Charles University in Prague

Faculty of Physical Education and Sport

Multilevel Structural Analysis of the Attitudes toward Physical Education: Methodological Considerations

An Extended Abstract of the Doctoral Thesis

written by

Mgr. Ondřej Pecha

supervised by

Prof. PhDr. Petr Blahuš, DrSc.

Prague, Czech Republic

2009

ABSTRACT

The present thesis 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

SOUHRN

Předložená disertační práce je metodologickou studií, ve které je prezentováno víceúrovňové strukturální modelování (VSM) jako vhodná metodologie v případech, ve kterých se výzkumník zabývá identifikací teoretických konceptů u závislých objektů pozorování. V našem případě se jednalo o měření postojů k tělesné výchově (PTV) u středoškolských studentů (n = 1157). Studenti dále patřili do skupin (m = 87), které měly společnou výuku tělesné výchovy. Při dvojúrovňové revalidizaci PTV bylo zjištěno, že faktorové struktury na vnitroskupinové (studenti) a meziskupinové úrovni jsou různé. Na vnitroskupinové úrovni byla potvrzena Wearova čtyřfaktorová teorie PTV, nicméně výsledky rovněž naznačují přítomnost dvou dalších faktorů, které souvisí s pozitivní a negativní orientací výroků v inventáři PTV. Na meziskupinové úrovni neexistovala teorie o struktuře inventáře PTV, protože ten byl původně vyvinut pouze pro měření PTV na vnitroskupinové úrovni. Výsledný třífaktorový model vykazoval na meziskupinové úrovni jen částečnou shodu s daty. Analýza dat byla rozdělena na čtyři případy: (1) data jsou jednoúrovňová a intervalová, (2) data jsou jednoúrovňová a ordinální, (3) data jsou víceúrovňová a intervalová, (4) data jsou víceúrovňová a ordinální. Odhadnuté parametry a jejich střední chyby byly porovnány a bylo zjištěno, že oba přístupy s ordinálními proměnnými produkují vyšší faktorové zátěže, vyšší střední chyby jejich odhadů a nižší jedinečnosti. Navíc zejména čtvrtý zmiňovaný přístup, VSM s ordinálními proměnnými, se ukázal být velmi výpočetně náročným nedoporučujeme jej aplikovat u složitých mnohofaktorových víceúrovňových modelů. V příloze disertační práce se nalézají dvě separátní metodologické studie, které navazují nebo rozšiřují zjištěné výsledky. První z nich se zabývá problematikou tzv. skrytých faset při hodnocení obsahové validity položek PTV. Druhá studie se pak zabývá tříúrovňovou analýzou konzumace alkoholu u stejných 1157 studentů.

Klíčová slova: Kinantropologie, postoje k tělesné výchově, strukturální modelování, víceúrovňová analýza, adolescenti, teorie zobecnitelnosti, konzumace alkoholu

INTRODUCTION

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. 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. 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.

Multilevel modeling

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. Let us 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.).

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).

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.

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.

Theory 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; 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).

In most kinanthropological 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. 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). 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 a 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. 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

***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.

METHOD

Sample

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. 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.

Instrument

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. 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.

All items have got three basic aspects. The summary of the aspects of all items is provided in Table 1. All items are listed in the same order as in the Appendix.

Table 1. 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).

Analysis Methods

In the present section, we introduce single-level SEM and especially CFA, multilevel extension of CFA, and SEM and CFA for ordinal variables. The ordinal variables paradigm within SEM framework is discussed in both single-level and two-level cases. 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. 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).

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).

The multilevel confirmatory factor analysis (MCFA) model assumes that sampling occurs at two levels: between groups and within groups. Total population covariance matrix$(\boldsymbol{\Sigma}_T)$contains both the between-group covariance matrix$(\boldsymbol{\Sigma}_B)$and the within-group covariance matrix$(\boldsymbol{\Sigma}_W)$. The process of MCFA involves five successive steps (Muthén, 1994):

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

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).

RESULTS

As a first step, basic univariate descriptive statistics were computed. 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) and kurtosis ranged from -1.39 (item 8) to 2.71 (item 23).

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). All four single-factor models exhibit acceptable fit. Only social domain of ATPE 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 2.

Table 2. 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 2, 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).

The estimation of the structural part of the final model at the within-group level is depicted in Figure 1. The χ^2 was 811, AIC was 1031, and SRMR was 0.032.

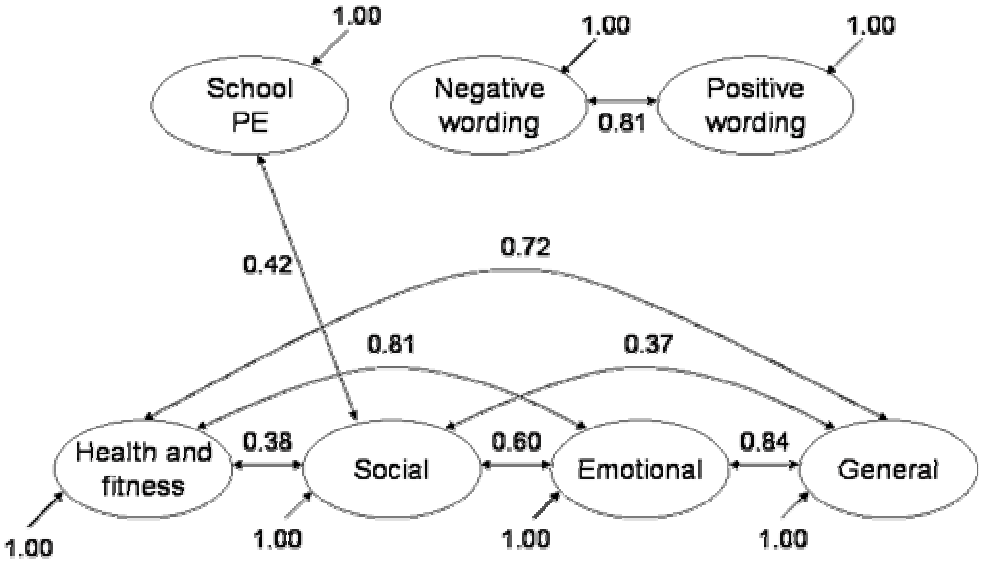

Figure 1. Structural part of the final model at the within-group level

At the between-group level, all attempts to fit any CFA models failed. Therefore, we have decided to conduct exploratory factor analysis (EFA) to have some notion about the factor structure and their number. 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 2.

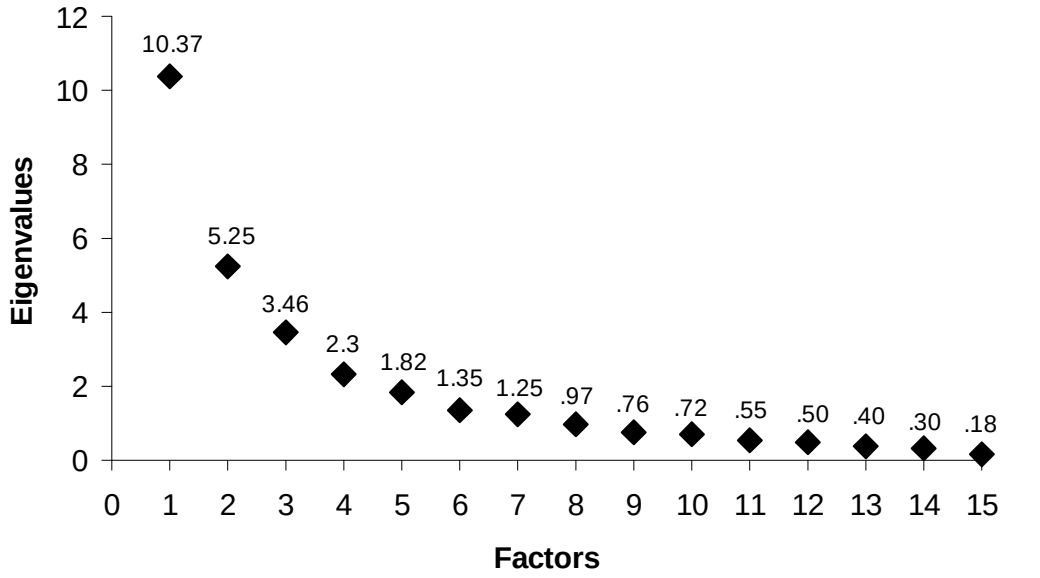

Figure 2. Scree plot of the first 15 factors

We might examine this plot in Figure 2 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.

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. 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.

As the last step in MCFA with continuous variables, the simultaneous analysis of the between- and within-group structures was conducted. This two-level model is depicted in Figure 3.

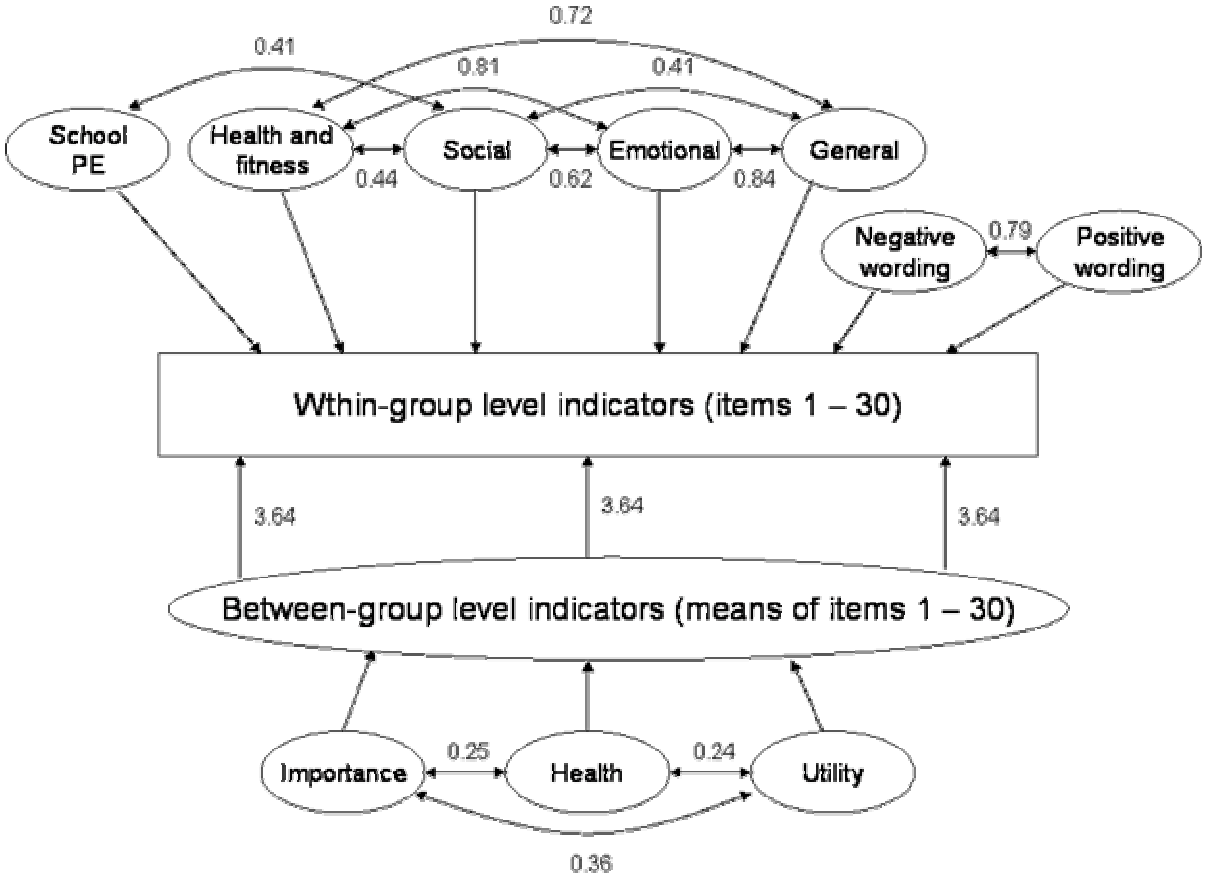

Figure 3. Structural part of the two-level model (all factor variances are fixed to 1)

The ICCs ranged from 0.02 to 0.11with mean value of 0.051. 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.

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. To make a comparison with other three approaches possible, the easier model has been estimated. The multilevel model with four factors at the within-group level and the one factor at the between-level was estimated.

Table 3. 4-factor model : 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.

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. Slightly higher standard errors are observed when parameters are estimated within the multilevel CFA for ordinal variables framework (see the last column of the Table 3).

The estimated factor variances, covariances and respective correlations are presented in Table 4. 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 4. Covariances between factors in 4-factor model using four different approaches

Sub-construct vs. sub-construct	**Factor loadings** (Standard errors)			
	Single-level continuous	**Multilevel continuous**	**Single-level ordinal**	**Multilevel ordinal**
H&F vs. H&F [a]	0.41 (0.06) *1.00*	0.39 (0.06) *1.00*	0.17 (0.08) *1.00*	0.72 (0.16) *1.00*
H&F vs. Social	0.38 (0.04) *0.81*	0.36 (0.04) *0.81*	0.19 (0.06) *0.80*	0.96 (0.15) *0.91*
H&F vs. Emo	0.40 (0.04) *0.90*	0.37 (0.04) *0.90*	0.18 (0.07) *0.89*	0.92 (0.13) *0.98*
H&F vs. General	0.60 (0.06) *0.82*	0.57 (0.06) *0.84*	0.23 (0.10) *0.79*	0.72 (0.16) *0.92*
Social vs. Social [a]	0.52 (0.06) *1.00*	0.51 (0.06) *1.00*	0.33 (0.11) *1.00*	1.53 (0.20) *1.00*
Social vs. Emo	0.39 (0.04) *0.79*	0.38 (0.04) *0.80*	0.22 (0.08) *0.79*	1.21 (0.15) *0.88*
Social vs. General	0.60 (0.05) *0.72*	0.58 (0.05) *0.74*	0.30 (0.11) *0.73*	1.52 (0.25) *0.85*
Emo vs. Emo [a]	0.48 (0.06) *1.00*	0.44 (0.06) *1.00*	0.24 (0.14) *1.00*	1.21 (0.20) *1.00*
Emo vs. General	0.71 (0.06) *0.89*	0.66 (0.06) *0.91*	0.30 (0.15) *0.88*	2.29 (0.31) *0.96*
General vs. General [a]	1.31 (0.11) *1.00*	1.19 (0.11) *1.00*	0.50 (0.30) *1.00*	4.71 (0.68) *1.00*

Note.
Standard errors are in parentheses, italic entries are correlations.
[a] Variances

First two approaches are again similar with respect to the variance and covariance estimations. Compared to these two approaches, the third approach (fourth column of the Table 4) provides roughly half values, whereas the values are about four times higher using the last approach (last column of Table 4). As was expected, both ordinal variable approaches produce generally smaller estimations of unique variances.

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.

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, $\boldsymbol{H_1}$**.**

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). 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).

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

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

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

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.

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

Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park (CA): SAGE Publications.

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.

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.

Furr, R.M., & Bacharach, V.R. (2008). *Psychometrics: An introduction*. Los Angeles: Sage Publications.

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

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.

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*

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.

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.

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.

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

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

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

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.

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

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

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.

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.

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

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

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

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.

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.

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.

APPENDIX

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.