Int J Public Health (2015) 60:59–67
DOI 10.1007/s00038-014-0628-9

SSPH+ SWISS SCHOOL OF PUBLIC HEALTH +

ORIGINAL ARTICLE

Test–retest reliability of selected physical activity and sedentary behaviour HBSC items in the Czech Republic, Slovakia and Poland

Daniela Bobakova · Zdenek Hamrik · Petr Badura · Dagmar Sigmundova · Hania Nalecz · Michal Kalman

Received: 25 April 2014 / Revised: 3 November 2014 / Accepted: 5 November 2014 / Published online: 4 December 2014

Abstract
Objectives Better assessment of the reliability of the physical activity and sedentary behaviour items across countries in all WHO regions is highly needed. The aim of the study was to examine the test–retest reliability of selected physical activity and sedentary behaviour items of the HBSC questionnaire in Czech, Slovak and Polish adolescents.
Methods We obtained data from 693 Czech, Slovak and Polish (50.9 % boys) primary school pupils, grades five (mean age = 11.08; SD = 0.45) and nine (mean age = 15.12; SD = 0.45), who participated in a test–retest study in 2013. We used the single measures of Intraclass Correlation Coefficients (ICC) and Cohen's Kappa statistic to estimate the test–retest reliability of all selected items within the sample and stratified by gender, age group and country.
Results Both physical activity items (VPA and MVPA) and most of the sedentary behaviour items showed moderate agreement (ICC 0.41–0.60) and a similarly moderate correlation (Cohen's Kappa 0.3–0.5) after dichotomization.
Conclusions The physical activity and sedentary behaviour items of the HBSC questionnaire seem to be at the borderline of reliability to be used in adolescents.

Keywords Test–retest reliability · Physical activity · Sedentary behaviour · HBSC · The Czech Republic · Slovakia · Poland

D. Bobakova (✉)
Faculty of Medicine, Health Psychology Unit, Institute of Public Health, P.J. Safarik University in Kosice, Tr. SNP 1, Košice 040 66, Slovak Republic
e-mail: bobakova.daniela@gmail.com; Daniela.Bobakova@upjs.sk

D. Bobakova
Olomouc University Social Health Institute, Palacky University in Olomouc, Univerzitni 22, Olomouc 77111, Czech Republic

Z. Hamrik · P. Badura · D. Sigmundova · M. Kalman
Faculty of Physical Culture, Department of Recreation and Leisure Studies, Palacky University in Olomouc, Tr. Miru 115, Olomouc 77111, Czech Republic

H. Nalecz
Faculty of Physical Culture, Department of Pedagogy, The Jozef Pilsudski University of Physical Education in Warsaw, Marymoncka 34 Str., 00-968 Warsaw, Poland

Introduction

Test–retest reliability studies aiming at the physical activity and sedentary behaviour items in frequently used questionnaires for children and adolescents (GSHS, HBSC or the HELENA study) provide significant scientific evidence for the development of valid and reliable tools for measuring an important spectrum of behaviours (Hallal et al. 2012) as well as for future research (Taylor et al. 2013). Low reliability tends to mask the real prevalence and important relationships, and this creates difficulties or leads to the improper development of relevant policies, programmes and practices for young people (Liu et al. 2010). However, in epidemiological studies the use of self-reported questionnaires is often the only feasible method (Kohl et al. 2000).

In HBSC, both Vigorous Physical Activity—VPA and Moderate to Vigorous Physical Activity—MVPA (Prochaska et al. 2001) questions dealing with adolescents' physical activity are included. A test–retest study conducted in Finland (Vuori et al. 2005) showed the acceptable reliability of the MVPA and VPA items, with

interclass correlation coefficients from 0.6 to 0.8. Test–retest reliability was also evaluated in a Chinese study, with an intraclass correlation of 0.82 (95 % CI 0.74–0.88) for the past 7 days and 0.74 (0.64–0.82) for a usual week for MVPA questions (Liu et al. 2010). Torsheim et al. (1995) and Samdal et al. (2007) found similar results when dealing with VPA questions.

In terms of sedentary behaviour only few studies have reported the reliability and validity of the measures used in children and adolescence, and no single self-report measure (including HBSC measures) was identified as having both acceptable reliability and acceptable validity (Lubans et al. 2011). Furthermore, the television items were tested in Belgium (Vereecken et al. 2006), with ICCs (95 % CI) of 0.76 (0.63–0.85) for boys and 0.81 (0.69–0.88) for girls. In the study of Liu et al. (2010) ICCs were moderate for use of a computer for playing games (0.54 weekdays, 0.69 weekends) and high for TV viewing (0.72 weekdays, 0.74 weekends). Computer use (excluding games) had lower ICCs: 0.33 for weekdays and 0.50 for weekends.

As the WHO HBSC questionnaire is currently used by 43 countries and its results are widely used for health-promoting strategies and policies (Currie et al. 2012), there is a strong need for better assessment of the reliability and validity of the physical activity and sedentary items across countries in all WHO regions. Unfortunately, test–retest reliability studies focusing on the physical activity of school-aged children and adolescents are rarely published in Central and Eastern European countries (Chinapaw et al. 2010). Moreover, several performed studies are burdened with methodological problems, especially an inappropriate test–retest period might have skewed the results (Lubans et al. 2011). Different repeatability is most likely due to actual changes in PA behaviour which has been shown to be more prevalent with longer test–retest intervals (Trinh et al. 2009).

Thus, the aim of the study was to examine the test–retest reliability of selected physical activity and sedentary behaviour items of the HBSC questionnaire in Czech, Slovak and Polish adolescents.

Methods

Sample and procedure

This test–retest study is based on the international Health Behaviour in School-aged Children (HBSC) study and is consistent with its methodology. More detailed information about HBSC methodology can be found in a paper by Roberts et al. (2009). The testing and re-testing procedures were conducted in November and December 2013 in the Czech Republic, Slovakia and Poland. We contacted 15 larger and smaller elementary schools (100 % response rate) located in rural as well as in urban areas in the Olomouc and Pardubice region in the Czech Republic (7 schools), the Kosice region in Slovakia (5 schools) and the Warsaw region in Poland (3 schools). The schools were chosen randomly. In the Czech Republic and Slovakia questionnaires were administered in the 5th and 9th grades by trained research assistants in the absence of a teacher during regular class time. Questionnaires were administered in the presence of a teacher only in Poland due to legal requirements.

Table 1 Demographic characteristics of the sample (Czech Republic, Slovakia and Poland, 2013)

	11 years	15 years	11 years		15 years	
	n	*n*	Mean age	SD	Mean age	SD
Slovakia	114	113	10.93	0.34	14.92	0.37
Czech Republic	195	158	11.02	0.45	15.11	0.46
Poland	53	60	11.60	0.30	15.50	0.33
Total	362	331	11.08	0.45	15.12	0.45

In the first part of data collection (Test) we obtained data from 406 adolescents in the Czech Republic (response rate: 83.20 %), 258 adolescents in Slovakia (response rate: 74.14 %) and 134 adolescents in Poland (response rate: 84.33 %). Non-response was primarily due to illness and parental disapproval of the participation of their children.

In the Czech Republic and Slovakia the second part of data collection (Retest) was conducted 4 weeks after the test. Since we were aiming to test how study design and differences in the research procedure can influence the obtained data, the retest in Poland was conducted 1 week after the test. According to the literature review, we hypothesised that the shorter time between the test and retest would lead to outwardly higher ICC coefficients (Rangul et al. 2008; Liu et al. 2010).

In the retest, 53 of the adolescents who participated in the first part of data collection (Test) in the Czech Republic dropped out, 31 adolescents in Slovakia dropped out and in Poland 21 dropped out. Thus, the final sample consisted of 353 Czech (50.0 % boys), 227 Slovak (52.9 % boys) and 113 Polish (49.6 % boys) elementary school pupils, grades five and nine (Table 1).

The study was approved by the relevant ethics committees of the participating universities under the projects: GACR—excellence, APVV 0032 11, and SKE 01 13/2013. Parents in Slovakia and Poland were informed about the study via the school administration and could opt out if they disagreed. Schools in the Czech Republic

have a general permission granted at the beginning of the school year by all parents. Participation in the study was voluntary and anonymous in all countries involved in the study, with no explicit incentives provided for participation.

Measures

MVPA was measured by asking the respondents the following question: "Over the past 7 days, on how many days were you physically active for a total of at least 60 min per day? Please add up all the time you spent in physical activity each day." with possible responses ranging from 0 to 7 (Prochaska et al. 2001; Currie et al. 2012). The question was preceded by explanatory text: "Physical activity is any activity that increases your heart rate and makes you get out of breath some of the time. Physical activity can be done in sports, school activities, playing with friends, or walking to school. Some examples of physical activity are running, brisk walking, rollerblading, biking, dancing, skateboarding, swimming, soccer, basketball, football, & surfing." Further, we dichotomised all items according to the generally used dichotomisation and WHO recommendations to create binary variables. The cutoff point for MVPA was at least 60 min of physical activity every day over the past week.

VPA was measured by asking the respondents the following question: "Outside school hours: How often do you usually exercise in your free time so much that you get out of breath or sweat?" with possible responses: Every day/4 to 6 times a week/2 to 3 times a week/Once a week/Once a month/Less than once a month/Never (Currie et al. 2012). The cut off point for dichotomization of VPA was at least 2 or 3 times a week.

TV use was measured by asking the respondents the following question: "How many hours a day, in your free time, do you usually spend watching TV, videos (including YouTube or similar services), DVDs, and other entertainment on a screen? Please tick one box for weekdays and one box for weekend." with possible responses: None at all/About half an hour a day/About 1 h a day/About 2 h a day/About 3 h a day/About 4 h a day/About 5 h a day/About 6 h a day/About 7 or more hours a day (Currie et al. 2012). The cutoff point for dichotomization of TV use was two or more hours every weekday.

Computer use was measured by asking the respondents the following question: "How many hours a day, in your free time, do you usually spend using electronic devices such computers, tablets (like iPad) or smart phones for other purposes, for example, homework, emailing, tweeting, facebook, chatting, surfing the internet? Please tick one box for weekdays and one box for weekend." with possible responses: None at all/About half an hour a day/About 1 h a day/About 2 h a day/About 3 h a day/About 4 h a day/About 5 h a day/About 6 h a day/About 7 or more hours a day (Currie et al. 2012). The cutoff point for dichotomization of computer use was two or more hours every weekday.

Sitting time was measured by asking the respondents the following question: "Outside school hours: How many hours a day do you usually spend time sitting in your free time (for example, watching TV, using a computer or mobile phone, travelling in a car or by bus, sitting and talking, eating, studying)? Please be aware that if activities take place at the same time, these only count once. Please tick one box for weekdays and one box for weekend." with possible responses: None at all/About half an hour a day/About 1 h a day/About 2 h a day/About 3 h a day/About 4 h a day/About 5 h a day/About 6 h a day/About 7 or more hours a day. The cutoff point for dichotomization of sitting time was two or more hours every weekday.

Statistical analyses

Due to the differences in test–retest period between the countries we analysed the Polish data separately. First, descriptive statistics (prevalence rates and means) for background characteristics among Czech and Slovak adolescents were computed (Table 2). Next, we used the single measure of ICC to estimate the test–retest reliability of all selected items in the Czech Republic and Slovakia stratified by gender, age group and country (Table 3). We performed the same analyses on the Polish data (not shown). As the results in Poland differ largely from those in the Czech Republic and Slovakia they are briefly described in the result section. According to Landis and Koch's subjective guidelines (1997), the strength of test–retest agreement for an ICC greater than 0.81 is considered almost perfect agreement; 0.61–0.80 was considered to be substantial agreement; 0.41–0.60 was considered moderate agreement; 0.21–0.40 was considered fair agreement; and an ICC below 0.20 was considered poor. We also computed the proportion of the respondents who answered the question identically or whose response shifted in one, two or three and more categories in the test and retest (Fig. 1). Further, we dichotomised all items according to the generally used dichotomisation and WHO recommendations to create binary variables (dummy variables). Here Cohen's Kappa statistic was computed for the Czech Republic and Slovakia as a measure of agreement to assess whether the created variables would offer us greater stability in the time from the test to the retest study (Table 5). A correlation greater than 0.5 was considered large, 0.3–0.5 moderate, 0.1–0.3 small and less than 0.1 trivial (Cohen 1988). All data were analysed using IBM SPSS 20 for Windows (IBM Corp. Released 2011).

Table 2 Distribution of examined variables in test and retest by country (Czech Republic, Slovakia and Poland 2013)

	Czech Republic		Slovakia		Poland	
	Test Mean (SD)	Retest Mean (SD)	Test Mean (SD)	Retest Mean (SD)	Test Mean (SD)	Retest Mean (SD)
Physical activity						
Moderate to vigorous Physical Activity (MVPA)	4.73 (1.95)	4.64 (1.93)	4.98 (2.08)	4.78 (2.12)	5.01 (2.00)	5.03 (1.92)
Vigorous physical activity (VPA)	3.09 (1.49) (1.67)	3.15 (1.47)	2.75 (1.67)	2.96 (1.69)	3.31 (1.66)	3.27 (1.71)
Sedentary behaviour						
TV use—weekdays	4.07 (1.59)	4.33 (1.63)	4.43 (2.06)	4.57 (1.99)	3.55 (1.49)	4.00 (2.02)
TV use—weekend	5.01 (1.95)	5.42 (1.99)	5.10 (2.27)	5.41 (2.13)	5.13 (1.88)	5.33 (2.17)
Computer use—internet—weekdays	3.97 (1.97)	4.04 (1.91)	4.44 (2.28)	4.32 (2.04)	3.76 (1.95)	4.16 (2.29)
Computer use—internet—weekend	4.64 (2.44)	4.86 (2.35)	4.83 (2.36)	5.06 (2.30)	5.10 (2.27)	5.13 (2.39)
Sitting time—weekdays	5.30 (2.26)	5.43 (2.28)	5.19 (2.37)	5.50 (2.23)	4.26 (1.88)	4.47 (2.15)
Sitting time—weekend	5.28 (2.17)	5.55 (2.18)	5.25 (2.40)	5.57 (2.28)	5.35 (2.40)	5.25 (2.37)

Results

An overview (means and standard deviations) of the examined variables in the test and retest for the Czech Republic, Slovakia and Poland can be seen in Table 2. In Table 3 the ICCs for HBSC items regarding physical activity and sedentary behaviour are shown by country. The ICCs varied from 0.54 ("TV use—weekdays") to 0.66 ("Computer use—Internet—weekend"). Slovakia reached greater agreement in most of the items compared with the Czech Republic. The results in Poland differ largely from those in the Czech Republic and Slovakia. The lowest but still substantial agreement (0.66) was observed in the item screen-based activities over the week. In all other variables the ICCs varied from 0.80 to 0.98, which indicates almost perfect agreement.

In Table 4 the ICCs for HBSC items regarding physical activity and sedentary behaviour are shown by gender, age group and country in the Czech Republic and Slovakia. The most substantial agreement was observed in the items Electronic devices over the week and over the weekend. In all other variables the ICCs varied from 0.51 to 0.55, which indicates moderate agreement. Physical activity items showed greater agreement in boys than in girls. Sedentary behaviour items, on the contrary, showed greater agreement in girls than in boys, with the exception of "Sitting time-weekdays". Substantial agreement was observed in only three items among girls. The ICCs in 11-year-olds varied from 0.45 ("Sitting time—weekend") to 0.59 ("Computer use—Internet—weekend") and in 15-year-olds from 0.42 ("TV use—weekend") to 0.58 ("VPA"), which represents moderate agreement.

The proportion of respondents who answered the questions identically varied from 24 to 40 % in the Czech Republic and from 26 to 43 % in Slovakia. The proportions of maximum one-category response shift between the test and retest varied from 57 to 81 % in the Czech Republic and from 54 to 78 % in Slovakia (Fig. 1). In contrast, 81 to 89 % answered the question identically in Poland (Fig. 1). In the Czech Republic and Slovakia the most stable answers were given on the question regarding vigorous physical activity.

To strengthen the data we dichotomised all items according to the generally used dichotomisation, and using WHO recommendations we created binary variables for further analyses. Table 5 shows Cohen's Kappa for HBSC items regarding physical activity and sedentary behaviour by gender, age group and country. Mostly we observed a moderate correlation between the test and retest in all of the examined variables. However, the creation of binary variables brought greater agreement between the test and retest variables in Slovakia. A large correlation (greater than 0.5; $p < 0.001$) was observed in both of the physical activity items and a half of the sedentary behaviour items.

Discussion

The aim of the study was to examine the test–retest reliability of selected physical activity and sedentary behaviour items of the HBSC questionnaire in Czech, Slovak and Polish adolescents. The physical activity items showed moderate agreement and the sedentary behaviour items showed moderate to substantial agreement. No significant gender and age differences were observed regarding reliability. However, large country differences were observed between Poland and the other two countries, which were substantially caused by applying a different period between the test and the retest study. After dichotomisation we observed a moderate correlation between the test and retest in almost all of the examined variables. Interestingly, we observed a large correlation in both the

Table 3 ICC for HBSC items regarding physical activity and sedentary behaviour by country (Czech Republic, Slovakia and Poland, 2013)

Items	All ($n = 693$)		Slovakia ($n = 227$)		Czech ($n = 353$)		Poland ($n = 113$)	
	ICC	95 % CI	ICC	95 % CI	ICC	95 % CI	ICC	95 % CI
Physical activity								
Moderate to vigorous physical activity (MVPA)	0.60	0.55–0.64	0.51	0.40–0.60	0.53	0.45–0.60	0.98	0.97–0.99
Vigorous physical activity (VPA)	0.62	0.57–0.66	0.62	0.53–0.70	0.49	0.41–0.57	0.90	0.86–0.93
Sedentary behaviour								
TV use—weekdays	0.54	0.49–0.60	0.54	0.44–0.63	0.47	0.39–0.55	0.66	0.53–0.76
TV use—weekend	0.58	0.52–0.63	0.54	0.44–0.63	0.51	0.42–0.59	0.88	0.83–0.92
Computer use—internet—weekdays	0.64	0.59–0.68	0.62	0.53–0.70	0.59	0.52–0.65	0.80	0.72–0.86
Computer use—internet—weekend	0.66	0.61–0.70	0.61	0.52–0.69	0.62	0.55–0.68	0.88	0.83–0.92
Sitting time—weekdays	0.61	0.55–0.65	0.61	0.52–0.69	0.51	0.42–0.58	0.91	0.87–0.94
Sitting time—weekend	0.60	0.54–0.64	0.58	0.48–0.67	0.50	0.41–0.57	0.92	0.88–0.94

physical activity items and a half of the sedentary behaviour items in Slovakia. In the Czech Republic the correlations were much lower, particularly in the physical activity items.

Both physical activity items (VPA and MVPA) showed moderate agreement and a similarly moderate correlation after dichotomisation. However, this was observed only in the Czech Republic and Slovakia. Greater reliability was observed in Poland, where the test–retest period was much shorter. Also, some other studies have found better agreement in physical activity items caused by the study procedure (Rangul et al. 2008; Liu et al. 2010). For example, Rangul et al. (2008) found substantial reliability, with ICCs of 0.71 for frequency and 0.73 for duration (Rangul et al. 2008). In this study the test–retest period was only 8–12 days. Booth et al. (2001) found results similar to ours, while the period between the test and retest was again shorter (2 weeks). Thus, one of the possible explanations is that the period between the test and retest was substantially shorter in these studies, as in our study in Poland. This short period might have caused respondents to simply remember their previous answers, and thus reliability seems to be higher. According to Allen and Yen (1979) a very short time interval could cause the carryover effects due to memory, practice, or mood more likely, whereas a longer interval increases the risk of a change in the conditions. There are also studies, such as the one from Beijing, China (Liu et al. 2010), where reliability of the selected items was higher and the period between the test and retest was comparable (3 weeks). When comparing our results with these studies, we need to take the distinct sociocultural differences between Central Europe and China into account. These might have caused the reliability of the physical activity items to be higher in some other countries. Considering the physical activity levels across countries, there is a need to analyse the number of physical education classes by week as well as other culturally determined facts which could affect physical activity levels and the answering of repeated questions in the test–retest procedure. Also, the seasonable character of some sports activities, current morbidity (seasonal epidemics) and other specific circumstances, such as different levels of demands placed on students during the school year might have influence on obtained results. Countries with a more stable climate and weather could note greater agreement in any season because this element is eliminated and does not interfere with the results. There are also other cultural differences referring to the organisational part of research, such as class discipline, the number of studies conducted at schools, openness and awareness of the importance of scientific research and their reliability, level of cooperation with researchers, etc. No distinct change in reliability was observed after dichotomisation of the items. Both dichotomised physical activity items maintained a moderate level of correlation. However, dichotomisation of the physical activity items in Slovakia enhanced their correlation up to a large level.

Most of the sedentary behaviour items showed moderate agreement and a similarly moderate correlation after dichotomisation. But again, this was observed only in the Czech Republic and Slovakia, as greater reliability was observed in Poland. In addition, previous findings on the reliability of these items suggest acceptable reliability (Schmitz et al. 2004; Utter et al. 2003; Liu et al. 2010). The test–retest correlations found by Utter et al. (2003) in the examined items was higher. However, the period between the test and retest in this study was shorter (2 weeks). Liu et al. (2010) also found higher correlation coefficients regarding TV use. On the other hand, the correlation coefficients regarding computer use found by Liu et al. (2010) were much lower but as with the physical activity items, we need to take sociocultural country differences

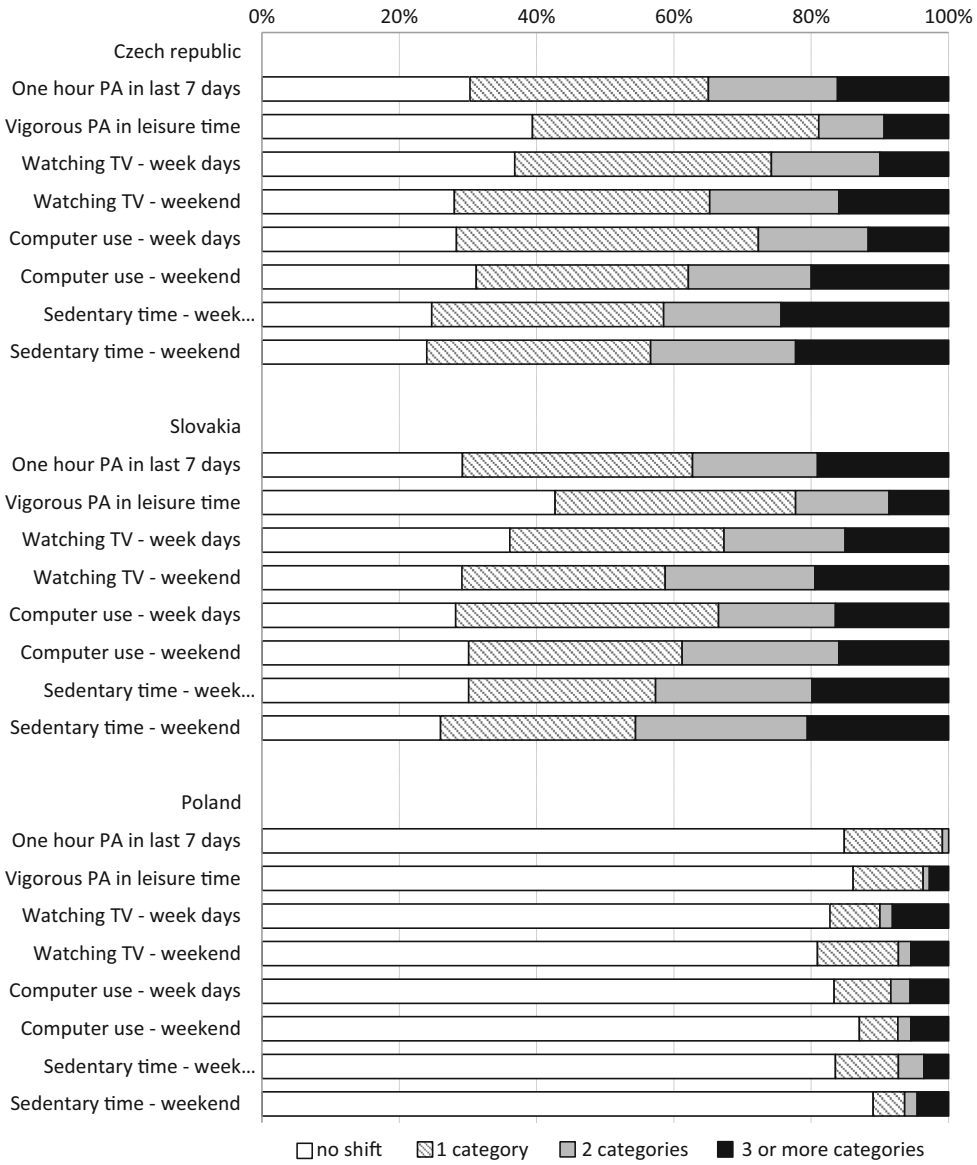

Fig. 1 Percentages of test-retest response shifts in physical activity and sedentary behaviour items counted separately for Czech Republic, Slovakia and Poland

into account. It is also necessary to take into account the changes in the structure of screen-time activities over the past decade (Sigmundova et al. 2011).

Assessing the reliability of the sedentary behaviour items is arguable and depends on different strictness of criteria suggested by different authors. There is no rigid recommendation of sufficient level of ICC. It depends on the area of research. In a number of studies, different levels of ICC are considered acceptable (Ngo et al. 2010; Evenson and McGinn 2005; Bland and Altman 1986; Landis and Koch 1977; Cohen 1988; Lubans et al. 2011). The overall findings of this study suggest that the HBSC questionnaire is an acceptable instrument for measuring physical activity and sedentary behaviour among adolescents (Landis and Koch 1977; Cohen 1988). Our results suggest that all of the examined HBSC items offer sufficient reliability in both genders in 11- and 15-year-old adolescents. The results of Cohen's Kappa after dichotomisation mostly duplicate the results from the ICCs, which indicates that fewer categories bring identical or comparable results. On the other hand, according to a systematic review of Lubans et al. (2011), stricter criteria should be applied here. An ICC and Kappa between 0.60 and 0.69 should be considered borderline and an ICC greater than 0.70 should be considered acceptable. Taking this into account, the reliability of the examined HBSC items found in the Czech Republic and Slovakia is unacceptable. However, the results are satisfying seeing that approximately 60–80 % of respondents answered in the same or an adjoining category. There are 7 (VPA), 8 (MVPA) and 9 (sedentary behaviour) response categories. For youths, whose actual physical activity, resp. time spent

Table 4 ICC for HBSC items regarding physical activity and sedentary behaviour by gender and age group (Czech Republic and Slovakia, 2013)

Items	All ($n = 580$)		Boys ($n = 297$)		Girls ($n = 283$)		11 years ($n = 309$)		15 years ($n = 271$)	
	ICC	95 % CI	ICC	95 % CI	ICC	95 % CI	ICC	95 % CI	ICC	95 % CI
Physical activity										
Moderate to vigorous physical activity (MVPA)	0.52	0.46–0.58	0.53	0.45–0.61	0.51	0.41–0.59	0.52	0.43–0.60	0.52	0.42–0.60
Vigorous physical activity (VPA)	0.55	0.49–0.61	0.56	0.48–0.64	0.53	0.44–0.61	0.52	0.44–0.60	0.58	0.50–0.66
Sedentary behaviour										
TV use—weekdays	0.51	0.45–0.57	0.50	0.41–0.59	0.52	0.43–0.60	0.55	0.46–0.62	0.44	0.33–0.53
TV use—weekend	0.52	0.46–0.58	0.45	0.35–0.54	0.60	0.52–0.67	0.58	0.50–0.65	0.42	0.31–0.52
Computer use—internet—weekdays	0.61	0.55–0.66	0.57	0.48–0.64	0.64	0.57–0.71	0.58	0.50–0.65	0.51	0.42–0.60
Computer use—internet—weekend	0.62	0.56–0.67	0.57	0.48–0.64	0.67	0.59–0.73	0.59	0.51–0.66	0.52	0.43–0.60
Sitting time—weekdays	0.55	0.48–0.60	0.55	0.46–0.63	0.54	0.45–0.62	0.51	0.42–0.59	0.47	0.36–0.56
Sitting time—weekend	0.53	0.47–0.59	0.51	0.42–0.60	0.55	0.46–0.62	0.45	0.35–0.54	0.52	0.43–0.61

Table 5 Cohen's Kappa for HBSC items regarding physical activity and sedentary behaviour by gender, age group and country (Czech Republic and Slovakia, 2013)

Items	All ($n = 580$) Cohen's Kappa	Boys ($n = 297$) Cohen's Kappa	Girls ($n = 283$) Cohen's Kappa	11 years ($n = 309$) Cohen's Kappa	15 years ($n = 271$) Cohen's Kappa	Slovakia ($n = 227$) Cohen's Kappa	Czech ($n = 353$) Cohen's Kappa
Physical activity							
Moderate to vigorous physical activity (MVPA)	0.44*	0.46*	0.43*	0.42*	0.48*	0.52*	0.38*
Vigorous physical activity (VPA)	0.41*	0.37*	0.43*	0.40*	0.42*	0.50*	0.36*
Sedentary behaviour							
TV use—weekdays	0.45*	0.43*	0.47*	0.40*	0.46*	0.51*	0.41*
TV use—weekend	0.41*	0.35*	0.47*	0.44*	0.30*	0.35*	0.45*
Computer use —internet—weekdays	0.49*	0.45*	0.54*	0.47*	0.39*	0.51*	0.48*
Computer use —internet—weekend	0.51*	0.44*	0.57*	0.47*	0.39*	0.51*	0.51*
Sitting time—weekdays	0.42*	0.51*	0.34*	0.44*	0.26*	0.43*	0.42*
Sitting time—weekend	0.39*	0.35*	0.42*	0.33*	0.34*	0.39*	0.38*

* $p < 0.001$

being sedentary, is somewhere near borderline of the particular category, it might be difficult to accurately distinguish the difference between the adjoining ones. Response shift of one category, therefore, can be only the result of this.

We suggest reducing the number of categories to ease the filling out of the questionnaires. Simplification of these kinds of questions, where respondents are asked to count means from often irregular activities, might bring even greater reliability. We also highly recommend careful consideration of the test–retest period.

The Czech Republic, Slovakia and Poland are socio-culturally quite similar; they are all Visegrad, post-communist, Central European countries, with very similar climates and conditions. The results of 2012 Eurostat also indicate nearly the same economic levels in these countries; therefore, we believe that the source of the differences lies in the study planning and procedures.

Strengths and limitations

The strength of our study is that it is the first to assess the test–retest reliability of the physical activity and sedentary behaviour items of the HBSC questionnaire in the Central/Eastern Europe. Furthermore, the period before administration of the retest in the Czech Republic and Slovakia was sufficiently long to avoid the retention of previously chosen answers and short enough to evade changes in lifestyle

patterns. Concerning this issue, it would be very helpful if the methodology of similar studies became uniform, thus subsequently enabling easier comparisons.

Several limitations should be taken into account concerning the present study. First, as mentioned above, seasonal influences are assumed to affect the results. This applies especially to the physical activity items, since children are more physically active during the summer months (Loucaides et al. 2003). Second, the construction of the sedentary behaviour items does not take into account the fact that children could be engaging in multiple activities at the same moment, e.g. watching TV while playing computer games. In connection with this, doubts might arise whether such activities should belong to one or another category or to both of them, which might then be reflected in the results. Third, this study focused on the reliability of the selected items but did not investigate their validity; therefore, we suggest further studies using a combination of the questionnaire and devices for physical activity monitoring, such as pedometers or accelerometers. The survey took place only in specific regions of the Czech Republic, Slovakia and Poland, and therefore, it might be complicated to interpret the data for other regions or countries. In this study the reliability was only analysed using ICC and Cohen's Kappa and proportional bias was not addressed.

Conclusion

Both physical activity items (VPA and MVPA) and most of the sedentary behaviour items showed moderate agreement and a similarly moderate correlation after dichotomisation. These findings suggest that the acceptability of physical activity and sedentary behaviour items used in the HBSC and other questionnaires as instruments to measure physical activity and sedentary behaviour among adolescents is disputable. In addition, interpretation of the results is even more complicated since there are no globally used recommendations to determine which values of ICC level are acceptable and which are not. Thus, more research using objective measures is needed to clarify the reliability of these instruments. These items in the current form used in HBSC questionnaires seem to be at the borderline of reliability and these results were probably highly influenced by test–retest procedures. Simplification of these kinds of questions by reducing the number of categories should ease the counting, which might subsequently lead to greater stability of these items.

Acknowledgments This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0032-11. This work was supported by the IGA UP research project "Leisure Time in School-aged Children-HBSC Study", reg. No. FTK_2013_020.

References

Allen MJ, Yen WM (1979) Introduction to measurement theory. Brooks/Cole, Monterey

Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–310

Booth ML, Okely AD, Chey T, Bauman A (2001) The reliability and validity of the physical activity questions in the WHO health behaviour in schoolchildren (HBSC) survey: a population study. Brit J Sport Med 35(4):263–267

Chinapaw MJM, Mokkink LB, van Poppel MNM, van Mechelen W, Terwee CB (2010) Physical activity questionnaires for youth: a systematic review of measurement properties. Sports Med 40(7):539–563. doi:10.2165/11530770-000000000-00000

Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Erlbaum Associates, Hillsdale

Currie C, Zanotti C, Morgan A et al (2012) Social determinants of health and well-being among young people. HBSC International Report from the 2009/2010 Survey. Health Policy for Children and Adolescents No. 6. WHO Regional Office for Europe, Copenhagen

Evenson KR, McGinn AP (2005) Test–retest reliability of a questionnaire to assess physical environmental factors pertaining to physical activity. Int J Behav Nutr Phys Act 2:7–13. doi:10.1186/1479-5868-2-7

Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund F (2012) Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 380(9838):247–257. doi:10.1016/S0140-6736(12)60646-1

Kohl HW, Fulton JE, Caspersen CJ (2000) Assessment of physical activity among children and adolescents: a review and synthesis. Prev Med 31(2):54–76. doi:10.1006/pmed.1999.0542

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174

Liu Y, Wang M, Tynjälä J, Lv Y, Villberg J, Zhang Z, Kannas L (2010) Test–retest reliability of selected items of health behaviour in school-aged children (HBSC) survey questionnaire in Beijing, China. BMC Med Res Methodol 10:73. doi:10.1186/1471-2288-10-73

Loucaides CA, Chedzoy SM, Bennett N (2003) Pedometer-assessed physical (ambulatory) activity in Cypriot children. Eur Phys Educ Rev 9(1):43–55. doi:10.1177/1356336X03009001179

Lubans DR, Hesketh K, Cliff DP, Barnett LM, Salmon J, Dollman J, Morgan PJ, Hills AP, Hardy LL (2011) A systematic review of the validity and reliability of sedentary behaviour measures used with children and adolescents. Obes Rev 12(10):781–799. doi:10.1111/j.1467-789X.2011.00896.x

Ngo T, Stupar M, Côté P, Boyle E, Shearer H (2010) A study of the test–retest reliability of the self-perceived general recovery and self-perceived change in neck pain questions in patients with recent whiplash-associated disorders. Eur Spine J 19(6):957–962. doi:10.1007/s00586-010-1289-x

Prochaska JJ, Sallis JF, Long B (2001) A physical activity screening measure for use with adolescents in primary care. Arch Pediat Adol Med 155(5):554–559. doi:10.1001/archpedi.155.5.554

Rangul V, Holmen TL, Kurtze N, Cuypers K, Midthjell K (2008) Reliability and validity of two frequently used self-administered physical activity questionnaires in adolescents. BMC Med Res Methodol 8:47. doi:10.1186/1471-2288-8-47

Roberts C, Freeman J, Samdal O, Schnohr C, Looze M, Gabhainn Nic S, Iannotti R, The HBSC Methods Development Group (2009) The Health Behaviour in School-aged Children (HBSC) study: methodological developments and current tensions. Int J Public Health 54(2):140–150

Samdal O, Tynjälä J, Roberts C, Sallis JF, Villberg J, Wold B (2007) Trends in vigorous physical activity and TV watching of adolescents from 1986 to 2002 in seven European Countries. Eur J Public Health 17(3):242–248. doi:10.1093/eurpub/ckl245

Schmitz KH, Harnack L, Fulton JE, Jacobs DR, Gao S, Lytle LA, Van Coevering P (2004) Reliability and validity of a brief questionnaire to assess television viewing and computer use by middle school children. J School Health 74(9):370–377. doi:10.1111/j.1746-1561.2004.tb06632.x

Sigmundova D, AnsariW El, Sigmund E, Frömel K (2011) Secular trends: a ten-year comparison of the amount and type of physical activity and inactivity of random samples of adolescents in the Czech Republic. BMC Public Health 11:731–743. doi:10.1186/1471-2458-11-731

Taylor NJ, Crouter SE, Lawton RJ, Conner MT, Prestwich A (2013) Development and validation of the Online Self-Reported Walking and Exercise Questionnaire (OSWEQ). JPhys Act Health 10(8):1091–1101

Torsheim T, Wold B, Samdal O, Haugland S (1995) Test–retest reliability of survey indicators measuring adolescent health and health behaviour. Research Centre for Health Promotion, University of Bergen, Bergen

Trinh OT, Do Nguyen N, van der Ploeg HP, Dibley MJ, Bauman A (2009) Test–retest repeatability and relative validity of the Global Physical Activity Questionnaire in a developing country context. J Phys Act Health 6(1):46–53

Utter J, Neumark-Sztainer D, Jeffery R, Story M (2003) Couch potatoes or french fries: are sedentary behaviors associated with body mass index, physical activity, and dietary behaviors among adolescents? J Am Diet Assoc 103(10):1298–1305. doi:10.1053/S0002-8223(03)01079-4

Vereecken CA, Todd J, Roberts C, Mulvihill C, Maes L (2006) Television viewing behaviour and associations with food habits in different countries. Public Health Nutr 9(2):244–250. doi:10.1079/PHN2005847

Vuori M (2005) Reliability of physical activity items in the HBSC study Pilot study among Finnish 11, 13 and 15-year-old school children. Research Centre for Health Promotion, University of Jyväskylä, Jyväskylä

Vuori M, Ojala K, Tynjälä J, Villberg J, Valimaa R, Kannas L (2005) The stability of physical activity survey items in the HBSC study. Liik Tiede 42(6):39–46