

The reliability and validity of the physical activity questions in the WHO health behaviour in schoolchildren (HBSC) survey: a population study

M L Booth, A D Okely, T Chey, A Bauman

Abstract

Objective—To assess the test-retest reliability and validity of the physical activity questions in the World Health Organisation health behaviour in schoolchildren (WHO HBSC) survey.

Methods—In the validity study, the Multi-stage Fitness Test was administered to a random sample of year 8 (mean age 13.1 years; n = 1072) and year 10 (mean age 15.1 years; n = 954) high school students from New South Wales (Australia) during February/March 1997. The students completed the self report instruments on the same day. An independent sample of year 8 (n = 121) and year 10 (n = 105) students was used in the reliability study. The questionnaire was administered to the same students on two occasions, two weeks apart, and test-retest reliability was assessed. Students were classified as either active or inadequately active on their combined responses to the questionnaire items. Kappa and percentage agreement were assessed for the questionnaire items and for a two category summary measure.

Results—All groups of students (boys and girls in year 8 and year 10) classified as active (regardless of the measure) had significantly higher aerobic fitness than students classified as inadequately active. As a result of highly skewed binomial distributions, values of kappa were much lower than percentage agreement for test-retest reliability of the summary measure. For year 8 boys and girls, percentage agreement was 67% and 70% respectively, and for year 10 boys and girls percentage agreement was 85% and 70% respectively. **Conclusions**—These brief self report questions on participation in vigorous intensity physical activity appear to have acceptable reliability and validity. These instruments need to be tested in other cultures to ensure that the findings are not specific to Australian students. Further refinement of the measures should be considered.

(*Br J Sports Med* 2001;35:263–267)

Keywords: physical activity; exercise; measurement; health behaviour in schoolchildren; adolescence

Regular participation in physical activity is associated with a range of physical and mental health benefits in adults, including reduced risk

for cardiovascular disease, non-insulin-dependent diabetes mellitus, overweight, hypertension, osteoporosis, some cancers, and anxiety and depression.¹ Because there have been few long term studies, the relation between physical activity during childhood and adolescence and health during adulthood is less clear. However, evidence is accumulating to support the theory that improved blood lipid profiles, blood pressure, body composition, glucose metabolism, bone strength, and psychological health are causally associated with regular participation in physical activity.²

The first step toward understanding and encouraging participation in physical activity and enhanced fitness involves the provision of representative data to assess and monitor the population prevalence and distribution of physical activity participation. It is important to know not only what proportion of children and adolescents are active, but also if there are differences across the population by sex, among diverse cultural groups, or according to socioeconomic backgrounds or place of residence. Identification of these subgroups will allow public health and health promotion resources to be more equitably distributed.

Few systems exist to monitor child and adolescent health. The best known is the World Health Organisation health behaviour in schoolchildren (WHO HBSC) surveys, carried out in Europe for the past 15 years.³ Initial participation was by only a few countries in the 1980s increasing to 123 000 young people from 26 countries in the most recent 1997/98 survey.⁴ These school based surveys used standardised protocols and core questions across countries to measure a range of adolescent health and lifestyle issues, including physical activity and sedentary behaviours such as time spent watching television or using computers.

Kohl and his colleagues⁵ have recently conducted a comprehensive review of measures of physical activity participation among children and adolescents, including the strengths and limitations of the various methods of assessing their validity and reliability. They concluded that, at present, administration of self report questionnaires is the only practical method of collecting a broad range of data from a large number of children and adolescents, but it is still important that the questionnaire items provide acceptably accurate data. Consequently, a central question for self reported population measures is whether they

NSW Centre for the Advancement of Adolescent Health, Department of Paediatrics and Child Health, University of Sydney at The Children's Hospital at Westmead, New South Wales, Australia
M L Booth

Physical and Health Education Unit, Faculty of Education, University of Wollongong, Wollongong, New South Wales, Australia
A D Okely

Epidemiology Unit, South Western Sydney Area Health Service, Liverpool, New South Wales, Australia
T Chey
A Bauman

Correspondence to:
Dr Booth, NSW Centre for the Advancement of Adolescent Health, The Children's Hospital at Westmead, Locked Bag 4001, Westmead, NSW 2145, Australia
michaeb4@chw.edu.au

Accepted 4 May 2001

are known to have good measurement properties, especially test-retest reliability and validity.

The two studies reported here include the HBSC physical activity questions. In the validity study, a field measure of aerobic fitness (the Multistage Fitness Test (MFT))⁶ was also administered to the students who completed the questionnaire, allowing partial validation of the self report instrument. In the second study (the reliability study), the questionnaire was administered to the same students on two occasions, two weeks apart, and test-retest reliability was assessed. The data were stratified by sex and school year (years 8 and 10) before analysis. This pair of studies represents one of the first assessments of the reliability and validity of the WHO HBSC physical activity questionnaire items.

Methods

The validity and reliability studies were conducted separately and involved different schools and students. The two studies used the same survey instrument, which was administered under the same conditions—that is, in class groups under the supervision of at least one research officer. The survey instrument sought a broad range of self reported information, including: demographics; participation in physical activity using different measures, including the physical activity assessment items from the WHO HBSC; the amount of time spent in sedentary activities; questions about school physical education; and, individual factors associated with sport and other physical activity participation based on Social Cognitive Theory.⁷ These studies were approved by the University of Sydney human ethics committee, and informed consent was provided by all students before their participation.

QUESTIONNAIRE ITEMS

The WHO HBSC physical activity questionnaire asked students to report the frequency and total amount of time spent exercising vigorously outside school hours. The frequency question asked students “Outside school hours: How often do you usually exercise in your free time, so much that you get out of breath or sweat?” The response alternatives were: “Once a month or less”, “Once a week”, “2–3 times a week”, “4–6 times a week”, and “Every day”. The duration question asked students “Outside school hours: How many hours do you usually exercise in your free time, so much that you get out of breath or sweat?” The response alternatives were: “None”, “About half an hour per week”, “About one hour per week”, “About 2–3 hours per week”, “About 4–6 hours per week”, and “About 7 hours per week”.

DATA HANDLING

For the validity analyses, the responses to the two questions were summarised in two ways, based on guideline 2 of the *Physical activity guidelines for adolescents*.⁸ The criterion provided by guideline 2 was selected as the basis for categorisation because it is likely to be used as a cut off point for estimation of the

prevalence of sufficient activity among young people, and it is important to know whether those identified as more active on the basis of self report also have, on average, greater aerobic fitness. Firstly, responses were dichotomised for each question separately. For the frequency question, those who reported being active for once a week or less were considered inadequately active, and those who reported being active “2–3 times per week” or more were considered to be active. For the duration question, those who reported that they were active for “half an hour a week” or less were considered inadequately active and those who reported that they were active for “an hour a week” or more were considered active. The *Physical activity guidelines for adolescents* of Sallis and Patrick⁸ identify three times a week as the minimum frequency of participation to meet the criterion of being “active”, so the criterion used here of “2–3 times per week” will incorrectly categorise some students who participate only twice per week as “active”. This is an unavoidable limitation to this study and a significant shortcoming of the questionnaire.

Secondly, a dichotomous summary measure was created. Students who reported that they were vigorously physically active “2–3 times per week”, “4–6 times per week”, or “every day” and that they were vigorously active for “about 1 hour per week” or more were classified as active. Students who reported their frequency of activity as “once a month or less” or “once a week” or who reported their duration of activity as “none” or “about half an hour per week” were classified as inadequately active. The uncategorised responses to the two questionnaire items and the two category summary measure were used for the reliability analyses. The responses to the two questionnaire items were analysed separately in an attempt to identify whether one was more or less reliable than the other.

VALIDITY STUDY

Sample selection

Data relevant to the validation analyses were collected as part of the NSW Schools Fitness and Physical Activity Survey, 1997, the details of which have been published elsewhere.^{9,10} Briefly, 44 high schools were selected at random from the three education sectors (Department of School Education, Catholic Education Commission, and Independent Schools) in proportion to the number of students enrolled in each sector. The likelihood of a school being selected in each stratum was proportional to the size of the student enrolment. Within each school, one class was chosen at random from years 8 and 10. Assuming that classes would be about equal in size, this method provides a self weighted sample in which each student has an approximately equal chance of selection. The data were collected during February/March 1997 (summer).

Aerobic fitness

Aerobic fitness was assessed using the MFT (also known as the 20 m Shuttle Run Test, the

Table 1 Percentage of year 8 and year 10 boys and girls in each category of each measure of physical activity for the validity study

	Year 8		Year 10	
	Boys (n=557)	Girls (n=515)	Boys (n=525)	Girls (n=429)
Frequency*				
Active	78.6	68.6	78.1	60.9
Inadequately active	21.4	31.4	21.9	39.1
Duration†				
Active	81.4	73.3	86.2	70.2
Inadequately active	18.6	26.7	13.8	29.8
Two categories‡				
Active	71.4	60.8	74.7	54.7
Inadequately active	28.6	39.2	25.3	45.3

*Those who reported being active once a week or less were considered inadequately active and those who reported being active "2–3 times per week" or more were considered to be active.

†Those who reported that they were active for "half an hour a week" or less were considered inadequately active and those who reported that they were active for "an hour a week" or more were considered active.

‡Those who reported that they were vigorously physically active "2–3 times per week", "4–6 times per week", or "every day" and that they were vigorously active for "about 1 hour per week" or more were classified as active. Those who reported their frequency of activity as "once a month or less" or "once a week" or who reported their duration of activity as "none" or "about half an hour per week" were classified as inadequately active.

Table 2 Mean (SEM) number of laps for year 8 and year 10 boys and girls for the frequency of participation question, the duration of participation question, and for responses to the two questions categorised into two categories of participation

	Year 8		Year 10	
	Boys (n=557)	Girls (n=515)	Boys (n=525)	Girls (n=429)
Frequency				
Active	48.0 (1.73)	33.3 (1.99)	66.2 (2.38)	39.7 (2.10)
Inadequately active	41.4 (2.41)*	27.5 (1.64)*	58.1 (2.55)*	30.0 (2.64)*
Duration				
Active	47.8 (1.69)	33.0 (1.89)	65.7 (2.26)	38.2 (2.27)
Inadequately active	40.3 (2.46)*	27.1 (1.68)*	56.6 (3.01)*	30.3 (2.74)*
Two categories				
Active	48.2 (1.75)	33.9 (2.04)	66.4 (2.41)	40.0 (2.21)
Inadequately active	41.7 (2.00)*	27.6 (1.58)*	58.6 (2.33)*	30.9 (2.53)*

*Significantly different from active ($p < 0.01$).

Beep Test, or PACER); it was first described by Leger and Lambert⁶ and identified in a recent review as a reliable and valid field test for use among children and adolescents.¹¹ Students are required to run between two lines 20 m apart (one "lap"), starting at 8.5 km/h and increasing by 0.5 km/h every two minutes, in synchrony with a cadence tape. Students were tested in groups of about 15, and the test was supervised by at least two of the field team. The number of laps completed was determined by the student failing to keep pace with the cadence tape on two consecutive laps or voluntarily withdrawing. The stage and level achieved by each student was written on the back of their hands with a water soluble marker and recorded when all students had completed the test. Stage and level were converted into number of laps completed (laps) for the analysis. We recognise that aerobic fitness is, at best, only an indirect method of self report validation and that

Table 3 Percentage of year 8 and year 10 boys and girls in both categories of the two category summary measure at test 1 (T1) and test 2 (T2) for the reliability study

	Year 8				Year 10			
	Boys (n=63)		Girls (n=58)		Boys (n=75)		Girls (n=30)	
	T1	T2	T1	T2	T1	T2	T1	T2
Two categories								
Active	75.4	75.4	76.8	67.9	62.7	53.3	48.3	44.8
Inadequately active	24.6	24.6	23.2	32.1	37.3	46.7	51.7	55.3

collection of accelerometer data may be a preferable approach to validation. Unfortunately, this study lacked the necessary resources to administer accelerometers to the students.

RELIABILITY STUDY

Five high schools were selected at random from all high schools in the southern half of the Sydney metropolitan region and two high schools from a regional city were approached to participate in the study. The questionnaire was administered on two occasions, two weeks apart, in each of the schools during November and December 1998. One year 8 and one year 10 class in each school was selected by school staff to participate in the study.

STATISTICAL ANALYSIS

For the validity study, the mean and standard error of the number of laps on the MFT were calculated for the two questions separately, and for the summary measure. Standard errors were adjusted for the design effects. Comparisons of the mean number of laps between the active and inadequately active categories were carried out using multiple regression with dummy variables. Significance levels were adjusted for the design effects (resulting from schools and classes being the units of random selection) using SUDAAN.¹² Results of analysis of square root transformed data were not different from the results of analyses using crude data.

Test-retest reliability was assessed for agreement beyond chance using unweighted (for 2×2 contingency tables) and weighted (for multi-way contingency tables) kappa statistics with 95% confidence intervals. Skewed binomial data can give rise to erroneously low values of kappa,^{13 14} so percentage agreement was also calculated as the number of students classified as active at both test 1 and test 2 plus the number of students classified as inadequately active at both test 1 and test 2 as a proportion of the total number of students.

Results

VALIDITY STUDY

Of the 1072 year 8 students, 48% were girls and the mean age was 13.1 years; 45% of the 954 year 10 students were girls and the mean age was 15.1 years. The response rates were 86.6 and 83.5% for year 8 boys and girls respectively and 80.5% and 70.8% for year 10 boys and girls respectively.

Table 1 shows the proportion of year 8 and year 10 boys and girls in each category of each of the measures of physical activity, and table 2 shows the mean number of laps completed on the MFT for each category of each measure for boys and girls in years 8 and 10. With regard to the separate frequency and duration questions, the active group had significantly higher aerobic fitness than the inadequately active group for both sexes and both years. Similarly, for the dichotomous summary measure, aerobic fitness was significantly greater for the active than for the inadequately active students, for boys and girls in both years. The differences

Table 4 Values of kappa (K) (95% confidence interval) and percentage agreement (A) (95% confidence interval) for test-retest reliability for year 8 and year 10 boys and girls for the frequency of participation question, the duration of participation question, and for responses to the two questions categorised into two categories of participation

	Year 8		Year 10	
	Boys (n=63)	Girls (n=58)	Boys (n=75)	Girls (n=30)
Frequency	K=0.36 (0.20 to 0.51)* A=41% (29 to 54)	K=0.44 (0.25 to 0.64)* A=56% (42 to 69)	K=0.60 (0.48 to 0.72)* A=57% (45 to 69)	K=0.57 (0.34 to 0.80)* A=59% (39 to 77)
Duration	K=0.22 (0.05 to 0.39)* A=35% (23 to 48)	K=0.26 (0.09 to 0.44)* A=28% (17 to 41)	K=0.58 (0.47 to 0.70)* A=44% (33.56)	K=0.52 (0.29 to 0.75)* A=50% (31 to 69)
Two categories	K=0.12 (0.00 to 0.38)† A=67% (54 to 79)	K=0.25 (0.00 to 0.52)† A=70% (56 to 81)	K=0.70 (0.54 to 0.86)† A=85% (75 to 92)	K=0.38 (0.04 to 0.71)† A=70% (51 to 85)

*Weighted kappa.

†Unweighted kappa.

were statistically significant for all groups of students.

RELIABILITY STUDY

Of the 121 year 8 students, 48% were girls and the mean age was 13.7 years; 29% of the 105 year 10 students were girls and the mean age was 15.7 years. Table 3 shows the proportion of year 8 and year 10 boys and girls who were active and inadequately active at test 1 and test 2, and table 4 shows the values of kappa (95% confidence interval) and percentage agreement for each of the measures for year 8 and year 10 boys and girls separately. For the two questionnaire items, the values of kappa and percentage agreement were generally similar and all were 0.6 or lower, suggesting poor to moderate reliability. However, these items used five and six point response scales, making high agreement less likely, even when agreement is assessed using weighted kappa.

In contrast, the two category summary measure generally showed low values of kappa and higher percentage agreement values. Among year 8 boys and girls, a much higher proportion of students were active than inadequately active, resulting in a skewed binomial distribution and, consequently, paradoxically low values of kappa, despite percentage agreement being close to 70%. A similar, but less pronounced, phenomenon occurred among year 10 students, for whom the distributions were less skewed. Given the characteristics of these data, we consider that percentage agreement provides the better indication of test-retest reliability.

Discussion

The results of this study indicate that, at least among Australian high school students, the HBSC physical activity questions had acceptable test-retest reliability among year 8 students and acceptable to good reliability among year 10 students. However, these findings should be viewed with some caution until they are confirmed in other populations. That the reliability of self report physical activity measures generally improves with the age of the students is a further issue which has also been reported by Sallis and his colleagues.¹⁵

The data also indicate that the HBSC instrument has at least partial validity, as higher scores on the MFT differentiated the groups by self reported activity level. Aerobic fitness is an indirect measure of validity because other factors, in addition to physical activity, also

influence aerobic fitness. That is, a measure of aerobic fitness is, at best, only partially related to the differences between self report physical activity categories. It should also be kept in mind that physical activity recommendations for children and adolescents is a vexed issue and one for which consensus has not been achieved. For example, Pate and his colleagues¹⁶ presented a set of physical activity recommendations that did not include specific reference to a quantum of vigorous activity. Despite these limitations, data on aerobic fitness make a useful contribution to a comprehensive understanding of the characteristics of self report measures, particularly those that focus on vigorous intensity physical activity, as the HBSC items do. A better understanding of the validity of these questionnaire items would require the use of other approaches to validation, such as the use of accelerometers.

Some speculation on the reasons why better reliability was not seen may assist in the development of more reliable and valid self report measures of physical activity participation for use with young people. Firstly, the poorer reliability for year 8 than year 10 students is possibly because year 8 students are at that stage of adolescence during which authority is challenged and few things are taken seriously—that is, we suspect that the year 8 students may have taken less care in responding to the questions. Of course, this may vary from culture to culture, and the results achieved with this Australian sample may not be replicated in other countries.

Secondly, it may be that the questions are too simple and brief to yield accurate responses from many students. The questions do not include a reference period—for example, last or usual week—so students may consider different time periods in formulating their answer on different occasions. The questions also do not make reference to different types of activity—for example, organised sports and non-organised activities—so students may have been unclear about which activities to include in their report. Finally, students' participation may vary substantially with the season and, as the questions do not make reference to the season, students may be unclear about which activities to report.

The findings of this study suggest that the HBSC questionnaire items have acceptable reliability and validity among year 8 students and acceptable to good reliability and validity among year 10 students, but there is clearly

room for improvement. However, the poorer performance of the measure among the younger students, a finding that is consistent with other studies, raises doubts about the use of these measures with even younger adolescents. Similar studies should be conducted on the same questions in several other countries to confirm that the findings are not specific to Australian students. If lesser levels of agreement are found in other cultures, thought should be given to further development of the HBSC self report physical activity questions.

The study was supported by grants from the NSW Department of Education and Training, the NSW Department of Health, and the National Professional Development Program.

- 1 US Department of Health and Human Services. *Physical activity and health: a report of the Surgeon General*. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Centre for Chronic Disease Prevention and Health Promotion, 1996.
- 2 Riddoch C. Relationships between physical activity and health among young people. In: Biddle S, Sallis J, Cavill N, eds. *Young and active? Young people and health-enhancing physical activity: evidence and implications*. London: Health Education Authority, 1998:17-49.
- 3 World Health Organisation. *The health of youth. A cross national study*. Copenhagen: WHO Regional Publications, European Series No 69, 1996.
- 4 World Health Organisation. *Health and health behaviour among young people*. Copenhagen: WHO Regional Publications, 2000.
- 5 Kohl HW, Fulton JE, Casperson CJ. Assessment of physical activity among children and adolescents: a review and synthesis. *Prev Med* 2000;31:S54-76.
- 6 Leger LA, Lambert J. A maximal multistage 20-m shuttle run test to predict VO_2 max. *Eur J Appl Physiol* 1982;49:1-12.
- 7 Bandura A. *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice-Hall, 1969.
- 8 Sallis JF, Patrick K. Physical activity guidelines for adolescents: consensus statement. *Pediatric Exercise Science* 1994;6:302-14.
- 9 Booth M, Macaskill P, McLellan L, et al. *NSW schools fitness and physical activity survey 1997*. Sydney: NSW Department of School Education, 1997.
- 10 Booth ML, Macaskill P, Phongsavan P, et al. Methods of the NSW schools fitness and physical activity survey, 1997. *J Sci Med Sport* 1998;1:111-24.
- 11 Freedson PS, Curetan KJ, Heath GW. Status of field-based fitness testing in children and youth. *Prev Med* 2000;31: S77-85.
- 12 Shah BV, Barnwell BG, Bieler GS. *SUDAAN users manual, Release 7.5*. Research Triangle park, NC: Research Triangle Institute, 1997.
- 13 Feinstein AR, Cicchetti DV. High agreement but low kappa. I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43:543-9.
- 14 Feinstein AR, Cicchetti DV. High agreement but low kappa. II. Resolving the paradoxes. *J Clin Epidemiol* 1990;43:551-8.
- 15 Sallis JF, Buono MJ, Roby JJ, et al. Seven-day recall and other physical activity self-reports in children and adolescents. *Med Sci Sports Exerc* 1993;25:99-108.
- 16 Pate R, Trost S, Williams C. Critique of existing guidelines for physical activity in young people. In: Biddle S, Sallis J, Cavill N, eds. *Young and active? Young people and health-enhancing physical activity: evidence and implications*. London: Health Education Authority, 1998:162-76.

Take home message

The items of the HBSC physical activity questionnaire have acceptable reliability and validity among Australian students, but should also be evaluated among students from other cultures. They provide limited information, so more comprehensive questions should be considered for development.

BASEM Merchandise 2001

Ties	Multi motif	£6 + £1.50 p&p
New stock to order		
Sweaters	Lambswool fine knit, V-neck or round neck with small motif. Machine washable. State colour and chest size required.	£32 + £3 p&p
Sweatshirts	With small motif. State colour and chest size required.	£25 + £3 p&p
Polo shirts	With small motif. State colour and chest size required.	£23 + £3 p&p

Send orders to John H Clegg JP BSc (Hons) DipEcon (Open) LDS RCS Eng, Hon Secretary, Birch Lea, 67 Springfield Lane, Eccleston, St Helens, Merseyside WA10 5HB, UK. (Tel and Fax: 01744 28198)