COMPREHENSIVE TEST IN SCHOOL PHYSICAL EDUCATION AT SECONDARY SCHOOLS IN THE CZECH REPUBLIC – STANDARDIZATION AND VERIFICATION

Jana Vašíčková, Filip Neuls, Karel Frömel

Faculty of Physical Culture, Palacký University, Olomouc, Czech Republic

Submitted in February 2010

BACKGROUND: Physical education is at present oriented to the practical realization of physical activities (PA); less time is devoted to the acquisition of certain knowledge about the importance of PA and other necessary information about a healthy lifestyle and nutrition.

OBJECTIVE: The main aim of the study is to penetrate the knowledge sphere of students in the field of a physically active lifestyle and health and thus to standardize the comprehensive test in a Czech environment. Further, to compare test results among the types of schools, classes and gender.

METHODS: For testing knowledge about health and PA, we designed a comprehensive test that was distributed at 10 secondary schools in the Czech Republic (two vocational schools and eight grammar schools). We obtained tests from 192 boys and 224 girls (15 to 19 year olds) who completed the test twice: as a pre-test and after five weeks as a post-test in 2008–2009. In total, we gathered 832 tests.

RESULTS: The reliability coefficient (half split method) was \( r_a = 0.527 \) (p < 0.01). Cronbach’s alpha of inner consistency was \( r = 0.627 \). Neither boys nor girls reached 50% of the correct answers in the test. No difference has been found between gender in the pre-test; in the post-test a significant difference was found, which was confirmed also by the effect size (\( d = 0.35 \)). Grammar school students (\( d = 0.87 \)) and students from higher classes (\( d = 0.47 \)) had better knowledge about health and PA when compared to vocational school students and first year students.

CONCLUSIONS: The comprehensive test can be used mainly for research, comparison and as a part of a complex scientific survey. Grammar school students and older students scored better on the test, we did not find any gender differences. Students reached the lowest score in the nutritional and educational sphere. Practical use of the comprehensive test can encourage pupils and students to think about their own healthy lifestyle. In order to change pupils’ physical behavior we may involve pupils more and use, for instance, pedometers for step monitoring.

Keywords: Reliability, content validity, physical activity, gender, grammar school, vocational school, knowledge.

INTRODUCTION

Testing (pedagogical measurement) is a sophisticated method of pedagogical diagnosis that has for many years been used in western countries. An association between the level of physical activity/fitness and knowledge about an active lifestyle is not investigated very often as a research question. Despite this fact some studies exist that refer to the importance of the connection between PA knowledge (including, for instance, proper nutrition) with lifestyle or intentions and attitudes towards PA particularly in children and youth (Murray, Suprapiboonchai, Wilson, Rodriguez, & Eldridge, 2008; Palmer, Graham, & Elliott, 2005; Rutkowski & Connolly, 2010; Tuuri et al., 2006) but also in adults and seniors (Grant & Kluge, 2007; Hui & Morrow, 2001). Neither foreign nor Czech research methods deal with standardized testing of students’ knowledge in the field of physical activities, health and fitness. One known test transferred into software that tests students’ knowledge in the basic fitness concept at the secondary school level is the American FitSmart test (Zhu, Safrit, & Cohen, 1999). This test is the result of the extensive effort of physical education teachers and experts to develop a standardized knowledge test for the physical fitness sphere. Their test can be applied as it is or can be used as a part of an educational program aimed at improving health and students’ well being. It is based on research findings that point at the relationship between a physical fitness concept and a healthy lifestyle.

Our research team also strived to design a similar tool that would test the sphere of health and PA and at the same time it would educate. We think it is necessary to teach students not only the physical principles of fitness, but also cognitive ones, either in physical education (PE) or in any similar module (Human beings and health – school education program). An essential part of the instruction should be inter-subject integration as a necessity of a modern concept and contemporary educational trends. Students that have got information about physical fitness should also know the benefits and drawbacks of regular physical activity. The more
students know about the risks of physical inactivity and the benefits of PA, the more likely they will be to accept a lifestyle with physical activity or exercising or eventually to accept the idea of regular participation in organized sport as it is assumed in the so called “health belief model” (Strecher & Rosenstock, 1997).

Because health and PA are not taught concentratedly in one school subject, the comprehensive test serves as summary survey of knowledge in this sphere. The applied comprehensive test is designated for investigating the situation about the knowledge students have about health and PA at a certain moment. Its application allows for getting to the heart of the students’ knowledge sphere in the fields of human biology, nutrition and their knowledge about PE and sport disciplines. Further, we obtained information necessary for analyzing really measured PA possibly for the comparison of data from international questionnaires (e.g. IPAQ). The comprehensive test is personalized and meant for 15 year old and older youth which means for the broader population. It was developed mainly for research needs for pupils and students at elementary, secondary schools and universities. It is a test that tries to find status and is not for the evaluation of knowledge level or whether some educational aims have been reached. Unfortunately, the test results cannot be compared with any classification; biology and PE are subjects where a part of the knowledge could be obtained. With respect to the fact that a mark or grade is a private matter of each student, it cannot be publicized unofficially and that is why we have to find out the level of their knowledge of health and PA in a different way. The developed comprehensive test has got a standardized Polish version (Vašíčková, Chmelík, Frömel, & Neuls, 2009) and an English version.

Aim

The main objective of the study was to carry out the standardization of the comprehensive test at secondary schools. A secondary aim was to find out the status of knowledge in secondary school students of health and physical activity. If differences exist between grammar and vocational schools, are they significant? Do gender differences exist in students’ knowledge? Do the differences between first year students and students from higher classes exist?

METHODS

The comprehensive test as it is designed contains 32 questions that are divided into four dimensions (conditions, energetics, nutrition, educational), with eight questions in each dimension. The concept and dimensions were based on questions believed to relate to overall knowledge of physical activity and health that are discussed in the Framework Educational Program and are included in the sphere of “Human Health” (Výzkumný ústav pedagogický, 2007). During the construction of the test, we consulted about all items with experts from the fields of pedagogy, nutrition, physical fitness and a healthy lifestyle. Items mostly agreed on by experts were inserted into the test. Some questions had to be rephrased according to the experts’ opinion. Each question offers a multiple choice of four possible answers from A–D. Respondents are supposed to choose only the correct answers according to their own knowledge. The test was distributed twice with a five weeks break in between. First the respondents filled in basic information – school, class, gender, name and surname (or code), height and weight and date of completion, so that the tests could be paired as pre-test and post-test. The time necessary for completing the test was approximately 20 to 30 minutes. The test is evaluated by the number of correct answers (maximum 32 points, ordinal data) and in each dimension (maximum 8 points in each dimension).

Standardization was carried out as part of an extensive survey in first year classes at secondary schools (vocational and grammar schools) that were selected on the basis of good experience by graduates in their place of residence. These schools were mostly from the Moravian region. Further, comprehensive pre-tests and post-tests from the second, third and fourth class students entered into the standardization procedure. A special database in MS Access software was created for test assessment where basic information about the respondent and answers to each question are to be inserted. (It is necessary to fill in only one answer to each question, neither can an answer to a question be missing nor are two or more answers possible. Such a test must be eliminated from analysis.) It is possible to export from this database onto an MS Excel sheet according to selected criteria. Data can be, later on, transferred into the statistical software IBM SPSS 18.0 (SPSS Inc., Chicago, IL). The result of each test is a total point score and a score in each dimension in a particular respondent (ordinal data).

For the standardization of a comprehensive test we determined the reliability coefficient and Cronbach’s alpha coefficient of inner consistency. Content validity was assessed by a team of experts from university and teachers from secondary schools. The differences between pre-test and post-test in total score and in each dimension score were assessed by the nonparametric Wilcoxon test for dependent variables to which we determined effect size (d) (Cohen, 1988). This effect size d was calculated from the Z score: $d = 2 \times Z \times (n_1 + n_2)^{-1/2}$. Further we determined the correlation coefficient (Spearman) for nonparametric data. Taking a group of respondents
We quantified both median and inter-quartile range for the total score and for each dimension and we did an item analysis for each question. The difference in total score was tested using the nonparametric Mann-Whitney U test for independent variables between types of schools (grammar versus vocational schools), classes (first versus higher classes) and gender (boys versus girls) and for these coefficients we quantified effect size (d). For illustration of the differences among dimensions in different groups we used mean values and Friedman’s ANOVA.

RESULTS

Standardization of the comprehensive test

We found the stability coefficient by the repeated method, that means between the pre-test and post-test, was found to be $r_{tt} = 0.527$ (significant with a $p < 0.01$), on the level of dimensions it ranges from $r_{tt} = 0.234$ to $0.445$ ($p < 0.01$) (TABLE 1). When testing the statistical significance of the stability coefficient, we found out that the correlation coefficient predicates the relationship between two tests and this relationship can be characterized as moderately positive. Intra-class correlation is expressed as Cronbach’s alpha coefficient and is $r = 0.627$. When we divided each test into odd and even questions, the reliability was found to be $r_{sb} = 0.514$ ($p < 0.01$). When we divided test questions into version “A” (the first 16 questions) and version “B” (the second 16 questions: 17–32) the result was $r_{sb} = 0.494$ ($p < 0.01$).

Despite the low correlational coefficients, when we tested the significance of these coefficients we proved a moderate relationship between both tests. Similarly it was found in the case of the lack of effect of size coefficients in that they do not reach any effect in any of the tested dimensions. The differences in mean score in each dimension as well as in total score show only small differences that are caused by the great number of analyzed comprehensive tests. In future, any tested group with smaller number of respondent should take this fact into account.

The Spearman coefficient of rank correlations among the results scores in each dimension and the total score on the comprehensive test (TABLE 2) reached middle to high dependency (in the educational dimension) (Chráska, 2007).

Comparison of comprehensive test results

Characteristics of our sample and basic statistical values are stated in TABLE 3. We analyzed a total of 832 tests, so that we were able to create 416 pairs of pre-test and post-tests. Each group of tests was divided according to the type of school (grammar and vocational school), gender (boys and girls) and class (first

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Comparison of the test results in the pre-test and post-test (n = 416) in particular dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>Mean number of points</td>
</tr>
<tr>
<td></td>
<td>Pre-test</td>
</tr>
<tr>
<td>Condition dimension</td>
<td></td>
</tr>
<tr>
<td>Energetic dimension</td>
<td></td>
</tr>
<tr>
<td>Nutrition dimension</td>
<td></td>
</tr>
<tr>
<td>Educational dimension</td>
<td></td>
</tr>
</tbody>
</table>

Legend: ** $p < 0.01$ – statistically significant level, Z – Wilcoxon test, d – effect size

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Reliability in each dimension with regard to the total score (**p &lt; 0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman’s coefficient</td>
<td>Condition dimension</td>
</tr>
<tr>
<td>Total</td>
<td>0.661**</td>
</tr>
</tbody>
</table>
class students and 2nd–4th class students). (We created a summary group from the students of the second, third and fourth classes because of the small number of tests in each class.)

Among the types of school we found out in both tests statistically significant differences with better scoring on the part of students from grammar schools and this difference was also confirmed by the effect size coefficient. Students from higher classes scored significantly better than first year students. In terms of gender we found out only one difference in the post-test where girls scored better than boys.

Fig. 1 illustrates the difference between the types of schools in each dimension only in the pre-test. From the statistical point of view the differences in dimensions were significant on the level of p < 0.01 (condition dimension: $F = 39.08$; energetic: $F = 9.21$; nutrition: $F = 66.86$; educational: $F = 40.89$).

When comparing pre-tests in two groups according to classes (Fig. 2), it is evident that 2nd–4th classes’ students scored better in their mean number of points than first year students. It is due to the fact that the education program in the third class is oriented to human biology and knowledge of it helps the students to answer several questions. That is why the differences in mean values reached in each dimension are statistically significant apart from the energetic dimension (conditions: $F = 9.44$; energetics: $F = 2.27$ – insignificant; nutrition: $F = 19.16$; educational: $F = 12.44$; p < 0.01).

Fig. 3 shows the mean number of points in each dimension in the pre-test according to gender. The total mean score was, in girls, 15.90 points (there were 49.69% correct answers), respectively 15.30 points (47.81%) in boys ($F = 2.81$; p = 0.094). Girls have got better knowledge about health and physical activity, the biggest statistically significant difference was found in the nutrition dimension ($F = 4.18$; p < 0.05).

**DISCUSSION**

Results of the standardization of the comprehensive test on health and physical activity is that this test can be used mainly for research purposes, particularly for comparison as one of the methodological instruments applied within a more complex research survey. Such a test could also serve for the identification of strengths and weaknesses in the knowledge sphere of students, as a motivational tool for enhancing physical activity interest or for the evaluation of certain interventional programs aimed at the increasing of PA level. Its application could stretch students’ imaginations about their own lifestyle. The reliability coefficient tested by the split half method is 0.514. A higher reliability coefficient was found by the test-retest method and it can be said that the post-test was realized under conditions comparable to those of the pre-test. If we apply the post-test after 5 weeks, students’ information about the testing area might change a little bit, their mood, the atmosphere in the class and other significant factors may influence the final result. We have to take this into account when discussing the reliability. This reliability coefficient indicates that the test can be used only as an approximate and additional criterion for the comparison of groups.

**TABLE 3**

Comprehensive test results in pre-test and post-test surveys and a comparison according to the type of school, class and gender

<table>
<thead>
<tr>
<th>Type of school</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mdn  IQR</td>
<td>Z   p   d</td>
</tr>
<tr>
<td>Grammar school (n = 257)</td>
<td>17  4</td>
<td>8.97 0.00 0.87c</td>
</tr>
<tr>
<td>Vocational school (n = 159)</td>
<td>14  4</td>
<td>4.84 0.00 0.47b</td>
</tr>
<tr>
<td>Class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First (n = 285)</td>
<td>15  5</td>
<td>1.37 0.17 0.13b</td>
</tr>
<tr>
<td>Second to fourth (n = 131)</td>
<td>17  4</td>
<td>3.75 0.00 0.35c</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (n = 192)</td>
<td>16  5</td>
<td></td>
</tr>
<tr>
<td>Female (n = 224)</td>
<td>16  4</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1
The mean number of points in each dimension of the comprehensive test according to the type of school

Fig. 2
Mean number of points in each dimension of the comprehensive test according to classes

Fig. 3
Mean number of points in each dimension of the comprehensive test according to gender
The critical disadvantage of any test generally is its result possibly being influenced by motivation. It is quietly assumed that the motivation of each student to score the best is relatively high. But if we carry out testing twice (pre-test and post-test) and the test itself is demanding, it can be reflected in the reliability coefficient. The criterion validity of the comprehensive test in the Czech Republic could not be tested because we did not have any suitable external criterion to which we could compare test results. Thus we used only content validity based on discussion with a team of experts as described in the Method section. Empirical validity was found in Polish youth in reference to their marks from biology and resulted in \( r = 0.277 \) (\( p = 0.067 \)), (Vašíčková et al., 2009). During the designing of the comprehensive test we took into account all general rules that must be followed during testing (Zimmerman, Sudweeks, Shelley, & Wood, 1990). The creation of any comprehensive or knowledge test is multiculturally and educationally (systematically) determined and thus very difficult.

From the total score of the test we found out that students do not have sufficient knowledge about health and physical activity in that they reached on an average less than 50% of the correct answers. Only grammar school students and students from higher classes (52.5–53.5%) scored above half of the correct answers both on the pre-test and on the post-test. It indicates that the test is very demanding despite the fact that its content is in harmony with the 9 class curriculum of elementary school. This problem warrants further research. On the other hand it gives evidence about the necessity of including information about health and physical activity in education programs for high schools in order to support the understanding of students of the benefits of a healthy lifestyle and how to achieve it or to influence alternatively, according to this knowledge, their effort to accomplish positive changes in behavior (Dunn, Thomas, Green, & Mick, 2006). In previous studies (Contento, Manning, & Shannon, 1992) the authors state that for mastering knowledge 10–15 hours of intervention are necessary and for positive change in behavior towards a healthy lifestyle up to 50 hours of intervention are needed.

Comparing the test results according to the type of school, we found statistically significant differences between grammar and vocational school students. These differences result from the different orientation of schools. Secondary grammar schools are more universally oriented and thus these students have got better knowledge about health and physical activity than vocational schools which are more specialized in certain branches and the tested knowledge is not part of their curriculum. Nevertheless, with regard to the future preparation of the next productive generation, it will be suitable to include and effectively disseminate knowledge about a healthy lifestyle and the benefits of regular physical activities and proper nutrition for instance into the physical education curriculum (or into any similar program comprising PE).

Evaluation of the differences between two groups according to classes confirmed that mastery of knowledge is positively manifested in the comprehensive test because human biology is in the curriculum for higher classes. It is interesting that the final results were not negatively influenced by vocational school students where biology is not in the curriculum. The lower mean score in the test of first year students can be explained by transfer from different basic schools, where biology is part of the curriculum (human biology is taught in the third class) and where there is a different level of interest in physical activity. Chiefly, in the last mentioned sphere, the intentional research on changes in physical behavior with a pedometer application should be aimed at.

We have found a greater general knowledge about the given problem in girls, which can be caused by the fact that girls are more interested in how they look and how to achieve health harmlessly (without drastic diets). On the other hand, Lynn, Irwing, and Cammock (2001) found that boys are significantly better than girls regarding knowledge about physical health and recreation (biology, games and sport). The differences we found in the pre-test were minimal (no effect) and in the post-test the effect was moderate.

We analyzed also the score in particular dimensions where the highest score was achieved in the education dimension (5.4 points from 8 possible, which means 67.5%) and the lowest score in the nutrition dimension (3.3 points, which means 41.25%). Another analysis was carried out in individual questions; the least correct answers were to the questions: no. 3 “What are the most nutritious sources of energy for the human body?” (11%), no. 14 “What caloric expenditure approximately does an hour of faster walking represent in a moderately heavy (60–70 kg) and tall individual (170–180 cm)?” and no. 23 “What is the most important factor for health – what level of physical and mental condition should there be in relation to a person’s weight?” (13%). The best answered questions were no. 11 “At what intervals does the healthiest eating occur?” (92%), no. 28 “What can be the main danger of regular but only one sided PA (e.g. only running) in the development of what illness?” (87%), and no. 12 “What population groups should perform physical activity?” (80%).

**CONCLUSIONS**

The main aim of this study was to find out the reliability coefficient, which is \( r = 0.527 \) (test-retest Spearman coefficient). Students from secondary school do not have sufficient theoretical knowledge about physical
activity and health because they scored on average less than 50% of the correct answers. Despite this fact we think that the comprehensive test is the only method for finding out students’ knowledge about health and physical activity, and it can serve for the verification of the effect of experimental research, intervention programs and their long lasting effect, for comparative studies and detail analysis of different factors in various groups of students divided according to test results. The results obtained differ according to the type of school (grammar school students scored better than vocational school students) and classes (students from higher classes scored better than first year students, this is due to the content of their respective educational programs). Gender differences were not statistically significant. The field of tested knowledge about health and PA should be taken into account and deserve greater attention in education programs for students because according to our experience their practical impact on a qualitatively spent life is far bigger that has been expected.

ACKNOWLEDGEMENT

The study has been supported by the research grant from the Ministry of Education, Youth and Sports of the Czech Republic (No. MSM 6198959221) “Physical Activity and Inactivity of the Inhabitants of the Czech Republic in the Context of Behavioral Changes”.

REFERENCES


VÝDOMOSTNÍ TEST

VE ŠKOLNÍ TĚLESNÉ VÝCHOVĚ

NA STŘEDNÍCH ŠKOLÁCH V ČR – STANDARDIZACE A VERIFIKACE

(Souhrn anglického textu)

VÝCHODISKA: Tělesná výchova je v současné době dobře orientována na praktickou realizaci pohybových aktivit (PA), méně často se věnuje předávání příslušných vědomostí o významu realizace PA a dalších nezbytných informacích o zdravém životním stylu a výživě.

CÍLE: Hlavním cílem studie je proniknout do vědomostní sféry studentů, která zasahuje zeměpisná oblast biologie člověka, výživy a znalostí z oblasti tělesné výchovy a sportu, a tak standardizovat vědomostní test o problematice zdraví a pohybové aktivity v českých podmínkách. Dílečím cílem je pak porovnat výsledky vědomostního testu z hlediska typu škol, tříd a pohlaví.

VÝSLEDKY: Koeficient reliability metodou půlení testů byl zjištěn \( r_p = 0,527 \) (p < 0,01). Cronbachov alfa vnitřní konzistence bylo \( r = 0,627 \). Chlapci ani děvčata nedosahují 50 % správných odpovědí na otázky vědomostního testu. V pre-testu nebyl zjištěn žádný rozdíl mezi pohlavími; v post-testu byl zjištěn významný rozdíl, který byl také potvrzen koeficientem „effect size“ (\( d = 0,35 \)). Gymnaziální studenti (\( d = 0,87 \)) a studenti vyšších ročníků (\( d = 0,47 \)) mají větší znalosti o problematice zdraví a PA než studenti středních odborných škol a prvních ročníků.

ZÁVĚRY: Vědomostní test je využitelný zejména pro výzkumné účely; komparaci a jako součást rozsáhlého výzkumného šetření. Studenti gymnázií a starší studenti dosahovali lepších výsledků v testu než studenti středních škol a nižších ročníků; mezi pohlavím nebyly zjištěny žádné rozdíly. Prakticky použití vědomostního testu by mohlo vést k zamyšlení se nad vlastním životním stylem. Pro změny pohybového chování by pak bylo vhodné více zainteresovat žáky a využít např. monitorování PA krokoměrem.

Klíčová slova: reliabilita, obsahová validita, pohybová aktivita, pohlaví, gymnázium, střední škola, vědomosti.

Mgr. Jana Vašíčková, Ph.D.

Faculty of Physical Culture
Palacký University
tr. Míru 115
771 11 Olomouc
Czech Republic

Education and previous work experience
1998 – Palacký University, Faculty of Physical Culture (Mgr.);
1998–2001 – doctoral student; Palacký University, Faculty of Physical Culture;
2002 – Palacký University, Faculty of Physical Culture (Ph.D.);
Since 2007 – research and pedagogical assistant at Palacký University, Faculty of Physical Culture, Institute of Active Lifestyle, Olomouc, Czech Republic.

Scientific orientation
Research in Kinanthropology; focus on Physical Education at different types of schools. Teaching methodology and pedagogy within Faculty undergraduate program. Focus on data analysis, statistics, sport pedagogy and teaching PE and analysis of physical activity.

First-line publications