VALIDITY AND RELIABILITY OF PHYSICAL ACTIVITY MEASURES IN GREEK HIGH SCHOOL AGE CHILDREN

Eugenia C. Argiropoulou, Maria Michalopoulou, Nikolaos Aggeloussis and Andreas Avgerinos

Democritus University of Thrace, Department of Physical Education & Sport Sciences, Komotini, Greece

Received: 12 February 2004 / Accepted: 09 June 2004 / Published (online): 01 September 2004

ABSTRACT
The aim of this study was to determine the validity and reliability of 3 physical activity questionnaires in Greek high school children. Forty children participated in the study aged M = 13.73 (SD 0.8 years). The validation study was conducted by comparing an accelerometer (MTI/CSA Model 7164) to 3 questionnaires: a) Three-day Physical Activity Record (3DPAR), b) Four by One-Day Recall Physical Activity Questionnaire (4BY1RPAQ) and c) Physical Activity and Life Style Questionnaire (PALQ). Validity of the 3 self-report questionnaires was assessed against the MTI/CSA accelerometer by comparing the scores obtained by each instrument on the first week of measurement. Reliability was assessed with two consecutive measurements performed two weeks apart. The measures of reliability were assessed by Intra Class Correlation, Typical Error and Limits of Agreement. A two-way ANOVA for repeated measures was performed. Repeated measures were week and day; in order to determine differences between the two scores obtained with the two measurements for MTI/CSA, 3DPAR and 4BY1RPAQ. A paired Student’s t-test was performed for the two scores obtained with the PALQ. Post-hoc multiple comparisons were performed using the Bonferroni test. Significance for all parts of the analysis was determined at an alpha level of \( p < 0.05 \). A paired Student’s t-test was performed for the two scores obtained with the PALQ. Results of this study indicated that reliability measured by intra class correlations (ICC) were for MTI/CSA (ICC = 0.52, \( p < 0.05 \)), 3DPAR (ICC = 0.97, \( p < 0.01 \)), 4BY1RPAQ (ICC = 0.70, \( p < 0.01 \)), and PALQ (ICC = 0.52, \( p < 0.01 \)). Significant Pearson product moment correlation coefficients (r) were observed between MTI/CSA and the other instruments, as a measure of validity: 3DPAR (r = 0.63, \( p < 0.01 \)), 4BY1RPAQ (r = 0.62, \( p < 0.01 \)), and PALQ (r = 0.53, \( p < 0.01 \)). The reliability of the four instruments used in this study was acceptable. Validity correlations were also significant for the three self-report instruments used in this study.

KEY WORDS: Physical activity, activity monitors, energy expenditure, children, validity, reliability.

INTRODUCTION
Physical activity appears to have a pervasive effect on health among adults and children (Baranowski et al., 1992). Many of the risk factors for coronary artery disease, hypertension, non-insulin dependent diabetes and osteoporosis seem to arise in childhood and early adolescence (Williams et al., 1981). Several studies have suggested that physical activity in childhood is a determinant of physical activity in adulthood (Dennison et al., 1988; Kuh and Cooper, 1992). Scientific evidence supports the hypothesis that moderate and high levels of physical activity are recognized as behaviours that lead to positive health benefits (Blair et al., 1989). In a longitudinal study performed over a period of 20 years physical activity behaviour was reported for 230 healthy men and women between 13 and 33 years of age (Kemper et al., 2002). Women increased participation in sport and light physical activities other than sport (4-7
to self report measures are used with children. The major sources of error identified are: the human cognitive processes, the definition of the desired variables, inadequate length of assessment and failure to account for weekday versus weekend and seasonal variations (Baronowski, 1988; Cale, 1994). Language and cultural differences may also impose additional constraints when assessing physical activity in population groups that differ from the ones that the instruments were originally validated for. Self-report methods are considered a cost advantages over other measures, but they also have limitations. Those limitations are evident when self report measures are used with children. The major sources of error identified are: the human cognitive processes, the definition of the desired variables, inadequate length of assessment and failure to account for weekday versus weekend and seasonal variations (Baronowski, 1988; Cale, 1994). Language and cultural differences may also impose additional constraints when assessing physical activity in population groups that differ from the ones that the instruments were originally validated for. Self-report methods are considered a cost effective assessment approach that is convenient to administer and feasible when testing large populations (Sallis et al., 1993).

To reduce the errors associated with self report methods in children, objective methods such as motion sensors have been developed. The Computer Science and Applications Inc. (MTI/CSA) uni-axial activity monitor (WAM 7164; Computer Science and Applications Inc., Shalimar, FL) is one example (Ekelund et al., 2001). The validation of the MTI/CSA monitor provides a precise tool for measuring changes in acceleration in laboratory settings (Metcalf et al., 2001). The MTI/CSA monitor can be useful in a field situation (Sirard et al., 2000). The validation of the MTI/CSA monitor certifies this monitor as valid, reliable and useful device for the assessment of physical activity in children (Trost et al., 1998; Puyau et al., 2002). Its small size and lightness creates an advantage when compared to other direct methods of estimating physical activity such as heart rate monitoring. Other advantages that accelerometers have when compared to other methods of measuring physical activity include storage of movement data for long periods of time, objective estimation of frequency, intensity and duration of physical activity and continuous recording. The continuous recording of physical activity data for future recall excludes problems related to subjective recall that physical activity questionnaires demand (Bassett et al., 2000).

Despite the advantages of accelerometers, the energy cost of an activity that includes muscular activity resulting from isometric contractions, from movements of the upper body, from additional load bearing (carrying, lifting, and pushing), and from movement on inclined or soft surfaces will not result in an increase in the number of counts recorded by any uni-axial accelerometer device (Bouten et al., 1994; Hendelman et al., 2000). Additionally accelerometers cannot be used while swimming or during any other activities where water is involved, and their effectiveness is limited in recording activities such as cycling and weight lifting (Bassett et al., 2000). Most motion sensors are sensible to velocity changes but not to gradient changes. This lack of sensitivity may be important in laboratory settings but in field studies variations in duration, frequency and intensity which are the most important components of physical activity are still recorded (Montoye et al., 1983).

Four days of measurement, including weekdays and weekend days, should provide acceptable correlations ensuring a representative measure of a child’s habitual physical activity in a large field based study (Janz, 1994; Janz et al., 1995; Trost et al., 2000). In the present study children wore the MTI/CSA monitors for 7 consecutive days during the two periods of measurement in order to achieve greater reliability in recording physical activity (Trost et al., 2000; Matthews et al., 2002).

The aim of the present study is to report on the reliability of three questionnaires 3DPAR (Bouchard et al., 1983), 4BY1RPAQ (Cale, 1993) and PARQ (Avgerinos, 2000) when used to record physical activity of Greek High School children. The validity of these instruments was assessed by using the MTI/CSA accelerometer as the criterion standard.

**METHODS**

**Subjects**

Forty children (23 boys and 17 girls) ranging in age from 13 to 14 years (mean 13.73 ± 0.80 years) participated in this study. Subjects were randomly selected and recruited from 7 high schools with the help of the physical education teachers who worked...
at the schools. The children and their parents were informed about the scope and the procedure that would be followed in this study and all parents signed an informed consent form.

Both body mass and height were recorded for all subjects. Height was measured without shoes by using a wall-mounted tape measure, and body mass was measured using a digital laboratory scale (Seca, Model 770, Olney, MD). Body mass index (BMI) was computed as weight in kilograms divided by height in square meters.

**Instruments**

*Motion Detector MTI/CSA 7164*

The MTI/CSA (7164) activity monitor is a small (5.1 x 3.8 x 1.5cm) lightweight (43-45g) uni-axial activity monitor. It is designed to detect acceleration ranging in magnitude from 0.05 to 2.0G with a frequency response from 0.25 to 2.50 Hz (CSA Inc., Shalimar, FL 1995). These parameters allow the detection of normal human motion while rejecting high frequency motion encountered outside these ranges. The filtered acceleration signal is digitized and the magnitude is summed over a user-specified epoch interval. At the end of each interval, the summed value or activity “count” is stored in memory. The stored data can be downloaded to a computer and the integrator can be reset. For the current study a one minute time interval was used. Further technical specification and performance properties have been described elsewhere (Janz, 1994; Melanson and Freedson, 1995; Freedson et al., 1998; 2000; Nichols et al., 2000; Swartz et al., 2000; Ekelund et al., 2000; 2001).

The MTI/CSA accelerometer has been well validated in both children and adolescents against a wide range of outcomes (Freedson et al., 1998; Trost et al., 1998; Freedson and Miller, 2000; Welk et al., 2000; Ekelund et al., 2000; 2001; Brage, et al., 2003). It has been validated against energy expenditure measured by indirect calorimetry and it was found to be a valid tool for quantifying energy expenditure in children and adults during treadmill running and walking. Correlations ranging from \( r = 0.50 \) to \( r = 0.74 \) have been reported between the MTI/CSA monitor and heart rate telemetry of children in field settings (Janz, 1994). Additionally the MTI/CSA monitor provided more accurate estimations of energy expenditure when compared to TriTrac and other accelerometers (Welk, et al., 2000). In studies with children, a significant correlation was observed between MTI/CSA activity counts and all energy expenditure estimates using the Doubly Labeled Water method (Ekelund et al., 2001).

**Three-day physical activity record**

This three-day activity record divides each of the three days (one must be a weekend day) into 96 periods each 15 minutes long. The responder records his/her energy expenditure for each 15 minute period using a scale from 1 (sleep) to 9 (vigorous physical activity and sport). The values of this scale correspond to a range of 1.0 to 7.8 METs and higher. These categories are explained to the participant in material given to him or her for personal use during the completion of the 3DPAR. Approximate median energy cost for each of the nine categories in (Kcal·kg\(^{-1}·15\) min\(^{-1}\)) was used to compute the daily energy expenditure for each individual in kcas (Bouchard et al., 1997). The reliability of this questionnaire has been tested for both children and adults (ICC = 0.86 - 0.95, \( p < 0.01 \)) and its validity has been also reported for children (\( r = 0.80, p<0.01 \)) and adults (\( r = 0.54, p < 0.01 \)) (Bouchard et al., 1983).

The four by one-day recall physical activity questionnaire

This interviewer administered self-report measure was designed for children aged 11 years and older and is used to gather four days of activity information (2 school days and two weekend days). It consists of two forms: school day and weekend day, the form is segmented into parts: morning, afternoon and evening and contains checklists of activities. The questionnaire measures four dimensions of physical activity: a) physical activity at school, b) sport at school, c) physical activity during leisure time and d) sport during leisure time. Additionally the questionnaire provides an objective scoring system and measure of physical activity in terms of: a) average daily energy expenditure (Kcal·kg\(^{-1}·day^{-1}\)), b) time spent on moderate physical activity, c) time spent on hard and very hard activity and d) bouts of “huff” and “puff” activity. The final score assigns the responder to one of the following categories: very inactive, inactive, moderately active and active. The reliability and validity of this questionnaire have been tested for 11 to 14 year-old children and have been reported as \( r = 0.62 \) (\( p < 0.05 \)) (reliability) and \( r = 0.61 \) (\( p < 0.01 \)) (validity) (Cale, 1994).

**Physical activity and lifestyle questionnaire**

This self-administered questionnaire is designed to assess the physical activity of youngsters, ages 11 to 18 years. The questionnaire is made up of two parts: Part 1 records participation in physical activity during leisure time in which the responder must record how frequently s/he participates in any of the 27 physical activities listed in the questionnaire. Part 2 assesses the responder’s participation in physical activities during the last 7 days. PALQ can be answered in 15 to 25 minutes depending on the age.
of the responder. The scoring procedure results in a total score of energy expenditure for each responder that can be calculated using the Compendium of Physical Activities (Ainsworth et al., 1993). Subjects are assigned to one of four categories according to their score: very inactive, inactive, moderately active and active. Additionally daily time of participation in vigorous and moderate physical activities and sport can be calculated from the data collected. Only data collected for the second part was used in this study.

Selection of questionnaires used in this study

No self-report and/or interviewer-administered instruments that assess physical activity have been validated for use with Greek High School children (Avgerinos, 2000). The researchers adapted the 3DPAR and the 4BY1RPAQ questionnaires in order to: a) check the children’s understanding of the questions and the material included in the translated versions of the two questionnaires, and b) determine the time required to read and complete them. The adaptation from English to Greek language did not require any changes in either the number or meaning of the questions (both questionnaires). Additionally, by using 3DPAR and 4BY1RPAQ we would able to use and provide data, for Greek children, that is comparable with data from studies conducted in other countries and assess how well these instruments can be adapted and used in Greece.

4BY1RPAQ uses the Compendium of Physical Activities (Ainsworth et al., 1993), which is also used by PALQ, a new instrument that has been designed to assess the physical activity of 12-18 year old Greek students. PALQ is the first questionnaire devised in Greece that provides native researchers and Physical Education instructors with the means to assess the habitual physical activity of Greek high school students.

Standard methods were used to translate and adapt the 4BY1RPAQ and the 3DPAR questionnaires. Two pilot studies were conducted in order to: a) check the children’s understanding of the questions and the material included in the translated versions of the two questionnaires, and b) determine the time required to read and complete them. The subjects in these pilot studies were Greek high school students, \( n_1 = 6 \) boys and 4 girls aged 13.57 ± 0.70 years and \( n_2 = 8 \) boys and 7 girls aged 13.64 ± 0.70 years). Where the subjects had difficulties in understanding the material given to them, additional information was included in the final versions of the questionnaires.

Study Design

The study was designed to compare the criterion method, the MTI/C SA, against a single-interviewer administered questionnaire and two self-report methods of assessing physical activity. The study entailed 14 days of children’s involvement. In order to provide a measure of reliability and/or behaviour stability, each questionnaire was administered twice and for a two week interval. Additionally, the subjects wore the MTI/C SA accelerometer twice for seven consecutive days (2 non consecutive weeks) during the same days that they responded to the questionnaires. Activity monitoring data was collected for 14 days.

All interviews were conducted in an environment familiar to the subject (school or home). Subjects were encouraged to maintain their daily physical activity schedule and in case of illness the study was discontinued. The MTI/C SA accelerometer was placed on a carrying pouch and participants were shown how to place the pouch on a belt at waist level on the right anterior axillary line. The investigator checked the functionality of the accelerometer every morning during the period of the study. Consistent and proper placement was emphasized, and subjects were told to wear the monitor at all times (day and night), even while sleeping, and to remove it only if the monitor would get completely wet, such as when showering or swimming. The investigators trained the subjects how to attach and detach the accelerometer. During both pilot studies, subjects would repeatedly forget to put on the MTI/C SA monitor as they were getting ready to go to school in the morning. This observation was the reason behind the decision to advise the children to continuously wear the MTI/C SA monitor. As a result, MTI/C SA data was collected for a period of 24 hours each day, which was also the case for the data from the questionnaires used in this study. Compliance was monitored by investigators every morning as the children arrived at school. The MTI/C SA was initialized according to the manufacturer’s specifications. The epoch interval was set for 1 minute. At the end of the 7-day recording period, the investigator used a reader interface unit to download the MTI/C SA data (counts) to a desktop computer. The data was then stored in an Excel file for further analysis.

The 3DPAR was completed by the children and returned to the investigator at the end of the third day. The 4BY1RPAQ was administered by the investigator for two weekdays and two weekend
days and referred to the physical activity undertaken by the subject the previous day. The third questionnaire (PALQ) was completed just after the end of the seven-day period during which the subjects wore the MTI/CSA accelerometer. Data collection was completed within a period of one month since changes in weather conditions over long periods of time could affect the children’s habitual physical activity. The possible effects due to weather changes were observed during the pilot study. The above procedure ensured that the questionnaires were recorded during identical periods of time, their guidelines were satisfied and activity recall from the completion of another questionnaire was minimized. Moreover, this design enabled the researchers to comparatively evaluate the data of all the measures of physical activity used in this study. (However, comparative evaluation is not discussed in this paper).

**Statistical Analysis**

The physical characteristics of the children were summarized using descriptive statistics. For reliability, intraclass correlation coefficients (ICC) were calculated for the repeated administrations of the MTI/CSA, the 3DPAR, the 4BY1RPAQ and the PALQ, using appropriate two-way ANOVA models (Baumgartner, 1989). A two-way ANOVA with repeated measures (week x day) was performed in order to determine differences between scores obtained during the two measurements for MTI/CSA, 3DPAR and 43BY1RPAQ. Post-hoc multiple comparisons were performed using the Bonferroni test. Significance for all parts of the analysis was determined at an alpha level of p<0.05. A paired Student’s t-test was performed for the two scores obtained with the PALQ. Two additional measures of reliability were also computed for the MTI/CSA, the 3DPAR, the 4BY1RPAQ and the PALQ: the typical error (Hopkins, 2000) and the limits of agreement (LOA) (Bland and Altman, 1986). Pearson product moment correlation coefficients between the MTI/CSA and the three questionnaires were used to assess their validity.

**RESULTS**

<table>
<thead>
<tr>
<th></th>
<th>CSA (counts·min⁻¹)</th>
<th>3DPAR (Kcal)</th>
<th>4BY1RPAQ (Kcal)</th>
<th>PALQ (Kcal)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>400 (168)</td>
<td>2 292 (658)</td>
<td>2 036 (492)</td>
<td>2 113 (566)</td>
</tr>
<tr>
<td>n = 33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B Data</strong></td>
<td>465 (215)</td>
<td>2 285 (503)</td>
<td>2 125 (426)</td>
<td>2 169 (408)</td>
</tr>
<tr>
<td>n = 18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Participants Physical Characteristics

Boys’ (n = 23) body mass (mean ± SD) was 58.7 ± 12.2 kg, their height was 1.66 ± 0.11m and their BMI was 21.3 ± 3.54. Girls’ (n = 17) body mass (mean ± SD) was 52.1 ± 10.8 kg, their height was 1.60 ± 0.07m and their BMI was 20.3 ± 3.3. From the BMI readings obtained in this study, boys resulted between the 75th and 85th percentile distribution and girls resided between the 50th and the 75th percentile. (National Center for Health Statistics, USA, 2000), both groups being between the normal range (Himes and Deitz, 1994).

The mean ± SD for the 4 instruments obtained during the first and second measurement period is presented in Table 1. Original 4BY1RPAQ and PALQ data was transformed from METs to Kcal.
were 38.84 ± 3139.68 counts·min⁻¹. The typical error and the LOAs for the separate days were for:
Saturday 265.68 and 55.68 ± 758.97 counts·min⁻¹,
Sunday 129.15 and -33.82 ± 368.96 counts·min⁻¹,
Monday 121.64 and -11.03 ± 347.48 counts·min⁻¹,
Tuesday 259.10 and -28.58 ± 740.17 counts·min⁻¹,
Wednesday 207.78 and 41.45 ± 593.58 counts·min⁻¹,
Thursday 221.78 and -26.98 ± 633.56 counts·min⁻¹ and Friday 299.32 and 42.03 ± 835.08 counts·min⁻¹.
No significant (week x day) interaction was reported (F1,17 = 0.003, p > 0.05), and no significant main effect for the factors week, (F 1,17 = 0.009, p > 0.05) and day (F1,17 = 0.577, p > 0.05).

3D Physical Activity Record.
The ICC for the two 3day periods of measurement was (ICC = 0.97, p < 0.01). When ICC was calculated for each day of measurement separately, weekdays gave higher ICC readings (weekdays 1 and 2 ICC = 0.97, p < 0.01) than the weekend day (ICC =0.88, p < 0.01). The typical error and the LOAs for 3DPAR between the two 3day periods of measurement were 382.51 and -375.30 ± 1092.72 Kcals respectively. The typical errors for each of the weekend days were equal to 4.77 METs and for each of the two weekdays were respectively 1.25 and 2.25 METs. No significant (week x day) interaction was reported (F1,20 = 2.746, p > 0.05). The total scores between the 2 weeks of measurement were not significantly different (F1,20 = 3.178, p > 0.05). A significant main effect was reported for the factor: day (F 1,20 = 9.172, p<0.01). A Bonferroni post-hoc test revealed that the weekend day scores (Sunday) of the 2 weeks of measurement were significantly different.

Physical Activity and Lifestyle Questionnaire
The ICC for the two periods of measurement was (ICC = 0.52, p < 0.05). The typical error and the limits of agreement, for PALQ, between the two periods of measurement were respectively 2.39 and -1.88 ± 6.82 METs. The physical activity score for the first week was significantly different from the score of the second week (t = 2.547, p < 0.05). Average energy expenditures (METs) recorded with 4BY1RPAQ per day and PALQ for the first (A Data) and the second measure (B Data) were presented in Figure 3.

4BY1 Recall Physical Activity Questionnaire
The ICC for the two 4day periods of measurement was (ICC = 0.70, p < 0.01). When ICC was calculated for each day of measurement separately weekdays had higher correlation coefficients (weekday 1 ICC = 0.83, p < 0.01, & weekday 2 ICC = 0.89, p < 0.01,) than weekend days (Saturday ICC = -0.30, p > 0.05, and Sunday ICC = 0.47, p > 0.05). The typical error and the LOAs for 4BY1RPAQ between the two 4day periods of measurement were 7.94 and -4.37 ± 22.67 METs respectively. The typical errors for each of the weekend days were equal to 4.77 METs and for each of the two weekdays were respectively 1.25 and 2.25 METs. No significant (week x day) interaction was reported (F1,20 = 0.020, p > 0.05). No differences were reported between the total scores of the 2 weeks of measurement (F1,20 = 10.108, p > 0.01). A significant main effect was reported for the factor: day (F 1,20 = 4.624, p < 0.05). A Bonferroni post-hoc test revealed that Sunday scores of the 2 weeks of measurement were significantly different.

Validity of the Physical Activity Instruments
MTI/CSA vs. 3D Physical Activity Record.
Table 2. Validity: Pearson product moment correlations (r) of 3DPAR, 4BY1RPAQ, and PALQ with MTI/CSA.

<table>
<thead>
<tr>
<th></th>
<th>Weekend day 1</th>
<th>Weekend day 2</th>
<th>Week day 1</th>
<th>Week day 2</th>
<th>Average scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DPAR</td>
<td>.56**</td>
<td>.09</td>
<td>.64**</td>
<td>.63**</td>
<td></td>
</tr>
<tr>
<td>4BY1RPAQ</td>
<td>.38**</td>
<td>.54**</td>
<td>.46**</td>
<td>.65**</td>
<td>.62**</td>
</tr>
<tr>
<td>PALQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.53**</td>
</tr>
</tbody>
</table>

The correlation between 3DPAR (kcal) and MTI/CSA (counts·min⁻¹) for the respective days was moderate (r = 0.63, p<0.01) (Table 2). The correlations of each of the three individual days between the 3DPAR and the MTI/CSA monitor were significant for weekend day 2 (r = 0.56, p<0.01) and for the weekday 2 (r = 0.64, p<0.01).

MTI/CSA vs. 4BY1RPAQ Recall Physical Activity Questionnaire
The average correlation between the 4 days of 4BY1RPAQ (METs) and the MTI/CSA counts·min⁻¹ was moderate (r = 0.62, p < 0.01, Table 2). The correlations of each of the four days between the 3DPAR and the MTI/CSA monitor were also significant for both weekend days (Saturday r = 0.38, p < 0.01 and Sunday r = 0.54, p < 0.01) and weekdays (Monday r = 0.46, p < 0.01 and Tuesday r = 0.65, p < 0.01).

MTI/CSA vs. Physical Activity and Lifestyle Questionnaire
The average correlation between the 7 days of PALQ (METs) and the respective days of MTI/CSA (counts·min⁻¹) was moderate (r = 0.53, p < 0.01, Table 2).

DISCUSSION
Promotion of physical activity in children and adolescents and the understanding of different aspects of physical activity state the need for valid and reliable assessment instruments. If data gathered is not accurate then valuable information may be lost and progress in understanding physical activity behaviour and its relationship to health behaviour may be hindered.

Reliability
The results in this study suggest MTI/CSA is a reliable instrument to monitor physical activity in children. Moderate but acceptable reliability (ICC = 0.52) was observed between the first and the second week of measurement. This finding is in accordance with the results of other reliability studies conducted with heart rate monitors and motion sensors in children and youngsters (Freedson and Evenson, 1991; DuRant, et al., 1992; 1993; Janz et al., 1992; 1994; 1995). In general though, the reliability studies that used the test re-test paradigm presented medium to high correlations.

In this study, the term reliability refers to the consistency of scores of measurement differentiating between inter-instrument reliability. Different experimental designs (test re-test reliability measured up to 12 times in one day and 3-6 months later) and statistical analysis (Intraclass Correlation Coefficient, Pearson product moment correlation coefficient and ANOVA) employed in studies assessing the reliability of one instrument limit drastically the possibilities of making meaningful comparisons between their findings.

Reliability of the three questionnaires was tested for the first and the second week of measurement and all reliability measures reported high correlations for the 3DPAR (ICC = 0.97) medium correlation for the 4BY1RPAQ (ICC = 0.70), and low correlation for the PALQ (ICC = 0.52). Even though energy expenditure scores of the weekend day were significantly higher for the first week of measurement when compared to the second week, high correlations were provided for the weekdays (Monday, ICC = 0.97 and Tuesday ICC = 0.97) as well as between the weekend day (Sunday ICC = 0.88). These findings are in accordance with the results of the study conducted by Bouchard and his colleagues (1983) (r = 0.96).

Reliability was moderate for 4BY1RPAQ between the first and the second week of measurement. However, lower correlations were observed between the separate days of the first and second week. For weekdays reliability scores were higher (Monday ICC = 0.83 and Tuesday ICC = 0.89) than weekend days (Saturday ICC = -0.30 and Sunday ICC = 0.47). Physical activity scores were not different between the two weeks of measurement but physical activity scores of the first Sunday were significantly higher when compared to Sunday scores of the second week. This difference in physical activity was also reported with data collected with the 3DPAR. The lower physical activity scores recorded for the second Sunday may be due to the fact that with repeated activity interviews subjects are inclined to underreport
physical activity (Kemper et al., 1983). This effect may be more apparent in a weekend day during which no structured activities which take place - are easier for children to recall, are included in their daily schedule.

The PALQ demonstrated the lowest reliability of all the instruments used in this study and according to the results physical activity recorded during the first week was significantly higher when compared to the second week. Scores were not produced for weekend days and weekdays since this instrument does not provide data for separate days within the week.

Test-re-test reliability of children’s physical activity questionnaires varies from $r = 0.20$ (Andersen and Haraldsdottir, 1993) to $r = 0.98$ (Weston et al., 1997). Age has been positively related with the reliability of questionnaires with higher correlations being recorded with older children and youngsters when compared with children about under the age of 10 years. A significant disadvantage of these instruments refers to limitations related to accuracy of recall and subjective interpretations of the questions (Sallis, 1991). In the present study when the PALQ was used subjects had difficulty recalling events and activities that had already been recorded with the other 2 questionnaires. Evidence for memory decay (Baranowski, 1988) comes also from a study that included the comparison of 7-day self reported activity of children at a diabetic camp with a diary of their activity maintained by their camp counselor. The children could with reasonable accuracy recall the activity of the previous day, but had great difficulty with day’s further back in time (Wallace et al., 1985). Daily logs – diaries are not used very often with children and youngsters (Bouchard et al., 1983; Freedson and Evenson, 1991). This has probably to do with the effort needed to consistently complete the daily log. (Children in the present study had similar difficulties in completing the 3DPAR).

Reliability is affected by the duration between repeated measurements of physical activity. Relatively low correlations were presented for the 7-day Physical Activity Recall (Sallis et al., 1993) when data collection was repeated 4 to 6 days later. A different correlation was obtained when a third set of data was collected 2 to 3 days later. Studies longer in duration (Aaron et al., 1993; Andersen and Haraldsdottir, 1993) would probably be useful in determining a subject’s behavioural consistency related to physical activity. Even though an instrument may be reliable behavioural changes may weaken the statistical interpretation of the reliability data (Kohl et al., 2000).

Validity
One goal of this study was to examine the validity of using the MTI/CSA as a criterion standard of the physical activity questionnaires used in this study.

Validation studies with objective measures of physical activities (motion sensors) in children and youngsters presented low, moderate (Klesges and Klesges, 1987; Ballor et al., 1989; Mukeshi et al., 1990; Noland et al., 1990; Sallis et al., 1990; Freedson and Evenson, 1991; Bray et al., 1994; Janz, 1994; Welk and Gorbin, 1995) and high correlations (Trost et al., 1998).

Self reports are the most commonly checked instruments for validity in children and adolescents. The criteria of convergent validity most commonly used are the methods of doubly labeled water (DLW), direct observation and electronically and mechanical monitoring of physical activity (Noland et al., 1990; Sallis et al., 1990; 1993; 1996; Simons-Morton et al., 1994; Janz et al., 1995; Weston et al., 1997). In their review Sirard and Pate (2001) considered 3 types of measures of physical activity in children and adolescents: primary, secondary and subjective measures. Direct observation, doubly labeled water and indirect calorimetry were considered the primary standards for assessment of physical activity. Heart rate monitors, pedometers and accelerometers were considered secondary measures because they provide an objective assessment of physical activity. Surveys, self – report questionnaires, interviews, proxy-reports and diaries were considered as subjective techniques (Sirard and Pate, 2001).

When motion sensors were used for the validation of self report measures with children 10 years and older, correlation coefficients ranged form $r = 0.03$ (Janz et al., 1994) to $r = 0.88$ (Weston et al., 1997). In the present study correlation between the MTI/CSA and the three questionnaires was moderate, $(r = 0.63, r = 0.62$ and $r = 0.53, p<0.01)$ for the 3DPAR, 4BY1RPAQ and PALQ, respectively. This range in variability is possibly due to differences in the design and philosophy of the questionnaires used and in the duration of monitoring physical activity.

The relationship between accelerometry and energy cost is highly dependent on the type of activity being performed. Therefore it may be in appropriate to apply equations based on laboratory tasks or locomotion to free living situations in attempts to quantify energy expenditure resulting from physical activity or to classify activity levels (Hendelman et al., 2000). No single regression equation appears to accurately predict energy expenditure based on acceleration scores for all activities (Basset et al, 2000). The vertical
acceleration of the body can be measured quite accurately, but the relationship with measured energy expenditure (METs) differs depending on the type of physical activity performed (Basset et al., 2000). Moreover converting accelerometer counts to units of energy expenditure may provide inaccurate estimates because of the additional measurement error (Sirard and Pate, 2001).

A validation study conducted by Sallis and colleagues (1993) included 4 instruments used by 66 children of the fourth grade in classroom settings allowing the comparison between questionnaires of different format and the Caltrac activity monitor. Correlations were consistently lower for the second administration of each instrument. With relatively low correlations the authors suggested some combination of monitoring devices and self-report may be a prudent step to increase validity (Sallis et al., 1993). Different methods of assessing physical activity are not measuring identical properties or components. Total physical activity is a function of the type of stimulus (mode of exercise), the intensity at which the stimulus is performed, and the duration of a single episode. Over an extended period, the frequency with which an exercise is performed is also important. For example, one type of electronic monitor may not measure the intensity of physical activity as well as a recall instrument or a diary may, but the monitor may more accurately measure duration. If this true, researchers may need to use multiple methods to more completely assess all components of physical activity (Kohl et al., 2000). Because of the limitations of any of the practical field methods of assessing energy expenditure and physical activity, there must be attempts to combine approaches in an effort to improve validity (Montoye et al., 1996).

The main findings of this study were the relatively high correlation of MTI/CSA and 3DPAR and the moderate correlations observed between MTI/CSA and the other instruments tested in this study. In a study by Sirard and colleagues (2000) the 3DPAR and MTI/CSA assessed similar patterns of physical activity in young adults.

In a previous study in which 3DPAR was compared to CALTRAC and 4 other physical activity questionnaires no significant correlations were found (Miller et al., 1994). One of the explanation that the authors suggested was that 3DPAR uses 9 physical activity categories that were insufficiently described and subjects were forced to interpret of each category resulting in misunderstandings In our study, this last element was present but its effect was minimized since the investigator met frequently with the subjects in order to complete the 4BY1RPAQ and provided further clarification when needed.

The participants in this study reported that unstable weather conditions (cold and rain) throughout short periods of this study affected their habitual physical activity. Difficulties were reported concerning the use of the physical activity index, which is included in the 3DPAR and when completing the PALQ, subjects were forgetting events and activities that had already been recorded with the other two questionnaires. The procedure of completing three questionnaires during the same period of time disturbed the participants since they had to meet with the researcher that “checked their schedule” on four consecutive days in order to complete the 4BY1RPAQ. The time children went to bed the night before the first weekday was not recorded by the 4BY1RPAQ. Even though the MTI/CSA monitor is a small and lightweight device, subjects reported having to wear specific underwear and clothing in order to keep it properly attached to their body.

Despite the limitations of accelerometers, objective monitoring of physical activity in population based samples of children and adolescents appears to be a feasible alternative to traditional self-report methods (Trost et al., 2002). The accelerometers provide the necessary objectivity that paper and pencil techniques lack and are not as expensive as the doubly labeled water technique (Sirard et al., 2000). The DLW method measures total energy expenditure over longer periods and therefore provides a good estimate of average daily energy expenditure. However, the high cost of the stable isotopes and sophisticated analysis techniques that are required by this method limit its usefulness in epidemiological studies (Ekelund et al., 2001). Additionally, the DLW method provides no data regarding brief periods of peak energy expenditure (Montoye et al., 1996). Lastly, accurate dietary records must be obtained during the measurement period for the energy expenditure calculations (Sirard and Pate, 2001).

CONCLUSIONS

In conclusion, the assessment of physical activity in children and adolescents faces several limitations. A predominant characteristic of human behaviour is interpersonal and intrapersonal variability. Only a combination of the available instruments would be able to respond to this variability when assessing physical activity in children and adolescents. Self-report instruments and accelerometers are probably able to quantify only gross fluctuations in physical activity. Additionally, continuous monitoring over long periods of time may disturb habitual physical activity. The results of this study lead to the conclusion that all 4 instruments were valid and
reliable in recording physical activity when used with children, since the instruments were able to detect changes in physical activity and the respective energy cost.

REFERENCES


Ekelund, U., Sjostrom, M., Nygve, A., Poortvliet, E., Nilsson, A., Froberg, K., Wederkopp, N. and


Validity, reliability of physical activity measures

Medicine and Science in Sport and Exercise 22, 698-703.


AUTHORS BIOGRAPHY

Eugenia C. ARGIROPOULOU
Employment
Democritus University of Thrace, Dept. of Physical Education & Sport Sciences, Komotini, Greece.
Degree
PhD
Research interests
Physical activity in youth.
E-mail michal_07030@yahoo.com

Maria MICHALOPOULOU
Employment
Ass. Prof. Democritus University of Thrace Dept. of Physical Education & Sport Sciences, Komotini, Greece.
Degree
PhD
Research interests
Physical activity and health promotion in the elderly, motor learning.
E-mail michal_07030@yahoo.com

Nikolaos AGGEOUSSI
Employment
Ass. Prof. Democritus University of Thrace Dept. of Physical Education & Sport Sciences, Komotini, Greece.
Degree
PhD
Research interests
Biomechanics, research methods.

Andreas Avgerinos
Employment
Loughborough University of Technology, UK
Degree
PhD
E-mail: avgerino@phed.auth.gr
KEY POINTS

- The PALQ demonstrated a moderate reliability (0.52) and validity (0.53) in recording physical activity.
- A relatively high correlation was observed between the MTI/CSA and 3DPAR and a moderate correlation was observed between MTI/CSA and the 4BY1RPAQ tested in this study.
- Only a combination of the available instruments would be able to respond to the interpersonal and intrapersonal variability when assessing physical activity in children and adolescents. Self-report instruments and accelerometers are probably able to quantify only gross fluctuations in physical activity.
- All 4 instruments used in this study were valid and reliable in recording physical activity when used with children, since the instruments were able to detect changes in physical activity and the respective energy cost.

Dr. E.C. Argiropoulou
Democritus University of Thrace, Dept. of Physical Education & Sport Sciences, Komotini 69100 GREECE.