

Measuring Physical Activity in Children With Pedometers: Reliability, Reactivity, and Replacement of Missing Data

David A. Rowe, Matthew T. Mahar, Thomas D. Raedeke,
and Joanna Lore

The study was undertaken to evaluate (a) the reliability of pedometer data and reactivity of children to wearing a pedometer, (b) the effectiveness of a missing data replacement procedure, and (c) the validity of the Leisure Time Exercise Questionnaire (LTEQ). Six days of pedometer data were collected from 299 middle-school children, followed by administration of the LTEQ. Six days of pedometer data were found to be adequately reliable for research into habitual physical activity ($R_{xx} = .79$) and no reactivity occurred. Inclusion of weekday and weekend scores is recommended where possible. The individual-centered data-replacement procedure did not adversely affect reliability, so this data-replacement method offers great promise to physical activity researchers who wish to maintain statistical power in their studies. The LTEQ does not appear to measure physical activity similarly to pedometers ($r = .05$), and researchers should use the LTEQ with caution in children until further research explains this discrepancy.

The need to develop appropriate physical activity measures for use with children is a recognized concern (12). Welk, Corbin, and Dale (30) described the unique nature of children's physical activity that necessitates separate validity and reliability evidence for instruments used in this population. A common method of objectively measuring physical activity is the pedometer. Pedometers offer several attractive features such as unobtrusiveness, simplicity of operation, and low cost, all of which make the pedometer especially suitable for large-scale studies investigating children's school-time and leisure-time physical activity (27). Total pedometer steps correlate moderately well with oxygen consumption in children. Eston, Rowlands, and Ingledew (10) investigated validity of the pedometer, uniaxial accelerometer, triaxial accelerometer, and heart rate monitor during various regulated, (e.g., jogging, walking) and unregulated (e.g., playing catch, crayoning) activities in 30 children. Standard errors of estimate for predicting oxygen consumption were lower for the pedometer than for the uniaxial accelerometer and heart rate monitor, but higher than for the triaxial accelerometer.

Rowe, Mahar, and Raedeke are with the Department of Exercise and Sport Science, East Carolina University, Greenville NC 27858. Lore is with the YWCA, 1012 Oberlin Rd., Raleigh NC 27605.

Because physical activity habits of children vary from day to day, it is important to determine the number of days of data needed to obtain reliable (representative) scores. Trost (26) recommended that accelerometer data be collected over 4 or 5 days in children (Grades 1–6) and over 8 or 9 days in adolescents (Grades 7–12) in order to obtain reliability coefficients above $R = .80$. From data in a broad range of ages (7–15 years), Janz, Witt, and Mahoney (17) recommended 4 days of data collection to meet the same standard when using accelerometers.

Another reliability concern with children's pedometer data that has been discussed extensively in the literature but rarely investigated is the occurrence of reactivity. Reactivity arises when the testing process influences behavior. In the context of pedometer data collection, it has been suggested that children will increase steps above their normal levels during the initial testing period, for example by jumping up and down or walking and running more than usual, to see the effect on the step-count display. This would be reflected in a higher step count on early days compared with subsequent days. On the basis of one study, it appears that reactivity does not occur when pedometers are sealed so that children cannot see the display (28). If daily data are to be recorded, however, unsealing and resealing pedometers places an added burden on the data collector. To date, no researchers have investigated the reactivity of children to wearing unsealed pedometers.

The use of motion sensors in children outside of controlled laboratory settings poses logistical problems such as losing, breaking, or forgetting to wear the motion sensor, forgetting to record the score, discomfort, and public embarrassment (7). This could result in large amounts of missing data. Methods of replacing missing data have received limited attention in the physical activity literature. In addition, the question of identifying and dealing with outliers has not been addressed in this literature to our knowledge.

The challenges associated with collecting objective data over several days often lead researchers to use retrospective self-report questionnaires in children. Although more convenient for the researcher, the accuracy of this method has been questioned because of the cognitive challenge involved, especially for younger children. The Leisure Time Exercise Questionnaire (LTEQ), developed by Godin and Shephard (13), was originally designed for use with adults but has been used in several studies of children's habitual physical activity (5,11,14,16,21). Although there is extensive evidence of the validity of the LTEQ for use in adults, only limited published validity evidence exists for its use in children (9,24).

This study had three purposes. The first was to determine an appropriate measurement schedule (i.e., minimum number of days needed for acceptable reliability) for collecting whole-day pedometer data in middle-school children. The second was to assess the effect of a missing data replacement method on the reliability of pedometer data. The third was to evaluate convergent validity evidence for the LTEQ in middle-school children.

Method

Participants

Participants were recruited from Grade 5 ($n = 43$), Grade 6 ($n = 84$), Grade 7 ($n = 84$), and Grade 8 ($n = 88$) at two middle schools in eastern North Carolina (i.e., children ranging from 10 to 14 years of age). School A had an enrollment of 442

students, drawn from a rural town with a population of 3,538, and School B had an enrollment of 896, drawn from two rural towns with populations of 4,433 and 2,307. Informed-consent forms were sent to parents of all children in School A and approximately half of all classes in School B, and children who returned affirmative parental consent were invited to participate. The resulting sample sizes were 148 from School A and 151 from School B (i.e., approximately a 33% response rate across both schools). Children also gave verbal assent before being included in the study and were paid \$10 after they had finished participating. The university Institutional Review Board and the local school board approved all procedures before any data was collected.

Instruments

The Yamax Digi-Walker SW-200 pedometer (Yamax Corp., Tokyo, Japan) was used throughout the study. This small (2.0 in. x 1.5 in. x .75 in.), lightweight (.75 oz.) pedometer is worn on a waistband or belt. The SW-200 measures total steps and has a button for resetting counts to zero. A visual display shows current step count. Of the various makes and models of pedometers available, the Yamax Digi-Walker is widely recognized as being one of the most accurate and reliable. In a series of studies by Bassett and colleagues (4,8,25), the Yamax Digi-Walker was found to be among the most accurate (correctly counting the number of steps and estimating distance walked) and reliable (high interinstrument agreement) of 15 different makes and models of pedometer. In a controlled shake test, the Yamax Digi-Walker exhibited less than 3% error, confirming the quality criterion for technical accuracy applied by its manufacturer (29). Additionally, in a school setting we have determined that the interinstrument reliability of the Yamax was high ($R = .96$) over 5 days of data collection in elementary schoolchildren (2). Yamax pedometers have thus gained general acceptance among physical activity researchers and appear to be the most commonly used pedometers in recently published studies.

Godin and Shephard (13) developed the LTEQ as a simple questionnaire measure of habitual exercise behavior in adults. Respondents indicate the number of discrete 15-min bouts of exercise of varying intensity (mild, moderate, strenuous) engaged in during a typical 7-day period. A composite score (LTEQ METs) is then calculated using the following formula: $(\text{mild} \times 3) + (\text{moderate} \times 5) + (\text{strenuous} \times 9)$. This formula assumes energy-expenditure values of 3, 5, and 9 METs for mild, moderate, and strenuous exercise, respectively.

Procedures

Data were collected during the spring semester between February and April. In order to facilitate a manageable data-collection procedure, two to three classes were tested at a time. For each class, the pedometer-testing period lasted one week, from Wednesday to Wednesday. Pedometers were given out on the first Wednesday morning, with an explanation of how to operate the reset button and how to wear the pedometers, and children were allowed to shake the pedometers to see how they worked. Data-recording procedures were also explained at that time.

Two slightly different data-collection procedures were used, to suit the different organization structure in the two schools. In School A, children self-recorded their pedometer steps each school morning in the homeroom. Every child had a

record card that was kept by the homeroom teacher. Homeroom teachers gave out the pedometer record cards every morning and instructed children to record their pedometer count, press the reset button, and return the pedometer record card. In School B, children reported to the school gym every morning where physical education teachers recorded the pedometer scores on each child's record card. On weekends, children in both schools took home a weekend pedometer record card; they were instructed to record their pedometer score first thing on Saturday morning and first thing on Sunday morning in order to replicate the timing of the weekday data collection. These procedures resulted in 7 days of pedometer data (Wednesday through Tuesday), although only the last 6 days of data were used in the analysis. Pedometer data from the first day (Wednesday) were not analyzed because the pedometers were distributed at varying times during the morning. Consequently, the Wednesday data did not represent a full day of steps.

If a child forgot to wear his or her pedometer, he or she was instructed to record a zero on the record card for that day, and to not wear the pedometer at all for the rest of the day (i.e., do not put it on at home later in the day.) In this situation, he or she would record his or her score the following day. During data entry, these record card entries were corrected to reflect the correct day that the pedometer was worn. All children were given reminder stickers to put on their nightstand, bathroom mirror, and exit door of their house. A simple instruction sheet was also given to parents, and brief instructions were provided on the weekend pedometer record card.

Within 1 week following completion of their pedometer data collection, each class completed the LTEQ in a controlled classroom environment. Research assistants were available to assist children who had questions or difficulties. The first author read out the introductory instructions for the LTEQ, asking the children to think about a typical week. They were then asked to estimate how many separate bouts of physical activity lasting at least 15 min they typically engage in during a week. The first author then slowly read out the descriptions for strenuous, moderate, and mild exercise. Sufficient time was allowed for children to respond to each section before moving on (i.e., children were asked if they were ready to move to the next section). The most common question that the children asked was whether prolonged bouts of exercise (e.g., playing baseball for 1 hour) should count as a single bout or multiple bouts (e.g., four bouts in the example provided). After the first few times this question was asked, an explanation was added to the verbal introductory instructions that longer bouts of exercise should still be counted as a single bout.

Treatment of Data

All data screening, manipulation, and analysis were conducted using Microsoft® Excel® 2000 (Microsoft Corporation, Redmond, WA) and SPSS® version 11.5 (SPSS, Inc., Chicago, IL). After data entry, all pedometer data-record cards and LTEQ sheets were checked against the data file and corrections were made to four record-card entries and 11 LTEQ entries. Frequencies were then run on all data to screen for outliers. Although outliers (extremely low or high pedometer scores) are a common practical problem in children's pedometer data, identification and treatment of outliers has never been addressed in the literature. Consequently, there are no current guidelines for dealing with outlying data. Unusually extreme scores can occur for many reasons, including instrument error; error in data collection,

recording, or entry; and sampling error (leading to inclusion of participants from another population, for example, a teacher, an over-age sibling, or a visiting student). An extreme score could also be a valid score obtained from an extreme member of the population (i.e., an extremely active or athletic child, or an extremely sedentary child). When outlying scores are identified, researchers can correct the score (if it is possible to determine the error), remove the score (leading to missing data), or remove and subsequently replace the score (3,15). In this study, outlying scores were removed and subsequently treated the same as missing data, that is, replaced.

Outliers can be identified either by expert or logical judgment of what constitutes an unreasonable or highly unlikely score, or by statistical methods. We arrived at upper and lower cutoffs for identifying outliers via three methods. Initially, the first and second authors discussed what would be a reasonable range of scores, based on prior data-collection experience and hypothetical scenarios of a “least active child” and “most active child.” Second, we decided that elimination of at least the most extreme 1% of data points was a practical and reasonable objective criterion, but that eliminating 5% of the data would be unreasonable, and could lead to exclusion of realistically possible (i.e., valid) scores. Because of the asymmetrical distribution of pedometer data, this criterion was determined via percentile ranks rather than by calculation of a confidence interval. Finally, for the lower limit we determined the minimum number of steps obtained in a previous study where school-day steps were obtained in a rigorously controlled school setting (2) and assumed that the least active child would add very few steps beyond those obtained in school. All three methods converged quite closely (within a few hundred steps), and we decided to use the cutoffs of 1,000 steps and 30,000 steps because these values represented the best agreement among the three methods and because rounding to the nearest 1,000 steps provided easier values to remember for use in future studies. Consequently, all pedometer scores below 1,000 and above 30,000 were deleted and treated as missing data. This resulted in the deletion of 17 low outlying scores and 13 high outlying scores, constituting 1.7% of the data.

In order to replace the missing pedometer data, a separate pedometer data file was created. Following procedures described by Kang et al. (18), data were first transposed in Microsoft Excel (i.e., rearranged so that children were columns and days were rows). The data were then copied to an SPSS data file that was split into weekdays and weekend days. Missing weekday data were replaced with the mean of a child’s own weekday counts, using the SPSS missing-data-replacement command. A similar procedure was followed for weekend data. In a simulation study, this procedure was shown to replace data more accurately than traditional group-centered methods such as replacement with the group mean or regression (18), even though some cases were missing several data points. In the current study, we did not omit any child from data replacement even if he or she had only 1 weekend day or 1 weekday of data. This was similar to the approach used by Kang et al., in which no person was omitted regardless of number of days of missing data. When data are replaced for group analysis, Kang et al. showed that individual-centered data-replacement methods were more accurate than group-centered methods. We discourage the use of a single data point, however, to replace missing data for purposes of evaluating an individual.

Simple frequencies were also used to assess the incidence of data that was missing initially. Reliability of daily pedometer counts was estimated using intraclass correlation coefficients, calculated from the two-way ANOVA model (Days \times Participants). Mean comparisons across adjacent days were made using a

repeated-measures one-way ANOVA and post hoc Fisher’s LSD with Bonferroni correction for multiple comparisons (resulting in a per-comparison alpha of $.05/k$, where k = the number of pairs of adjacent days). Convergent validity evidence between the LTEQ and pedometer scores was assessed using Pearson interclass correlation coefficients. A scattergram was also used to visually evaluate the relationship between the pedometer and LTEQ scores.

Results

From initial data screening, it was determined that there were no significant [$t(208) = |.16|$ to $|.88|$, $p \geq .05$] differences in mean step counts between the two schools for any day. There were differences, however, in the incidence of missing data. At School A, where children self-reported pedometer steps, 79 of 888 possible days of data (i.e., 9% of days) were missing. At School B, where the physical education teachers recorded pedometer scores, 49 of 906 possible data points (i.e., 5% of all days) were missing. Expressed as the percentage of children with incomplete data (fewer than 6 days of data), 49 of 148 children (33%) at School A had incomplete data, and 30 of 151 children (19%) at School B had incomplete data. Based on similar means and similar reliability analyses, the data from the two schools were combined for all subsequent analyses.

Descriptive statistics for pedometer data are presented in Table 1. Mean daily steps were more than 2,000 steps per day higher than those from one study with children of similar ages (28), and more than 2,000 steps per day lower than those from another study with children of similar ages (27). After deleting three LTEQ scores above the 99th percentile, mean LTEQ bouts were 16.6 ($SD = 10.7$, range = 0 to 65), and mean LTEQ METs were 91.0 ($SD = 63.3$, range = 0 to 405).

Reliability

Starting with the first 2 complete days (Thursday and Friday), and subsequently adding 1 day of data at a time, reliability of successive combinations of days was

Table 1 Descriptive Statistics for Daily Pedometer Steps—Originally Complete Data ($N = 210$) and Combined Data (Originally Complete Plus Replaced Data, $N = 288$)

Days	Complete data only			Complete plus replaced data		
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range
Thursday	9,713	3,955	1,869–24,149	9,618	4,189	1,051–24,149
Friday	10,693	5,290	1,001–29,278	10,752	5,417	1,001–29,278
Saturday	9,741	5,074	1,036–28,281	9,741	5,229	1,036–29,781
Sunday	8,269	5,024	1,012–26,759	8,642	5,172	1,012–29,781
Monday	8,995	3,805	1,077–21,882	9,053	4,025	1,077–21,882
Tuesday	8,619	3,583	1,044–22,212	8,770	3,910	1,044–22,212

Table 2 Pedometer Reliability Over 6 Consecutive Days of Data Collection—Originally Complete Data ($N = 210$) and Combined Data (Originally Complete Plus Replaced Data, $N = 288$)

Days	Complete data only		Complete plus replaced data	
	R_{xx}	CI ₉₅	R_{xx}	CI ₉₅
Th/F	.69	.59–.76	.72	.64–.77
Th/F/Sa	.71	.64–.78	.76	.70–.80
Th/F/Sa/Su	.75	.69–.80	.79	.75–.83
Th/F/Sa/Su/M	.77	.72–.82	.81	.77–.84
Th/F/Sa/Su/M/Tu	.79	.74–.83	.83	.80–.86
Weekdays	.76	.70–.81	.81	.77–.84
Weekend days	.59	.47–.69	.68	.59–.74

estimated. These results are presented in Table 2. As would be expected, reliability increased as the number of days of data increased. Nunnally and Bernstein (22) defined the criterion value for minimally acceptable reliability as $R \geq .70$. In the original data (prior to data replacement), this standard was attained with 3 consecutive days of data. Minimally acceptable reliability, however, is not synonymous with a desirable level of reliability. When data are intended for group analyses (e.g., research studies for examining correlations in groups, or comparing group means), Nunnally and Bernstein suggested that $R \geq .80$ is a more appropriate standard. This level of reliability was almost achieved ($R = .79$) in the original data from 6 days of data.

Reliability analyses were conducted separately for weekday and weekend days because, in some studies, physical activity behavior has systematically differed on weekdays as compared with weekend days. Mean weekday steps ($9,504 \pm 3,200$) were compared with mean weekend steps ($9,005 \pm 4,258$), using a repeated-measures one-way ANOVA and were determined to be marginally significantly different ($F[1, 209] = 3.99, p < .05$). The standardized difference—Cohen's $d = 0.13(6)$ —however, indicated a small effect for type of day (weekday vs. weekend) and that the significant result was because of a high level of statistical power in this large sample. Reliability of the 2 weekend days in the original data was $R = .59$, and for the 4 weekdays was $R = .76$.

Regarding the question of reactivity, a significant repeated-measures one-way ANOVA omnibus test ($F[5, 1045] = 13.22, p < .05$) indicated that there was at least one significant mean difference among the 6 days of data. Mean steps increased significantly ($p < .05$) from Thursday to Friday, and decreased significantly ($p < .05$) from Saturday to Sunday, but no other adjacent means were significantly different. In summary, although there were significant changes between some days, there was no consistent pattern reflecting the hypothetical reactivity of children (i.e., initial decrease followed by leveling off). In comparison to a previous study

reporting no reactivity with sealed pedometers (28), the range of day-to-day differences in the current study (377 to 1,472 steps) was similar to the difference in the previous study (210 to 1,800 steps).

Missing Data Replacement

In the original data set, 128 data points (7.1% of the data) were missing and 30 outlying data points were removed (1.7%), resulting in 158 of 1,794 missing data points (8.8% of the data). Following the individual-centered missing data replacement, all except 11 children had complete data (these 11 children were either missing both weekend days or all weekdays, preventing the missing data replacement procedure). Two analyses were conducted to determine whether the replacement of the data altered the reliability or means of the data. First, the mean step count for children whose missing data were replaced ($n = 78$, $M \pm SD = 9,675 \pm 4,234$) was not significantly different [$t(286) = .47$, $p \geq .05$] from the mean for the children with complete data ($n = 210$, $M \pm SD = 9,338 \pm 3,155$), nor was the mean difference clinically meaningful (Cohen's $d = 0.09$). Means and standard deviations of the combined data ($n = 288$) were similar to those of the originally complete data ($n = 210$; see Table 1). In addition, reliability estimates for replaced data were slightly higher than for complete data, which would be expected because replaced data points were estimated from the remaining data points. Combining the replaced data with the complete data ($n = 288$) also increased the reliability estimates slightly (see Table 2).

Convergent Validity

Pedometers and questionnaires are often used as indicators of habitual (typical) physical activity. In order to assess the question of whether the LTEQ and pedometers were measuring a similar construct, Pearson correlations were calculated between 6-day step count and LTEQ METs. This analysis was conducted on 208 children with complete pedometer and LTEQ data. The correlation was low [$r(206) = .05$, $p \geq .05$]. Although we did not expect a high correlation, the extremely low correlation value was surprising. Calculating a correlation in two populations simultaneously (e.g., boys and girls) can result in a low correlation if the slope of the correlation is different in the two populations, so Pearson correlations were also calculated separately for boys and girls. The results were similar: $r(85) = .08$ and $r(119) = .03$, $p \geq .05$ in boys and girls, respectively.

Visual inspection of the scattergram of LTEQ METs and 6-day step total confirmed the weak association between the two measures and showed no bivariate outliers unduly affecting the correlation estimate. Reanalysis in the combined data set (complete data plus replaced data) resulted in similar correlations (all correlations were lower than $r = .08$).

Discussion

We investigated three aspects of pedometer data collection in middle-school children, namely (a) the reliability of pedometer data, (b) the effectiveness of a missing data replacement procedure, and (c) convergent validity with the LTEQ.

Reliability and Reactivity

Based on our results, the minimal measurement schedule for obtaining reliable pedometer data from middle-school children is 6 consecutive days, assuming that this has been preceded by a familiarization day and includes both weekend days and weekdays. This is similar to the recommendation by Vincent and Pangrazi (28), based on a different data-collection schedule (i.e., sealed pedometers, data collected on 8 weekdays over 2 consecutive weeks). In their study, 5 days of data were necessary to obtain $R \geq .80$ in children in Grades 2, 4, and 6. In another study by the same authors, 3 days of pedometer data yielded a reliability of $R = .70$ in children 6 to 12 years old (27). Three days yielded similarly reliable data in our study.

Reactivity does not appear to occur in this study even when children are able to see the pedometer display. Vincent and Pangrazi (28) also found no evidence of reactivity, but their pedometers were sealed so that children were unable to see the display. Indeed, whereas Vincent and Pangrazi reported a nonsignificant decrease from Day 1 to Day 2 of over 1,000 steps in their Grade 6 children (S.D. Vincent, personal communication, February 7, 2003), in the current study steps increased by almost 1,000 steps from the first to second day. Our findings with unsealed pedometers have important practical implications because sealing and resealing pedometers on a daily basis is not feasible in many settings with large samples.

Treatment of Missing Data and Outliers

We found that 79 out of 299 children (26%) had incomplete data over a 6-day data-collection period. The problem of missing data is pervasive when collecting motion-sensor data from children, as has been reported previously. For example, 25% of children had missing pedometer data over a 4-day data-collection period in a study of 6- to 12-year-old children (27), 26% of 7- to 11-year-old girls had missing Caltrac data over a 6-day period (24), and 44% of children in Grades 8 through 12 had missing Caltrac data (defined as fewer than 5 days of data) over a 7-day data-collection period (20). We also developed reasonable practical criteria for identifying outliers in children's daily step counts, namely fewer than 1,000 and more than 30,000 steps. Informal inspection of children's pedometer data sets from other researchers indicated that these cutoffs show promise for future use in pedometer studies in children. These outlying data points could then be treated as missing data, and replaced accordingly.

Researchers with missing data and/or outliers are faced with the choice of either omitting data during data analysis or using some method for estimating missing and deleted scores. Three problems associated with omitting the data are the consequent waste of resources, decrease in statistical power, and the possibility that children with missing data may be systematically different from children who have complete data sets (e.g., may be less motivated or have different activity patterns). Conversely, replacing missing data using a group mean or group regression approach presents a different problem in that information on a group is used to determine the score attributed to an individual (thus calling into question the appropriateness of using a child's data for individual interpretation). In the current study we demonstrated that an individual-centered data-replacement procedure resulted in slightly more reliable data, and that when this method was applied,

children who had missing data were similarly active to children who had complete data. Additionally, replacement of missing pedometer data did not alter the relationship between pedometer data and LTEQ scores. This finding has particularly important implications for physical activity researchers. The opportunity to replace missing data would increase sample size, improve the representativeness of population estimates, and increase the power of statistical significance tests.

Validity of the Leisure Time Exercise Questionnaire

In response to a previous call to validate the LTEQ with objective measures in children (23), we found that the LTEQ does not appear to measure the same construct as a pedometer. This may be because of the different reference period (actual physical activity during 6 prior days as compared with recall of a typical week), the cognitive demand of understanding the task described in the LTEQ, or the inability of middle-school children to recall physical activity accurately. Alternatively, it might simply reflect that children, who tend to be active very intermittently and for very short periods of time (1), cannot relate to the concept of 15-min activity bouts.

Only two other published studies have directly addressed this question. In one, a low correlation was obtained ($r = .10$) between LTEQ scores and 6 days of Caltrac data in a mixed group of 7- to 11-year-old female gymnasts and nongymnasts (24). In the second, Eisenmann, Milburn, Jacobsen, and Moore (9) reported a correlation with 12-hr Caltrac scores of $r = .50$ in 31 fifth-grade students. The higher convergent validity coefficient obtained by Eisenmann et al. might be explained by the fact that the LTEQ was completed on the same day as the Caltrac data were collected, whereas in the current study children completed the LTEQ within 1 to 7 days. Other convergent validity coefficients for the LTEQ were found in studies addressing other questions. For example, a correlation of $r = .24$ was reported with 7 days of Caltrac data in children in Grades 4–6 by Kowalski, Crocker, and Faulkner (19). In a similar study involving adolescents, Kowalski, Crocker, and Kowalski (20) reported a correlation of $r = .19$ with 7 days of Caltrac data. Overall, except for one study, the body of evidence indicates a minimal association between LTEQ scores and objective measures of physical activity in children and adolescents.

Limitations of the study design are a factor that should be taken into consideration when interpreting these results. Unlike other studies that mainly involved weekday data collection, we started data collection midweek and collected data over the weekend. This might make comparisons with prior studies tenuous (it should be noted that although there was a marginally significant mean difference between weekday average and weekend average, this difference was clinically unimportant). As with any convenience sample, these data may not be representative of all children. Both schools were in a semirural area of eastern North Carolina, the return rate of consent forms was approximately 33%, and volunteers may be more consistent or compliant than nonparticipants.

Conclusions and Recommendations

In conclusion, the use of unsealed pedometers with middle-school children should yield reliable data over 6 or more days, although it should be noted that a practice

day occurred before data collection in the current study and that these data included 2 weekend days. When data are missing, individual-centered data replacement does not substantially alter the nature of the group data and offers advantages both for group analyses and individual interpretation of pedometer scores. Future research is needed to further investigate the function of the LTEQ in children, and in the meantime, interpretation of LTEQ data obtained from children should be treated with caution.

References

1. Bailey, R.C., J. Olson, S.L. Pepper, J. Porszaz, T.L. Barstow, and D. M. Cooper. The level and tempo of children's physical activities: an observation study. *Med. Sci. Sports Exerc.* 27:1033-1041, 1995.
2. Barfield, J.-P., D.A. Rowe, and T. Michael. Interinstrument consistency of the Yamax Digi-Walker in elementary school children. *Meas. Phys. Educ. Exerc. Sci.* 8:109-116, 2004.
3. Barnett, V., and T. Lewis. *Outliers in Statistical Data*. 3rd ed. New York: Wiley, 1994.
4. Bassett, D.R., Jr., B.E. Ainsworth, S.R. Leggett, C.A. Mathien, J.A. Main, D.C. Hunter, et al. Accuracy of five electronic pedometers for measuring distance walked. *Med. Sci. Sports Exerc.* 28:1071-1077, 1996.
5. Chung, M.H., and D.A. Phillips. The relationship between attitude toward physical education and leisure-time exercise in high school students. *Phys. Educ.* 59(3):126-138, 2002.
6. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.
7. Crocker, P.R.E., D.R. Holowachuk, and K.C. Kowalski. Feasibility of using the Tritrac motion sensor over a 7-day trial with older children. *Pediatr. Exerc. Sci.* 13:70-81, 2001.
8. Crouter, S.E., P.L. Schneider, M. Karabalut, and D.R. Bassett, Jr. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Med. Sci. Sports Exerc.* 35:1455-1460, 2003.
9. Eisenmann, J.C., N. Milburn, L. Jacobsen, and S.J. Moore. Reliability and convergent validity of the Godin Leisure-Time Exercise Questionnaire in rural 5th-grade school-children. *J. Hum. Mov. Stud.* 43:135-149, 2002.
10. Eston, R.G., A.V. Rowlands, and D.K. Ingledeu. Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children's activities. *J. Appl. Physiol.* 84:362-371, 1998.
11. Falk, B., S. Portal, R. Tikitsky, Y. Weinstein, N. Constantini, and U. Martinowitz. Anaerobic power and muscle strength in young hemophilia patients. *Med. Sci. Sports Exerc.* 32:52-57, 2000.
12. Freedson, P.S., and E.L. Melanson. Measuring physical activity. In: *Measurement in Pediatric Exercise Science*. D. Docherty (Ed.). Champaign, IL: Human Kinetics, 1996, pp. 261-283
13. Godin, G., and R.J. Shephard. A simple method to assess exercise behavior in the community. *Can. J. Appl. Sport Sci.* 10:141-146, 1985.
14. Hagger, M., B. Ashford, and N. Stambulova. Russian and British children's physical self-perceptions and physical activity participation. *Pediatr. Exerc. Sci.* 10:137-152, 1998.

15. Hawkins, D.M. *Identification of Outliers*. London: Chapman and Hall, 1980.
16. Hunt, J.D. *The Impact of a Daily Physical Education Program on Students' Attitudes Towards, and Participation in, Physical Activity*. Eugene, OR: Microform Publications, International Institute for Sport & Human Performance, University of Oregon, 1995.
17. Janz, K.F., J. Witt, and L.T. Mahoney. The stability of children's physical activity as measured by accelerometry and self-report. *Med. Sci. Sports Exerc.* 27:1326-1332, 1995.
18. Kang, M., W. Zhu, C. Tudor-Locke, and B.E. Ainsworth. An experimental determination of the best missing-value recovery method in assessing physical activity using pedometers. (Abstract). *Res. Q. Exerc. Sport.* 74(1, Suppl.):A25-A26, 2003.
19. Kowalski, K.C., P.R.E. Crocker, and R.A. Faulkner. Validation of the Physical Activity Questionnaire for older children. *Pediatr. Exerc. Sci.* 9:175-186, 1997.
20. Kowalski, K.C., P.R.E. Crocker, and N.P. Kowalski. Convergent validity of the Physical Activity Questionnaire for adolescents. *Pediatr. Exerc. Sci.* 9:342-352, 1997.
21. McGuire, M.T., D.R. Neumark-Sztainer, and M. Story. Correlates of time spent in physical activity and television viewing in a multi-racial sample of adolescents. *Pediatr. Exerc. Sci.* 14:75-86, 2002.
22. Nunnally, J., and I.H. Bernstein. *Psychometric Theory*. 3rd ed. Boston: McGraw-Hill, 1994.
23. Sallis, J.F., M.J. Buono, J.J. Roby, F.G. Micale, and J.A. Nelson. Seven-day recall and other physical activity self-reports in children and adolescents. *Med. Sci. Sports Exerc.* 25:99-108, 1993.
24. Scerpella, T.A., P. Tuladhar, and J.A. Kanaley. Validation of the Godin-Shepard questionnaire in prepubertal girls. *Med. Sci. Sports Exerc.* 34:845-850, 2002.
25. Schneider, P.K., S.E. Crouter, O. Lukajic, and D.R. Bassett, Jr. Accuracy and reliability of 10 pedometers for measuring steps over a 400-m walk. *Med. Sci. Sports Exerc.* 35:1779-1784, 2003.
26. Trost, S.G. Objective measurement of physical activity in youth: Current issues, future directions. *Exerc. Sport Sci. Rev.* 29:32-36, 2001.
27. Vincent, S.D., and R.P. Pangrazi. An examination of the activity patterns of elementary school children. *Pediatr. Exerc. Sci.* 14:432-441, 2002.
28. Vincent, S.D., and R.P. Pangrazi. Does reactivity exist in children when measuring activity levels with pedometers? *Pediatr. Exerc. Sci.* 14:56-63, 2002.
29. Vincent, S.D., and C.L. Sidman. Determining measurement error in digital pedometers. *Meas. Phys. Educ. Exerc. Sci.* 7:19-24, 2003.
30. Welk, G.J., C.B. Corbin, and D. Dale. Measurement issues in the assessment of physical activity in children. *Res. Q. Exerc. Sport.* 71(2, Suppl.):59-73, 2000.

Acknowledgments

We owe special thanks to Debbie Keel and Jami Bendle, physical education teachers, for their cooperation and help with the data collection. We also acknowledge the support of David Bassett, Greg Welk, and Robert Pangrazi who allowed us access to their pedometer data sets to help in evaluating the outlier criteria. Funding was provided by the University of North Carolina Institute of Nutrition and the Centers for Disease Control and Prevention.

Copyright of Pediatric Exercise Science is the property of Human Kinetics Publishers, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.