



Test–retest reliability of the scale of participation in organized activities among adolescents in the Czech Republic and Slovakia

Lucia Bosakova · Peter Kolarcik · Daniela Bobakova · Martina Sulcova · Jitse P. Van Dijk · Sijmen A. Reijneveld · Andrea Madarasova Geckova

Received: 27 February 2015 / Revised: 2 July 2015 / Accepted: 24 September 2015
© Swiss School of Public Health (SSPH+) 2015

Abstract

Objectives Participation in organized activities is related with a range of positive outcomes, but the way such participation is measured has not been scrutinized. Test–retest reliability as an important indicator of a scale’s reliability has been assessed rarely and for “The scale of participation in organized activities” lacks completely. This test–retest study is based on the Health Behaviour in School-aged Children study and is consistent with its methodology. **Methods** We obtained data from 353 Czech (51.9 % boys) and 227 Slovak (52.9 % boys) primary school pupils, grades five and nine, who participated in this study in 2013. We used Cohen’s kappa statistic and single measures of the intraclass correlation coefficient to estimate the test–retest

reliability of all selected items in the sample, stratified by gender, age and country.

Results We mostly observed a large correlation between the test and retest in all of the examined variables (κ ranged from 0.46 to 0.68). Test–retest reliability of the sum score of individual items showed substantial agreement (ICC = 0.64).

Conclusions The scale of participation in organized activities has an acceptable level of agreement, indicating good reliability.

Keywords Organized activities · Adolescents · HBSC · Test–retest · Reliability

L. Bosakova (✉) · P. Kolarcik · D. Bobakova · J. P. Van Dijk · A. M. Geckova
Department of Health Psychology, Medical Faculty, P. J. Safarik University in Kosice, Kosice, Slovak Republic
e-mail: lucia.bosakova@upjs.sk

L. Bosakova
Department of Quantitative Methods, Faculty of Business Economy in Kosice, University of Economics in Bratislava, Bratislava, Slovak Republic

L. Bosakova · P. Kolarcik · D. Bobakova · A. M. Geckova
Olomouc University Social Health Institute (OUSHI), Palacky University in Olomouc, Olomouc, Czech Republic

M. Sulcova
Department of Recreation and Leisure Studies, Faculty of Physical Culture, Palacky University in Olomouc, Olomouc, Czech Republic

J. P. Van Dijk · S. A. Reijneveld
Department of Social Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Introduction

The leisure time of our children is usually devoted to some kind of organized activities, which should provide them with meaningful content. Participation in organized activities has been found to be positively related with a range of positive outcomes, but the way such participation is measured has not been scrutinized. While important, the reliability of the scale measuring participation in organized activity has rarely been reported. Using test–retest technique, this study attempts to provide such information.

Organized activities are characterized by structure, certain rules, regular and scheduled meetings, adult supervision and an emphasis on skill building (Larson 2000; Mahoney et al. 2005). In general, adolescents who are regularly involved in some kind of organized community report better health, are more likely to have healthier lifestyles and are less likely to behave in a risky manner (Zambon et al. 2010). Moreover, organized activities, such as community-based activities, are associated with both

short- and long-term indicators of positive development and are also positively correlated with achievement, educational aspirations, self-esteem, ability to overcome adversity and leadership qualities among young people (Fatusi and Hindin 2010). Though high levels of participation in such activities might not have solely positive effects on individual well-being, it may also have some negative effects (Leversen et al. 2012). This is known as an overscheduling hypothesis assuming that if the children and adolescents are overburdened with activities, it supposedly threatens their well-being (Mahoney et al. 2008). However, this hypothesis has very limited empirical backup (Bohnert et al. 2010; Mahoney et al. 2006) and some studies even support the point that non-participation is of higher concern than over-scheduling (Badura et al. 2015; Mahoney et al. 2006).

Thus, it is important to assess the participation of adolescents in organized activities to provide a solid basis for precisely targeted interventions on how to reduce risk time, increase opportunities for appropriate use of time and increase the amount of time adolescents spend on situations leading to healthy development (Leversen et al. 2005). Several studies exist focussing on participation in organized activities among young people, monitoring the association with youth development (Eccles et al. 2003; Fredrics and Eccles 2006), examining the association between adolescents' participation in organized activities and their health (Youniss et al. 1999; Rhodes and Spencer 2005; Zambon et al. 2010) or examining different influences and incentives of youth participation in organized activities (Persson et al. 2007; Fawcet et al. 2009).

Regarding measurement, a self-reported questionnaire has been the most commonly used tool, and it seems at present to be the only practical method for collecting a broad range of data from a large number of children and adolescents (Kohl et al. 2000). However, the use of a scale is not uniform and varies between studies (King et al. 2006; Zambon et al. 2010). We focused on the organized activities scale developed within the WHO HBSC study. As the HBSC questionnaire is currently used by 43 countries and its results are widely used for health-promoting strategies and policies (Currie et al. 2012), and at the same time a central requirement for all self-reported population measures is to have good measurement properties, there is a strong need for evaluation of the reliability and validity of the organized activities item.

Reliability refers to the overall consistency of a measure which provides similar results of the measure under consistent conditions. Several ways of estimation of test reliability exist in psychometrics. First and most widely used is the internal consistency of the scale or questionnaire, usually expressed as the Cronbach alpha coefficient. Another type is the split-half method, where internal

consistency of the test is expressed by the correlation coefficients between two halves of the test's items. In some cases, where test development produced more versions of the test, it is possible to use parallel test forms to estimate the test reliability (Nunnally and Bernstein 1994; DeVellis 2003; Downing 2004; Cook and Beckman 2006). The last way for estimation of the test reliability is the test–retest method which aims to directly assess the degree to which test scores are consistent between two measurements in time. Test–retest reliability is usually expressed as intra-class correlation coefficient (continuous variables) or Cohen's kappa coefficient (categorical variables) (Nicker-son 1997; Feder 2008).

Based on our knowledge, there are several studies dedicated to the test–retest reliability of adolescents' physical activities (Booth et al. 2002; Liu et al. 2010; Hong et al. 2012) and recreation and leisure-time activities (King et al. 2004; Bult et al. 2010), some of which are also related to organized activities (Booth et al. 2002; King et al. 2004; Bult et al. 2010; Hong et al. 2012), though rather marginally. The study of Booth et al. (2002) was aimed at adolescents' physical activities and assessed the 2-week test–retest reliability and validity of the Adolescent Physical Activity Recall Questionnaire (APARQ) among 13- and 15-year-old Australians, including two main components: participation in organized and non-organized physical activities. The kappa values were 0.60 or higher (with the exception of boys in grade 8 in summer activities) for both the three-category (vigorous, adequate, inactive) and two-category (active, inactive) measures as well as for summer and winter organized activities. The ICC calculated for energy expenditure due to organized activities ranged from 0.39 (boys, grade 8), 0.50 (girls, grade 8) and 0.52 (boys, grade 10) to 0.87 (girls, grade 10) in summer activities and from 0.64 (boys, grade 8), 0.41 (girls, grade 8) and 0.40 (boys, grade 10) to 0.63 (girls, grade 10) in organized winter activities. Overall, the findings of this study suggest that the APARQ has acceptable to good reliability, which tends to be better among older students and among girls compared with boys. The study of King et al. (2004) reports evidence of the test–retest reliability, internal consistency and factor structure of the Children's Assessment of Participation and Enjoyment (CAPE) and its companion measure, Preferences for Activities of Children (PAC), providing information about six dimensions of participation (i.e. diversity, intensity, where, with whom, enjoyment and preference) and two categories of recreation and leisure activities: (i) formal and informal activities; and (ii) five types of activities (recreational, active physical, social, skill-based and self-improvement). The test–retest reliabilities for CAPE activity type participation intensity scores ranged from 0.72 to 0.81, indicating sufficient test–retest reliability, along with internal consistency

reliabilities for the PAC formal, informal and activity type preference scores, which were also acceptable (ranging from 0.67 to 0.84). The study of Bult et al. (2010) examined the reliability and validity of the Dutch language version of the CAPE/PAC and indicated it to be a reliable and valid instrument for measuring participation in recreation and leisure activities for children with and without physical disabilities aged 6–18 years. The ICC test–retest reliability values varied from 0.61 (social activities), 0.68 (physical activities), 0.71 (recreational activities) and 0.72 (self-improvement activities) to 0.78 (skill-based activities). The study of Hong et al. (2012) aimed to assess the reliability and validity of the Vietnamese Adolescent Physical Activity Recall Questionnaire (V-APARQ), where students were asked to report their usual organized and non-organized physical activities during a normal week. Reliability of the V-APARQ was assessed using a 2 week test–retest. The ICC ranged from 0.37 for moderate physical activity (MPA) to 0.62 for vigorous physical activity (VPA) and 0.57 for moderate-to-vigorous physical activity (MVPA), and was higher in boys than in girls. Weighted kappa coefficients ranged from 0.33 (MPA) and 0.52 (VPA) to 0.75 (MVPA) and were also higher in boys than in girls. Thus, the reliability of the V-APARQ varied from low to fair, but according to the authors it is comparable to other self-report physical activities questionnaires used among adolescents.

However, it seems that no test–retest reliability study focused solely on organized activities has yet been conducted, and the reliability of such measures should be known. Thus, the aim of the study was to examine the test–retest reliability of the organized activities scale of the HBSC questionnaire in Czech and Slovak adolescents. Based on the above-mentioned studies, we expected this measure to reach acceptable but not excellent levels of reliability.

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 the HBSC methodology can be found in a paper by Roberts et al. (2009). The testing and retesting procedures were conducted in November and December 2013 in the Czech Republic and Slovakia. We contacted 12 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) and the Kosice region in Slovakia (5 schools). The schools were

chosen randomly. Questionnaires were administered in the fifth and ninth grades by trained research assistants in the absence of a teacher during regular class time.

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

The second part of data collection (retest) was conducted 4 weeks after the test. According to Allen and Yen (1979), the first testing may influence the second as there is a time and potential to learn, and carry over, leading to obtain recall effects. The test–retest period may influence and thus affect the estimates of reliability of the examined measures. In general, 1–4 weeks period is recommended to examine test–retest reliability (Kurpius and Stafford 2006). However, the test–retest period needs to be sufficiently long to avoid the retention of previously given answers and short enough to evade changes in lifestyle patterns (Bobakova et al. 2015). We therefore decided to choose a 4-week period as we considered it to be the most appropriate and it has also been used in other similar test–retest studies (Trinh et al. 2009; Bult et al. 2010; Bobakova et al. 2015).

In the retest, 53 of the adolescents who participated in the first part of data collection (test) in the Czech Republic and 31 adolescents in Slovakia dropped out. Thus, the final sample consisted of 353 Czech (50.0 % boys) and 227 Slovak (52.9 % boys), grades five and nine.

Respondents who participated only in the first part of data collection (test) and then dropped out did not differ from the others regarding gender, grade and parental education and employment status.

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

Our survey used a battery of questionnaires and scales, and for the purposes of this paper we selected demographical items (gender, grade and nationality) and a scale of participation in organized activities.

The scale of participation in organized activities was measured by six items adopted from the HBSC study. These items were designed to monitor the type of

organized activities that students participated in. The scale asked the following question: “Are you doing any of these organized activities in your free time?” with dichotomized responses: yes or no. Organized activity was explained as an activity that respondents do in a sports club or another type of club or organization. Six categories of organized activities were presented, and respondents were asked to indicate the type of activities that they participated in. The six categories were:

- (a) Organized team sport activities (e.g. football, basketball and volleyball).
- (b) Organized individual sport activities (e.g. tennis, gymnastics and karate).
- (c) Attending a basic art school (e.g. playing a musical instrument, dance, theatre).
- (d) Children and youth organizations (e.g. scouting),
- (e) Attending a recreational course in a leisure-time centre or at school (e.g. chess, model-building and debate course).
- (f) Meeting in church and singing in the church choir (e.g. Salesians).

We computed the total score from the positive answers of the scale to create a composite variable. Furthermore, we dichotomized the composite variable to create a binary variable. The cutoff point for dichotomization was engaging in at least one activity.

Statistical analysis

In the first step, we computed the frequencies of the background characteristics and the proportion of positive responses on the individual items of “The scale of participation on organized activities” (see Tables 1, 2). In the second step, we computed Cohen’s kappa coefficients (κ) for each item of the scale from the test and retest data collection, stratified by country, gender and age group. Next, we used the single measure of the intraclass correlation coefficient (ICC) to estimate the test–retest reliability of the sum score of the individual items. Moreover, the sum score was also tested by the Wilcoxon test for paired samples. In the final step, we computed Cohen’s kappa

coefficients (κ) for the dichotomized sum score of activities.

According to Landis and Koch’s subjective guidelines (Landis and Koch 1977), a strength of test–retest agreement for an ICC greater than 0.81 is considered to be almost perfect agreement; 0.61–0.80 substantial agreement; 0.41–0.60 moderate agreement; 0.21–0.40 fair agreement; and an ICC below 0.20 poor. In Cohen’s kappa statistic, 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).

Results

Descriptive statistics of the sample and distribution of the examined variables in the test and retest for the Czech Republic and Slovakia can be seen in Tables 1 and 2, respectively. Table 3 shows Cohen’s Kappa for items regarding organized activities by country, gender and grade. We mostly observed a large correlation between the test and retest in all of the examined variables. Kappa coefficients for the whole sample ranged from 0.46 to 0.68. Only one item (leisure centres) was lower than 0.5, but we consider this to be acceptable according to Cohen (1988). The kappa coefficient in both genders showed large agreement in all items except “visiting leisure centres” and participation in religious groups (only in boys). Regarding age differences, a lower level of agreement (moderate) was observed only in the “individual sport activities” item in the ninth grade (15-year-olds) and in the “visiting leisure centres” and “religious groups” items in the fifth grades (11-year-olds). It follows that all of the items seem to have good reliability except the item “leisure centres”, which despite all the stratification was weak.

We computed a total score from the positive answers of the scale to create a composite variable. We used the single measure intraclass correlation coefficient (ICC) to estimate the test–retest reliability of the sum score of the individual items which showed substantial agreement. After stratification, only boys showed a moderate level

Table 1 Descriptive statistics of the sample (Czech Republic and Slovakia, 2013)

	Total		Boys		Girls		5th grade (11-year-olds)		9th grade (15-year-olds)	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Total	580		296	51.1	283	48.9	309	53.3	271	46.7
Slovaks	227	39.1	120	52.9	107	47.1	114	50.2	113	49.8
Czechs	353	60.9	176	50.0	176	50.0	195	55.2	158	44.8
Boys							158	53.4	138	46.6
Girls							150	53.0	133	47.0

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

	Test		Retest	
	<i>N</i>	%	<i>N</i>	%
Team sport activities				
Yes	280	48.8	288	49.8
Individual sport activities				
Yes	182	31.7	180	31.1
Art school				
Yes	158	27.5	156	27.0
Youth organization				
Yes	73	12.7	63	10.9
Leisure centres				
Yes	160	27.9	162	28.0
Religious groups				
Yes	70	12.2	66	11.4
	Mean (SD)	<i>N</i>	Mean (SD)	<i>N</i>
Sum of activities				
Min = 0; max = 6	1.61 (1.23)	574	1.58 (1.21)	578
	<i>N</i>	%	<i>N</i>	%
Dichotomized sum of activities				
No activity	92	16.0	103	17.8
1 or more activities	482	84.0	475	82.2

of agreement between test and retest. According to the Wilcoxon test for paired samples, there is no significant difference between test and retest regarding this composite variable.

Furthermore, we dichotomized the composite variable to create a binary variable. In the total sample, as well as in both countries, genders and grades, we observed a large correlation between the test and retest.

Discussion

The aim of the study was to examine the test–retest reliability of the organized activities scale of the HBSC questionnaire in Czech and Slovak adolescents. Referring to Landis and Koch (1977), the results of this study indicate that the HBSC organized activities scale has acceptable (from substantial to moderate) test–retest reliability which remains acceptable even in analyses stratified by country, gender or age. Moreover, the composite variable also showed an acceptable level of reliability.

The stability of the individual items might depend on several aspects related to a particular organized activity, such as financial investment, commitment to it and

monitoring of attendance. For example, the lowest reliability according to the Kappa coefficient was found in the leisure centres variable. One possible explanation might be that leisure centre activities are offered for a symbolic fee, partially supported via governmentally issued vouchers, so the small investment made is not usually enough to motivate parents to keep their children participating in this kind of activity. Thus, the test–retest reliability of this type of activity may be lower. On the other hand, organized sport activities or art school are connected with significantly higher costs for parents, perhaps perceived by them as an investment, which might contribute to their motivation to keep their children participating in such activities. Moreover, the rules in leisure centres are not as strict, and regular attendance is not required, as, for example, it is in organized sports activities. This might explain the higher fluctuation in participation of children in leisure centres compared with other organized activities. In contrast, the highest reliability was found for the art school variable. Art schools in the Czech Republic and Slovakia are set up under the Ministry of Education and the evaluation system is similar to a common elementary school, with a certificate given after the final examinations. Thus, it is very likely that

Table 3 Results of testing the test–retest reliability represented by Cohen’s kappa coefficients for the total sample and for gender, age and country subsamples (Czech Republic and Slovakia, 2013)

	Total		Slovaks		Czechs		Boys		Girls		5th grade		9th grade	
	κ	SE	κ	SE	κ	SE	κ	SE	κ	SE	κ	SE	κ	SE
Team sport activities	0.63	0.03	0.56	0.06	0.67	0.04	0.62	0.05	0.54	0.05	0.58	0.05	0.67	0.05
Individual sport activities	0.50	0.04	0.52	0.06	0.48	0.05	0.50	0.06	0.50	0.06	0.50	0.05	0.48	0.06
Art school	0.68	0.04	0.71	0.05	0.66	0.05	0.64	0.07	0.66	0.05	0.68	0.05	0.65	0.06
Youth organization	0.57	0.05	0.51	0.09	0.61	0.07	0.58	0.08	0.56	0.07	0.57	0.06	0.51	0.11
Leisure centres	0.46	0.04	0.43	0.07	0.49	0.05	0.47	0.06	0.45	0.06	0.40	0.06	0.54	0.06
Religious groups	0.51	0.06	0.50	0.07	0.44	0.10	0.36	0.09	0.61	0.07	0.47	0.07	0.56	0.11
	ICC	95 % CI	ICC	95 % CI	ICC	95 % CI	ICC	95 % CI	ICC	95 % CI	ICC	95 % CI	ICC	95 % CI
Sum score of activities	0.64	0.59–0.69	0.61	0.53–0.69	0.65	0.59–0.71	0.54	0.46–0.62	0.72	0.65–0.77	0.62	0.54–0.68	0.62	0.54–0.69
	κ	SE	κ	SE	κ	SE	κ	SE	κ	SE	κ	SE	κ	SE
Dichotomized sum of activities	0.61	0.05	0.59	0.08	0.62	0.06	0.63	0.07	0.59	0.06	0.52	0.08	0.65	0.06

children who, for example, play a musical instrument for several years will continue their activity regularly to gain their certificate. Next, we found that levels of reliability according to the Kappa coefficients differed between boys and girls for participation in religious groups. The higher reliability in girls might be explained by typical gender differences in religiosity (Smith et al. 2002). It seems that girls are more religious compared with boys (Smith et al. 2002) and this may have caused the higher test–retest reliability regarding participation in religious groups found in girls.

As it is important to reduce risk time and to increase the opportunities for appropriate use of time and the amount of time adolescents spends on situations leading to their healthy development, it is essential to assess their participation in organized activities, to provide a solid basis for precisely targeted interventions (Leversen et al. 2005). Adolescents’ participation in organized activities has already been examined or discussed from different perspectives in several studies (Youniss et al. 1999; Eccles et al. 2003; Rhodes and Spencer 2005; Fredrics and Eccles 2006; Persson et al. 2007; Fawcett et al. 2009; Zamboni et al. 2010). However, research on the reliability and validity of the organized activities measures is still scarce. Thus, comparing and interpreting our findings with prior results is a challenge, as previous studies are mainly focused on physical activities or recreation and leisure activities rather than solely on organized activities. Based on our knowledge, only a few studies have examined test–retest of adolescents’ participation in organized activities within the broader set of questions related to physical or leisure-time activities (Booth et al. 2002; King et al. 2004; Bult et al. 2010; Hong et al. 2012). Their findings are similar to ours, as all of the above-mentioned studies confirmed good reliability of the examined questionnaires.

Strengths and limitations

The strength of our study consists in the fact that, based on our knowledge, it is the first to assess the test–retest reliability of participation in organized activities. Another strength of our study is the period before administration of the retest, which was sufficiently long to avoid the retention of previously chosen answers and short enough to evade changes in lifestyle patterns. However, it is possible that some participants may have altered their pattern of activities over this period, which can be seen as a limitation of our study. This might also explain why the test–retest levels of agreement, although interpreted as acceptable, were not higher than those observed. Thus, the scale of participation in organized activities can be recommended for use in population studies, but in intervention studies it should be used with caution.

Another limitation should also be taken into account concerning the present study. The survey took place only in specific regions of the Czech Republic and Slovakia, and therefore our results should be viewed with some caution until they are confirmed in other populations. This study focused on the reliability of the selected items but did not investigate their validity; this would be an issue for future research.

Conclusion

The scale of participation in organized activities has an acceptable level of agreement, which indicates the good reliability of this measurement. Considering recent findings from the 4-H study of positive youth development (Bowers et al. 2014), it is important to consider both the breadth and the depth of activity involvement. Using the individual items of this scale enables measurement of the variety of engagement in organized structured activities (also cluster respondents) and more specificity with regard to their impact on developmental outcomes (Agans et al. 2014). On the other hand, the composite variable (the dichotomized or trichotomized number of activities respondents reported being engaged in) and not only individual items of this scale, offers us satisfactory reliability, but the composite variable also seems to be substantially stable. Therefore, this scale might be recommended as a brief measurement of adolescents' participation in organized activities. Future research and practice should continue to develop instruments that have better precision to reduce or overcome the methodological bias of the data.

Acknowledgments This work was supported by the Slovak Research and Development Agency under Contract No. APVV-0032-11 and by the IGA UP research project “Leisure Time in School-aged Children—HBSC Study”, Reg. No. FTK_2013_020, by the Grant Agency of the Czech Republic through the project “Social determinants of health in socially and physically disadvantaged and other groups of population” (Reg. No. CZ.1.07/2.3.00/20.0063) and by the ECOP project “Strengthening the scientific potential of research teams in promoting physical activity at Palacky University” Reg. No. CZ.1.07/2.3.00/20.0171.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

- Agans JP, Champine RB, DeSouza LM, Mueller MK, Johnson SK, Lerner RM (2014) Activity involvement as an ecological asset: profiles of participation and youth outcomes. *J Youth Adolesc* 43:919–932
- Allen MJ, Yen WM (1979) Introduction to measurement theory. Brooks/Cole, Monterey
- Badura P, Madarasova Geckova A, Sigmundova D, van Dijk JP, Reijneveld SA (2015) When children play, they feel better: organized activity participation and health in adolescents (**Under revision in BMC Public Health**)
- Bobakova D, Hamrik Z, Badura P, Sigmundova D, Nalecz H, Kalman M (2015) Test–retest reliability of selected physical activity and sedentary behaviour HBSC items in the Czech Republic, Slovakia and Poland. *Int J Public Health* 60:59–67
- Bohnert A, Fredricks J, Randall E (2010) Capturing unique dimensions of youth organized activity involvement: theoretical and methodological considerations. *Rev Educ Res* 80(4):576–610
- Booth ML, Okely AD, Chey T, Bauman A (2002) The reliability and validity of the Adolescent Physical Activity Recall Questionnaire. *Med Sci Sports Exerc* 34:1986–1995
- Bowers EP, Geldhof GJ, Johnson SK, Lerner JV, Lerner RM (2014) Special issue introduction: thriving across the adolescent years: a view of the issues. *J Youth Adolesc* 43:859–868
- Bult MK, Verschuren O, Gorter JW, Jongmans MJ, Piskur B, Ketelaar M (2010) Cross-cultural validation and psychometric evaluation of the Dutch language version of the Children's Assessment of Participation and Enjoyment (CAPE) in children with and without physical disabilities. *Clin Rehabil* 24(9):843–853
- Cohen J (ed) (1988) Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates, New Jersey
- Cook DA, Beckman TJ (2006) Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 119:166.e7–166.e16
- Currie C, Zanotti C, Morgan A, Currie D, de Looze M, Roberts C, Samdal O, Smith ORF, Barnekow V (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
- DeVellis RF (2003) Scale development: theory and applications, 2nd edn. Sage Publications, Thousand Oaks
- Downing SM (2004) Reliability: on the reproducibility of assessment data. *Med Educ* 38:1006–1012
- Eccles JS, Barber BL, Stone M, Hunt J (2003) Extracurricular activities and adolescent development. *J Soc Issues* 59:865–889
- Fatusi AO, Hindin MJ (2010) Adolescents and youth in developing countries: health and development issues in context. *J Adolesc* 33:499–508
- Fawcett LM, Garton AF, Dandy J (2009) Role of motivation, self-efficacy and parent support in adolescent structured leisure activity participation. *Aust J Psychol* 61:175–182
- Feder M (2008) Test–retest reliability. Encyclopedia of survey research methods. In: Lavrakas PJ (ed) SAGE Publications, Inc, Thousand Oaks, pp 888–890
- Fredricks JA, Eccles JS (2006) Extracurricular involvement and adolescent adjustment: impact of duration, number of activities, and breadth of participation. *Appl Dev Sci* 10:132–146
- Hong TK, Trang NHHD, van der Ploeg HP, Hardy LL, Dibley MJ (2012) Validity and reliability of a physical activity questionnaire for Vietnamese adolescents. *Int J Behav Nutr Phys Act* 9:1–7
- King GA, Law M, King S, Hurley P, Hanna S, Kertoy M, Rosenbaum P, Young N (2004) Children's assessment of participation and enjoyment (CAPE) and preferences for activities of children (PAC). Harcourt Assessment Inc, San Antonio
- King GA, Law M, King S, Hurley P, Hanna S, Kertoy M, Rosenbaum P (2006) Measuring children's participation in recreation and leisure activities: construct validation of the CAPE and PAC. *Child Care Health Dev* 33(1):28–39
- Kohl HW, Fulton JE, Caspersen CJ (2000) Assessment of physical activity among children and adolescents: a review and synthesis. *Prev Med* 31:54–76
- Kurpius SER, Stafford ME (2006) Testing and measurement: a user-friendly guide. Sage Publications Inc, California

- Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
- Larson RW (2000) Toward a psychology of positive youth development. *Am Psychol* 55:170–183
- Leveresen I, Rodriguez CM, Munoz-Tinoco V, Samdal O, Wold B (2005) Health behaviour in school-aged children: a WHO cross-national study. Research protocol for the 2005/2006 Survey. pp 1–15
- Leveresen I, Danielsen AG, Birkeland MS, Samdal O (2012) Basic psychological need satisfaction in leisure activities and adolescents' life satisfaction. *J Youth Adolesc* 41:1588–1599
- 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):1–9
- Mahoney JL, Larson R, Eccles J, Lord H (2005) Organized activities as developmental contexts for children and adolescents. In: Mahoney J, Larson R, Eccles J (eds) *Organized activities as contexts of development: extracurricular activities, after-school and community programs*. Lawrence Erlbaum Associates, New Jersey
- Mahoney JL, Harris AL, Eccles JS (2006) Organized activity participation, positive youth development, and the over-scheduling hypothesis. Social policy report. Giving child and youth development knowledge away. A publication of the society for research in child development, Vol XX, Number IV
- Mahoney JL, Harris AL, Eccles JS (2008) The over-scheduling myth. Research-to-results brief. Child Trends Inc, Washington, DC
- Nickerson CAE (1997) A note on “a concordance correlation coefficient to evaluate reproducibility”. *Biometrics* 53(4):1503–1507
- Nunnally JC, Bernstein IH (1994) *Psychometric theory*, 3rd edn. McGraw-Hill, New York
- Persson A, Kerr M, Stattin H (2007) Staying in or moving away from structured activities: explanations involving parents and peers. *Dev Psychol* 43:197–207
- Rhodes J, Spencer R (2005) Someone to watch over me: mentoring programs in the after-school lives of children and adolescents. In: Mahoney J, Larson R, Eccles J (eds) *Organized activities as contexts of development. extracurricular activities, after-school and community programs*. Lawrence Erlbaum Associates, Mahwah, New Jersey
- Roberts C, Freeman J, Samdal O, Schnohr C, de Looze ME, Nic Gabhainn S, Iannotti R, Rasmussen M, International HBSC Study Group (2009) The health behaviour in school-aged children (HBSC) study: methodological developments and current tensions. *Int J Public Health* 54(Suppl 2):140–150
- Smith C, Lundquist Denton M, Faris R, Regnerus M (2002) Mapping American adolescent religious participation. *J Sci Study Relig* 41(4):597–612
- Trinh OT, Nguyen ND, 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):S46–S53
- Youniss J, McLellan J, Su Y, Yates M (1999) The role of community service in identity development: normative, unconventional, and deviant orientations. *J Adolesc Res* 14:248–261
- Zambon A, Morgan A, Vereecken C, Colombini S, Boyce W, Mazur J, Lemma P, Cavallo F (2010) The contribution of clubs participation to adolescent health: evidence from six countries. *J Epidemiol Community Health* 61:89–95