coefficients for different number of raters within each group

Number of Teachers	Number of Students	Number of Investigators	Total Number of Raters	Composite Relative Error Variance	Composite Generalizability Coefficient
2	2	5	9	.3147	.65
2	3	4	9	.3077	.64
2	4	3	9	.3105	.62
2	5	2	9	.3254	.60
3	2	4	9	.2691	.71
3	3	3	9	.2714	.69
3	4	2	9	.2845	.67
4	2	3	9	.2450	.75
4	3	2	9	.2576	.72
5	2	2	9	.2693	.75

Note. Number of occasions equal to 1 for all performed D-studies.

Discussion

Three research questions were addressed in the present extending study. First, the influence of occasions on ratings of item relevance has been examined. Generalizability is always lower, if we generalize to the broader universe. The universe where raters are only one source of error is restrictive, because generalization is made only across raters, and not over occasions and/or groups. If we ignore the occasion facet as a source of error, the reliability is overestimated about .1 (see Table A3) in the present study.

Second research question addressed a similar problem, where the effect of other existing competent groups of raters on reliability has been investigated. To which extent is the reliability overestimated, if a scale developer includes only the raters from one of all competent groups. It is obvious that different research domains may have different number of groups of raters which are competent to evaluate the item relevance. In the present study, three competent groups of raters have been identified. The influence of the groups on reliability seems to be higher than the influence of occasions. The reliability is overestimated about .25 (see Tables A2 and A3), if the two other fixed groups are not taken into account.

Generally speaking, the first two questions are related to the topic of the hidden facets. Two hidden facets were considered in the present study – the occasions and groups of raters. Their common influence decreases reliability about .35. For the present data collection design, the *G* coefficient decreases from .86 to .51, if we added the information about the contribution of these two facets to the total variability. This reduction in coefficient size is

caused by expansion of very restrictive to much broader universe, in which generalization of scores was realized not only across raters, but also across occasions and groups.

The last purpose of this study was to examine the variability of importance of particular group. When raters from all groups are involved in a study, the question is how many raters from each group are needed to obtain maximal composite generalizability. In the current study, the composite *G* coefficient can be improved from .69 to .75 (see Table A5), if one more physical educational teacher instead of one student is used in a panel of experts on 1 occasion.

Based on the present findings, respective scale developers in the stage of designing the ratings of item relevance should answer the following questions:

1. Is it reasonable to consider the influence of the occasions to be trivial?
2. Are really all competent groups of raters included in the panel of judges?
3. If all competent groups are present at the panel, is it necessary to have the same number of raters within each group?

FOLLOW-UP STUDY

Drinking and smoking among adolescents in Prague

Together with the ATPE, alcohol consumption and cigarette smoking have been investigated conducting a joint questionnaire survey. The ATPE inventory was supplemented by the questions related to the students´ smoking and drinking behavior. We do not intend this study to present complete and rigorous account of the alcohol and smoking problems. Instead, the present research should serve as useful information for interested investigators. Therefore, the results of this study are presented in the *preliminary communication* style. In scientific journals, the preliminary communications are short articles describing essential findings, which should be published rapidly. Hence, the sections like methods, discussion and conclusion are not incorporated here. We believe that the current findings can be still considered as actual. Furthermore, the World Health Organization has prioritized continuous monitoring of health problems attributable to alcohol over the last years (WHO, 2008).

For these reasons, we have decided to include the current follow-up study in the Appendix of the present thesis. The results are divided into two parts. First, basic descriptive statistics of drinking and smoking behavior are presented. Subsequently, the multilevel analysis of alcohol consumption is *outlined* in the second part.

Descriptive statistics

The alcohol consumption has been investigated using some parts of the ESPAD (European Survey Project on Alcohol and other Drugs) methodology (ESPAD, 2007). Here, the amount of alcohol within one typical drink can be expressed using following *approximate* equation:

1 "drink" = 0.5l of beer = 0.2l of wine = 0.5l of alcopops = 0.05l of spirits

The most fundamental question was how many total abstainers were involved in the study. The ratio of consumers to abstainers is depicted in Figure B1. A person is considered as abstainer if he or she had not consumed one whole "drink" during last 3 months. As we can see, it makes really sense to deal with this issue, because the ratio of consumers is substantial.

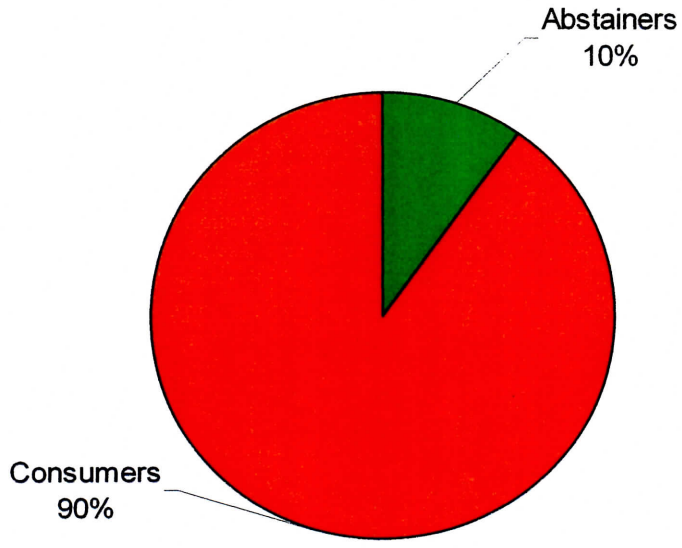


Figure B1. Alcohol "drinks" during last 3 months (N=1157)

The alcohol consumption during the last week (until the day of survey) is depicted in Figure B2. Respondents were guided to specify only those days, when they started with drinking (if it persisted all over the night). As we can see, the respondents drank most frequently on Saturday and Friday. This finding is not surprising, because people are usually drinking alcohol during the weekend. Subsequently, the frequency of drinking during the last week is showed in Figure B3.

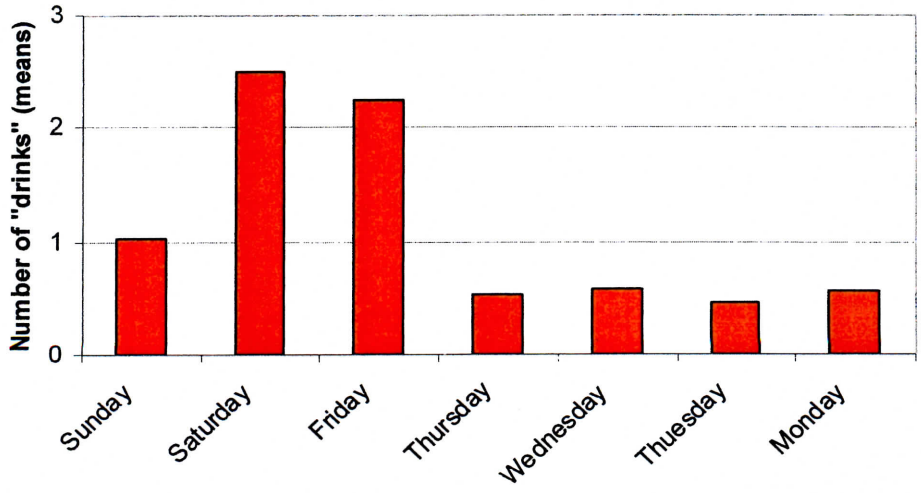


Figure B2. Average alcohol consumption sorted by days in the last week (N=1157)

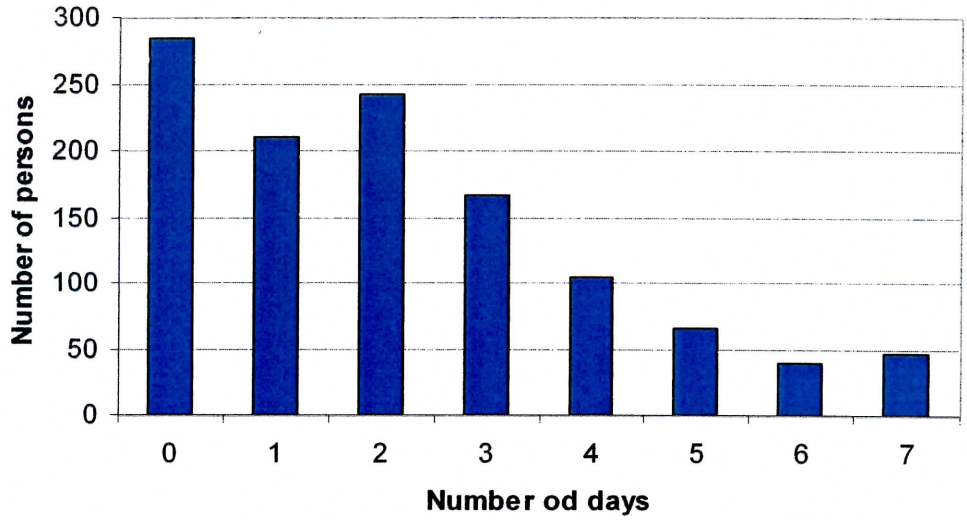


Figure B3. Frequency of alcohol consumption: number of drinking days during the last week (N=1157)

Next two figures represent mean number of "drinks" consumed during last week before the survey. Firstly, the mean weekly alcohol consumption sorted by age is depicted in Figure B4. Subsequently, the mean weekly consumption is sorted by gender and depicted in Figure B5. The findings indicate that gender and age might be relevant predictors of alcohol behavior in the multilevel regression analysis presented later on.

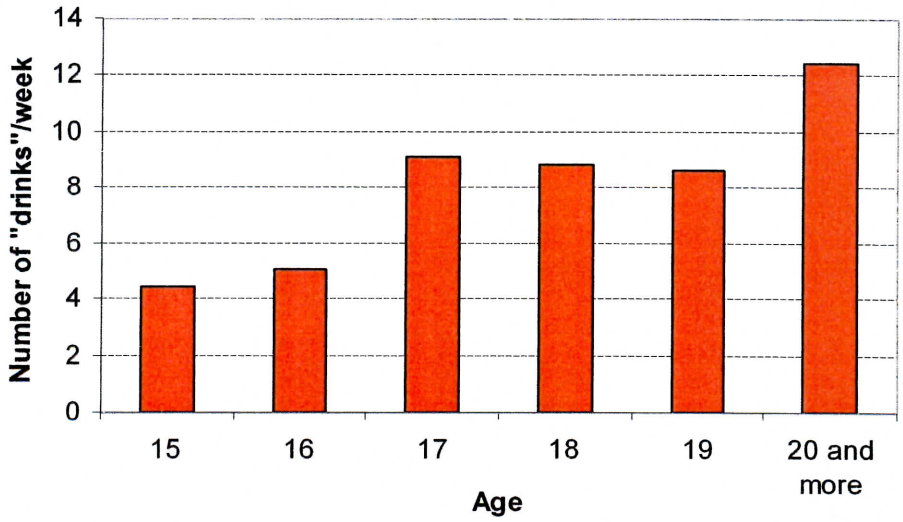


Figure B4. Mean weekly alcohol consumption sorted by age (N=1157)

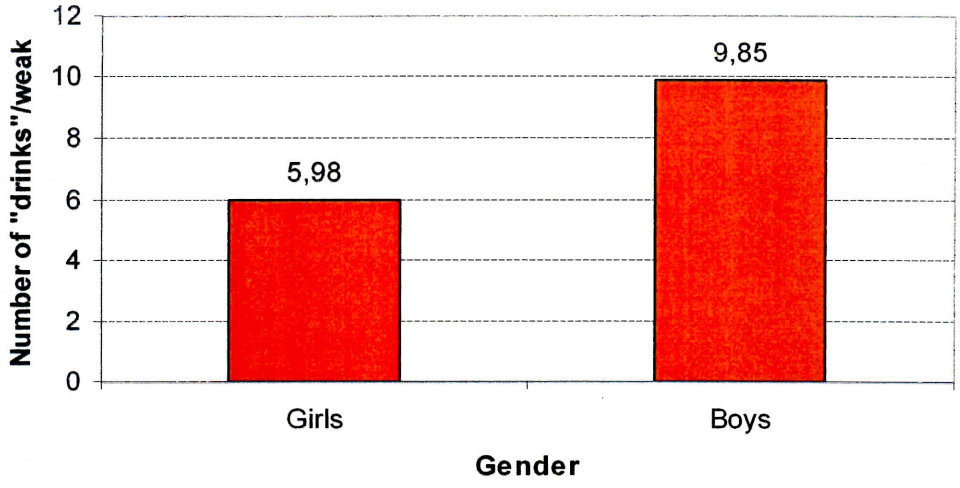


Figure B5. Mean weekly alcohol consumption sorted by gender (N=1157)

Basic statistics related to smoking are presented in following four figures. Smoking of the commonly accessible tobacco cigarettes was investigated here. First, the total percentage of smokers and non-smokers is pictured in Figure B6. Next, the ratios of smokers to non-smokers sorted by gender and age are illustrated in Figures B7 and B8, respectively.

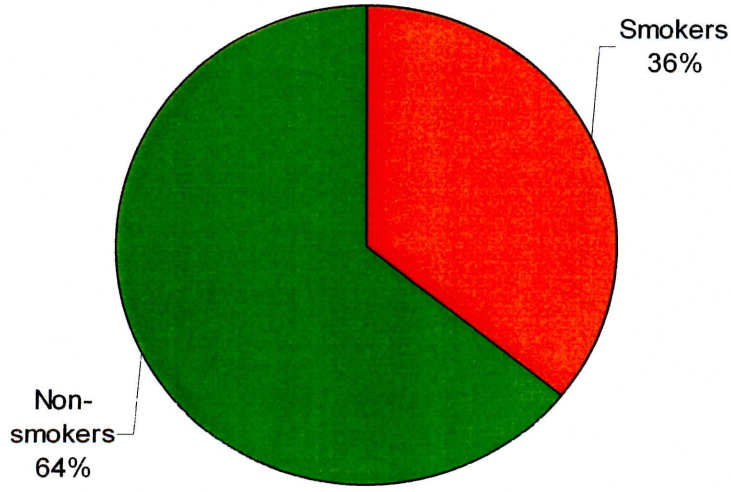


Figure B6. Total percentage of smokers and non-smokers (N=1157)



Figure B7. Percentage of smokers and non-smokers sorted by gender (N=1157)

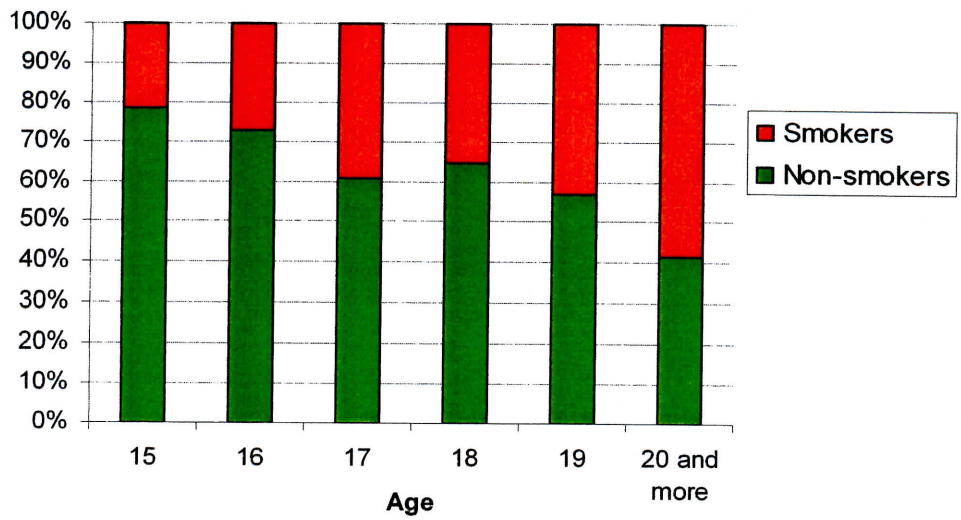


Figure B8. Percentage of smokers and non-smokers sorted by age (N=1157)

It might seem that age is a significant predictor of smoking and drinking behavior. However, we must still keep in mind that the design of this study is cross-sectional and not longitudinal. Hence, this relationship is contaminated by the inter-individual variability. Finally, the severity of tobacco smoking among all 413 smokers is presented in Figure B9. The response categories were chosen as follows: a) 1 to 4 cigarettes per day; b) 5 to 10 cigarettes per day; c) 11 to 20 cigarettes per day; d) more than 20 cigarettes per day.

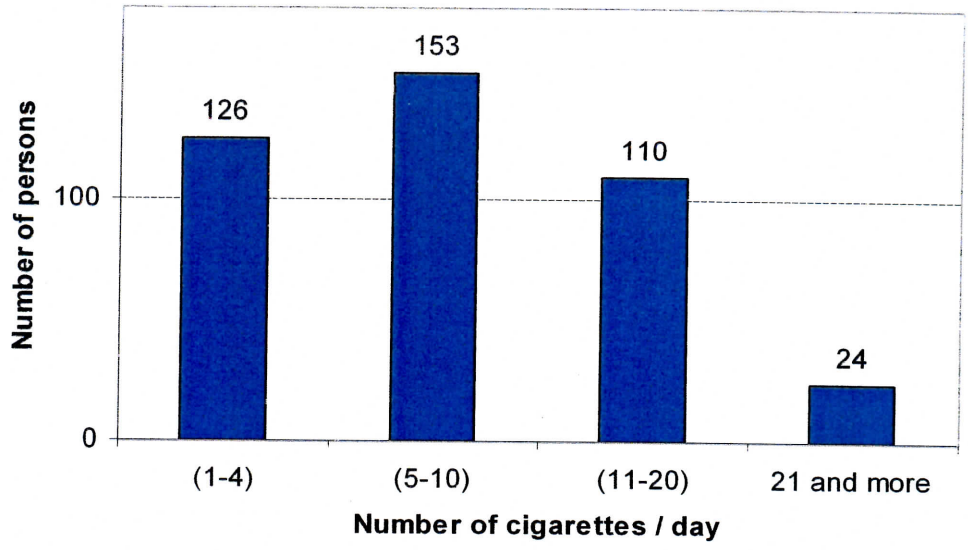


Figure B9. Number of cigarettes smoked per day by smokers (N=413)

Multilevel analysis of alcohol drinking

Up until now, only two-level structure has been taken into account in the main study. However, the data structure is three-level. There is an outcome variable at the first (the lowest) level. In principle, there may be also independent variable at all levels. In our case, the levels and respective variables are given as follows:

Level 1:

Units: 1157 students

Variables: weekly alcohol consumption (outcome variable), gender, smoking

Level 2:

Units: 58 classrooms

Variables: averaged age (classrooms means)

Level 3:

Units: 11 schools

Variables: percentage of students accepted by some university

The students are nested within classrooms (not within PE classes now) that are further nested within schools. All analyses were conducted using HLM software (Raudenbush et al., 2004). Three-level analysis includes two consecutive steps: 1. fully unconditional model, 2. conditional models.

1. Unconditional model

The simplest three-level model is fully unconditional. This means that no independent variables are specified at any level. Such a model represents how variation in an outcome measure is allocated across the three different levels (student, classroom, and school). We model weekly alcohol consumption of each student as a function of a classroom mean plus a random error:

$$Yi_{jk} = \pi_{0jk} + e_{ijk}\,, \tag{48}$$

where Yi_{jk} is the number of "drinks" per week consumed by student i in classroom j and school k, π_{0jk} is the mean weekly alcohol consumption of classroom j in school k, and e_{ijk} is the deviation of the ijk's score from the classroom mean. The indices i, j, and k denote students, classrooms, and schools where there are

$i = 1, 2, ..., n_{jk}$ students within classroom j in school k
$j = 1, 2, ..., J_k$ classrooms within school k
$k = 1, 2, ..., 11$ schools.

Each classroom mean, π_{0jk}, is considered as an outcome varying randomly around some school mean:

$$\pi_{0jk} = \beta_{00k} + r_{0jk}\,, \tag{49}$$

where β_{00k} is the mean weekly alcohol consumption in school k and r_{0jk} is the deviation of classroom jk's mean from the school mean. Subsequently, we view the school means, β_{00k}, as varying randomly a grand (overall) mean:

$$\beta_{00k} = \gamma_{000} + u_{00k}, \tag{50}$$

where γ_{000} is the grand mean and u_{00k} is the deviation of the school k's mean from the grand mean.

The results of the fully unconditional model are presented in Table B1. There is only one fixed effect (grand mean γ_{000}) and three random effects (e_{ijk}, r_{0jk}, and u_{00k}) The variances of these three random effects are of our interest now. In a two-level case, the intraclass correlation is estimated. Analogously, the variance partitioning among all levels is accomplished in our three-level case.

Table B1. Estimation of fixed and random effects of the unconditional three-level model

Random effect	**Variance**	**df**	χ^2	**P-value**	**Variance decomposition** *(Percentage by level)*
e	88.5	--	--	--	87.8 (level 1)
r_0	8.1	47	130	0.00	8.0 (level 2)
u_{00}	4.2	10	34	0.00	4.2 (level 3)
Fixed effect	**Coefficient**	**S.E.**	**T-ratio**	**P-value**	
γ_{000}	7.73	0.79	9.73	0.00	

Note. df = degrees of freedom, S.E. = standard error, χ^2 = chi-square

The largest percentage of variance (87.8%) lies at level 1 (students). Smaller but still substantial percentage of variance (8.0%) lies at level 2 (classrooms). The smallest percentage of variance (4.2%) is situated at the level 3 (schools). The amount of variance at each level was found to be substantial (and also statistically significant ($p < 0.01$)). Therefore it has sense to estimate the variance of outcome by independent variables listed above at all three levels.

2. Conditional models

Part of the variability at each level can be estimated[4] by measured variables at each level. Firstly, the "gender" is added as an independent variable to the model (48). Consequently, we obtain:

$$(\text{ALCOHOL})_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{GENDER})_{ijk} + e_{ijk}, \qquad (51)$$

[4] We don't suggest using the term *explained,* because it could indicate causal relationship that cannot be proven using cross-sectional research design.

where π_{0jk} is the regression intercept, π_{1jk} is the gender slope, and e_{ijk} is residual regression variance. The coefficients π_{0jk} and π_{1jk} were allowed to vary randomly over classrooms but were not estimated by classroom- or school-level variables. Thus, the model at level 2 is:

$$\begin{aligned} \pi_{0jk} &= \beta_{00k} + r_{0jk} \\ \pi_{1jk} &= \beta_{10k} + r_{1jk} \end{aligned} \quad . \tag{52}$$

Similarly, the model at level 3 is:

$$\begin{aligned} \beta_{00k} &= \gamma_{000} + u_{00k} \\ \beta_{10k} &= \gamma_{100} + u_{10k} \end{aligned} \quad . \tag{53}$$

Overall, two fixed effects (γ_{000} and γ_{100}) and five random effects (e_{ijk}, r_{0jk}, r_{1jk}, u_{00k}, and u_{10k}) are introduced here. The estimates of these coefficients are given in Table B2.

Table B2. Estimation of fixed and random effects of the three-level model with *gender* as independent variable at level 1

Random effect	**Variance**	**df**	χ^2	**P-value**
e	85.1	--	--	--
r_0	7.9	40	110	0.00
r_1	15.6	40	82	0.00
u_{00}	4.3	9	31	0.00
u_{10}	1.4	9	8	>0.50
Fixed effect	**Coefficient**	**S.E.**	**T-ratio**	**P-value**
γ_{000}	7.69	0.80	9.65	0.00
γ_{100}	2.65	0.93	2.85	0.02

Note. df = degrees of freedom, S.E. = standard error, χ^2 = chi-square

Two important results are presented in Table B2. First, the gender slope (γ_{100}) is statistically significant ($p = 0.02$) independent variable at level 1 and varies substantially over classrooms but not over schools. The interpretation of this parameter is that boys are drinking alcohol more than girls and the difference between boys and girls is stable over schools but not over classrooms.

Secondly, the amount of variance of the outcome estimated by gender at level 1 can be computed as follows:

$$\% \text{ of reduction in variance} = \frac{88.5 \text{ (see Table B1)} - 85.1 \text{ (see Table B2)}}{88.5 \text{ (see Table B1)}} = 3.8\%$$

This means that only about 3.8% of variance situated at level 1 has been "explained" by gender. Therefore, more level-one variables are needed to account for the remaining residual variance. For illustrative purposes, the "smoking" variable has been added to the model (51). Consequently, we obtain:

$$(\text{ALCOHOL})_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{GENDER})_{ijk} + \pi_{2jk}(\text{SMOKING})_{ijk} + e_{ijk} \qquad (54)$$

At level 2 we have:

$$\begin{aligned} \pi_{0jk} &= \beta_{00k} + r_{0jk} \\ \pi_{1jk} &= \beta_{10k} + r_{1jk} \\ \pi_{2jk} &= \beta_{20k} + r_{2jk} \end{aligned} \qquad (55)$$

The random effect u_{10k} is not included in the following level-3 model, because we know from the last estimation that the gender slope doesn't vary over schools:

$$\begin{aligned} \beta_{00k} &= \gamma_{000} + u_{00k} \\ \beta_{10k} &= \gamma_{100} \\ \beta_{20k} &= \gamma_{200} + u_{20k} \end{aligned} \qquad (56)$$

The results of this three-level model are presented in Table B3.

Table B3. Estimation of fixed and random effects of the three-level model with *gender* and *smoking* as independent variables at level 1

Random effect	**Variance**	**df**	χ^2	**P-value**
e	69.53	--	--	--
r_0	9.64	39	139	0.00
r_1	12.88	49	91	0.01
r_2	2.24	39	66	0.00
u_{00}	3.03	10	32	0.00
u_{20}	0.05	10	13	0.20
Fixed effect	**Coefficient**	**S.E.**	**T-ratio**	**P-value**
γ_{000}	7.71	0.72	10.63	0.00
γ_{100}	3.12	0.78	4.09	0.00
γ_{200}	3.54	0.33	10.82	0.00

Note. df = degrees of freedom, S.E. = standard error, χ^2 = chi-square

The results again suggest that some simplification of the level-3 model is in order. The random effect of the "smoking" slope, u_{20}, is not significantly different from zero. The estimated smoking slope coefficient (3.54) indicates that smokers tend to drink more alcohol than non-smokers. This difference between smokers and non-smokers with respect to alcohol

drinking vary over classrooms but not over schools. In addition, some reduction of the level-1 residual variance was achieved by adding smoking into the level-1 model. The amount of estimated variance by two independent variables is now computed as:

$$\% \text{ of reduction in variance} = \frac{88.50 - 69.53}{88.50} = 21.4\%.$$

In the context of percentage of reduction of variance at level 1, similar analyses can be conducted at both other levels. Firstly, the "mean age" variable is added to the model (49) at level 2. Consequently, the percentage of school-leavers accepted by some university ("accepted" for short) is added to the model (50) as independent variable at level 3. The model (49) can be now written in following form:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(\text{MEAN AGE})_{jk} + r_{0jk}. \qquad (57)$$

The regression coefficients, β_{00k} and β_{01k} are expressed as outcomes at level 3. We obtain:

$$\begin{aligned} \beta_{00k} &= \gamma_{000} + u_{00k} \\ \beta_{01k} &= \gamma_{010} + u_{01k} \end{aligned}. \qquad (58)$$

The estimated parameters of this model are presented in Table B4.

Table B4. Estimation of fixed and random effects of the three-level model with *mean age* variable at level 2

Random effect	**Variance**	**df**	χ^2	**P-value**
e	88.59	--	--	--
r_0	6.62	36	108	0.00
u_{00}	2.10	10	38	0.00
u_{01}	0.12	10	3	>0.50
Fixed effect	**Coefficient**	**S.E.**	**T-ratio**	**P-value**
γ_{000}	7.73	0.79	9.80	0.00
γ_{010}	0.97	0.19	5.20	0.00

Note. df = degrees of freedom, S.E. = standard error, χ^2 = chi-square

The random effect u_{01} is negligible, which means that the *mean age* slope, γ_{010}, doesn't vary over schools. The amount of variance "explained" by *mean age* variable at level 2 can be calculated as:

$$\% \text{ of reduction in variance} = \frac{8.10 - 6.62}{8.10} = 18.3\%.$$

Following the same way of thinking, the *accepted* variable is added to the model (50) at level 3 as follows:

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(\text{ACCEPTED})_k + u_{00k} \,. \tag{59}$$

The results of the model above are presented in Table B5.

Table B5. Estimation of fixed and random effects of the three-level model with *accepted* variable at level 3

Random effect	**Variance**	**df**	χ^2	**P-value**
e	88.54	--	--	--
r_0	8.04	47	130	0.00
u_{00}	2.03	9	20	0.02
Fixed effect	**Coefficient**	**S.E.**	**T-ratio**	**P-value**
γ_{000}	10.47	1.34	7.84	0.00
γ_{001}	-0.05	0.02	-2.92	0.02

Note. df = degrees of freedom, S.E. = standard error, χ^2 = chi-square

As we can see, the estimation of the *accepted* slope coefficient, γ_{001}, is negative. Thus, the average weekly alcohol consumption decreases when the percentage of school-leavers accepted by some university increases. Furthermore, the percentage of variance "explained" by the *accepted* variable at level 3 is now:

$$\%\text{ of reduction in variance} = \frac{4.20 - 2.03}{4.20} = 51.7\% \,.$$

Finally, the most complex three-level model including all four independent variables can be formulated in this manner:

Level 1 (equation (54)):

$$(\text{ALCOHOL})_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{GENDER})_{ijk} + \pi_{2jk}(\text{SMOKING})_{ijk} + e_{ijk}$$

Level 2:

$$\begin{aligned} \pi_{0jk} &= \beta_{00k} + \beta_{01k}(\text{MEAN AGE})_{jk} + r_{0jk} \\ \pi_{1jk} &= \beta_{10k} + \beta_{11k}(\text{MEAN AGE})_{jk} + r_{1jk} \\ \pi_{2jk} &= \beta_{20k} + \beta_{21k}(\text{MEAN AGE})_{jk} + r_{2jk} \end{aligned} \tag{60}$$

Level 3:

$$\begin{aligned}
\beta_{00k} &= \gamma_{000} + \gamma_{001}(\text{ACCEPTED})_k + u_{00k} \\
\beta_{01k} &= \gamma_{010} \\
\beta_{10k} &= \gamma_{100} \\
\beta_{11k} &= \gamma_{110} \\
\beta_{20k} &= \gamma_{200} \\
\beta_{21k} &= \gamma_{210}
\end{aligned} \tag{61}$$

The estimated fixed and random effects of this model are presented in Table B6.

Table B6. Estimation of fixed and random effects of the final version of three-level model with *gender* and *smoking* as independent variables at level 1, *mean age* as independent variable at level 2, and *accepted* as independent variable at level 3

Random effect	**Variance**	**df**	χ^2	**P-value**
e	69.39	--	--	--
r_0	8.16	38	131	0.00
r_1	12.73	48	91	0.00
r_2	2.09	48	66	0.04
u_{00}	1.36	9	21	0.01
Fixed effect	**Coefficient**	**S.E.**	**T-ratio**	**P-value**
γ_{000}	7.59	0.63	12.11	0.00
γ_{001}	-0.05	0.01	-3.40	0.01
γ_{010}	0.97	0.18	5.39	0.00
γ_{100}	3.18	0.63	5.06	0.00
γ_{110}	0.47	0.47	0.99	0.32
γ_{200}	3.47	0.31	11.14	0.00
γ_{210}	-0.40	0.18	-2.20	0.03

Note. df = degrees of freedom, S.E. = standard error, χ^2 = chi-square

The results indicate that the γ_{110} coefficient is close to zero. This means that varying difference between boys and girls with respect to weekly alcohol consumption can not be "explained" by mean *age* variable at level 2. On the other hand, the relationship between smoking and drinking, γ_{210}, can be "explained" (to a certain degree) by mean *age*. This finding belongs to one of the most interesting results of the present follow-up study. The negative value of this coefficient (-0.40) has straightforward interpretation. As was mentioned above, smokers consumed more alcohol per week than non-smokers (expressed by γ_{200}=3.47). Increasing average *age* decreases this difference. This could be explained by the fact that the non-smokers tend to start with regularly drinking later than smokers.

Finally, let us recapitulate the variance of weekly drinking estimated by independent variables at each level. These values are summarized in Table B7.

Table B7. Percentage of estimated ("explained") variance of drinking by independent variables at each level

Level	Variables	% of variance "explained"
1	gender, smoking	21.4
2	mean age	18.3
3	% of accepted	51.7

The present follow-up study will continue with analyzing of relationships between ATPE, smoking, and alcohol consumption. However, such an analysis cannot be conducted using the same procedure presented above. Hierarchical linear regression modeling proposed by Raudenbush and Bryk (2002) assumes that all variables are measured without measurement error. Nevertheless, this assumption becomes unrealistic in the ATPE case. As was shown in the main study of the present thesis, the amount of error associated with all sub-constructs of ATPE is substantial. Therefore, the data will presumably be analyzed within the SEM framework using some SEM program. In all cases, this is a challenge for the future investigations.