Published in final edited form as:
Meas Phys Educ Exerc Sci. 2011 ; 15(4): 257–273. doi:10.1080/1091367X.2011.594361.

Measuring physical activity in preschoolers: Reliability and validity of The System for Observing Fitness Instruction Time for Preschoolers (SOFIT-P)

Shreela Sharma[1], **Ru-Jye Chuang**[1], **Katherine Skala**[1], and **Heather Atteberry**[2]

[1] Michael and Susan Dell Center for Healthy Living, The University of Texas School of Public Health, Houston, TX USA

[2] Michael and Susan Dell Center for Healthy Living, The University of Texas School of Public Health Austin Regional Campus, Austin, TX USA

Abstract

The purpose of this study is describe the initial feasibility, reliability, and validity of an instrument to measure physical activity in preschoolers using direct observation. The System for Observing Fitness Instruction Time for Preschoolers was developed and tested among 3- to 6-year-old children over fall 2008 for feasibility and reliability (Phase I, *n*=67) and in fall 2009 for concurrent validity (Phase II, *n*=27). Phase I showed that preschoolers spent >75% of their active time at preschool in light physical activity. The mean inter-observer agreements scores were ≥.75 for physical activity level and type. Correlation coefficients, measuring construct validity between the lesson context and physical activity types with and with the activity levels, were moderately strong. Phase II showed moderately strong correlations ranging from .50 to .54 between the System for Observing Fitness Instruction Time for Preschoolers and Actigraph accelerometers for physical activity levels. The System for Observing Fitness Instruction Time for Preschoolers shows promising initial results as a new method for measuring physical activity among preschoolers.

Keywords

direct observation; measurement; early childhood; preschool; physical activity

INTRODUCTION

Obesity has been on the rise in the United States, with national data indicating that, in 2003–2006, 24.4% of children, ages 2 to 5 years were overweight or obese (Lutfiyya, Garcia, Dankwa, Young, & Lipsky, 2008; Ogden, Carroll, & Flegal, 2008). Physical activity plays an important role in maintaining a healthy weight (Dugan, 2008; Hill & Wyatt, 2005; Rippe & Hess, 1998; Roberts, 2000;Warburton, Nicol, & Bredin, 2006), and preschoolers are at an age where they are beginning to develop their activity habits (Birch, 2008; Timmons, Naylor, & Pfeiffer, 2007).

Evidence supports an inverse relationship between physical activity levels in preschoolers and adiposity (Hill & Wyatt, 2005). The National Association for Sport and Physical Education (NASPE) recommends preschoolers spend at least 60 min daily in structured physical activity, 60 min and up to several hours of unstructured physical activity every day, and should not be sedentary for more than 60 min at a time unless sleeping (NASPE, 2002). It is important to determine what physical activity preschoolers are involved in, so as to determine if they are meeting the recommendations. Currently, there are a few guidelines for

physical activity in preschools across the United States, and those that do exist vary considerably by state (Kaphingst & Story, 2009). This is because the literature on physical activity levels for preschool children, especially those from low-income minority populations, is limited. One of the reasons for this paucity in literature could be the challenges in measurement of physical activity of this age group. Preschoolers' physical activity types and levels are different than other age groups. Furthermore, it is not plausible to do self-report with this age group, making measurement more expensive and time consuming.

Direct observation is a salient method of measuring physical activity among children, including preschoolers. One advantage of direct observation is that it allows the researcher to gather valuable contextual information, including activity pattern and type during the activity time (Pate, O'Neill, & Mitchell, 2010). The System for Observing Fitness Instruction Time for Preschoolers (SOFIT-P), adapted from the System for Observing Fitness Instruction Time (SOFIT) instrument (McKenzie, Sallis, & Nader, 1991a), is one such new method of measuring physical activity level, type, and context in preschoolers. Currently there are various measures of direct observation of physical activity in preschoolers. However, the constructs measured and the reliability and validity of these instruments is varied. The Children's Physical Activity Form (CPAF) measures physical activity intensity and has been validated using heart rate monitors in older children ages 8 to 10 years and does not have any reliability data available (O'Hara, Baranowski, Simons-Morton, Wilson, & Parcel, 1989) . The Children's Activity Rating Scale (CARS) has been validated using accelerometry in 2- to 5-year-old children but measures only physical activity intensity and not context (Finn & Specker, 2000; Noland, Danner, DeWalt, McFadden, & Kotchen, 1990; Puhl, Greaves, Hoyt, & Baranowski, 1990). The Children's Activity Timesampling Survey (CATS) measures physical activity intensity, environment, and child interaction with others. However, the validity of this instrument remains to be seen (Klesges, Eck, Hanson, Haddock, & Klesges, 1990). The Behaviors for Eating and Activity for Child Health: Evaluation System (BEACHES) measures only global environmental information for physical activity in preschools or home settings with no child activity level data for analysis (McKenzie et al., 1991a). Finally, the Observational System for Recording Activity in Children–Preschool Version (OSRAC-P), which served as a model for refining the SOFIT-P, measures the physical activity type, intensity, and contexts for preschoolers (Brown, Pfeiffer, McIver, Dowda, Almeida, & Pate, 2006). However, the validity of this method against a criterion standard has not yet been published. Validating an instrument, especially in the age group in which it is intended for use, is critical. Furthermore, using an acceptable criterion standard for validity, such as accelerometry, is important to ensure the accuracy of the instrument. The purpose of this article is to (a) describe the development of the SOFIT-P, (b) present the results from initial feasibility and reliability testing conducted in two Head Start centers and, (c) present the results of the validity testing of the SOFIT-P conducted using accelerometers in one Head Start Center.

METHODS

The SOFIT-P was developed and validated as part of the Coordinated Approach to Child Health in Underserved Populations (CATCH UP) study. The purpose of CATCH UP was to develop and pilot test a preschool-based nutrition and physical activity program called CATCH Early Childhood (CEC) among 3 to 6 year old children enrolled in Harris County Department of Education (HCDE) Head Start centers. HCDE Head Start has 16 centers in Harris County, Texas. The population is 49.6% Hispanic, 41.4% African-American, 3.5% Bi-racial/Multi-racial, 3.3% White, and 2.2% other ethnicities (Head Start Director, personal communication, June 20, 2008). All children in the HCDE Head Start are economically disadvantaged.

The CATCH UP study was conducted in two phases. Phase I consisted of testing the CEC program for feasibility and was conducted in two HCDE Head Start centers for six weeks, from October to December 2008. Data were collected at pre-test (prior to implementation of intervention) and post-test (end of six weeks). Details are available elsewhere (Sharma, Chuang, & Hedberg, 2011). The initial feasibility, reliability, and construct validity testing of the SOFIT-P was conducted during this phase. Phase II consisted of testing CEC in the 2009–2010 school year with a larger sample of HCDE Head Start children and parents using a group-randomized controlled trial design (n=6 centers, 3 intervention, 3 control centers; 296 parent/child dyads). The validation of the SOFIT-P using the criterion standard Actigraph GT3X accelerometers (Pensacola, Florida, USA, www.actigraph.com) was conducted as part of this phase during baseline measurements in fall 2009 on 27 children from two classrooms in 1 Head Start center (19 SOFIT-P observations).

Active parental consent was obtained for all preschoolers to participate in the CATCH UP study. This study was approved by the University of Texas Health Science Center at Houston, Committee for Protection of Human Subjects (HSC-SPH-04-0414), as well as the HCDE Head Start Institutional Review Board.

PHASE I (fall 2008)

Adaptation of System for Observing Fitness Instruction Time for Preschoolers (SOFIT-P)—The SOFIT-P is a modified version of the SOFIT instrument (McKenzie et al., 1991a; McKenzie, 2005). The SOFIT is a validated direct-observation instrument that measures the percent of time students spend in moderate-to-vigorous physical activity (MVPA) during physical education (PE) time along with lesson context (McKenzie et al., 1991a) at school. Observers randomly select four children from each class, and the children are observed for 20 sec each for a total of 4-min rotating blocks. These four children are considered to be representative of the physical activity for their entire class. Thus, group, and not individual, data are reported. The SOFIT's five physical activity codes have been validated with heart-rate monitoring for elementary and middle school children (R=.80–.91, $p < .01$) (Rowe, Schuldheisz, & van der Mars, 1997), using CALTRAC accelerometers with fifth graders (r=.74) (McKenzie, Sallis, & Armstrong, 1994) and TriTrac accelerometers with third through fifth graders, during PE (r=0.61) (Pope, Coleman, Gonzalez, & Heath, 2000). The instrument also has high inter-observer reliability ratings, >90% for energy expenditure rate estimates, total energy expenditure estimates, and time spent in MVPA (McKenzie, Marshall, Sallis, & Conway, 2000).

The SOFIT requires two days of training and has ten coding definitions for two environmental contexts (McKenzie et al., 1991a). It has been used extensively in prior studies (Luepker et al., 1996; Perry et al., 1990, 1997), as well as by the investigative team, with great success. The investigative team initially conducted an extensive literature review on other methods of direct observation of physical activity among preschoolers (Baranowski, Thompson, DuRant, Baranowski, & Puhl, 1993; Brown, 2005; DuRant et al., 1993; DuRant, Baranowski, Johnson, & Thompson, 1994; Jago, Baranowski, Thompson, Greaves, & Baranowski, 2005; Puhl et al., 1990; McKenzie et al., 1991b). The OSRAC-P served as a model for the codes used for refining the SOFIT-P. The preliminary version of the SOFIT-P included a candidate list of physical activity levels, types, and contexts (see Appendix A). The physical activity level, type, and context are coded every 30 sec usingmomentary time sampling (20 sec observation, 10 sec recording) similar to the original SOFIT instrument. Four observers, all university-educated females with previous experience in the SOFIT, pre-tested this first version of the instrument in the four classrooms at two non-study HCDE Head Start centers during active times in the preschool day (8 observations, 4 indoor and 4 outdoor, 2 observations/classroom). Active times were when

the students participated in free play (indoors or outdoors with no teacher instruction), circle time (instructor led activity time), music and movement time or learning centers (time spent at specific stations, including art, blocks, reading, etc.). Nap times and meal times were not considered as active times. These active times were identified from the classroom daily schedules with feedback from the teachers. The observers preselected one student randomly at a time and documented their level, activity type, and context throughout the active time using the codes from the candidate list that they found most applicable. After finishing the fourth student, if the activity was still in session, the observer started with the first student again. Observers were instructed to include comments at the end of each observation. Further modifications were made to the instrument based on this pre-testing, which included eliminating codes that were least applicable to this population based on ≥90% agreement from the observers. A final list of physical activity level, type, and context was developed to accurately reflect and capture the physical activity of a preschooler (see Appendix B). In the SOFIT-P, for activity level, codes 1 to 4 describe the body position of the student (lying, sitting, standing, walking), and code 5 identifies when the student is being more active than walking (fast/very active). The activity types are classified based on the energy expenditure related to each activity (S—lie down/sit/stand/squat, C—climb/crawl, W—walk/ride, P—push/pull/throw, K—rock/swing, D—dance/jump/skip, R—run/roll/rough and tumble). Finally, the indoor and outdoor contexts provide information on the lesson context, i.e., under what context did the activity take place. An iPod having the observing and recording prompts was used by the observer to pace and record the physical activity level and physical activity type.

Training—The investigative team developed a protocol for the SOFIT-P providing detailed guidelines for trained observers to select students, perform the observation, summarize data collected, perform reliability checks, and calculate reliabilities. The protocol includes definitions and coding specifics of activity level, activity type, indoor context, and outdoor context (see Appendix B for sample definitions and coding specifics). For this phase, four research staff, two of whom had prior experience with using the SOFIT, participated in a one-day classroom training session conducted by study investigators, which included an overview of the SOFIT-P momentary time sampling coding procedure using video recordings of examples, role playing, practice with the SOFIT training videos, and overview of the new coding definitions. After the classroom training, research staff practiced using the SOFIT-P at the two non-study HCDE Head Start centers for feasibility of implementation and coding. During this phase, trained research staff were asked to write down any problems or questions they had regarding the use of the tool, which were then discussed and clarified.

Reliability testing—Following the adaptation and training, reliability of the SOFIT-P was measured in four classrooms at two participating HCDE Head Start centers (centers 1 and 2) at all “active” times during preschool hours. All observations used were between 20 to 50 min in length. A daily activity schedule was obtained from each of the participating classrooms to determine when to conduct the measurements. On the day of the measurement, trained observers were present at the Head Start center for the entire duration of the school day (7.30 A.M. to 2.30 P.M.), and SOFIT-P measures were conducted during the above mentioned “active” times. Measurements were conducted separately for each classroom, each of which had no more than 23 children. A total of 30 observations were conducted at baseline and 16 observations at post-test (including double observations conducted for reliability). Double observations were conducted on 35% of all observations. Inter-observer reliability was conducted by two research staff observing the same four preschoolers during the same observation period. The primary observer who had prior experience using the SOFIT selected the preschoolers and led the observation. The two

observers shared the same iPod with the pacing track, so they were observing and recording the same preschooler simultaneously.

PHASE II (fall 2009)

Validation using criterion standard—For Phase II, in fall 2009, SOFIT-P observations were validated against the Actigraph GT3X accelerometers. Accelerometers provide objective measures of the amount, time, intensity, and frequency of activity, as well as step counts, and they are considered the criterion standard for measurement of physical activity in children (Pate, Almeida, McIver, Pfeiffer, & Dowda, 2006). Trained research staff that conducted the reliability testing in Phase I participated again in this phase. A booster training was conducted for the staff as per protocol prior to measurements. These project staff was also trained in how to put the accelerometers on the children at the start of the school day. Twenty-seven children from two classrooms in one Head Start center participated. Children wore the accelerometers throughout the entire school day. Epochs were set so that data were collected every 5 sec. The accelerometers were worn at the waist over the child's right hip over his/her clothing and attached with an elastic belt. All children were given the opportunity to familiarize themselves to the accelerometer prior to data collection so that the actual measurements will reflect physical activity. Children and teachers were adequately instructed not to remove the accelerometers during the measurement period. Project staff or teachers removed the accelerometers at the end of the school day prior to the child's going home. Data were downloaded directly from the accelerometer into the study database and activity counts were compiled. Five-second epochs were integrated to 60-sec intervals. Cut points for preschool children as determined by Sirard, Trost, Pfeiffer, Dowda, and Pate (2005) were used to determine physical activity levels. Sedentary, light, moderate, and vigorous levels of physical activity were determined from these activity counts. As in Phase I, SOFIT-P measurements were conducted at predetermined active times during the day. A total of 19 SOFIT-P observations were collected and validated against the corresponding accelerometer data collected at these active times.

Data Analysis—All analyses were performed using SPSS version 17.0 (SPSS Inc., Chicago, Illinois, USA), and statistical significance was set at $p < .05$.

The mean percent time spent in each physical activity level was calculated after data were collected. Light physical activity was calculated by collapsing percent time spent at level 1, level 2, and level 3. Moderate physical activity was calculated by percent time spent at level 4. MVPA was calculated by percent time spent at level 5. MVPA was calculated by collapsing percent time spent at level 4 and level 5. Inter-observer agreement (IOA) for physical activity level and activity type were calculated at pre-test (prior to intervention implementation) and post-test (end of six weeks) for each set of reliability observations to determine if observers can reliably code the preschoolers' activity. An IOA score at or above .80 indicates that independent observers frequently agree on the physical activity level and activity type codes (McKenzie et al., 1991a). Any score below .80 indicates that retraining is required.

The kappa statistic was calculated to determine how different the observed agreement is from the expected agreement. All kappa scores above .80 were considered to be very good agreement, while scores between .80 and .60 were considered moderate agreement. Any score below .60 was considered poor agreement (Viera & Garrett, 2005). The higher the kappa score, the higher the consistency among observers for recording preschoolers' physical activity level and activity type. The physical activity level and activity type IOA,

kappa mean scores, standard deviations, and ranges were calculated for reliability at pre-test and post-test.

Construct validity refers to the extent to which operationalization of a construct measures what theory says it does. Construct validity of the SOFIT-P was determined by correlating the lesson contexts and the physical activity type with physical activity level using point biserial correlation coefficients (*rpb*).

Concurrent validity was determined by correlating the physical activity level (i.e., percent of time spent in sedentary, light, moderate, and vigorous physical activity) data from the SOFIT-P with that obtained from the accelerometers. For the accelerometers, activity counts obtained from the 5-sec epochs were integrated to 60 sec and cut points for children ages 3 to 5 years (Sirard et al., 2005), which were used to determine the physical activity level. Following this, percent time of sedentary was calculated by combining percent sedentary and light activity, and percent of MVPA was calculated by combining moderate, heavy, and very heavy. Pearson's correlation coefficients were used to determine the correlation between the SOFIT-P and accelerometer data.

RESULTS

Given that the study sample was different for Phases I and II, results for the two phases are presented separately. Table 1 presents the demographic characteristics of the study participants in Phases I and II. The population was primarily Hispanic and African American, and mean children's age was approximately 4 years in both phases.

PHASE I

A total of 24 observations at pre-test (i.e., prior to the intervention) and 10 observations at posttest (end of six weeks) were conducted using the SOFIT-P. Reliability check was conducted on 35% of the sample, resulting in a total of 30 pre- and 16 post-test observations for the study. Table 2 presents the IOA and kappa mean scores separately for pre-test and post-test by physical activity level and activity type. This was done to obtain data on improvements in the measurement over time by trained observers. The mean IOA for physical activity level at pretest was .78± .085 and at post-test was .92±.028. For physical activity type, the mean IOA at pre-test was .79±.072 and at post-test was .78±.097. For physical activity level, the mean kappa score for interrater reliability at pre-test was .69±. 106 and at post-test was .86±.040. For physical activity type, the mean kappa score at pre-test was 69±.125 and at post-test was .66±.140.

Table 3 shows the percent of active time spent in light, moderate, vigorous, and moderate–vigorous activity by center and classroom. This excludes the double observations conducted for reliability. Since the purpose of this study was to determine the feasibility and validity of using the SOFIT-P, data from the pre-test and post-test were collapsed for this analysis. Results showed that, of their most active time at preschool both indoors and outdoors, preschoolers spent the majority of it in light physical activity (77.1%) and only 22.9% of their time in MVPA. Given that most of the observations were conducted indoors, the percent time spent indoors and outdoors on various activities further analyzed separately (Table 4). This included 24 indoor observations and 10 observations outdoors. Results showed that while indoors, the preschoolers were engaged in light physical activity most of the time (86.4%). Only 13.7% of the preschoolers' active time was spent in MVPA while indoors. However, 45.2% of their time was spent in MVPA while outdoors.

Results on classroom context variables showed that while outdoors, preschoolers spent their time playing on fixed playground equipment or playing in open spaces. Indoors, the

preschoolers participated in the following context categories: related to art, book/ preacademic, group time, large blocks, manipulatives, music, sociodramatic, and transition time. By nature and structure, indoor classroom-based activities provide fewer opportunities for vigorous activity as compared to outdoor play. Table 5 shows the construct validity of the SOFIT-P by correlating the lesson contexts with the activity levels. Results showed that the outdoor context correlated negatively with sitting time (*rpb* =–.641, *p*=<.001) and percent of sedentary time (*rpb* =–.779, $p < .001$) and had strong positive correlations with walking (*rpb* =.581, $p < .001$) and fast/active (*rpb* =.801, $p < .001$) and percent MVPA time (*rpb* =.779, $p < .001$). Table 6 also shows the construct validity of the SOFIT-P instrument by correlating the activity types with the levels. Results show that for both indoor and outdoor, activity types such as lying down/sitting/standing correlate positively with percent sedentary time (*rpb* =.523, *p*=.005), while activity types such as run/roll/rough and tumble correlate positively with percent vigorous time (*rpb* =.455, *p*=.02). Also, both indoor and outdoor walking/riding lesson context showed strong correlations with percent walking time (*rpb* =.805, $p < .001$ for indoor context; *rpb* =.960, $p < .0019$ for outdoor context).

PHASE II

This phase consisted of validating the 19 SOFIT-P observations against the criterion standard using data from Actigraph GT3X accelerometers on 27 children. Results showed that the percent time in moderate activity determined using the SOFIT-P correlated positively with accelerometer data for percent time in moderate (Pearson's *r*=.506, *p*=.007) and percent time in MVPA (Pearson's *r*=.532, *p*=.004) and correlated negatively with percent time in sedentary activity (Pearson's *r*=–.532, *r*=.004). Similarly, the percent time in MVPA on the SOFIT-P correlated positively with accelerometer data for percent time in MVPA (Pearson's *r*=.541, *p*=.004) and percent time in moderate activity (Pearson's *r*=.530, *p*=.004) and negatively for percent time in sedentary activity (Pearson's *r*=–.541, *p*=.003). Finally, percent time in sedentary behavior on the SOFIT-P correlated positively with corresponding accelerometer data for percent time in sedentary activity (Pearson's *r*=.541, *p*=.004).

DISCUSSION

The SOFIT-P was adapted from the SOFIT to measure physical activity levels and type for children aged 3 to 6 using direct observation. Overall, the SOFIT-P proved to be a feasible, reliable, and valid method to use in a Head Start preschool environment.

The mean IOA scores for physical activity level and type were greater than .75, indicating that, overall, the SOFIT-P allows for observers to independently and reliably record preschoolers' physical activity levels and types at different centers and during different instruction period types. However, some of the IOA scores were as low at .63 and were lower at pre-test than at post-test. The observer comments indicated that the physical activity type codes, especially for distinguishing between codes S (lie down/sit/squat/stand) and C (climb/crawl), as well as between P (push/pull/throw) and R (run/roll/rough and tumble), were sometimes challenging. These results, along with those for the kappa scores, which were overall less than .80, indicate that additional modification of the training component with more time spent in the field training of the SOFIT-P would be recommended for this instrument. This can be further attested to by the improvement in the scores pre- to post-test. While regression to the mean could be one explanation of this change in the scores, the improvement in scores post-test was more likely a result of the booster training that was implemented between pre- and post-test for the observers.

The SOFIT-P also showed good construct validity between the lesson contexts and activity type with the activity levels. All active time observed indoors included classroom-based

activities, such as music and movement time or free play using large blocks or other games, and mainly involved light physical activity. On the other hand, outdoor time typically included free play in the designated play area or teacher-arranged activities mainly involving moderate or vigorous physical activity. As expected, the results showed that indoor context was strongly correlated with the light physical activity levels, while the outdoor context was correlated with moderate or vigorous physical activity levels. Additionally, there was consistency observed for the correlations between activity type and activity levels, with lower-energy activities, such as lying down, sitting, standing, correlating positively with sedentary time, and higher-energy activities correlating positively with percent MVPA.

The SOFIT-P showed that preschoolers at the two participating Head Start centers in Phase I spent a majority of their time in light physical activity, more indoors than outdoors. These results are consistent with the literature on physical activity levels of preschoolers measured using direct observation, which indicated that preschoolers spend a majority of their time in schools doing light physical activity (Brown, Pfeiffer, McIver, Dowda, Addy, & Pate, 2009; Oliver, Schofield, & Kolt, 2007; Pate, McIver, Dowda, Brown, & Addy, 2008). Other studies using pedometers have found similar results (Louie & Chan, 2003; Nevill, McKee, Boreham, & Murphy, 2005; Oliver, Schofield, Kolt, & Schluter, 2007; Pate, Pfeiffer, Trost, Ziegler, & Dowda, 2004), which alludes to the external validity of the SOFIT-P. The high percentage of light physical activity during active times at preschool also indicates a need to incorporate programs promoting physical activity, especially MVPA, in a preschool setting.

Feedback from the trained observers in Phase I showed that the SOFIT-P was relatively easy to use and feasible in an HCDE Head Start setting, both indoors and outdoors. The training component was not reported to be overwhelmingly cumbersome for the observers. None of the teachers or children refused to participate in the measurement, reacted negatively, or voiced any discomfort or unease at being observed for prolonged periods of time. Given the various “active” times during the day in a preschool, which might vary slightly for each classroom, it is recommended that the observers obtain the daily schedule from each classroom for planning of the measurement. Additionally, given the variability of the physical activity in each preschool, prior classroom and field training, which includes inter-observer reliability testing, in the preschool setting that is to be measured is highly recommended.

Finally, the validity of the SOFIT-P against the criterion standard of accelerometers showed promising results with moderately strong correlation coefficients between the physical activity levels determined by both methods. Actigraph accelerometers are a validated method of measurement of physical activity in the preschool population (Pate et al., 2006). The magnitude of the correlation coefficients seen in the current study between the two measures can be considered moderate. These results coupled with the results of the reliability testing indicate that the SOFIT-P shows promising initial results for validity; however, additional refining of the instrument, training protocol, and validating in a more diverse and larger sample may be warranted.

Limitations

One of the limitations of the study is the small number of observations, especially at post-test for the initial reliability testing. However, this was a pilot study providing results on the feasibility of using the SOFIT-P in a preschool setting as well as initial results for reliability and validity. Future studies employing a larger number of observations are needed. Second, the SOFIT-P currently does not collect data on instructor behavior, unlike the SOFIT. Instructor behavior was omitted because at Head Start centers, there is no official PE instruction time or an instructor, making it difficult to measure these characteristics. However, there are current considerations of including various instructor behaviors

constructs as appropriate to a preschool setting for the instrument. Finally, reliability and validity of physical activity contexts were not conducted as part of these two phases of the study but will be followed up in future studies.

CONCLUSIONS

In conclusion, the SOFIT-P tool does show promising preliminary results as a feasible and reliable method for collecting physical activity level data using direct observation for preschoolers during structured active times at school. Future plans include further refining of the instrument and training protocols and validating in a larger and diverse sample.

Acknowledgments

The authors would like to acknowledge the HCDE Head Start director, Ms. Venetia Peacock, the Health Services Manager, Ms. Gulshan Rahman, and the participating center teachers and managers. The authors would also like to thank Dr. Ann Marie Hedberg, Ms. Tamara Lewis, and Ms. Peggy Murray for their assistance with the study. Finally, the authors would like to acknowledge the Michael & Susan Dell Center for Healthy Living for providing administrative assistance with the study.

REFERENCES

Baranowski T, Thompson WO, DuRant RH, Baranowski J, Puhl J. Observations on physical activity in physical locations: Age, gender, ethnicity, and month effects. Research Quarterly for Exercise and Sport. 1993; 64(2):127–133. [PubMed: 8341835]

Birch LL. Development of food acceptance patterns in the first years of life. Proceedings of the Nutrition Society. 2008; 57(04):617–624. [PubMed: 10096125]

Brown, WH. Observational system for recording physical activity in children-preschool (OSRAC-P) training manual for observers. Columbia, SC: 2005.

Brown WH, Pfeiffer KA, McIver KL, Dowda M, Addy CL, Pate RR. Social and environmental factors associated with preschoolers' nonsedentary physical activity. Child Development. 2009; 80(1):45–58. [PubMed: 19236392]

Brown WH, Pfeiffer KA, McIver KL, Dowda M, Almeida M, Pate RR. Assessing preschool children's physical activity: The observational system for recording physical activity in children-preschool version. Research Quarterly for Exercise & Sport. 2006; 77(2):167–176. [PubMed: 16898273]

Dugan SA. Exercise for preventing childhood obesity. Physical Medicine & Rehabilitation Clinics of North America. 2008; 19(2):205–216. [PubMed: 18395644]

DuRant RH, Baranowski T, Johnson M, Thompson WO. The relationship among television watching, physical activity, and body composition of young children. Pediatrics. 1994; 94(4):449. [PubMed: 7936851]

DuRant RH, Baranowski T, Puhl J, Rhodes T, Davis H, Greaves KA, et al. Evaluation of the children's activity rating scale (CARS) in young children. Medicine & Science in Sports & Exercise. 1993; 25(12):1415. [PubMed: 8107551]

Finn KJ, Specker B. Comparison of actiwatch activity monitor and children's activity rating scale in children. Medicine and Science in Sports and Exercise. 2000; 32(10):1794–1797. [PubMed: 11039655]

Hill JO, Wyatt HR. Role of physical activity in preventing and treating obesity. Journal of Applied Physiology. 2005; 99(2):765. [PubMed: 16020440]

Jago R, Baranowski T, Thompson D, Greaves K, Baranowski J. Sedentary behavior, not TV viewing, predicts physical activity among 3-to 7-year-old children. Pediatric Exercise Science. 2005; 17(4): 364.

Kaphingst KM, Story M. Child care as an untapped setting for obesity prevention: State child care licensing regulations related to nutrition, physical activity, and media use for preschool-aged children in the united states. Preventing Chronic Disease. 2009; 6(1):A11. [PubMed: 19080017]

Klesges RC, Eck LH, Hanson CL, Haddock CK. Effects of obesity, social interactions, and physical environment on physical activity in preschoolers. Health Psychology. 1990; 9(4):435–449. [PubMed: 2373068]

Louie L, Chan L. The use of pedometry to evaluate the physical activity levels among preschool children in Hong Kong. Early Child Development and Care. 2003; 173(1):97–107.

Luepker RV, Perry CL, McKinlay SM, Nader PR, Parcel GS, Stone EJ, et al. Outcomes of a field trial to improve children's dietary patterns and physical activity: The child and adolescent trial for cardiovascular health (CATCH). JAMA. 1996; 275(10):768. [PubMed: 8598593]

Lutfiyya MN, Garcia R, Dankwa CM, Young T, Lipsky MS. Overweight and obese prevalence rates in African American and Hispanic children: An analysis of data from the 2003-2004 national survey of children's health. The Journal of the American Board of Family Medicine. 2008; 21(3):191.

McKenzie, TL. System for observing fitness instruction time (SOFIT) procedures manual. San Diego, CA: 2005.

McKenzie TL, Sallis JF, Armstrong CA. Association between direct observation and accelerometer measures of children's physical activity during physical education and recess. Medicine and Science in Sports and Exercise. 1994; 26(5)

McKenzie TL, Sallis JF, Nader PR. SOFIT: System for observing fitness instruction time. Journal of Teaching in Physical Education. 1991; 11:195–205.

McKenzie TL, Sallis JF, Nader PR, Broyles SL, NELSON JA. Anglo-and mexican-american preschoolers at home and at recess: Activity patterns and environmental influences. Journal of Developmental & Behavioral Pediatrics. 1992; 13(3):173. [PubMed: 1613112]

McKenzie TL, Sallis JF, Nader PR, Patterson TL, Elder JP, Berry CC, et al. BEACHES: An observational system for assessing children's eating and physical activity behaviors and associated events. Journal of Applied Behavior Analysis. 1991; 24(1):141–151. [PubMed: 2055797]

McKenzie TL, Marshall SJ, Sallis JF, Conway TL. Student activity levels, lesson context, and teacher behavior during middle school physical education. Research Quarterly for Exercise and Sport. 2000; 71(3):249–259. [PubMed: 10999262]

McKenzie TL, Sallis JF, Elder JP, Berry CC, Hoy PL, Nader PR, et al. Physical activity levels and prompts in young children at recess: A two-year study of a bi-ethnic sample. Research Quarterly for Exercise and Sport. 1997; 68(3):195–202. [PubMed: 9294873]

National Association for Sport and Physical Education. Active start: A statement of physical activity guidelines for children birth to five years. NASPE; Reston, VA: 2002.

Nevill AM, McKee DP, Boreham CAG, Murphy MH. Validation of the digiwalker pedometer for measuring physical activity in young children. Pediatric Exercise Science. 2005; 17(4):345.

Noland M, Danner F, DeWalt K, McFadden M, Kotchen JM. The measurement of physical activity in young children. Research Quarterly for Exercise and Sport. 1990; 61(2):146–153. [PubMed: 2094925]

Ogden CL, Carroll MD, Flegal KM. High body mass index for age among US children and adolescents, 2003-2006. JAMA. 2008; 299(20):2401. [PubMed: 18505949]

O'Hara NM, Baranowski T, Simons-Morton BG, Wilson BS, Parcel G. Validity of the observation of children's physical activity. Research Quarterly for Exercise and Sport. 1989; 60(1):42–47. [PubMed: 2489824]

Oliver M, Schofield GM, Kolt GS. Physical activity in preschoolers: Understanding prevalence and measurement issues. Sports Medicine. 2007; 37(12):1045–1070. [PubMed: 18027993]

Oliver M, Schofield GM, Kolt GS, Schluter PJ. Pedometer accuracy in physical activity assessment of preschool children. Journal of Science and Medicine in Sport. 2007; 10(5):303–310. [PubMed: 16911872]

Pate RR, Almeida MJ, McIver KL, Pfeiffer KA, Dowda M. Validation and calibration of an accelerometer in preschool children. Obesity (Silver Spring, Md.). 2006; 14(11):2000–2006.

Pate RR, McIver K, Dowda M, Brown WH, Addy C. Directly observed physical activity levels in preschool children. Journal of School Health. 2008; 78(8):438–444. [PubMed: 18651931]

Pate RR, Pfeiffer KA, Trost SG, Ziegler P, Dowda M. Physical activity among children attending preschools. Pediatrics. 2004; 114(5):1258. [PubMed: 15520105]

Pate RR, O'Neill JR, Mitchell J. Measurement of physical activity in preschool children. Medicine & Science in Sports & Exercise. 2010; 42(3):508–512. [PubMed: 20068498]

Perry CL, Sellers DE, Johnson C, Pedersen S, Bachman KJ, Parcel GS, et al. The child and adolescent trial for cardiovascular health (CATCH): Intervention, implementation, and feasibility for elementary schools in the United States. Health Education & Behavior. 1997; 24(6):716–735. [PubMed: 9408786]

Perry CL, Stone EJ, Parcel GS, Ellison RC, Nader PR, Webber LS, et al. School-based cardiovascular health promotion: The child and adolescent trial for cardiovascular health (CATCH). Journal of School Health. 1990; 60(8):406–413. [PubMed: 2255198]

Pope RP, Coleman KJ, Gonzalez BF, Heath EM. Validating SOFIT during physical education in a predominantly Hispanic community. Medicine & Science in Sports & Exercise. 2000; 32(5):328. Supplement.

Puhl J, Greaves K, Hoyt M, Baranowski T. Children's activity rating scale (CARS): Description and calibration. Research Quarterly for Exercise and Sport. 1990; 61(1):26–36. [PubMed: 2091164]

Rippe JM, Hess S. The role of physical activity in the prevention and management of obesity. Journal of the American Dietetic Association. 1998; 98(10):S31–S38. [PubMed: 9787734]

Roberts SO. The role of physical activity in the prevention and treatment of childhood obesity. Pediatric Nursing. 2000; 26(1):33–41. [PubMed: 12026314]

Rowe PJ, Schuldheisz JM, van der Mars H. Measuring physical activity in physical education: Validation of the SOFIT direct observation instrument for use with first to eighth grade students. Pediatric Exercise Science. 1997; 9(2):136–149.

Sharma SV, Chuang R, Hedberg AM. Pilot testing CATCH early childhood: A preschool-based program aimed at promoting healthy nutrition and physical activity among 3 to 5 year old children enrolled in head start. American Journal of Health Education. In press.

Sirard JR, Trost SG, Pfeiffer KA, Dowda M, Pate RR. Calibration and evaluation of an objective measure of physical activity in preschool children. J Phys Act Health. 2005; 2(3):345–357.

Timmons BW, Naylor PJ, Pfeiffer KA. Physical activity for preschool children-how much and how? Applied Physiology, Nutrition, and Metabolism. 2007; 32(S2E):122–134.

Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. Family Medicine. 2005; 37(5):360–363. [PubMed: 15883903]

Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: The evidence. Canadian Medical Association Journal. 2006; 174(6):801. [PubMed: 16534088]

Table 1

Characteristics of study participants in Phase I, fall 2008 (n=67) and Phase II, fall 2009 (n=27).

Variables	Phase I (N=67)	Phase II (N=27)
	N (%)	N (%)
Child gender		
Male	27 (40.3)	11 (40.7)
Female	40 (59.7)	16 (59.3)
Child's Ethnicity		
African-American	23 (34.8)	6 (23.1)
Hispanic	42 (63.6)	17 (65.4)
Other (White/Native American)	1 (1.5)	2 (7.7)
Child age		
3-4 years	11 (19.0)	17 (70.8)
4-5 years	33 (56.9)	7 (29.2)
	14 (24.1)	0
5+ years		
Child age (in months)		
Mean (SD)	52.72 (6.4)	46.04 (6.2)

Table 2

Pretest and posttest means, standard deviation and range for interobserver agreement (IOA) and kappa scores for physical activity level and type measured using SOFIT-P, Phase I (fall 2008).

	Pretest (30 observations)				Posttest (16 observations)			
	Mean	*SD*	*Min*	*Max*	*Mean*	*SD*	*Min*	*Max*
Physical Activity Level								
IOA	0.78	0.085	0.63	0.87	0.92	0.028	0.89	0.95
Kappa	0.69	0.106	0.49	0.81	0.86	0.040	0.83	0.91
Physical Activity Type								
IOA	0.79	0.072	0.68	0.74	0.78	0.097	0.65	0.90
Kappa	0.69	0.125	0.53	0.88	0.66	0.140	0.49	0.85

Table 3

Percent time spent in each of four physical activity intensity levels by center and classroom (34 observations), Phase I (fall 2008)

	Light	Moderate	Vigorous	Mod-Vig
Total	77.07	16.46	6.47	22.93
Center D	76.90	15.11	7.99	23.10
• Class 1	71.63	16.54	11.83	28.37
• Class 2	78.71	14.13	7.16	21.29
• Class 3	81.54	14.25	4.21	18.46
Center S	77.27	17.98	4.75	22.73
• Class 1	71.64	23.30	5.06	28.36
• Class 2	86.66	9.11	4.23	13.33

[a]number of observations excludes double observations conducted for reliability check.

Table 4

Percent time spent in each of four physical activity intensity levels by center and classroom; indoors and outdoors, Phase I (fall 2008)

	indoor (24 observations)				outdoor (10 observations)			
	Light	***Moderate***	***Vigorous***	***Mod-Vig***	***Light***	***Moderate***	***Vigorous***	***Mod-Vig***
Total	86.34	11.91	1.74	13.65	54.82	27.37	17.81	45.18
Center D	84.20	12.68	3.12	15.80	62.29	19.96	17.74	37.70
• Class 1	81.46	14.89	3.65	18.54	47.04	20.67	32.29	52.96
• Class 2	84.12	12.41	3.47	15.88	70.60	16.71	12.69	29.40
• Class 3	87.69	10.11	2.20	12.31	69.24	22.51	8.24	30.75
Center S	88.49	11.14	0.36	11.50	43.61	38.49	17.90	56.39
• Class 1	83.08	16.59	0.33	16.92	44.93	38.96	16.11	55.07
• Class 2	96.07	3.51	0.42	3.93	39.66*	37.07*	23.27*	60.34*

*only one observation

Table 5

SOFIT-P point biserial correlations (r_{pb}) among selected categories), Phase I (fall 2008)

Lesson Context	Category	r_{pb}	*p*
Indoor/Outdoor	% Lying down Time	-.133	NS
(1=indoor; 2=outdoor)	% Sitting Time	-.641	<0.001
	% Standing Time	.053	NS
	% Walking Time(Moderate)	.581	<0.001
	% Fast/very active Time (Vigorous)	.801	<0.001
	% MVPA Time	.779	<0.001
	% Sedentary Time	-.779	<0.001

Note. N=44 direct observations; NS= not significant; %=percent % Sedentary time= (% Lying down + % sitting + % standing) time

Table 6

SOFIT-P point biserial correlations (r_{pb}) between physical activity type and level, Phase I (fall 2008)

Activity Type	Category	r_{pb}	p
Indoor –Lie down/sit/squat/stand	% Sedentary Time	.523	0.005
	% Walking Time(Moderate)	-.551	0.003
	% Fast/very active Time (Vigorous)	-.149	NS
	% MVPA Time	-.523	0.005
Indoor –Walk/ride	% Sedentary Time	-.852	<0.001
	% Walking Time(Moderate)	.805	<0.001
	% Fast/very active Time (Vigorous)	.472	0.01
	% MVPA Time	.852	<0.001
Indoor –run/roll/rough & tumble	% Sedentary Time	-.297	NS
	% Walking Time(Moderate)	.163	NS
	% Fast/very active Time (Vigorous)	.455	0.02
	% MVPA Time	.297	NS
Outdoor –Lie down/sit/squat/stand	% Sedentary Time	.934	<0.001
	% Walking Time(Moderate)	-.777	0.001
	% Fast/very active Time (Vigorous)	-.596	0.019
	% MVPA Time	-.934	<0.001
Outdoor –Walk/ride	% Sedentary Time	-.740	0.002
	% Walking Time(Moderate)	.960	<0.0019
	% Fast/very active Time (Vigorous)	.042	NS
	% MVPA Time	.740	0.002
Outdoor –run/roll/rough & tumble	% Sedentary Time	-.604	0.017
	% Walking Time(Moderate)	.490	NS
	% Fast/very active Time (Vigorous)	.401	NS
	% MVPA Time	.604	0.017

Note. N=27 indoor, 15 outdoor; NS= not statistically significant; %=percent % Sedentary time= (% Lying down + % sitting + % standing) time