2009

Validation of the SOPLAY direct observation tool with an objective accelerometry-based physical activity monitor

Pedro Frederico De Saint-maurice Maduro
Iowa State University

Validation of the SOPLAY direct observation tool with an objective accelerometry-based physical activity monitor

by

Pedro De Saint-Maurice

A thesis submitted to the graduate faculty in partial fulfillment

of the requirements for the degree of

MASTER OF SCIENCE

Major: Kinesiology (Behavioral Basis of Physical Activity)

Program of Study Committee:
Gregory J. Welk, Major Professor
Philip Martin
Lorraine Lanningham-Foster

Iowa State University

Ames, Iowa

2009

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I cannot celebrate this accomplishment without acknowledging,

the great support from the research team involved in this project,

the wisdom from my POS committee members and Dr Welk,

my dear friends and family,

my mother, Maria Gabriela, for the unique opportunity, role model, and support,

and to Helga, my love, friend and confident, who was there every single night,

Thank You.

ABSTRACT

The System for Observing Play and Leisure Activities (SOPLAY) is a direct observation instrument designed to facilitate observation of groups and environmental contexts. To date, no field-based studies have been done to test validity of SOPLAY using objective criteria. **Purpose:** The purpose of this study was to test the validity of the SOPLAY with corresponding data from an accelerometry-based activity monitor (Biotrainer pro) in a sample of 9-12 years old children. **Methods:** Data was collected from 19 different physical activity sessions with a total of 160 children (mean age 10.5±0.8). The SOPLAY direct observation tool and activity monitors were used to evaluate the activity levels of the groups on two different occasions. The agreement between instruments and the influence of SOPLAY sampling rate on its validity were tested. The primary outcome measure for all comparisons was the percent of youth categorized as being “active” but different measures were used to reflect this variable. One measure (MVPA1) used the sum of “walking” + “very active” to reflect activity while another measure (MVPA2) was based on the percentage of youth coded as “very active”. **Results:** Difference between observed and recorded activity levels varied depending on what coding was used. There were large and significant differences when the standard scoring system was used for interpreting the SOPLAY (MVPA1: 50.55±26.41%, p-value<0.001). There was an overall better agreement (non-significant, p-value>0.01) when the alternative measure was used (MVPA2: 1.33±22.06%). The combination of “walking” and “very active” was found to have good correspondence when compared with a parallel measure from the accelerometer based on the sum of light and moderate to vigorous activity (MVPA3: -2.02±29.00%). Correlations for the different MVPA classifications followed the same pattern (MVPA1: 0.404, MVPA2: 0.562, MVPA3: 0.575). Evaluation of the impact of scan rate on validity was evaluated by comparing agreement with different number of scans. Although observation scans every 2, 4, 6, 8, 12 and 20 had systematic increases in error, there was a substantial absolute error (21.76%) associated when there was only 1 scan per session (20 minutes scans). **Conclusions:** Observations provide valid indicators of MVPA if coding is based on the percentage of youth classified as “very active”. The results demonstrate that more frequent scans can improve the validity of the estimations.

CHAPTER I – INTRODUCTION

The prevalence of obesity among children and adolescents has increased significantly from 1999-2004 (Ogden, Carrol, Curtin, McDowell, Tabak and Flegal, 2006). Recent data from the National Health and Nutrition Examination Survey (NHANES) have shown that 31.9% of children and adolescents (2-19 years old) were at or above the 85th percentile of the Centers for Disease Control and Prevention (CDC) 2000 BMI-for-age growth charts (Ogden, Carroll and Flegal, 2008). These troubling trends have led to increased interest in understanding and promoting physical activity behavior in children.

Accurate assessments of physical activity are important to advance research on children's physical activity (Welk, 2002). Measurement of physical activity in field conditions can be done using a variety of instruments but each has advantages and disadvantages (Welk and Wood, 2000). Self-report instruments are widely used but depend on responses from the child and therefore are limited by children's ability to accurately report details of their activity behaviors (Sallis and Saelens, 2000). Objective measures of physical activity (e.g. heart rate monitors, motion sensors, pedometers, and accelerometers) offer advantages over subjective methods but are more expensive and place a burden on participants and researchers. While these methods are also widely used, challenges in processing and interpreting this type of data limit the utility of these approaches. A method that avoids some limitations of both subjective and objective methods is the direct observation technique. This method is often viewed as the most effective (gold standard) technique for youth related research because behavior is directly observed (McKenzie, 2002).

The most widely used direct observation tool is called SOFIT (System for Observing Fitness Instructor Time) and this tool has been validated using a variety of methods (McKenzie, Sallis, Nader, Patterson, Elder, Berry, Rupp, Atkins, Buono and Nelson, 1991; Rowe, 1997). While the SOFIT tool has been widely used, a limitation is that it can only be used to evaluate activity behaviors of individual children. Because an observer can only code 1 person at a time, the cost and burden limit its utility for field based research. To address this limitation, an alternative tool called SOPLAY (System for Observing Play and Leisure Activities) was developed. This instrument was specifically designed to facilitate observation of groups and environmental contexts (McKenzie, 2002). The SOPLAY assesses physical activity levels and contextual factors (e.g., is the area under observation usable for physical activity?) using momentary group time sampling techniques. An observer scans a particular setting over a specific time interval (e.g. 5 minutes) and codes the number of youth at different levels of activity using a simple 3 point classification system (McKenzie, 2006). Pre-determined target areas are scanned and each individual's physical activity at that time is coded as sedentary (i.e., lying down, sitting or standing), walking or very active (these last two codes represent moderate-to-vigorous physical activity). Scans are done separately for boys and girls to allow for group comparisons. Contextual factors of the target area are also coded to facilitate interpretation of the data. Unique, pre-defined codes are used to assess whether the setting was accessible (e.g., not locked), usable, supervised, organized, and whether equipment was provided. SOPLAY also records the predominant type of activity being performed at the time of scanning (McKenzie, Marshall, Sallis and Conway, 2000).

The SOPLAY tool can capture detailed information about children's physical activity behavior in different settings. Surprisingly, limited research has been conducted on the reliability and validity of the SOPLAY tool. Interobserver reliability has been tested and considered acceptable for both SOPLAY contextual variables and activity counts (IOA=80%, R=.75) (McKenzie, Marshall et al., 2000). McKenzie et al. (2000) observed leisure-time physical activity in school environments and established the concurrent validity of SOPLAY with self-report measures. This study indicated a low correlation between the self-report measure and the SOPLAY for the after-school period (r=.35). Nevertheless, the authors suggested that this finding was most likely attributable to the self-report measurement error (McKenzie, Marshall et al., 2000). To date, no field-based studies have been done to test validity of SOPLAY using objective criteria. Direct validation of SOPLAY is still needed and accelerometers provide the most effective way to do this work since they can be temporally matched to the individual scanning periods.

Purpose of the Study

The purpose of this study was to test the concurrent validity of the SOPLAY with corresponding data from an accelerometry-based activity monitor in a sample of 9-12 year old children during an after-school physical activity program. The Biotrainer Activity monitor was selected for use in the present study to provide objective information about physical activity. The BioTrainer Activity monitor (Biotrainer, IM systems, Baltimore, MD), is a bidirectional accelerometer and has demonstrated high correlations with output from other accelerometers (Welk, Blair, Wood, Jones and Thompson, 2000). This instrument has been showed to provide valid and reliable indicators of energy expenditure in adults (Welk, Almeida and Morss, 2003) and it has also been previously calibrated for use with children – using direct observation techniques (Welk, Eisenmann, Schaben, Trost and Dale, 2007).

The overall project was conducted as part of an ongoing evaluation study of activity levels in an after school physical activity program for children in West DesMoines, IA. The SOPLAY tool was used in the project to provide information about physical activity in this time period. To evaluate the validity of the SOPLAY assessments, children in the after school period were provided with an activity monitor to wear during the observed physical activity sessions. By linking the observed SOPLAY data to the recorded physical activity levels from the activity monitor it is possible to directly validate the SOPLAY coding.

Specific Research Questions

The study specifically evaluated two research questions related to the validity of SOPLAY.

1. Does SOPLAY data accurately characterize the activity levels of groups of children participating in structured physical activity? It was hypothesized that there would be good overall agreement (evidenced by high correlations and non-significant differences with p-value>0.01) between observed and recorded activity levels.

2. Does the sampling rate for SOPLAY observations affect the validity of physical activity data collected during group activity? It was hypothesized that correlations with the Biotrainer would be higher when observation data are aggregated at 2 minute intervals rather than 4, 6, 8, 10, 12 or 20 minute intervals.

CHAPTER II – EXTENDED LITERATURE REVIEW

The study evaluated the agreement between observed physical activity levels (based on the SOPLAY recording system) and recorded activity assessed with an accepted, accelerometry-based activity monitor. Important topics relevant to the assessment of children's physical activity are summarized in this literature review to provide a background and justification of the study. Specific information is provided on the associations between physical activity and weight status in youth as well as the importance of school-based physical activity interventions for promoting physical activity. Detailed information on the types of physical activity tools used for evaluating activity in youth is then provided. A variety of measures are summarized and an example of a physical activity intervention is provided: CATCH, measuring physical activity (subjective measures: self-report questionnaires; secondary measures: accelerometers; criterion standards: direct observation methods) and future research focus (issues and concerns).

Overweight Status and Children's level of Physical Activity

The prevalence of obesity among children and adolescents has increased significantly from 1999-2004 (Ogden, Carrol et al., 2006). Ogden et al (2006) analyzed data from the National Health and Examination Survey (NHANES) on 3958 children aged 2-19 years old. The prevalence in female children population increased from 13.8% in 1999-2000 to 16% in 2003-2004 as in boys increased from 14.0% to 18.2%.

Overweight status in this survey was defined based on sex-specific BMI for age growth charts from the Centers for Disease Control and Prevention (CDCP). Children between the 85th and the 95th percentile were classified as at risk of overweight as a BMI over or at the 95th percentile would be classified as overweight. In general, both risk of overweight and overweight status prevalence increased significantly from 1999 to 2004. Also interesting is to note that Mexican-Americans and non-Hispanic black children were at increased risk for overweight when compared with non-Hispanic white children. It was concluded that prevalence of overweight and obesity in the United States remains high (Ogden, Carrol et al., 2006). In addition to these findings, data from the 2006 NHANES also suggested high levels of overweight children even if no significant changes were found between 2004 and 2006. This data suggests that 31.9% of children and adolescents were at or above the 85th percentile based on growth curves from CDCP (Ogden, Carroll et al., 2008).

Obesity is now considered to be a critical public health threat as its rate in the past three decades has more than doubled in preschool children (aged 2-5 years old) and adolescents (aged 12-19 years old) and tripled for children aged 6-11 years old (Koplan, Liverman and Kraak, 2005). It is estimated that approximately nine million children over 6 years old are obese. There is consensus that effective interventions strategies are needed to address this national health problem (Koplan, Liverman et al., 2005). While researchers agree that obesity is a priority health issue, there are still many unanswered questions about factors that influence activity and associations with weight status. A good example provided by Eisenmann et al (2008) is the relation between television, physical activity and childhood overweight. It seems that watching more than four hours of TV and performing less than two days a week of MVPA contributes for a significant increased risk for overweight in both males and females adolescents. Nevertheless this finding was not consistent for males when analyzing the relation of low levels of vigorous physical activity and high levels of watching TV behavior. As the literature suggests, physical activity and TV watching are two independent behaviors and by doing so both should be measured independently as indicators of activity and inactivity (Ekelund, Brage, Froberg, Harro, Anderssen, Sardinha, Riddoch and Andersen, 2006; Eisenmann, Bartee, Smith, Welk and Fu, 2008).

Studies have shown, a decrease in overall levels of physical activity (Allison and Adlaf, 2000) and also in bouts of five, ten and twenty minute from childhood through adolescence (Trost, Pate et al., 2002). The provision of structured physical activity during the day is important for youth activity promotion since it increases opportunities for youth to engage in regular physical activity (Allison and Adlaf, 2000). After-school programming is frequently targeted since it reflects an important part of the day where physical activity levels can be most easily modified (Stone, McKenzie et al., 1998).

The importance of School-based Physical Activity Interventions

Current recommendations suggest that children should engage in 60 minutes or more of daily moderate-or-vigorous physical activity (Services, 2008). Assessments of school playtime suggest that youth typically achieve only 40% of this recommendation while at school (Ridgers, Stratton and Fairclough, 2006). Therefore, additional interventions are needed, exploring the potential benefits of structured activity as for example playground-based interventions (e.g. organizing and leading games during playtime) (Ridgers, Stratton and Fairclough, 2006). Improving the quality of physical education programming is also important since evidence suggests that time spent in MVPA is often less than 50% of the total session time (Allison and Adlaf, 2000; Powers, Conway, McKenzie, Sallis and Marshall, 2002). McKenzie et al (1995) determined from a total of 293 physical education sessions observed in 95 different schools that children were only enrolled in MVPA or vigorous physical activity during 17.5% and 36.2% of the lesson time respectively. Notice that average duration of physical education sessions across all schools was approximately 30 minutes (McKenzie, Feldman, Woods, Romero, Dahlstrom, Stone, Strikmiller, Williston and Harsha, 1995). In another study, where 430 physical education lessons in 24 different middle-schools (6-8th grade) were observed, physical activity levels were also relatively low. In average, only 48.5% of the total session time (16.8 min) was spent in MVPA (McKenzie, Marshall, Sallis and Conway, 2000). These two studies mentioned provide some evidence that activity promotion during school time is not sufficient to accomplish the national objectives for physical activity (from the Department of Health and Human Services: Healthy People 2010) during physical education classes (children should be active during 50% of class time) (McKenzie, Feldman et al., 1995; McKenzie, Marshall et al., 2000). It is important that opportunities for children to engage in high intensity and organized activity are provided so children can have additional health benefits and physical activity recommendations can be fulfilled (Powers, Conway et al., 2002; Hoos, Kuipers, Gerver and Westerterp, 2004; Sigmund, De Ste Croix, Miklankova and Fromel, 2007).

Research to date supports the importance of school-based interventions as this setting offers the potential to reach large segments of the population and to promote activity at multiple time points (during physical education time, recess and during class breaks) (Dale, Corbin and Dale, 2000). Physical education classes have been considered to be insufficient to fulfill physical activity recommendations. Therefore, other strategies are needed to increase physical activity levels in the school setting (Ridgers, Stratton et al., 2006). The time after school has been increasingly targeted since it avoids scheduling constraints and pressures on teachers time. Coordinated after school programs can also address other social problems likely to occur during after-school hours (McKenzie, Marshall et al., 2000). School-based physical activity interventions seem to be of great importance to help achieving higher levels of physical activity in youth (Stone, McKenzie et al., 1998; Ridgers, Stratton et al., 2006). As the Centers for Disease Control and Prevention already stated, there is a need for extracurricular after-school physical activity programs in schools targeting the youth population (Prevention, 1997).

Examples of established school-based physical activity interventions include SPARK (Sports, Play & Active Recreation for Kids), PLAY (Promoting Lifetime Activity for Youth) and CATCH (Coordinated Approach to Child Health). These programs have been shown to be effective in school settings (Luepker, Perry, McKinlay, Nader, Parcel, Stone, Webber, Elder, Feldman, Johnson and et al., 1996; Sallis, McKenzie, Alcaraz, Kolody, Faucette and Hovell, 1997; Ernst and Pangrazi, 1999). The CATCH program (originally developed with the acronym of Child and Adolescent Trial for Cardiovascular Health), is a multicomponent, multicenter trial that targets diet, physical activity and non-smoking behaviors among children from third to fifth grade

(McKenzie, Nader, Strikmiller, Yang, Stone, Perry, Taylor, Epping, Feldman, Luepker and Kelder, 1996). It was originally developed to enhance quality of physical education classes but concepts and principles from the program have been adapted for use in a variety of settings including after-school activity programs (Coleman, Tiller, Sanchez, Heath, Sy, Milliken and Dzewaltowski, 2005). The CATCH design consisted of a development and application of classroom curriculum, school environmental change and family involvement programs, for each grade level and behavioral focus (e.g. physical activity). Physical activity goals were to promote children enjoyment and participation in moderate to vigorous physical activity (MVPA) during physical education classes. The CATCH program actually consists of four of the eight components that the Center for Disease and Control (CDC) has determined to prevent undesirable health outcomes and social problems: Health Education, Physical Education, Family Community Involvement, Nutrition Services, Health Services, Counseling Psychological Services, Healthy School Environment, and Health Promotion for Staff, and Family Community Involvement. This intervention is suitable for all the children either overweight or not and consists of five components: Go For Health Classroom Curriculum, CATCH Physical Education, Eat Smart School Nutrition Guide, CATCH Kids Club, and Family Team Activities. Special attention should be given to the CATCH Kids Club, an after-school portion of the CATCH program that contains a curriculum related with both physical activity and nutrition concepts (Coordinated Approach to Child Health, 2009).

To effectively evaluate physical activity interventions is important to have valid and reliable assessments of physical activity behavior. One can argue that the relation of physical activity with those intervention programs has been attenuated due to the limitations of physical activity instruments used in the past. Assessing physical activity in children is considerably more difficult than in adults due to limitations in recall and more sporadic intermittent activity patterns (Welk, Corbin and Dale, 2000).

A review of available techniques is provided in the next section to provide a background for the present study.

Measuring Physical Activity

Physical activity can be quantified and interpreted in a variety of ways. Caspersen et al. (1985) previously described physical activity as: "Any bodily movement produced by skeletal muscles that results in caloric expenditure", this definition has been accepted within the field of epidemiology (Caspersen, Powell et al., 1985). This concept is often associated and related with two primary domains, surveillance research and intervention research. The concept of physical activity is used when establishing relations with health outcomes, or when performing basic research (understanding of body response and adaptation to physical activity). Measurement of this construct has proven to be very challenging. To advance research on activity promotion it is important to develop more effective measurement tools (Welk, 2002). Sirard and Pate (2001) categorized physical activity instruments in three types: criterion standards, secondary measures and subjective measures (Sirard and Pate, 2001).

Criterion standard measures of physical activity are considered to be the most valid and reliable approach and should be used when precision is needed or when validating other assessment tools. Examples of criterion standard measures include direct observation, doubly labeled water and indirect calorimetry. Secondary measures provide an objective indicator of activity. Examples include heart rate monitors and motion sensors (pedometers, accelerometers). These approaches are typically validated against a primary standard and represent some of the most commonly used techniques (Sirard and Pate, 2001). Subjective techniques, such as survey methods, are considered subjective because they rely on responses from the child and by doing so some precautions should be taken when using this type of instruments. Details on the methods are provided below. Subjective methods are described first, followed by secondary measures and criterion measures. The limitations of subjective techniques justifies the need to use objective techniques with children. The focus in the secondary measures is on accelerometers and the focus in the criterion section is on direct observation since these methods were compared in the study

Subjective measures:

There are four main examples of subjective instruments: self-report questionnaires, interviewer-administered questionnaires, proxy-report questionnaires and diaries (Trost, Pate et al., 2002). Some limitations of these approaches are particularly relevant for studies involving children. Children have particular memory and recall skill limitations but even so, surveys and questionnaires are often used in this population (Sirard and Pate, 2001). When using these methods, ambiguous terms like physical activity and moderate intensity must be understood by researchers, as they can require cognitive skills beyond children capabilities (Sallis and Saelens, 2000). A study with 4th grade students compared the effect of video, verbal, and no instruction about the concept of physical activity. One hundred and twenty seven 127 children were randomly divided across these three groups, instructions were given about physical activity according to the different methods, and then children would fill a 17-item checklist testing their understanding of physical activity. The physical activity checklist asked students to classify 17 commonly performed activities in physical activity and not physical activity. Interestingly, only 35.6% of the students in the control group compared with 52.4% and 70% of the students in the verbal and video group were able to classify 15 or more of the checklist items correctly. This study indicates that both the generic verbal description and instructional video were effective in helping students understand the meaning of physical activity, with the instructional video being more effective than the verbal description. The authors concluded that fourth grade students display a limited understanding of the concept of physical activity, and brief instructions can help these students when trying to understand this concept (Trost, Morgan, Saunders, Felton, Ward and Pate, 2000). McKenna et al. (2004) explored qualitative interviews when recalling physical activity. They found in participants aged 8-16 years old, that when unassisted, participants gave vague descriptions of daily activities. It was found that a range of qualitative techniques can help one recall important elements of physical activity (Mckenna, Foster and Page, 2004). Sallis et al (1993), using a sample of 102 children from 5th to 11th grade, found that older subjects were more reliable than the young ones. An age of 10 years old seems to be the minimum to achieve adequate reliability and validity to use in research (Sallis, Buono, Roby, Micale and Nelson, 1993).

Another complicating issue is related with exercise intensity reproduction. Often participants have to indicate how many bouts of moderate or vigorous exercise they performed. Seventeen children aged 8-14 years old were compared with an adult group in their ability to execute prescriptive exercise using the Borg 6-20 rating of perceived exertion scale (RPE). Exercise was performed at four different RPE levels: 7, 10, 13 and 16. During pace controlled exercise (cycling), children had similar perceptions than adults when reproducing four incremental intensities but when performing self-paced tasks, children could only discriminate RPE 7 from the other levels. Adults were closer to the criterion measure in both types of tasks (established heart rate for RPE scale) than children (Trost, Ward, Moorehead, Watson, Riner and Burke, 1998).

Gender is another factor that can influence recall due to higher levels of physical activity performed by boys. Results from the 1990 Youth Behavior Survey indicated that boys perform more vigorous physical activity than girls, and this has been attributed to perceived confidence in overcoming barriers to physical activity and participation in community physical activity programs (Trost, Pate, Dowda, Saunders, Ward and Felton, 1996). Sallis et al (1993) also found that boys are usually more reliable than girls when reporting physical activity (Sallis, Buono et al., 1993). In addition to these factors, social desirability also seems to influence physical activity report - typically leading to over-reporting of physical activity levels. Even if not assessed in children, a study using 81 women reported that social approval was weakly associated with underestimation of physical activity when using a 24-hour physical activity recall questionnaire. Social desirability and social approval may influence self-reported physical activity on some survey instruments (Adams, Matthews, Ebbeling, Moore, Cunningham, Fulton and Hebert, 2005).

Besides children's ability to recall and gender differences, other limitations may emerge as for example, reliability of physical activity assessment (e.g., one day, one week, two weeks?). Researchers usually extrapolate these findings and assume physical activity in children has a regular pattern across a day or a week. Research has shown that typical activity patterns of children should be collected across several days (Stone, McKenzie et al., 1998). Note that a study from Baranowski et al. (1999), found that to achieve a reliability of 0.8 using a seven-day activity record, two weeks of daily assessment is required. The study had a sample of 165 teachers from 3 to 5-grade in 32 elementary schools and was done during late February to mid March time of the year, which can have some influence on the findings obtained. It was found that depending on the time of the year, levels of physical activity can possible be higher (Baranowski, Anderson and Carmack, 1998). Fisher et al (2005) studied 209 children who attend nursery in Glasgow to see if there was a significant seasonal variation in objectively measured habitual physical activity and sedentary behavior in young children. It was found that in preschool children, even with small differences, had greater levels of physical activity during the summer than in the fall (Fisher, Reilly, Montgomery, Kelly, Williamson, Jackson, Paton and Grant, 2005).

Secondary measures: Accelerometers

Due to the limitations already mentioned of subjective instruments, objective activity instruments are now used more frequently (Trost, 2001). Accelerometers as the name indicates are based on measures of acceleration which makes it possible to quantify human movement. The instruments are composed of piezoelectric bender elements that work as transducers. Whenever acceleration occurs, a charger is produced that is proportional to the force exerted. Further, accelerometers also include high and low-pass filters that exclude accelerations and signals that are outside the range of human movement. By doing so, high-pass filters remove low frequency signals (e.g. if a person is stationary a value of zero for the acceleration will be recorded) and low-pass filters remove high frequency accelerations and electrical interference (e.g. human body accelerations are usually less than 10 Hz and between -6 and 6 g even if forces higher than that may occur during high impacts and landings). Accelerometers measure limb acceleration instead of whole body acceleration, and by doing so, detects acceleration changes in the hip during each stride. Less prominent to bias than self-report measures, this type of instruments can be used to assess the frequency, intensity, duration, and energy expenditure of physical activity (Welk, 2002).

There are several factors that can influence accelerometer output and contribute to measurement error. The effect of positioning of the accelerometer should be consistent because if the position of the accelerometer changes during the study and across participants (e.g. waist vs hip vs ankle) one might expect significant error in the measure. Another limitation is that waist worn accelerometers can’t detect upper body movement (Welk, 2002). The use of a wrist placed monitor would me more appropriate but then locomotion movement would not be detected. The same is true for cycling, a monitor in the knee would most likely detect movement when compared with a waist mounted monitor. Nevertheless, the knee would not be a desirable place to assess other common activities. There have been research done that analyzed the use of multiple accelerometers, placed in different places of the body, but still the degree of improvement for predicting energy cost using those multiple accelerometers simultaneously does not seem to be of great advantage. The use of multiple accelerometers may not justify the increased cost of these instruments and also the burden that the participant may experience (Welk, 2002).

The last relevant factor that may influence accelerometer output is related with the instrument orientation. Accelerometers can detect movement in different planes of motion (e.g. picking up movement around a particular axis), recording movement proportionally to the direction of orientation. Considering that an accelerometer is designed to only measure movement in the vertical axis, it should be placed in a vertical position in the body (e.g. waist) so it can detect movement in that plane of motion. Three-dimensional accelerometers can attenuate this issue since they detect acceleration in three axis of motion. Nevertheless, depending on the predominant movement, a uniaxial accelerometer can be adequate to quantify movement, accessing movement on the Y direction (Eston, Rowlands and Ingledew, 1998; Welk, 2002).

There are other important issues related to data processing and issues associated as non-compliance with monitoring protocols Accelerometers can provide data for long periods of time (e.g. 7 days) without the need to recharge or download the data. Data is usually expressed in counts (considered to be the raw outcome from accelerometers) but these raw scores can be difficult to interpret particularly since physical activity guidelines are based on minutes of activity (Welk, 2002). Equations have been established to allow raw activity counts to be converted into time spent in different activity intensity categories and this has become common practice in field research. The problem with the use of established cut-points is that they may vary from monitor to monitor as data is processed in different ways and raw counts from accelerometers have different meanings (Welk, Corbin and Dale, 2000). Nevertheless there are already established cut-off points for different intensity thresholds (Welk, 2002). Freedson et al (1998) used a sample of 80 children aged 6-17 years old to differentiated the raw counts into categories of light (<3 METS), moderate (3-6 METS), hard (6-9 METS) and very hard (>9 METS) activity. Three different treadmill conditions were tested (walking and jogging pace) as oxygen consumption was measured and accelerometer was used. Data from the accelerometer and oxygen consumption was linked and predictors of METS were identified so a regression equation to predict energy expenditure could be generated (Freedson, Melanson and Sirard, 1998).

Even if these thresholds can give some meaning to the raw counts for public health research, there are still some issues associated with these techniques as for example, the transformation of a continuous outcome (raw scores) in a categorical outcome and also the inter-individual variability related with anthropometric measures (e.g. a taller person that would have a longer leg will generally have fewer movement counts for a standardized bout of activity). In addition, the assumption that transitions between the different thresholds are associated with a steady state exercise may also be misleading. Children for example have shown to have very sporadic activity patterns leading to a systematic underestimation of activity in this population. Depending on the time frame used to detect activity, the scores obtained may mislead physical activity and energy expenditure estimations (Welk, 2002; Nilsson, Brage, Riddoch, Anderssen, Sardinha, Wedderkopp, Andersen and Ekelund, 2008).

So far the potential for this instrument to assess physical activity seems to be well established. There are some excellent reviews available that address this matte r (Trost, Ward et al., 1998; Welk, 2002; Trost, McIver and Pate, 2005). Besides physical activity measures, some studies have tested the ability of activity monitors to measure energy expenditure but there seems that there still are some limitations when it comes to that purpose (Lamonte and Ainsworth, 2001; Leenders, Sherman, Nagaraja and Kien, 2001; Puyau, Adolph, Vohra, Zakeri and Butte, 2004; Corder, Brage, Wareham and Ekelund, 2005). There are several studies available that have been done to test the validity and reliability of these instruments in different settings (e.g. laboratory and field based research).

The validity of accelerometers in controlled settings was tested in a study with thirty Welsh children with mean age 9.2 years old. Eston et al (1998) compared the accuracy between heart rate monitoring, pedometry, triaxial accelerometer and uniaxial accelerometry for estimating oxygen consumption during typical children's activities. The authors found that all measures correlated significantly with oxygen consumption measured by on-line gas analysis every 30 seconds using Biokinetics, Bangor, UK instrument. A multiple regression equation included heart rate and CSA counts from triaxial accelerometer as better predictors than other measure alone (r^2=0.85). The best single measure was the triaxial accelerometer (r^2=0.83) (Eston, Rowlands et al., 1998).

Activities performed included: two walking speeds in a treadmill at 4 and 6 km/h and to running speeds, at 8 and 10 km/h. Also three non-regulated play activities were performed: playing catch, hopscotch and sitting and crayoning. The uniaxial accelerometer together with heart rate measures, assessed hopscotch more accurately than any other measure. The accuracy of each activity measure depended on the type of activity monitor used but even so, the triaxial accelerometer provided the best measure in the vertical direction when compared with CSA (uniaxial accelerometer). This study showed that the vertical direction is the most important when quantifying movement in children. But, still, a three dimensional measure for predicting energy expenditure was superior to CSA (only one dimension) (Eston, Rowlands et al., 1998). Further, a study with thirty children aged 10-14 years old, indicated that CSA was sensitive to changes in speed during 3, 4 and 6 mph in a treadmill with no inclination. CSA activity monitor was validated against indirect calorimetry (oxygen consumption). Values for intraclass correlation between two accelerometers used across different speeds were 0.87 and both CSA counts were significantly correlated with energy expenditure calculated by indirect calorimetry (0.86 and 0.87) (Trost, Ward et al., 1998). A study from Puyau et al (2002) supported this finding, showing that the CSA monitor is a valid instrument to measure activity energy expenditure and also able to distinguish between sedentary (<800 counts), light (<3200 counts), moderate (<8200) and vigorous (≥8200) physical activity. Twenty-six children aged 6 to 16 years' old performed different structured activities while using the CSA and the Mini-Mitter Actiwatch (MM) monitors. Energy expenditure was measured by room respiratory calorimetry, as activity by the microwave detector and heart rate by telemetry (Puyau, Adolph, Vohra and Butte, 2002).

Some other studies have analyzed the validity of accelerometers in field settings. Welk and Corbin (1995), demonstrated a moderate correlation between Tritrac monitor and heart rate (r=0.58) and high correlation with Caltrac monitor (r=0.88). A sample of thirty 9-11 years old children were studied and correlations between activity monitors and heart rate were higher during periods of free playing activity (lunch/recess, after school), and lower when activity was more limited (class time) (Welk and Corbin, 1995). Another example in field settings conditions was a study from Ekelund et al (2001), that used the doubly labeled water method to study the predictive validity of the CSA activity monitor to estimate physical activity and total energy expenditure under free-living conditions (Ekelund, Sjostrom, Yngve, Poortvliet, Nilsson, Froberg, Wedderkopp and Westerterp, 2001). From the 60 children selected randomly from four different schools, 26 participated in the study. Total energy expenditure and physical activity were assessed during 15 days and Pearson correlations coefficients were determined to analyze the linear relationship between activity counts (counts/min) and total energy expenditure, activity energy expenditure (both adjusted and unadjusted for gender and body weight) and physical activity level. Relation between activity counts and subjects physical characteristics (body weight, fat free mass, fat mass and height) were also analyzed. Trost et al (1998) equation was used to estimate daytime energy expenditure from CSA counts. Association were all significant between activity counts and energy expenditure components (p-value<0.05). Relations between CSA counts and energy expenditure were stronger after adjusting for weight and gender (p-value<0.01). There were no significant associations between subjects physical characteristics and CSA activity counts. Multivariate analysis indicated that gender, activity counts, and body weight better explained variation in total energy expenditure (60% of the variance). In addition, activity counts were the only predictor of activity energy expenditure (r^2=0.16). Correlation between these two variables was significant at 0.01 levels (r=0.54). The authors concluded that the equation develop from Trost et al (1998) underestimated total energy expenditure during free-living conditions and that this was explained by the fact that the equation used was developed under laboratory settings using a walking and running protocol. Nevertheless, the CSA was able to provide valid information in regard of the total amount of physical activity in 9 year old children (Ekelund, Sjostrom et al., 2001).

There are different type of accelerometers that differ in memory capacity (smaller or longer epochs used will also decide what the memory capacity available), software capabilities, type, size and sensitivity of the sensor, price (from $200 to $500) (Welk, 2002). Even if they differ in they characteristics and importantly in the way they process and report the accelerometry signals, several studies have shown that they can provide similar information (Welk and Corbin, 1995; Welk, Blair et al., 2000; Welk, 2002).

The BioTrainer Pro Activity monitor (Biotrainer, IM systems, Baltimore, MD), is a bidirectional accelerometer (positioned at 45º to vertical, sagital plane) that has demonstrated high correlations with output from other accelerometers (Welk, Blair et al., 2000; Welk, 2002). This instrument has been showed to give valid and reliable indicators of energy expenditure in adults (Welk, Almeida et al., 2003) and it has also been previously calibrated for use with children using direct observation techniques (Welk, Eisenmann et al., 2007). In this study, 30 children aged 8-12 years old were involved in two minutes activities as for example, sit, stand and dribble ball intermittently, walk continuously and jogging and also free-playing activity was also assessed. Using the

Biotrainer accelerometer, heart rate monitor and also a direct observation (videotaped) instrument (Behavioral Evaluation System and Taxonomy. BEST), cut points were generated using both a mixed-regression and the ROC curves approach. The BEST was used as the criterion method and showed that as energy intensity increased, both the Biotrainer and the heart rate values also increased. When comparing the cut points from different methods used, a cut point of 4 counts to differentiate between active and inactive behavior seemed to reflect the more reasonable specificity (93.2%) and sensitivity (60.9%) and also kappa agreement (0.58). this threshold does not differentiate between moderate and vigorous activity but provides a way to estimate levels of overall activity with reasonable accuracy. (Welk, Eisenmann et al., 2007).

Criterion standards: Direct Observation Methods

Direct observation methods are considered to be an objective (gold standard) technique because behavior is directly observed. Very importantly this method can be of major value when understanding behavior because environmental factors can also be assessed (McKenzie, 2002). Direct observation methods are considered to be a practical and appropriate criterion measure of physical activity even if they can require a high experiment burden and promote a potential reactivity of the study participant (Sirard and Pate, 2001). This method concerns with classifying free-living physical activity behaviors into different categories that can be quantified and analyzed. Knowing that physical activity is a contextual phenomenon, the setting where it takes place is of extreme relevance. When characterizing physical activity levels and settings of children this instrument becomes a possible first choice as it doesn't require any direct response from the child being assessed (McKenzie, 2002).

When using direct observation instruments there are some details that one should pay attention to in order to obtain accurate measures. The number of categories that serves to identify movement behavior can vary, but commonly, instruments use five or more independent categories or codes (e.g. sedentary, very active etc). Nevertheless some instruments code physical activity using different levels that are associated with one category (e.g. posture: lying down, standing, walking and very active, from the BEACHES instrument). The decision regarding the number of codes one instrument should use is fairly logical. The more codes one instrument has, the harder it will be to the observer to code physical activity, possibly promoting higher levels of fatigue decreasing reliability of the measures. Even if more codes would be associated with increases in precision of the measure, it is important to notice that the precision gained may not be necessary, depending on the research question being answered (McKenzie, 2002).

Further, observational periods are usually divided into short observe record moments and intervals are equally distributed between observing and recording time periods. The sampling method used will determine what subjects are to watch, when to watch and how to record their behavior. There are different sampling techniques that are used by different instruments: momentary time sampling (instantaneous or scan sampling), partial time sampling (i.e., recording the event if it occurs at any time during the observe interval) and whole interval sampling (the event is recorded only if it occurs through the whole interval). Sampling periods are usually well defined, either using stop watches or audiotape players with pre-recorded signals to initiate and end recording periods (McKenzie, 2002). One important issue and limitation with this type of instruments is the observer training, and if self-report measures are said to be dependent of the child being accessed, direct observation techniques strongly rely on the accuracy and skills of the observer to identify physical activity behaviors. Training the observers can guarantee an adequate data collection, but still, enough training should be provided. Usually, this training process is based on videotaped samples that assure certification and reassessment of observer skills. Studies that require long term follow-up periods should have a well defined protocol ensuring observer training as these keep changing over time. There are several different instruments that use observation to generate physical activity data from population (BEACHES: Behaviors of Eating and Physical Activity for Children’s Health; CARS: Children’s Activity Rating Scale; SOFIT: System for Observing Fitness Instructor Time; SOPLAY: System for Observing Play and Leisure Activity in Youth) (McKenzie, 2002).

The System for Observing Play and Leisure Activity (SOPLAY) is an example of an instrument that is based on direct observation to determine physical activity levels of a specific group of interest. This instrument uses momentary group time sampling (Placheck recording) and is suitable for all areas where physical activity may occur, before, during and after-school time periods. As opposed to the five codes mentioned above, SOPLAY uses only three categories to code physical activity (sedentary, walking and very active). This instrument is designed to indicate the percentage of people attending or involved in different activity categories (McKenzie, 2006). Notice that so far this instrument is unique as it gives a group indicator of physical activity, in an attempt to characterize a class or a subgroup level of physical activity during a specific part of the day (McKenzie, 2002). The recording process differentiates between boys and girls and also requires entries for the time of the day, observation starting times, temperature, area accessibility, area usability, presence of supervision, presence and classification of organized activity, equipment availability and also activity being performed (according to a existent SOPLAY checklist code for different activities). This description of the setting is extremely important as some settings can be more associated with higher levels of physical activity and vice-versa. Energy expenditure rates can also be calculated using already validated conversion rates for each category considered (McKenzie, 2006). The number of children counted as sedentary, walking and very active is multiplied by 0.051, 0.096 and 0.144 kcal/kg/min respectively (McKenzie, Marshall et al., 2000).

Stability reliability of SOPLAY was estimated in a study from McKenzie et al (2000) indicating a minimum of four days needed to achieve a reliability greater than 0.80 (McKenzie, Marshall et al., 2000). In addition, there seems to exist some evidence that supports the validity of energy expenditure coefficient codes used by SOPLAY (McKenzie, Sallis et al., 1991). In this study, 24 middle-schools in California were recruited to primarily investigate children physical activity levels and respective environment characteristics. Analyses were made for before school, lunch time and after-school periods. The second purpose of this study was to evaluate SOPLAY properties when measuring physical activity. After the selection of schools to participate in the study, all potential areas for leisure-time physical activity were identified for data collection purposes. There were 151 total target areas that included indoor and outdoor spaces or facilities. Physical activity levels were determined using the SOPLAY frequency counts of children in different behaviors and then converted to energy expenditure using the coefficients already described. MVPA was generated by adding the walking and very active categories. Observations were randomly conducted on three days at each school with a total duration of 72 days. Has relevant to the topic being discussed here, the after-school observations were made every 15, 45 and 75 minutes after classes were finish. Further, SOPLAY training was given to 4 observers that did all the observations, and had the duration of three consecutive days. Observers also had a retraining and assessment session at the 9th week of the study (McKenzie, Marshall et al., 2000).

Reliability testing was conducted in 14 of the 72 measurement days by rotating pairs of observers so different pairs could be compared. T-tests statistic indicated no differences between observers except for counts of very active boys. Interobserver intraclass correlation was 0.98, 0.95 and 0.76 to sedentary, walking and very active, respectively. Energy expenditure as expected followed the same pattern, were also significantly different in boys for very active category and intraclass correlations were higher (r=0.99). The five contextual variables had agreements of 95% for area accessibility, 97% for usability, 93% for presence of supervision, 96% for presence of organized activity and 88% for provision of equipment. Three days of consecutive measures using energy expenditure rates calculated by SOPLAY indicated an intraclass correlation of 0.88 and 0.75 for boys and girls, respectively. Concurrent validity was assessed with a self-report physical activity survey distributed to each school during the same semester. Those surveys were then randomly distributed to three classes in each school (1 class of each grade). A total of 1678 surveys returned and analyzed indicated a

Pearson correlation of observed and self-report MVPA during the after-school period of 0.35. Correlations for before school and lunch time were higher, 0.71 and 0.73 respectively. The authors stated that the low correlation coefficient verified for the after-school portion of the day was most likely due to measurement error in self-report measure than error in SOPLAY observations. Another factor that may have influenced the correlation between the two instruments was the reduced number of students participating in the after-school portion of the day (McKenzie, Marshall et al., 2000). Nevertheless, the authors stated that codes of the SOPLAY demonstrated to be valid. Mckenzie et al (1991) developed an integrated system for coding direct observations of children's eating and physical activity (Behaviors of Eating and Activity for Children's Health Evaluation System – BEACHES). The sample consisted of 42 children aged 4-8 years old and they're respective families. Observations were conducted in home (90 consecutive minutes) and school environment (lunch time during 20 minutes and recess during 30 minutes) for one day a week during 8 consecutive weeks.

The BEACHES is based on recorded image and is a tool designed to assess simultaneously eating and activity patterns in children as well as the setting characteristic. This instrument is composed by 10 different dimensions: Environment (alone, mother, father, siblings, peers, teacher, other adults, food available and views TV), Physical location (inside home, outside home, outside general, playground/play space, inside school, cafeteria, outside school, school play space), Activity level (lying down, sitting, standing, walking and very active), Eating behavior (ingests no food, ingests food), Interactor (alone, mother, father, siblings, peers, teacher, other adults), Antecedents (none during interval, prompts to increase, prompts to decrease, provides imitative model, child request) and Prompted event (not applicable, high-intensity activity, low-intensity activity and food). The target child behavior was coded if it had occurred during a 25 seconds interval. Each cycle of observation would last 1 minute and the observer would focus on a target child for 25 seconds and then spend 35 seconds to code the observed behavior. Interobserver reliability was tested during both home observations and during videotaped tests. Results from 19 home observations and 24 video observations (8 observers coded the same videotape on three different occasions at least 6 weeks apart; each observer score was compared with a criterion considered to be the assessment made by the investigators) indicated an average agreement of 94% and 93% respectively (with kappa range of 0.69 to 1 for the Home observations and 0.47 to 1 for the Videotaped observations) (McKenzie, Sallis et al., 1991).

This study also tested the validity of the coding system to estimate the energy expenditure associated with each activity. Nineteen children aged 4 through 9 were used to compare the heart rate (sampled each 30 seconds) during specific activities considered to be representative of the different activity codes. Energy expenditure rates were estimated from the heart rates using normal values for young children previously validated. Energy expenditure at rest was considered to be 6 ml/kg/min and each 10 beat-per-minute was estimated to represent an increase of 4.4 ml/kg/min in energy expenditure. Mean heart rates ranged from 99 to 153 beats per minute (lying down and very active category respectively). Mean heart rate for lying (lying watching TV) was 99 bpm which corresponded to a estimated 0.0029 kcal/kg/min, 107 bpm for the sitting activity (sitting watching TV, kneeling, easy swinging) which corresponded to a estimated energy cost of 0.047 kcal/kg/min, for the standing activity (standing and talking) was 110 bpm with a estimated energy cost of 0.051 kcal/kg/min, for walking (slow and easy walking, vigorous walking) was 130 bpm with a estimated energy cost of 0.096 kcal/kg/min, and for very active (cycling, running, hard swinging, sliding) was 153 bpm with a estimated energy cost of 0.144 kcal/kg/min (McKenzie, Sallis et al., 1991).

The SOFIT (System for Observing Fitness Instruction Time) is another observation tool that uses momentary sampling and an interval scanning recording system. This instrument uses pre-selected students and observes one at the time determining their level of physical activity. Each observation has the duration of 20 seconds and the coding is based on the observed activity of the target student at the moment the observation ends (McKenzie, Sallis et al., 1991). A validation study was conducted to test the SOFIT energy expenditure coefficients codes (5 categories), the exact same scale as the BEACHES. Further, the sample was composed by 173 volunteer students from 1-8th grade and Rowe et al (1997) used heart rate to concurrently validate the SOFIT scale. A 36 minutes protocol was used during a physical education session similar to the scale used to test BEACHES scale but in addition curl-ups were added to the protocol. As relevant information from this study, the fact that category one and two (lying down and sitting) did not differ significantly in terms of heart rate, and by doing so the authors concluded that these two could be combined in only one category (Rowe, 1997). Another studies have also tested the validity of the SOFIT with heart rate and accelerometers (McKenzie, Sallis et al., 1991; Scruggs, Beveridge and Clocksin, 2005). The SOPLAY uses this adapted three category scale derived from the BEACHES and SOFIT validation studies that use a five category scale.

Future Research: Issues and concerns

Some studies have used SOPLAY to measure PA and have claimed that validity was already established. However, the only study directly assessing the validity of SOPLAY used a self-report measure and this may not be precise enough to evaluate the SOPLAY tool. As the authors mentioned, the great margin of error associated with this type of instruments may have influenced the results, leading to an irrelevant comparison. In addition, individual measures of students using the self-report survey were not identified, and by doing so it becomes hard to match the SOPLAY data with the self-report data. One could argue that the students who filled the self-report survey were not the same being observed during the after-school period. In addition, no information was provided about the survey (self-report measure) used. There are several surveys questionnaires already validated to measure physical activity, but none of those seemed to be the one used in this study, which supports the belief that error from that instrument may have been considerable high.

So far SOPLAY validity hasn't been actually tested in its original format. Manuscripts using the SOPLAY have referred to the validity of the SOFIT and BEACHES instruments as supporting the validity of the SOPLAY tool but these may not be appropriate assumptions. The method used for observation differ considerably between the BEACHES and SOPLAY instruments, so the validity of the BEACHES physical activity codes may not translate to the SOPLAY. The BEACHES protocol for observation uses a partial time sampling technique (i.e., recording the event if it occurs at any time during the observe interval) but the SOPLAY uses a momentary time sampling (instantaneous or scan sampling). One can argue that these two different techniques can provide different scores when observing the same session. In addition, it is important to note that the protocols for observation differ for these two instruments. The BEACHES is based on an individual assessment behavior that is then assumed to be representative of the whole group. The SOPLAY instrument, in contrast, is designed to assess a whole group and registers the number of children enrolled in each category or behavior at a given time. Also, the BEACHES uses 5 different activity codes but only 3 codes are used in SOPLAY.

There is a need to conduct a formal evaluation of the SOPLAY instrument using objective criteria. Because the SOPLAY is designed to differentiate between inactive and active children it is important to evaluate it using a similar outcome measure. A previous study from McKenzie, Marshall et al., (2000) used estimates of energy expenditure from heart rate monitors to evaluate the energy expenditure of groups observed with SOPLAY. Energy expenditure from activity can be influenced by many factors including individual variability in age and body size. Efforts to calculate energy expenditures based on a group observation ignore this factor so this approach would likely lead to spurious results.

The SOPLAY instrument has utility for field based research in youth but more appropriate validation studies are needed. To date, studies using objective measures of physical activity have not been conducted to test the validity of this tool for assessing physical activity.

The purpose of this study is to test the concurrently validity of the SOPLAY with accelerometers in a sample of 9-11 years old children during an after-school physical activity program. Our research team is assisting a research team from Blank Children's Hospital in evaluating an after school physical activity program for children in West Des Moines. The SOPLAY tool is being used to provide information about physical activity in this time period. The proposed study would be conducted in conjunction with this study. Children in the after school period would be provided with an activity monitor during the after school period and data from the activity monitor would be directly related to the observation codes (by time) to evaluate the validity and reliability of the observation tool for capturing the overall activity level of the population. Interobserver reliability will also be tested in the study by comparing observation codes from two observers during the same periods.

CHAPTER III – METHODS AND PROCEDURES

Design of the Study

This study was conducted as part of an ongoing intervention project conducted by Blank Children's Hospital in Des Moines, Iowa. The intervention was designed to improve the quality of existing after school programming for promotion of physical activity and obesity prevention. Two after-school programs participated in the project, the PACES program in Perry, Iowa, and the KIDS WEST program in West Des Moines, Iowa. The programs both utilized the established CATCH after-school program and staff received training by certified CATCH professionals to ensure appropriate implementation of the program. The SOPLAY direct observation tool was used to evaluate the activity levels of the groups during the after school programming on two different occasions during the school year. Individual participants in the after school program were also monitored using the Biotrainer activity monitor to provide an additional indicator of physical activity. The present study complemented the existing project by evaluating the agreement between the group level activity assessment from SOPLAY and aggregated (group level) activity data from the Biotrainer activity monitor.

Participants

The study population included 160 children aged 9-12 years old (4th to 6th grade) from 9 different elementary schools from Des Moines, Iowa (79 males and 81 females). Before the start of the study approval from the Institutional Review Board of Iowa Health was obtained. Participants were informed about the purpose and procedures of the study and a parent sign informed consent was requested in order to participants be involved in the study.

Physical Activity Measures

SOPLAY

The System for Observing Play and Leisure Activity (SOPLAY) is designed to assess group levels of physical activity in different settings. This instrument uses momentary group time sampling and is suitable for assessing behavior in all areas where physical activity may occur. The SOPLAY uses three categories to code physical activity (sedentary, walking and very active) and separate records are created for boys and girls to facilitate group comparisons. The categories of walking and very active together are generally used to reflect participation in moderate-to-vigorous physical activity (MVPA). Past research tested the interobserver reliability of SOPLAY indicating no significant differences between observers except for counts of very active boys. Interobserver intraclass correlations coefficients were 0.98, 0.95 and 0.76 to sedentary, walking and very active, respectively. Concurrent validity assessed with a self-report physical activity survey indicated a Pearson correlation of observed and self-report MVPA during the after-school period of 0.35. Correlations for before school and lunch time were higher, 0.71 and 0.73 respectively (McKenzie, Marshall et al., 2000).

Biotrainer Pro

The BioTrainer Pro Activity monitor (Biotrainer, IM systems, Baltimore, MD), is a bidirectional accelerometer (positioned at 45º to vertical in sagital plane) that has demonstrated high correlations with output from other accelerometers (Welk, Blair et al., 2000; Welk, 2002). This instrument has been showed to provide valid and reliable indicators of energy expenditure in adults (Welk, Almeida et al., 2003) and it has also been previously calibrated for use with children using direct observation techniques (Welk, Eisenmann et al., 2007). A value of 3 counts per minute was found to provide the best cutpoint to differentiate between light activity and MVPA. This value was shown to have good sensitivity (Se = .78) and specificity (Sp = .78) for detecting MVPA when tested on an independent (cross validation) sample. To take into account the intermittent nature of children's activity the Biotrainer was initialized to collect physical activity data using 15 seconds epochs at a 40x resolution.

Study Procedures

Training of Observers

Three observers were trained to use the SOPLAY by our staff during the four weeks that preceded the beginning of the study. The staff member responsible for the training was previously trained and certified by the original developers of SOPLAY. The observers were trained using the training DVD provided by the original test developer and also using live physical activity sessions. Reliability was assessed using a videotaped physical activity session from a group of 12 children aged 8-12 years old. The videotaped session lasted 16 minutes and included similarly aged children engaging in structured physical activity. SOPLAY scans were performed every two minutes during the observation period. Test-retest reliability (three days apart) ranged from 0.70 to 0.86. Validity was assessed by evaluating agreement in coding the final assessment included in the training DVD (criterion measured provided by the original developer). This final assessment included 28 scans but 15 scans with varying difficulty and varying participants (3 to 12 individuals per scan) were used for the comparisons. Pearson correlations ranged from 0.80 to 0.92.

One of the three observers was selected as the criterion observer and collected the SOPLAY data from the beginning to the end of the study. Pre and post-test reliability (6 weeks apart) was tested for this observer using the same videotape to make sure consistency of coding was kept until the end of data collection (ICC= 0.74).

Data Collection

A total of nine schools were involved in the project and data were obtained from each school on two occasions (fall of 2008 and spring of 2009). The focus of the project was to evaluate active opportunities during the after school program but three additional observation assessments were conducted to observe less-active settings at several schools. Thus, data were obtained during 21 distinct scanning periods. Data collection during the Fall and Spring terms was conducted within a 5 week period. Reliability of SOPLAY was assessed during both fall and spring by having multiple observers code the same activity sessions. The research team arrived at every designated school 30 minutes before the physical activity session started. Children were again reminded about the study purpose and procedures and told they would be wearing an accelerometer and that one or two observers (depending if reliability was being tested) would be watching them during the session.

Colored jerseys were provided to participants to help facilitate the observation being done by the researchers. There were six different packets of colored jerseys (green, blue, purple, yellow, red and orange) each identified with numbers ranging from 1 to 12 in order to provide unique identification of the participants during the session (total of 72 jerseys). Each participant was given a colored jersey and respective accelerometer also identified with a matching colour and numbered label. This method allowed for individual tracking to avoid incorrect accelerometer and SOPLAY matching. If a participant had to leave earlier or leave the target zone for other reason, the time of the event would be recorded for future analysis. Boys were always identified with green, purple or blue jerseys and girls with yellow, red and orange jerseys.

Scans were performed at varying intervals (2, 3 or 5 minutes) during the fall semester but the process was standardized in the spring with scans being conducted every 2 minutes. Observer’s clocks were synchronized at the same second so every scan was done at the exact same time, starting from the left to the right of the observers, scanning girls first and boys second. In the schools where reliability was tested, observers stood approximately two meters apart to make sure no information was shared between them during the entire session. The staffs from the schools responsible for leading the sessions were told to maintain the usual structure and activities used during typical activity sessions in the after-school program.

During the session, the principal investigator was responsible for assuring that the accelerometers were fixed to the children’s waist in the correct position. The investigator recorded the start and stop time for each session and monitored the number of youth participating in the activity monitoring protocol. Once the session had finished, children were asked to give the colored jerseys and accelerometers to the research team members.

The activity sessions were conducted in the school gymnasium. In the Spring data collection period, observations were also conducted during three less active time periods (snack time). During this period the participants were often seated while eating or reading books. The less active period preceded the activity session with a brief period of transition. This transition period was not recorded.

Data Processing and Analysis

The analyses examined the agreement between the observed activity levels from the SOPLAY codes and the recorded activity from the Biotrainer activity monitor. The data were analyzed in two ways to examine measurement agreement at different levels. In Aim 1, data were aggregated by time to allow the validity of individual SOPLAY codes to be examined. In Aim 2, data were aggregated by school to examine the validity of the overall SOPLAY assessment for capturing activity levels in a group setting. The data processing and analyses for each aim are described below:

Analyses for Aim 1

The first analysis compared the observed versus recorded levels of physical activity on a minute by scan basis. This analysis was conducted using all available scans from the fall and spring data collection periods. The coding from the SOPLAY is designed to produce estimates of the number (percentage) of youth that are sedentary, somewhat active (walking) or very active. Guidelines are not available to interpret SOPLAY data so several different comparisons were made to determine the most appropriate way to express this data. Three specific comparisons were used reflecting different ways of expressing outcome data from the SOPLAY and the Biotrainer. The first method (MVPA1) compared estimates of MVPA from the SOPLAY (operationalized as the percentage of youth in the “walking” and “very active” categories) with estimates of MVPA from the Biotrainer (percentage in MVPA). The second measure (MVPA2) compared estimates of vigorous activity from the SOPLAY (operationalized as the percentage of youth in the “very active” category) with the same measure from the Biotrainer (percent in MVPA). The third measure (MVPA3) compared the estimate of MVPA from the SOPLAY (operationalized as the percentage of youth in the “walking” and “very active” categories) with estimates of Light Activity and MVPA from the Biotrainer (percent in light activity and MVPA).

The three comparisons reflect different ways of combining and interpreting the data from the two instruments. The comparisons between MVPA1 and MVPA2 use the same outcome measure of MVPA from the Biotrainer so differences between these two indicators will reveal which interpretation of data from the SOPLAY best corresponds with the an established, objective estimate of MVPA. The comparison between MVPA1 and MVPA3 use the same outcome measure from the SOPLAY (walking + very active) but use two different ways of expressing data from the Biotrainer. The analyses here will evaluate whether walking as defined by the SOPLAY corresponds with Light activity as assessed from the Biotrainer. Walking is generally used to define the threshold for MVPA so the three comparisons evaluated here will provide information about whether coded observations of walking really count as MVPA in youth.

To provide appropriate comparisons between the SOPLAY and Biotrainer data it was necessary to categorize activity assessed from the Biotrainer in a similar way. A previous study (Welk et al, 2007) established the cutpoint of 3 counts to represent an appropriate criteria or threshold to define MVPA. The threshold for light activity was set at the level of 1 count based on an examination of the data from the original validation study. Thus, the individual minute by minute Biotrainer data were coded into three categories that correspond (conceptually) to the three designations coded on the SOPLAY: Sedentary (counts < 1), Light (Counts between 1 and 3) and MVPA (≥ 3 counts/minute). The total number of youth active for a given minute was tracked to provide an indicator of the overall activity level for the group during that minute. The percentage of active youth was then computed for each minute by dividing the number of active youth by the total number of youth monitored for that minute. The processed Biotrainer data were merged with the corresponding SOPLAY data for each of the available minutes of scanning. Because each scan typically took about 15 seconds, the SOPLAY data were matched with the respective 15 second epoch counts from the accelerometers. Data were analyzed independently for each school first and then merged so all schools data could be analyzed together. Analyses used the total number of matched minutes of Biotrainer and SOPLAY data across all sites in order to maximize the available sample. This processing provides an appropriate comparison for the SOPLAY data which is designed to track the number of youth in the group that were active in a given minute. Because data were temporally linked by time it was possible to match the observed and recorded data regardless of what rates the scans were performed.

Standard measurement agreement analyses were conducted to examine the agreement between the observed and recorded data. The percent of youth categorized as active were used as the primary outcome of interest. Differences in observed and recorded activity rates were compared using standard t-tests. Correlation analyses (Pearson Product Moment) were computed to reflect overall associations. Bland-Altman plots were also used to examine agreement across the range of observed activity levels. Analyses were conducted separately for males and females as well as combined since the SOPLAY is designed to record activity in this manner.

Analyses for Aim 2

The second set analysis examined the validity of the SOPLAY tool for assessing group levels of physical activity during a typical after-school setting. The SOPLAY is designed to capture random snapshots of activity levels and these snapshots are presumed to reflect activity behavior of the group during the whole period of time. An unresolved question is how frequent scans have to be to accurately capture activity levels.

The influence of sampling rate was analyzed by evaluating site-level agreement in reported activity levels when observation data are aggregated at different intervals. These data were restricted to data obtained in the spring which were collected at 2 minute intervals. This restriction was needed to facilitate the aggregation of data into standardized intervals. For aim2, Biotrainer data were obtained every minute (average across 1 minute) but observation data were obtained every two minutes so the goal was to determine how many minutes of observation data are needed to accurately characterize the activity levels of a group activity session (as assessed by the recorded Biotrainer data). For these analyses, the serial SOPLAY data were aggregated by site to create an overall indicator of the activity level at that site. Several different aggregations were computed to determine the impact of scanning rate on the validity of the group (setting) estimation. To accomplish this, average activity levels (computed as % of youth in MVPA) were computed using data aggregated every 2, 4, 6, 8, 10, 12 and 20 minutes. For example, the 4 minute average was computed using data from every other 2 minute scan while the 6 minute average was computed using data from every third 2 minute scan. Therefore, there were different numbers of scans used to create the average activity levels for each school/site. Most sessions were 30 minutes so the 2 minute scan estimate was based on 15 values while the 4, 6, 8, 10, 12 and 20 minute estimates were based on 7, 5, 3, 3, 2, and 1 scans, respectively. The various SOPLAY-derived estimates were coded using distinct names and merged with the corresponding average Biotrainer values for the site. This merged data set made it possible to directly compare the estimates from the various set of SOPLAY-derived estimates with the objective data from the Biotrainer. Higher correspondence (less error) is expected for SOPLAY estimates derived from more frequent scans (2,4 minutes) compared to less frequent scan rates (12 minutes, 20 minute).

Similar measurement agreement analyses were conducted to examine the validity of the group level SOPLAY coding. The operational definitions for the MVPA2 comparison in Aim 1 were used to evaluating activity levels since this classification provided the best indicator of involvement in MVPA. Standardized t-tests were used to evaluate school level differences in observed and reported activity rates and correlations were used to reflect overall associations. To further evaluate differences due to scanning rates, absolute difference scores were computed for each school by computing the difference in the observed values from the recorded values. The average absolute difference score was computed for each scan rate to determine the impact of scan rate on validity.

Because the results are predicated on the accuracy of the observations, additional analyses evaluated the reliability of the observers collecting the SOPLAY data. Interobserver reliability was evaluated using intraclass correlation coefficients (ICC) across all matched minutes of observation.

For all analyses the level of significance was established at 0.01 to avoid the probability of type I error. All analyses were conducted using the Statistical Package for Social Sciences (SPSS) and SAS v9.1.

CHAPTER IV – RESULTS

Descriptive Statistics

Two visits were done to each school, and three of those nine schools that were visited during the fall had two sessions each on the same visit. Overall, there was total of 21 sessions observed. Preliminary examination of the data revealed some inconsistencies in the way the data were recorded or coded at two schools in the spring (problems with timing issues and incomplete activity codes). To provide a more appropriate evaluation of the after school setting, the data from these two schools were excluded from the analyses. The final sample included data from 19 sessions with a total of 160 children observed and monitored (79 males and 81 females) that were observed and concurrently wore accelerometers. There were no significant differences (p-value>0.05) between males and females for age (mean value= 10.42±0.8 years), height (mean value= 143.5±9.9 cm), weight (mean value= 38.3±10.3 kg) and BMI (mean value= 19.4±4.5 (table 1). A total of 29 individuals (14 males and 15 females) had incomplete information on height, weight and BMI but were included in the study since these are less important for the goals of the study.

Of the 160 children involved in the study, a total of 155 were assessed during the fall but, due to absences, data were obtained from only 73 participants during the spring (table 2). Data from 73 youth were obtained on both occasions but the lower attendance is not a limitation for the present study since the focus was on evaluating the overall activity level of the environment. A total of 416 activity scans were conducted during the fall and spring terms at the 19 schools (208 scans each for males and females). Some additional scans (n = 98) were conducted during less active periods in the program (49 each for males and females). To provide an indicator of overall activity rates, activity levels were also computed for a combined male and female sample. Thus, there were a total number of 771 scans analyzed for the project (624 active scans and 147 less active scans). The total number of participants per session ranged from 1 to 21 (table 3).

Analysis 1: Agreement between SOPLAY and Biotrainer

The focus of Aim 1 was to examine agreement between individual SOPLAY codes and temporally matched outcomes from the Biotrainer. Analyses were conducted using different SOPLAY indicators in comparison with different Biotrainer indicators in order to determine the best match (MVPA1 = proportion of individuals engaged in “walking” or “very active” behaviors; MVPA2 = proportion of individuals engaged in “very active” behaviors; and MVPA3 = proportion of individuals engaged in SOPLAY “walking” or “very active” and Biotrainer light or moderate-to-vigorous activity). The direct comparisons of these outcome measures provide useful information about the best way to interpret SOPLAY data.

Descriptive statistics for the observed levels of MVPA and the recorded Biotrainer data are provided in Table 4. Physical activity levels assessed by the Biotrainer and averaged across the total number of scans performed tended to be higher for boys than girls (p-value<0.01). This indicates that, on average, more boys will be active than girls even when exposed to the same environment. The data from the Biotrainer were compared against the MVPA indicators to determine the validity of the observed

SOPLAY scans. There were large and significant differences in rates of activity for comparisons with the MVPA1 indicator for both gender and also when combined (Difference$_{together}$= 50.55±26.41%; $_{males}$= 47.62±32.29%; $_{females}$= 54.38±33.23%). The difference score was significant indicating that the observed values were significantly different than recorded levels in all the three groups ($t_{together}$ = 27.60, p-value<0.001; t_{males} =21.32, p-value<0.001; $t_{females}$ =23.60; p-value<0.001). The difference scores were small for the alternative MVPA2 comparison in the three groups (Difference$_{together}$= 1.33±22.06%; $_{males}$= -1.77±29.92%; $_{females}$= 4.54±29.49%) with no significant differences between the three groups ($t_{together}$= 0.87, p-value<0.001; t_{males}= -0.85, p-value<0.001; $t_{females}$= 2.22; p-value<0.001). Small differences were also observed for the MVPA3 comparison (Difference$_{together}$= -2.02±29.00%; $_{males}$= -4.98±35.83%; $_{females}$= 1.2±37.73%) and non significant ($t_{together}$= -0.99, p-value=0.324; t_{males}= -2.01, p-value=0.046, $t_{females}$= 0.46, p-value=0.647). Supplementary analyses examined the consistency of these relationships across schools and the patterns were generally consistent across schools (data not shown).

Correlations between the observed and recorded levels of activity are provided in Table 5. Overall correlations were found to be positive and moderate, with a similar pattern in all the different comparisons of physical activity interpretations (MVPA1, MVPA2 and MVPA3). Associations were strongest for MVPA3 ($r_{together}$= 0.575, r_{males}= 0.498 and $r_{females}$= 0.471) and lower for MVPA1 ($r_{together}$= 0.404, r_{males}= 0.368 and $r_{females}$= 0.239). Pearson Correlations for MVPA2 were similar to MVPA3 ($r_{together}$= 0.562, r_{males}= 0.428 and $r_{females}$= 0.394). Correlations tended to be higher in the combined analysis (possibly because of the larger sample and more stable results).

Comparisons between genders revealed a tendency for correlations to be slightly higher in boys when compared with girls.

A scatter plot showing the association between these variables is provided in Figure 1 (for the combined boy and girl data). Individual data points were more linearly distributed in both MVPA2 and MVPA3. Matched data from MVPA1 was systematically above the graph diagonal reflecting overall SOPLAY overestimation of active individuals. The graphs also show that both MVPA2 and MVPA3 were identifying two distinct segments of the SOPLAY coding. The MVPA2 plots reflect lower activity levels since this coding only identified individuals in moderate-to-vigorous activity. In contrast, the MVPA3 plot shows higher activity levels for both instruments since the categorization includes light and moderate activity.

Bland-Altman plots showing the agreement across the range of activity levels are provided in Figure 2. The results for the original coding of SOPLAY (MVPA1) indicated a clear and systematic overestimation of activity across the whole range of activity (figure 2a). The results for both MVPA2 (figure 2b) and MVPA3 (figure 2c) revealed a tendency for non-systematic bias even if with a characteristic pattern. Dispersion of disagreement was lower in those two figures indicating better overall agreement. Nevertheless, all analyses show better agreement when percentage of active individuals was either low or very high. Disagreement was substantially higher when approximately half of the class was considered to be active. Despite this somewhat odd distribution, the systematic overestimation of SOPLAY observed in figure 1a for the MVP1 comparison is largely reduced for the alternative MVPA2 and MVPA3. Separate Bland-Altman plots for these comparisons are available in the Appendix (Panel A).

Supplemental analyses were conducted to evaluate other measurement issues with the SOPLAY. One analyses compared the validity for coding less active time periods. A total of 147 scans done during a inactive time period (snack) compared MVPA1 with MVPA2 and showed better agreement between the two instruments when classifying active individuals with MVPA2 (Difference $_{together}$ = -1.29±9.78%; Difference $_{males}$= -0.24±10.81% and Difference $_{females}$= -2.38±13.18%, t-test values ranged from -1.27 to -0.15 with p-value>0.01). Mean differences for MVPA1 were all significant (p-value<0.001) with t-test values ranging from 6.10 to 7.94 (Difference $_{together}$= 27.62±24.35%; Difference $_{males}$= 30.4±28.87% and Difference $_{females}$= 25.14±28.84%). Details are provided in Appendix (Panel B). Another set of analyses examined if the validity of the SOPLAY was dependent on the number of scans performed. The better classification of active participants analyzed in aim1 was used for this purpose (MVPA2). Data were combined from the activity and less active periods using only the absolute proportion of active participants ("together"). A total of 257 scans (from both activity and inactive time combined) were divided in three categories (category 1: scans with 4 to 7 participants; category 2: scans with 8 to 12 participants; category 3: scans with 13 to 21 participants).

Results in figure 3 indicate progressively stronger associations between the instruments from category 1 to category 3. Correlations were all positive and moderate to high (r category1= 0.494; r category2= 0.663 and r category3= 0.749). The higher correlations for category 2 and 3 indicate that the associations were stronger when more youth were being observed. The narrow line of data points in figure 3c as opposed to the wider interval in figure 3a also indicates better agreement (distribution of data points closer to a possible line of fit). More details are provided in Appendix (Panel C).

Analysis 2: Sampling rate influence on SOPLAY validity

The goal in the second analyses was to determine the impact of scanning rate on the validity of the group level SOPLAY estimates. These analyses were performed using MVPA3 since it was shown in analysis 1 to provide the better absolute agreement with the criterion selected.

Data used were from seven different sessions (7 schools) with a total of 252 scans and 73 participants observed (table 6). Results indicated that the mean absolute disagreement for schools when identifying active individuals was higher when data was aggregated using 20 minute scans (21.76±13.80%; equal to a total of 1 scan per session). Disagreement was lower for 6 and 2 minutes scans with absolute differences of 7.23±3.74% and 7.43±1.43% respectively. Absolute disagreement using other sampling rates (4, 8, 10 and 12 minute scans) ranged from 8.49±5.63% to 12.39±6.88%. Figure 5 shows the increasing error associated with the decrease in frequency of scans completed. Although there is some variability there is a progressive increase in overall error rates as scanning frequency is reduced. A figure illustrating school level variability is available in Appendix (Panel D).

Additional analyses were done to test the interobserver reliability. There was a total sub-sample of 87 scans collected in 7 different schools. Differences between observers was not significant (p-value=0.195) and the intraclass coefficient for the averaged measures was 0.874 (p-value<0.01).

CHAPTER V – DISCUSSION

This study evaluated the validity of the SOPLAY instrument compared with objectively measured physical activity data from an accelerometry-based activity monitor. The results generally support the validity of the SOPLAY for use in evaluating youth physical activity settings and provide valuable information about the appropriate interpretation of SOPLAY data.

Interpretation of Aim 1

The results indicated that the traditional interpretation of SOPLAY data may lead to overestimation of the actual percentage of youth that are active (as measured by the accelerometer). The likely reason for this is that youth observed to be walking (level 2 on the SOPLAY) are probably not active enough to be counted as achieving minutes of MVPA. While walking is often used to characterize moderate intensity activity, the intensity and duration must be sufficiently high to warrant being counted as “activity”. Brief and intermittent stepping may be coded as “walking” on the SOPLAY even though it may not constitute the sustained form of walking generally used to represent moderate intensity activity.. Previous research has indicated that walking can be characterized as both light-to-moderate or moderate-to-vigorous activity (Harrel, McMurray, Baggett, Pennel, Pearce and Bangdiwala, 2005).

Comparisons with alternative interpretation of the SOPLAY data show good agreement with the objectively monitored physical activity. There was good agreement for the MVPA2 comparison that used the percent classified as “very active” as the overall indicator of MVPA.Correlations were moderate and there were non-significant differences in the percentage of youth classified as being active. This provides good validity to support the validity of the SOPLAY instrument with this coding strategy. If more liberal definitions of activity are desired, then the sum of “walking” and “very active” may be an alternative indicator. We found good agreement with this MVPA3 comparison which essentially classified “walking” as light activity. There were moderate correlations and non-significant differences in percentages when compared with corresponding estimates of light and MVPA from the Biotrainer.

It is noteworthy that agreement was consistently worse among females. The alternative SOPLAY coding (MVPA3) provided higher increases in the correlations for females than males. This may suggest that the proportion of females engaged in walking behaviors were not enrolled in sufficient activity to be considered of moderate-to-vigorous activity by the Biotrainer. This finding can possibly be explained by the differences in activity levels between males and females (males were found to be more active than girls; p-value<0.01). Literature supports this finding, showing that 9-and 15 year old boys spend more time engaged in moderate physical activity than girls (Riddoch, Bo Andersen, Wedderkopp, Maarike, Klasson-HeggebØ, Sardinha, Cooper, Ekelund, 2004). Additional research is needed to determine if the subtle gender differences are real or due to artifact or random error.

Supplementary analyses examined the impact that the number of youth observed have on the validity of the SOPLAY assessment. Interestingly, validity increased for higher numbers of youth being scanned. This suggests that the number of participants per scan may have an impact in the accuracy of SOPLAY. It would seem logical for validity to be lower if observers had to observe more youth but it appears that this is not the case. It is possible that small errors or misclassifications in small groups have a larger overall impact on the estimates for the group. These findings suggest that misclassification of individuals enrolled in MVPA can have a greater impact when observing smaller groups and also that groups ranging from 13 to 21 participants can still be observed with good accuracy (See Appendix - Panel C, for more details).

Interobserver reliability in this study (ICC= 0.874) was lower than reported in the study by Mckenzie et al (2000) but still acceptable. The previous study used the number of subjects enrolled in each behavior category in contrast to this study where we used the absolute percentage of active children. This might explain some of the differences although the use of proportions was used in this study to minimize the impact of different sample size groups observed per session.

Interpretation of Aim 2

An additional goal of the study was to determine the impact that scanning intervals have on the validity of the SOPLAY.

The second analysis reflected the impact that different scanning rates or number of scans per session can have on the validity of SOPLAY. The results clearly showed that there is a substantial absolute error (21.76%) associated when there is only 1 scan per session (20 minutes scan). The data from the two minute scans was expected to have the least error compared to the Biotrainer data but this was not the case. For example, values for 2 minute scans were not that different when compared with values using 6 minutes scans (7.43% vs. 7.23% respectively). Further, scans every 2, 4, 6, 8, 12 and 20 had systematic increases in error possibly suggesting that there might exist an exponential pattern. To support this idea note that values using scan rates of 8 or 10 minutes were both based on three scans in a 30 minute period – yet there were differences in the reported estimates. The differences between these two sampling rates could be due to variability associated with the instant moment of observation. Nevertheless, the pattern was not consistent until the 12 minutes scans possibly suggesting some abnormalities or reduced sample size impact on the results (total number of schools= 7). Based on the results, a conservative interpretation is that a 30 minutes session requires at least 3 scans (10 minutes rate) to provide accurate estimates of the proportion of individuals enrolled in MVPA behaviors.

Conclusions

Physical activity is a complex behavior and this makes it difficult to assess. The Biotrainer activity monitor provided an objective indicator of physical activity (and served as a criterion measure in the study) but it also has limitations. Accelerometers provide measures of limb acceleration only and therefore can be misleading when upper body activities are performed (Welk, 2002). Further, the sporadic activity patterns of children can be difficult to accurately assess (Welk, 2002). Data analyzed by activity monitors are often a result of average measures of each minute. In addition, the transformation of continuous data to categorical data also has its own limitations. The choice of a cut point to identify different physical activity behaviors is based on some assumptions (continuous activity). By doing so, one can expect that levels of physical activity averaged for each epoch used are not exactly representative of the true behavior adopted. Averaging data across a specific period of time can often misclassify more vigorous activity, underestimating levels of physical activity (Welk, 2002). This study tried to minimize this source of error by using matched 15 seconds epochs with observed physical activity data.

Nevertheless, activity monitors are less prone to bias than self-report measures that were previously used in another study to test the validity of the SOPLAY (McKenzie, Marshall et al., 2000). Results from that study indicated a correlation of 0.35 for the after-school portion of the day which was possibly explained by the error associated with the self-report instrument used. Associations in this study were stronger in both alternative contrasts (MVPA2 r= 0.562; MVPA3 r= 0.575). It is likely that the objective data from the activity monitor were more effective at validating the observed behavior than a self-report instrument.

In conclusion, the findings from this study suggest that observations provide valid indicators of MVPA if coding is based on the percentage of youth classified as “very active”. The results demonstrate that more frequent scans can improve the validity of the estimations. The trends demonstrate that error rates increase as rates of scanning decrease but it is not possible to determine an optimal scanning rate for all research. In general, more frequent scans should be obtained if possible to improve the accuracy.

The SOPLAY was found to be a very useful instrument that can provide a description of the environment where active behaviors occur. Nevertheless, future research is needed to better understand the different sources of error associated with physical activity measures and explore the potential of direct observation tools. Thus this study provided a unique test of the validity of the SOPLAY instrument for assessing youth physical activity behavior. Strengths of the study include the use of temporally matched data from the Biotrainer and the processing that allowed directly comparable outcome measures.

REFERENCES

Adams, S. A., C. E. Matthews, et al. (2005). The effect of social desirability and social approval on self-reports of physical activity. Am J Epidemiol 161(4): 389-98.

Allison, K. R. and E. M. Adlaf (2000). Structured opportunities for student physical activity in Ontario elementary and secondary schools. Can J Public Health 91(5): 371-5.

Baranowski, T., C. Anderson, et al. (1998). Mediating variable framework in physical activity interventions. How are we doing? How might we do better? Am J Prev Med 15(4): 266-97.

Blair, S. N. (1984). Behavioral Health: A Handbook of Health Enhancement and Disease Prevention. How to assess exercise habits and physical fitness. S. M. W. a. J. A. H. J. D. Matarazzo. New York, Wiley: 424-447.

Caspersen, C. J., K. E. Powell, et al. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 100(2): 126-31.

Coleman, K. J., C. L. Tiller, et al. (2005). Prevention of the epidemic increase in child risk of overweight in low-income schools: the El Paso coordinated approach to child health. Arch Pediatr Adolesc Med 159(3): 217-24.

Corder, K., S. Brage, et al. (2005). Comparison of PAEE from combined and separate heart rate and movement models in children. Med Sci Sports Exerc 37(10): 1761-7.

Coordinated Approach to Child Health (2009), http://www.catchinfo.org/

Dale, D., C. B. Corbin, et al. (2000). Restricting opportunities to be active during school time: do children compensate by increasing physical activity levels after school? Res Q Exerc Sport 71(3): 240-8.

Eisenmann, J. C., R. T. Bartee, et al. (2008). Combined influence of physical activity and television viewing on the risk of overweight in US youth. Int J Obes (Lond) 32(4): 613-8.

Ekelund, U., S. Brage, et al. (2006). TV viewing and physical activity are independently associated with metabolic risk in children: the European Youth Heart Study. PLoS Med 3(12): e488.

Ekelund, U., M. Sjostrom, et al. (2001). Physical activity assessed by activity monitor and doubly labeled water in children. Med Sci Sports Exerc 33(2): 275-81.

Ernst, M. and R. Pangrazi (1999). Effects of a physical activity program on children´s activity levels and attraction to physical activity. Pediatric Exercise Science 11: 393-405.

Eston, R. G., A. V. Rowlands, et al. (1998). Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children's activities. J Appl Physiol 84(1): 362-71.

Fisher, A., J. Reilly, et al. (2005). Seasonality in Physical Activity and Sedentary Behavior in Young Children. Pediatric Exercise Science 17(1).

Freedson, P. S., E. Melanson, et al. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 30(5): 777-81.

Harrell, J. S., R. G. McMurray, et al. (2005). Energy costs of physical activities in children and adolescents. Med Sci Sports Exerc 37(2): 329-36.

Hoos, M. B., H. Kuipers, et al. (2004). Physical activity pattern of children assessed by triaxial accelerometry. Eur J Clin Nutr 58(10): 1425-8.

Koplan, J. P., C. T. Liverman, et al. (2005). Preventing childhood obesity: health in the balance: executive summary. J Am Diet Assoc 105(1): 131-8.

Lamonte, M. J. and B. E. Ainsworth (2001). Quantifying energy expenditure and physical activity in the context of dose response. Med Sci Sports Exerc 33(6 Suppl): S370-8; discussion S419-20.

Leenders, N. Y., W. M. Sherman, et al. (2001). Evaluation of methods to assess physical activity in free-living conditions. Med Sci Sports Exerc 33(7): 1233-40.

Luepker, R. V., C. L. Perry, et al. (1996). Outcomes of a field trial to improve children's dietary patterns and physical activity. The Child and Adolescent Trial for Cardiovascular Health. CATCH collaborative group. Jama 275(10): 768-76.

Mckenna, J., L. J. Foster, et al. (2004). Exploring Recall of Physical Activity in Young People Using Qualitative Interviewing. Pediatric Exercise Science 16: 5-14.

McKenzie, T. (2002). Use of Direct Observation to Assess Physical Activity. Physical Activity Assessments for Health-Related Research. G. Welk. Champaign,IL, Human Kinetics.

McKenzie, T. L. (2002). Use of direct observation to assess physical activity. Physical Activity Assessments for Health-Related Research. G. J. Welk. Champaign, IL, Human Kinetics: 179-195.

McKenzie, T. L. (2006). SOPLAY: System for Observing Play and Leisure Activity in Youth. Description and Procedures Manual. San Diego, CA: 1-7.

McKenzie, T. L., H. Feldman, et al. (1995). Children's activity levels and lesson context during third-grade physical education. Res Q Exerc Sport 66(3): 184-93.

McKenzie, T. L., S. J. Marshall, et al. (2000). Leisure-time physical activity in school environments: an observational study using SOPLAY. Prev Med 30(1): 70-7.

McKenzie, T. L., S. J. Marshall, et al. (2000). Student activity levels, lesson context, and teacher behavior during middle school physical education. Res Q Exerc Sport 71(3): 249-59.

McKenzie, T. L., P. R. Nader, et al. (1996). School physical education: effect of the Child and Adolescent Trial for Cardiovascular Health. Prev Med 25(4): 423-31.

McKenzie, T. L., J. F. Sallis, et al. (1991). BEACHES: an observational system for assessing children's eating and physical activity behaviors and associated events. J Appl Behav Anal 24(1): 141-51.

Nilsson, A., S. Brage, et al. (2008). Comparison of equations for predicting energy expenditure from accelerometer counts in children. Scand J Med Sci Sports.

Ogden, C. L., M. D. Carrol, et al. (2006). Prevalence of Overweight and Obesity in the United States, 1999–2004. JAMA 295(13): 1549-1555.

Ogden, C. L., M. D. Carroll, et al. (2008). High body mass index for age among US children and adolescents, 2003-2006. Jama 299(20): 2401-5.

Powers, H. S., T. L. Conway, et al. (2002). Participation in extracurricular physical activity programs at middle schools. Res Q Exerc Sport 73(2): 187-92.

Prevention, C. D. C. (1997). Guidelines for school and community programs to promote lifelong physical activity among young people. MMWR. 46: 1-32.

Puyau, M., A. Adolph, et al. (2004). Prediction of activity energy expenditure using accelerometers in children. Med Sci Sports Exerc 36(9): 1625-31.

Puyau, M. R., A. L. Adolph, et al. (2002). Validation and calibration of physical activity monitors in children. Obes Res 10(3): 150-7.

Ridgers, N. D., G. Stratton, et al. (2006). Physical activity levels of children during school playtime. Sports Med 36(4): 359-71.

Riddoch, Bo Andersen, et al. (2004). Physical Activity Levels and Patterns of of 9- and 15-yr Old European Children. Med Sci Sports Exerc 36(1): 86-92.

Rowe, P. J., Schuldheisz, J.M., Mars, H.V.D. (1997). Validation of SOFIT for Measuring Physical Activity of First- to Eighth-Grade Students. Pediatr Exerc Sci 9: 136-149.

Sallis, J. F., M. J. Buono, et al. (1993). Seven-day recall and other physical activity self-reports in children and adolescents. Med Sci Sports Exerc 25(1): 99-108.

Sallis, J. F., T. L. McKenzie, et al. (1997). The effects of a 2-year physical education program (SPARK) on physical activity and fitness in elementary school students. Sports, Play and Active Recreation for Kids. Am J Public Health 87(8): 1328-34.

Sallis, J. F. and B. E. Saelens (2000). Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport 71(2 Suppl): S1-14.

Scruggs, P., S. Beveridge, et al. (2005). Tri-Axial Accelerometry and Heart Rate Telemetry: Relation and Agreement with Behavioral Observation in Elementary Physical Education. Measurement in Physical Education and Exercise Science 9(4): 213-218.

Services, U. S. D. H. H. H. (2008). 2008 Physical Activity Guidelines for Americans. Washington: 15-20.

Sigmund, E., M. De Ste Croix, et al. (2007). Physical activity patterns of kindergarten children in comparison to teenagers and young adults. Eur J Public Health 17(6): 646-51.

Sirard, J. R. and R. R. Pate (2001). Physical activity assessment in children and adolescents. Sports Med 31(6): 439-54.

Stone, E. J., T. L. McKenzie, et al. (1998). Effects of physical activity interventions in youth. Review and synthesis. Am J Prev Med 15(4): 298-315.

Trost, S. G. (2001). Objective measurement of physical activity in youth: current issues, future directions. Exerc Sport Sci Rev 29(1): 32-6.

Trost, S. G., K. L. McIver, et al. (2005). Conducting accelerometer-based activity assessments in field-based research. Med Sci Sports Exerc 37(11 Suppl): S531-43.

Trost, S. G., A. M. Morgan, et al. (2000). Children´s Understanding of the Concept of Physical Activity. Pediatric Exercise Science 12: 293-299.

Trost, S. G., R. R. Pate, et al. (1996). Gender differences in physical activity and determinants of physical activity in rural fifth grade children. J Sch Health 66(4): 145-50.

Trost, S. G., R. R. Pate, et al. (2002). Age and gender differences in objectively measured physical activity in youth. Med Sci Sports Exerc 34(2): 350-5.

Trost, S. G., D. S. Ward, et al. (1998). Validity of the computer science and applications (CSA) activity monitor in children. Med Sci Sports Exerc 30(4): 629-33.

Welk, G. (2002). Use of Accelerometry-Based Activity Monitors to Assess Physical Activity. Physical Activity Assessments for Health-Related Research. G. J. Welk. Champaign, IL, Human Kinetics: 3-18.

Welk, G. J., J. Almeida, et al. (2003). Laboratory calibration and validation of the Biotrainer and Actitrac activity monitors. Med Sci Sports Exerc 35(6): 1057-64.

Welk, G. J., S. N. Blair, et al. (2000). A comparative evaluation of three accelerometry-based physical activity monitors. Med Sci Sports Exerc 32(9 Suppl): S489-97.

Welk, G. J. and C. B. Corbin (1995). The validity of the Tritrac-R3D Activity Monitor for the assessment of physical activity in children. Res Q Exerc Sport 66(3): 202-9.

Welk, G. J., C. B. Corbin, et al. (2000). Measurement issues in the assessment of physical activity in children. Res Q Exerc Sport 71(2 Suppl): S59-73.

Welk, G. J., J. C. Eisenmann, et al. (2007). Calibration of the biotrainer pro activity monitor in children. Pediatr Exerc Sci 19(2): 145-58.

Welk, G. J. and K. Wood (2000). Physical Activity Assessments in Physical Education - A Practical Review of Instruments and Their Use in the Curriculum. Journal of Physical Education, Recreation & Dance 71(1): 30-40.

TABLES AND FIGURES

Table 1. Descriptives by Gender.

	n	Age[1]	Height[2]	Weight[3]	BMI
males	70	10.5±0.8	144.1±8.9	37.1±9.4	18.9±3.6
females	68	10.4±0.8	142.9±10.9	39.5±11.2	20.04±5.3

[1]years
[2]centimeters
[3]kilograms

Table 2. Study Population Distribution.

		n	%
Gender			
	male	79	49.4
	female	81	50.6
Grade	4th	75	46.9
	5th	59	36.9
	6th	26	16.3
Observed			
	Fall	155	96.9
	Spring	73	45.6
	Both	73	46.9
School			
	WesternHills	16	9.5
	Crestview	12	7.1
	Clive	11	6.5
	Fairmeadows	9	5.3
	HillSide	24	14.2
	JordanCreek	26	16.3
	Perry	29	17.2
	Westridge	14	8.3
	CrossRoads	19	11.9

Total n= 160.

Table 3. Scan Distribution.

	Activity (n=624)		Less Active (n=147)[1]		Total (n=771)	
	n	%	n	%	n	%
males	208	33.3	49	33.3	257	33.3
females	208	33.3	49	33.3	257	33.3
together	208	33.3	49	33.3	257	33.3
Fall	309	49.5			309	40.1
Spring	315	50.5	147	100.0	462	59.9
WesternHills	69	11.1	15	10.2	84	10.9
Crestview	75	12.0	15	10.2	90	11.7
Clive	81	13.0	15	10.2	96	12.5
Fairmeadows	75	12.0	36	24.5	111	14.4
HillSide	93	14.9	12	8.2	105	13.6
JordanCreek	36	5.8			36	4.7
Perry	93	14.9	45	30.6	138	17.9
Westridge	75	12.0	9	6.1	84	10.9
CrossRoads	27	4.3			27	3.5
2 minutes scan	363	58.2	147	100	510	66.1
3 minutes scan	201	32.2			201	26.1
5 minutes scan	24	3.8			24	3.1
Alternate interval	36	5.8			36	4.7
[2]1 to 6 participants	202	32.4	50	34.0	252	32.7
7 to 10 participants	266	42.6	58	39.5	324	42.0
11 to 14 participants	100	16.0	9	6.1	109	14.1
15 to 21 participants	56	9.0	30	20.4	86	11.2

[1]All scans were done during the spring and all with 2 minute intervals.
[2] Participants per scan

Table 4. Mean Differences between Biotrainer and SOPLAY for the three classifications of Activity.

	n (scans)	SOPLAY	Biotrainer	difference[2]	sd[3]	t
			MVPA1			
together[1]	208	72.64	22.09	50.55	26.41	27.60**
males	208	73.18	25.55	47.62	32.29	21.32**
females	208	73.15	18.78	54.38	33.23	23.60**
			MVPA2			
together[1]	208	23.42	22.09	1.33	22.06	0.87
males	208	23.79	25.55	-1.77	29.92	-0.85
females	208	23.32	18.78	4.54	29.49	2.22
			MVPA3			
together[1]	208	72.64	74.62	-1.99	29.00	-0.99
males	208	73.18	78.16	-4.98	35.83	-2.01
females	208	73.15	71.95	1.20	37.73	0.46

[1] absolute percentage of active boys and girls
[2] SOPLAY - Biotrainer (% of active individuals)
[3]standard deviation
**significant at $p<0.001$
MVPA1 = "walk" + "very active" (SOPLAY) vs mvpa (Biotrainer)
MVPA2 = "very active" (SOPLAY) vs mvpa (Biotrainer)
MVPA3 = "walk" + "very active" (SOPLAY) vs light + mvpa (Biotrainer)

Table 5. Pearson Correlations between the Biotrainer and SOPLAY.

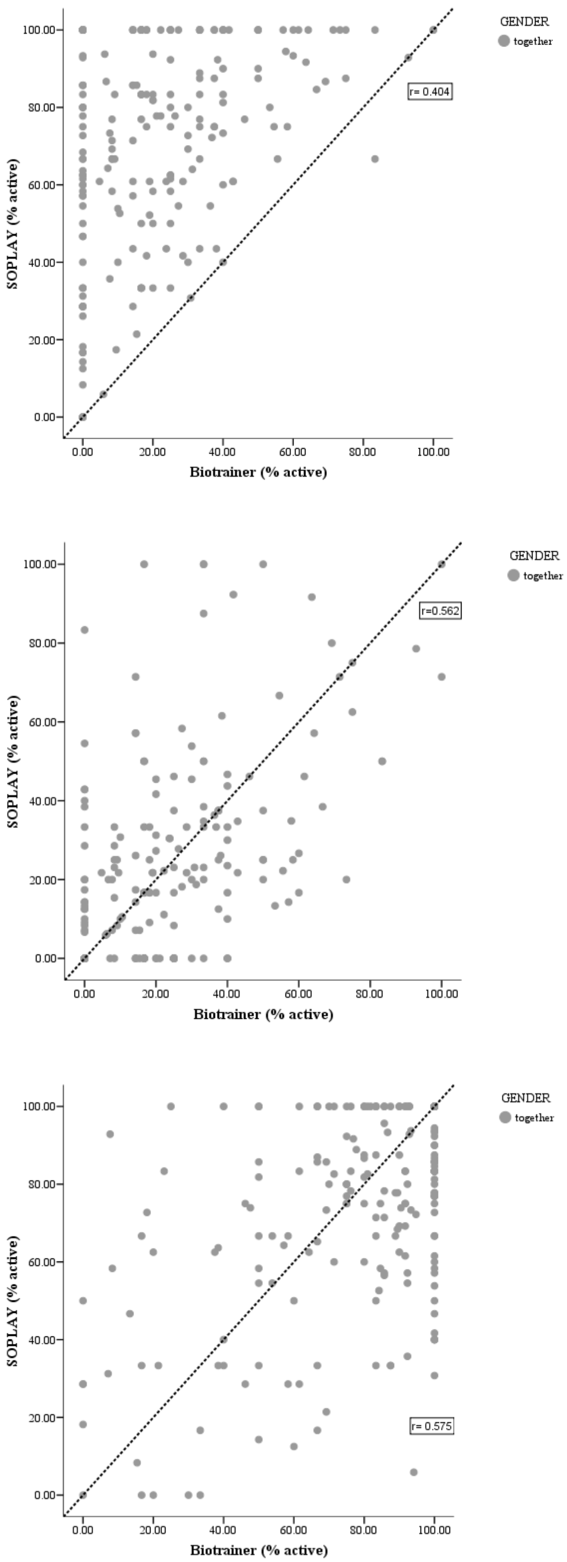
	together[1]	males	females
MVPA1	0.404**	0.368**	0.239*
MVPA2	0.562**	0.428**	0.394**
MVPA3	0.575**	0.498**	0.471**

[1] absolute percentage of active boys and girls
*significant at p<0.01
**significant at p<0.001
MVPA1 = "walk" + "very active" (SOPLAY) vs mvpa (Biotrainer)
MVPA2 = "very active" (SOPLAY) vs mvpa (Biotrainer)
MVPA3 = “walk” + "very active" (SOPLAY) vs light + mvpa (Biotrainer)

Table 6. Frequency of participants and scans for Analysis 2.

Participants distribution		n	%
Gender			
	male	43	58.9
	female	30	41.1
Observed			
	Spring	73	100.0
School			
	WesternHills	11	15.1
	Crestview	8	10.9
	Clive	6	8.2
	Fairmeadows	6	8.2
	HillSide	14	19.2
	Perry	21	28.8
	Westridge	7	9.6
Scans Distribution		n	%
Scan Frequency			
	2 minutes scan	105	41.67
	4 minutes scan	49	19.44
	6 minutes scan	35	13.89
	8 minutes scan	21	8.33
	10 minutes scan	21	8.33
	12 minutes scan	14	5.56
	20 minutes scan	7	2.78
TOTAL		252	100

Only data from schools with two minutes sampling intervals and 30 minutes of session were included in these analyses.



A)

B)

C)

Figure 1. Association between SOPLAY and Biotrainer for A) MVPA1, B) MVPA2 and C) MVPA3. The reference line (dashed line) represents a hypothetical perfect relation between the two instruments.

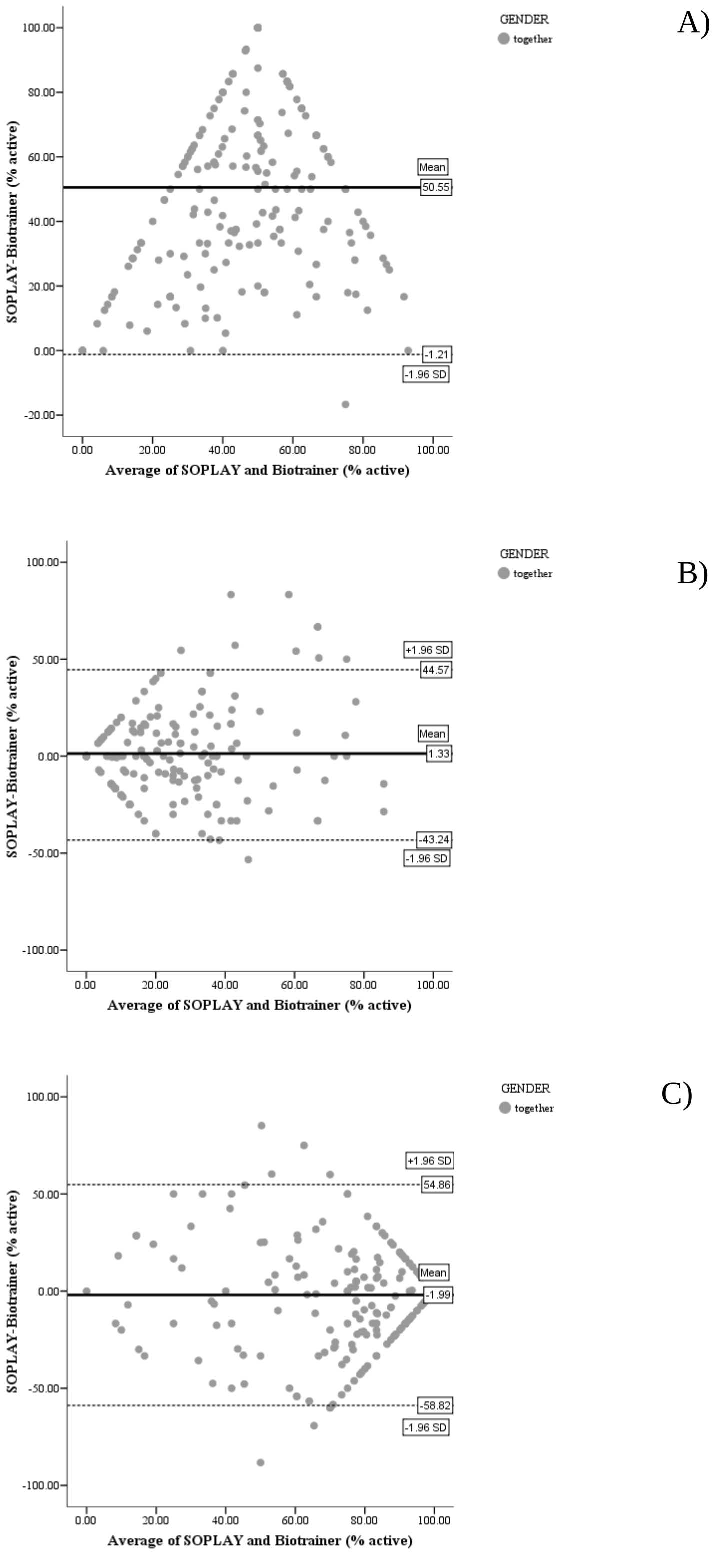


Figure 2. Bland-Altman plots for absolute proportion of active participants using A) MVPA1, B) MVPA2 and C) MVPA3.

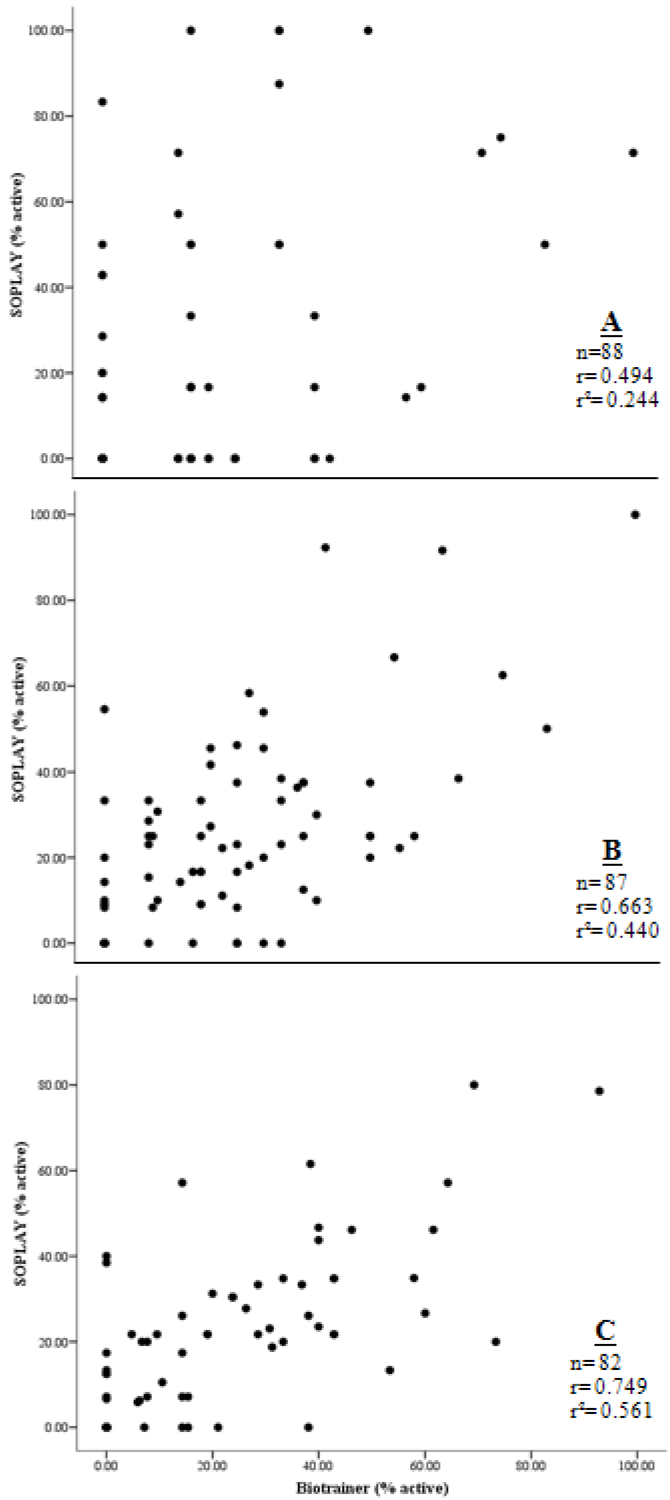


Figure 3 - Relation between agreement and number of participants per scan.
A) 4 to 7 participants
B) 8 to 12 participants
C) 13 to 21 participants

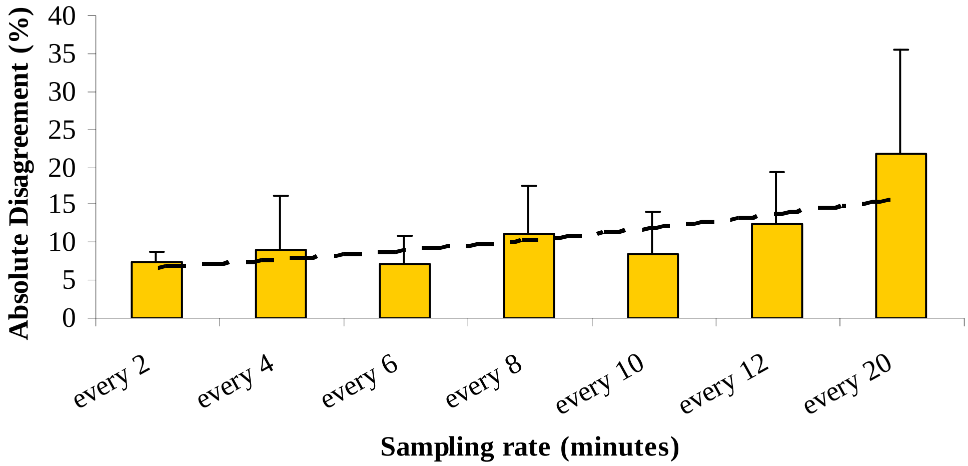


Figure 4 - Absolute disagreement variability when using different sampling rates of observations.

APPENDIX

Panel A

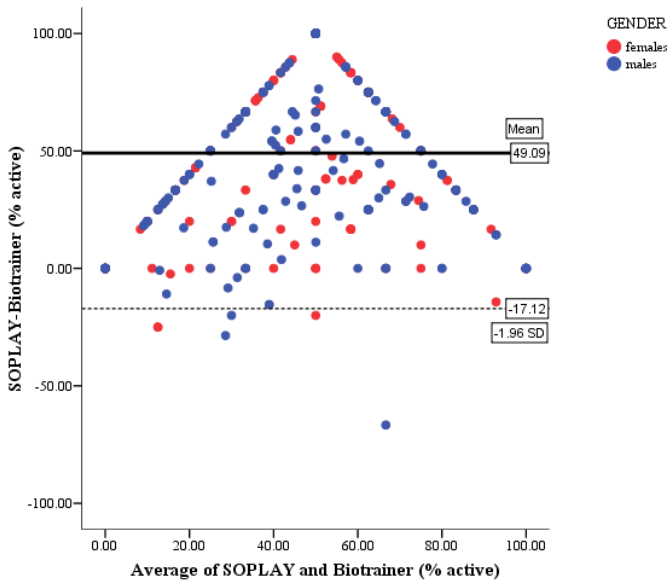


Figure A.1 – Bland-Altman plot for MVPA1 showing a systematic overestimation across the all range of activity levels.

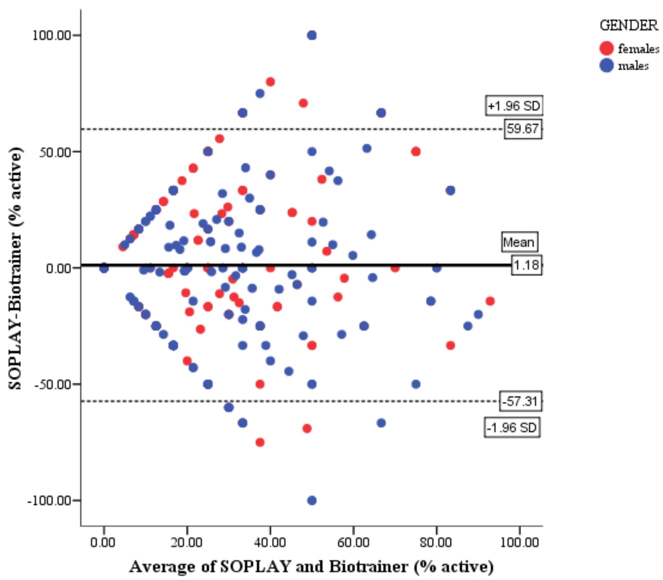


Figure A.2 – Bland-Altman plot for MVPA2 illustrating a symmetric disagreement across the all range of activity levels. Nevertheless, the distribution of the data seems to be shifted (left sided) indicating lower levels of physical activity when using this activity indicator.

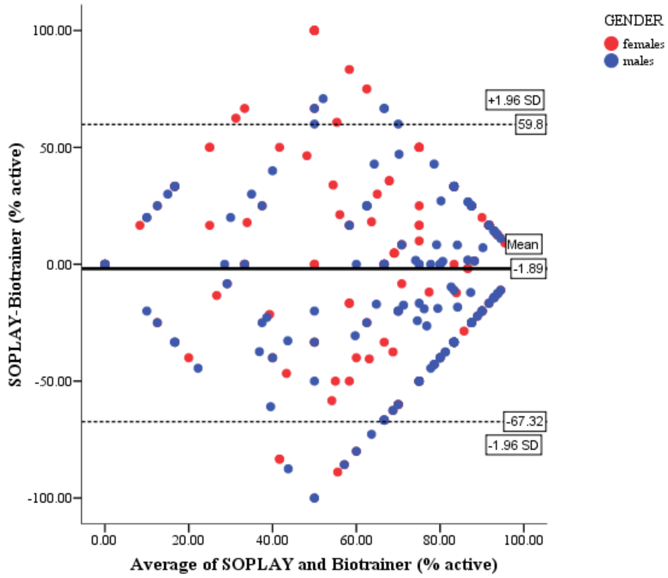


Figure A.3 – Bland-Altman plot for MVPA3 illustrating a symmetric disagreement across the all range of activity levels. Data points seem to be shifted to the right indicating higher levels of physical activity when using this activity indicator.

Panel B

Table B.1. Overall differences between the Biotrainer and the SOPLAY during the "less active period".

	n (scans)	SOPLAY	Biotrainer	difference[2]	sd[3]	t
				MVPA1		
together[1]	49	33.55	5.93	27.62	24.35	7.94**
males	49	36.79	6.39	30.4	28.87	7.37**
females	49	30.92	5.78	25.14	28.84	6.10**
				MVPA2		
together[1]	49	1.36	2.61	-1.24	9.78	-0.89
males	49	2.28	2.52	-0.24	10.81	-0.15
females	49	0.68	3.06	-2.38	13.18	-1.27

[1] absolute percentage of active boys and girls
[2] SOPLAY - Biotrainer (% of active individuals)
[3]standard deviation
**significant at $p<0.001$
MVPA1 = "walk" + "very active" (SOPLAY) vs mvpa (Biotrainer)
MVPA2 = "very active" (SOPLAY) vs mvpa (Biotrainer)

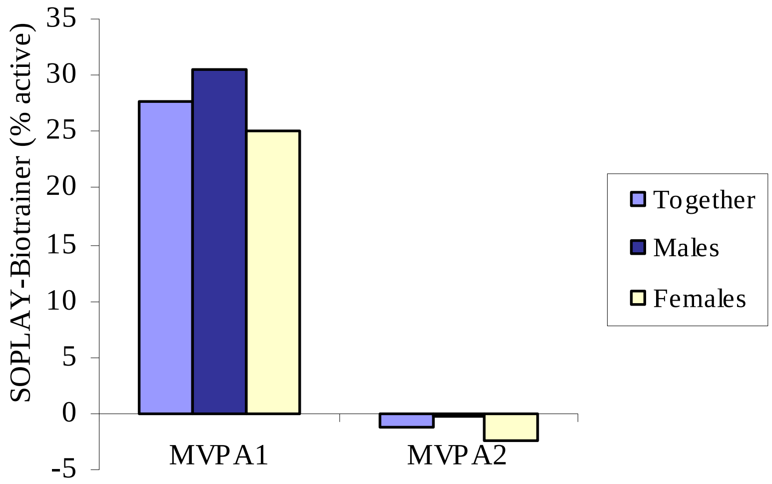


Figure B.1. Mean difference values during the less active period (MVPA1 vs MVPA2). Differences were significantly higher for MVPA1 (p-value<0.01).

MVPA1 = "walk" + "very active" (SOPLAY) vs mvpa (Biotrainer)
MVPA2 = "very active" (SOPLAY) vs mvpa (Biotrainer)

Panel C

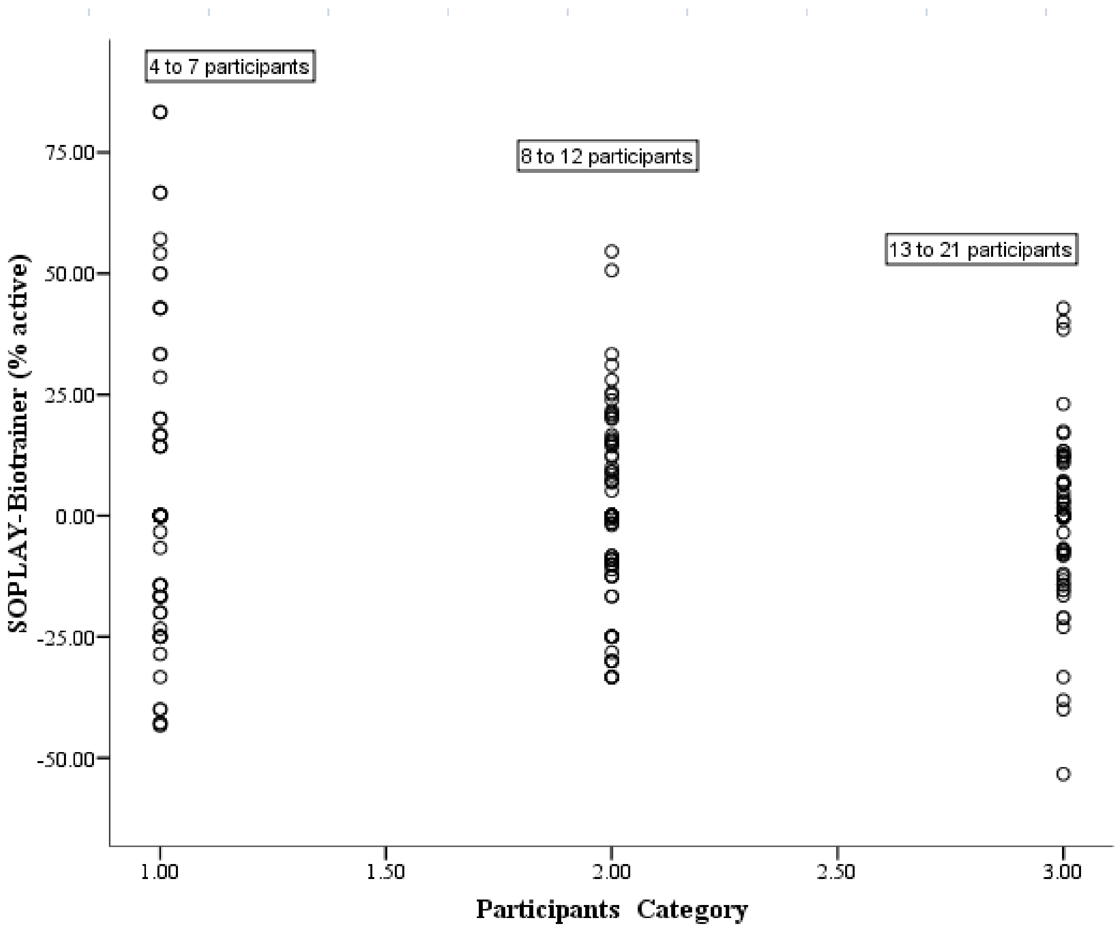


Figure C.1 – Average disagreement and dispersion of correct classification of active individuals. Category 3 (13 to 21 participants) had the individual points more concentrated around 0, suggesting better agreement.

Panel D

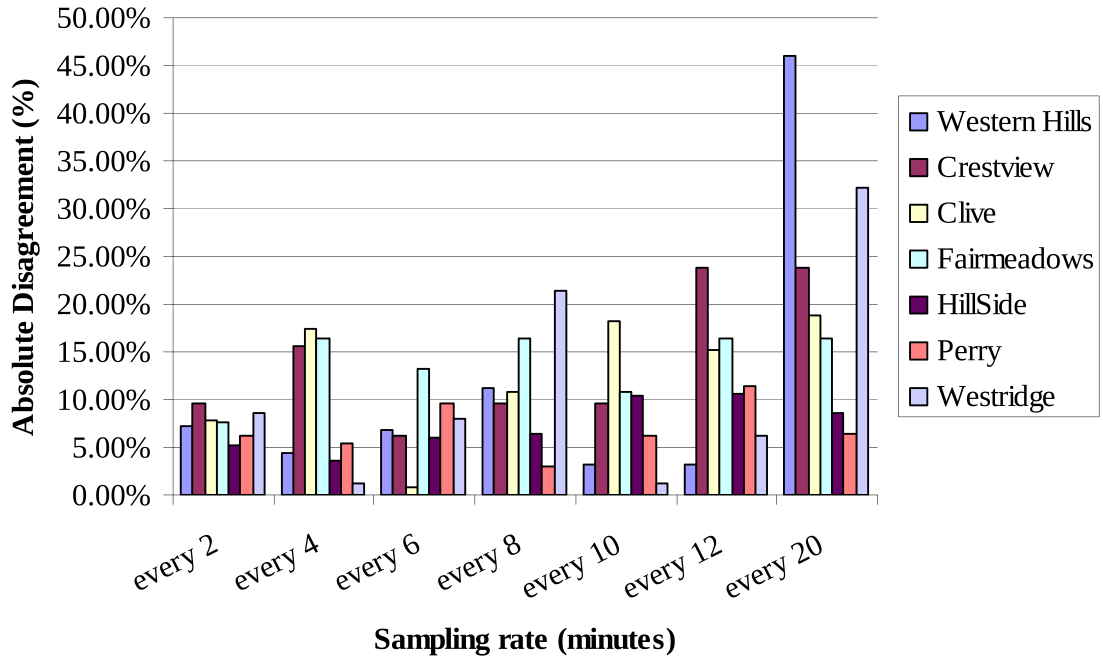


Figure D.1 – Variability in absolute disagreement among the 7 schools where data was collected. With exception of a few schools, overall pattern was similar within group of schools for the different scan rates. The highest differences were verified for the 20 minutes interval of observation.