Calibration and Validation of the Youth Activity Profile: The FLASHE Study

Pedro F. Saint-Maurice, PhD[1,2], **Youngwon Kim, PhD**[1,3], **Paul Hibbing, MS**[1], **April Oh, PhD, MPH**[4], **Frank M. Perna, PhD**[5], and **Gregory J. Welk, PhD**[1]

[1]Department of Kinesiology, Iowa State University, Ames, Iowa [2]School of Psychology, University of Minho, Braga, Portugal [3]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom [4]Health Communication and Informatics Research Branch, Behavioral Research Program, National Cancer Institute, Bethesda, Maryland [5]Health Behaviors Research Branch, Behavioral Research Program, National Cancer Institute, Bethesda, Maryland

Abstract

Introduction—This study describes the calibration and validity of the Youth Activity Profile (YAP) for use in the National Cancer Institute's Family Life, Activity, Sun, Health, and Eating (FLASHE) study. The calibrated YAP was designed to estimate minutes of moderate to vigorous physical activity (MVPA) and sedentary behavior (SB).

Methods—The YAP was calibrated/validated in adolescents (aged 12–17 years) using cross-sectional data from the FLASHE study. Participants wore a GT3X+ ActiGraph on the dominant wrist for 7 days and then completed the YAP. Calibration was conducted for school (*n*=118), out of school (*n*=119), weekend (*n*=61), and SB (*n*=116) subsections of the YAP and by regressing percentage time in MVPA/SB (%MVPA/%SB) on each respective YAP subsection score, age, and the interaction between these two. The final algorithms were applied to independent samples (*n*=39–51) to examine validity (median absolute percentage error, equivalence testing).

Results—The final algorithms explained 15% (school), 16% (out of school), and 12% (weekend) of the variability in GT3X+ %MVPA and 7% of the variability in GT3X+ %SB. The calibrated algorithms were applied to independent samples and predicted GT3X+ minutes of MVPA/SB with median absolute percentage error values ranging from 12.5% (SB section) to 32.5% (weekend section). Predicted values obtained from the YAP were within 10%–20% of those produced by the GT3X+.

Address correspondence to: Pedro F. Saint-Maurice, PhD, National Cancer Institute, 9609 Medical Center Drive Room-6E572, Rockville MD 20850. pedro.saintmaurice@nih.gov.

Conclusions—The YAP-predicted minutes of MVPA/SB resulted in similar group estimates obtained from an objective measure. The YAP offers good utility for large-scale research projects to characterize PA/SB levels among groups of youth.

INTRODUCTION

A variety of measures have been used to assess physical activity (PA) levels in youth but they each have inherent advantages and disadvantages.[1] Methods found to be the most accurate also tend to be the most expensive or complicated to use.[2] Therefore, this interplay among accuracy, feasibility, and scope of the measure typically determines the selection of the most appropriate tool for a given research application.[3–5]

Questionnaires measuring PA are known to be susceptible to a variety of sources of error, yet they provide key advantages for many types of PA research.[6–8] Questionnaires are particularly well suited for large-scale applications at the national or community level because they provide detailed profiles of activity behavior while capturing important contextual information (e.g., type, location). A vast array of PA questionnaires is available in the literature for assessing youth; however, only a few have been found to have acceptable measurement properties.[7,9] This can be partially explained by the combination of the unique patterns of youth activity[10] and also the limitations in the design of the instruments as well as the lack of context-related cues to facilitate accurate recall.[11] For example, a review from Biddle and colleagues[9] identified only three questionnaires of PA (of 89 total) as having potential for refinement and use in large-scale applications, and two of these (i.e., the Youth Risk Behavior Surveillance Survey and Physical Activity Questionnaire for Children/ Adolescents) include reference to context in their design (i.e., context-related items). A limitation of these two instruments though is that neither can be readily converted into meaningful outcome measures. The Physical Activity Questionnaire for Children/ Adolescents, for example, uses a simple 5-point scale with narrative choices ranging from *rarely* to *often* to capture levels of activity in different settings. The items capture useful information about PA levels but cannot be used to estimate time spent in moderate to vigorous PA (MVPA) or adherence to public health guidelines.

The authors developed the Youth Activity Profile (YAP) to specifically address these limitations. A unique advantage of the YAP is that it was designed to facilitate calibration of both reported PA and sedentary behavior (SB) into more-accurate and -useable estimates. The calibration of SB is a unique feature of the YAP, considering that the availability of related measures for youth populations are still quite limited.[12,13] Calibration is a standard step in the process used to obtain meaningful information from accelerometry-based activity monitors and it offers the same potential for improving the accuracy and utility of report-based measures. Through calibration, a crude estimate from a report-based measure such as the YAP can be re-scaled to replicate estimates of PA behavior obtained by more objective monitor-based methods.[5,8,14,15] Calibration of report-based measures is especially important for surveillance work and other large-scale applications because it helps to minimize the large disparities that have commonly been observed when comparing estimates with monitor-based measures.[16,17] The calibration approach has been tested with the YAP and demonstrated that predicted values and accelerometer estimates of MVPA were correlated

by 0.19–0.75 (depending on the period of the day) and were deemed equivalent at the 10%–30% level. However, this study was done with a sample of 291 youth from Midwest U.S.[18,19]; therefore, additional work is needed to determine if the YAP calibration approach holds in a different sample and independently of the activity monitor/measurement methodology used.

The present study addresses this need by evaluating the predictive utility of the YAP algorithms in a sample of adolescents involved in the National Cancer Institute's Family Life, Activity, Sun, Health, and Eating (FLASHE) study. The authors specifically calibrated the YAP against wrist-worn ActiGraph measures of MVPA and SB and validated the algorithms in an independent sample also enrolled in FLASHE.

METHODS

Design and Participants

The FLASHE study (http://cancercontrol.cancer.gov/brp/hbrb/flashe.html) was a cross-sectional national study focused on assessing diet and PA behaviors (and their correlates) among adolescent children (aged 12–17 years). The calibration/validation of the YAP was conducted in a randomly selected subsample of adolescents in FLASHE (*n*=628). Participants were asked to wear an ActiGraph GT3X+ on the dominant wrist for 7 full days (waking hours and sleep) and complete (self-administered) the YAP via a web-based format at the end of the measurement protocol so that data were temporally matched. The ActiGraph GT3X+ was mailed to participants along with a daily wear log that asked about school start and end times, and non-wear periods. The FLASHE was reviewed and approved by the U.S. Government's Office of Management and Budget, National Cancer Institute's Special Studies IRB, and Westat's IRB.[21,22]

Data Processing

Accelerometer data were processed as described below and later merged to provide a comprehensive examination of the YAP. The YAP includes 15 items and was designed to provide estimates of MVPA at school (five items) and both MVPA (five items) and SB (five items) out of school (Appendix 1). The MVPA estimate for out-of-school time is further segmented in MVPA accumulated during the week and weekend to provide additional information.

Data obtained from the GT3X+ were processed separately (i.e., school, out of school, and weekend) to capture MVPA/SB patterns for the same three segments used in the YAP: school, out of school, and weekend. These segments were defined based on individualized participant's schedule and extracted for processing and verification of compliance with the protocol. Activity at school measured by the GT3X+ included MVPA estimates accumulated between commuting to school (defined as 30 minutes before school start time) and commuting from school to home (i.e., 30 minutes after school end time). Activity out of school was defined as the period of time going from after school through the end of the evening (i.e., 10:00PM) and also included any activity that occurred before school, (i.e., between 6:00AM and commuting to school). Activity accumulated during the weekend was

defined to approximate the length of a weekday and included all activity that occurred between 6:00AM and 10:00PM. Time spent in SB measured during weekday out-of-school time was also computed.

Each 5-second epoch classified as being in SB (35 counts, 1.5 METs) or MVPA (361 counts, 3 METs)[23] was processed to compute relative time spent in SB ($SB_{\%/day}$) and MVPA ($MVPA_{\%/day}$). Non-wear bouts of activity were identified using the algorithm of Choi et al.[24] that was designed to flag consecutive zeroes (Window 1) during a 90-minute interval while allowing for 2 minutes of non-zero counts (i.e., artifactual movement). The algorithm also incorporates a second window of 30 minutes that looks for consecutive zeroes above and below any artifactual movement. This algorithm had not been validated for wrist location or for use among adolescents. However, a supplemental study was conducted to fill this gap and found that non-wear time using this algorithm was strongly correlated (i.e., r =0.86) with non-wear time obtained from the SenseWear Mini, a pattern recognition monitor that can automatically detect non-wear time (P Saint-Maurice et al, Iowa State University, unpublished observations, 2016). The authors also accounted for periods where the monitor was purposefully removed and that were reported by participants in the log provided in the mail. Most of these included water-based activities and sports-related events (e.g., baseball, softball) and were assigned energy expenditure values as provided in the compendium of physical activities for youth[25,26] and using standardized procedures as recommended in the literature.[27]

Key compliance criteria included having worn the accelerometer for 70% of the corresponding segment duration (e.g., school) on three different days (e.g., school days) However, the number of valid days required for the weekend was limited to 1. The separate compliance checks by segment revealed, for example, that eight participants did not provide school schedule information, 21 participants did not wear the accelerometer on at least 1 school day during 70% of the school time, and that 26 participants did not have three valid (wear time 70%) school days. The same compliance checks resulted in slightly higher exclusion rates for out-of-school and weekend segments.

Data collected from the YAP were processed similarly to the GT3X+ to obtain corresponding estimates of: (1) school (i.e., average of YAP Items 1–5); (2) out of school (i.e., average of YAP Items 6–8); and (3) weekend activity (i.e., average of YAP Items 9–10). Sedentary behavior was also computed using the average obtained from Items 11–15 (Appendix 1). Participants with missing scores on each of the sections or that did not complete the YAP within 2 days of completing the activity monitoring protocol were not included in the analyses.

The PA estimates recorded by the GT3X+ during school, out-of-school, and weekend time were merged with YAP composite scores for school, out-of-school, and weekend sections, respectively. Estimates of SB recorded during weekday out-of-school time were merged with YAP composite score for sedentary behaviors. The YAP was originally designed to characterize activity levels during the regular school year. As a considerable portion of the sample completed the PA measurement protocol during non-school days (e.g., summer break) or were not enrolled in school in the regular school system, these participants were

excluded from the study (*n* ranged from 231 to 261 participants). Preliminary analyses did not find any relevant differences between the included and excluded participants. The final samples with both valid YAP and accelerometer data resulted in 169, 170, and 130 with valid data at school, out-of-school, and weekend periods (equivalent to 16%–37% of the initial sample), respectively (Table 1).

Statistical Analysis

Calibration was first conducted in a randomly selected portion of the sample (i.e., 70% of the sample with valid data) and validated in the remaining 30% of the sample. This partition resulted in samples for calibration (validation) of 118 (51), 119 (51), and 91 (39) participants, for the school, out-of-school, and weekend segments, respectively. A distinction in methodology over past calibration work on the YAP is that it was not possible to directly calibrate each individual item prior to computing the overall indicators for school and out-of-school time because specific time periods for school periods were not available in FLASHE. In this case, the aggregated scores from the school items and out-of-school items were used to compute overall activity levels in the corresponding period. This provides a useful evaluation of whether this simplified scoring method can still provide reasonable estimates of group levels of MVPA in this population. All analyses were done using SAS, version 9.2 and the details of the calibration/validation analyses are provided below.

Calibration

Segmented YAP scores (i.e., school, out of school, and weekend) were calibrated separately for school, out of school, and weekend, resulting in three models for PA and one for SB. Each of the four models was defined by regressing age (centered at the group mean of 14.5 years), the respective YAP composite score (centered at the median value of 3.0), and the interaction between these two factors, on measured outcomes recorded by the GT3X+ $MVPA_{\%/day}$ or $SB_{\%/day}$. Because preliminary analyses demonstrated that gender was not a statistically significant predictor in any of the models, this variable was not included in the final models. The regression models were generated using quantile regression (i.e., set at the median or 50th percentile of the dependent variables) to minimize any deviation from normality in the distributions, and unstandardized β-weights ± SE associated with YAP scores were examined using Wald tests after age was entered in the model. CIs were obtained using bootstrapped SEs as a more robust measure of variability that can account for lack of normality. Statistical significance was set at α=0.05. The robustness of each model was examined by computing root mean square error (RMSE) and by examining the normality of residuals using the Shapiro–Wilk test.

Validation

The algorithms obtained from the calibration process were also validated separately for each of the school, out-of-school, and weekend segments. Agreement was examined for both individual and group level by converting YAP composite scores into weekly minutes of MVPA/SB (i.e., $MVPA_{minutes/week}$ and $SB_{minutes/week}$) using the algorithms developed in the calibration analyses (detailed procedures for computation of predicted YAP MVPA/SB in Appendix 2). The validity of each YAP section was first examined using Pearson product correlations and Appendix 3 includes additional plots to complement these results. The

authors first examined individual-level agreement using 95% limits of agreement (LOA) as described by Bland and Altman[28] and by computing the median absolute percent error (MAPE). Group-level agreement was determined using equivalence testing.[29] Agreement was established if the two 95% CI for predicted minutes of activity were within a 10% range (equivalence zone) of total week MVPA/SB values recorded by the GT3X+. When there was lack of evidence of equivalence at 10%, the equivalence zone was increased by 5% (i.e., 15%, 20%) until equivalence was reached.

RESULTS

The results are described separately for the calibration and validation phase, and for each of the key segments of the week: school, out of school, and weekend.

Calibration

Three participants were missing YAP sedentary scores and therefore the out-of-school sample used to calibrate this section and generate estimates of SB was reduced to 116 observations. The calibration model for the school segment explained 15% of the variance ($R^2=0.15$) and the RMSE was 4.8%. In other words, $MVPA_{\%/day}$ at school could be estimated with an average error of 4.8 units. The school activity score was a significant predictor of $MVPA_{\%/day}$ at school after entering age in the model ($\beta=1.24$ [SE=0.55], $p=0.02$) (Table 2). The calibration model for the out-of-school segment explained 16% of the variance in $MVPA_{\%/day}$ with an average error of 5.1 units (RMSE=5.1). The composite score obtained from YAP out-of-school items was also a significant predictor of $MVPA_{\%/day}$ after entering age in the model ($\beta=1.17$ [SE=0.58], $p=0.04$) (Table 2). The calibration model for $MVPA_{\%/day}$ during the weekend explained 12% of the variability in activity scores ($R^2=0.12$) and was estimated with an average error of 4.8 (RMSE=4.8). The weekend composite score was a significant predictor of $MVPA_{\%/day}$ recorded during the weekend ($\beta=1.30$ [SE=0.55], $p=0.02$) (Table 2).

The calibration model for SB explained 7% of the variability in the proportion of sedentary time measured by the GT3X+ ($R^2=0.07$) and with an error of 9.8 (RMSE=9.8). Contrary to the other sections, the composite score for the sedentary section was not a significant predictor of measured SB ($\beta=3.44$ [SE=2.15], $p=0.11$) (Table 2). Shapiro–Wilk tests revealed non-significance (p-values >0.05) for all the segment-specific calibration models, indicating normal distributions of residuals.

Validation

There was one participant that was missing one or more scores on the YAP SB, leaving 50 participants in the final sample. Predicted $MVPA_{minutes/week}$ at school was significantly correlated with MVPA obtained from the GT3X+ ($r=0.38$, $p<0.01$) (Appendix Figure 1A). The LOA for individual-level error ranged between −78.4% and 94.6% of GT3X+ activity measured during school time and the MAPE was 29.0%. Group-level estimates of $MVPA_{minutes/week}$ obtained from GT3X+ and YAP scores differed by −19.4 minutes or approximately −4 minutes/day (GT3X+, 240.0 [SE=24.0] minutes/week; YAP, 259.4

SE=12.4] minutes/week). The upper limit of the 95% CI for YAP scores (95% CI=247.0, 271.8) was deemed equivalent at 15% (Figure 1).

Predicted $MVPA_{minutes/week}$ accumulated during out-of-school time were moderately correlated with $MVPA_{minutes/week}$ obtained from the GT3X+ (r=0.52, p<0.01) (Appendix Figure 1B) and the LOA for individual-level error ranged between −82.3% and 84.5% of GT3X+ activity. The MAPE for this segment was 31.5%. Group-level estimates of $MVPA_{minutes/week}$ obtained from GT3X+ and YAP scores differed by −3.0 minutes or <1 minute/day (GT3X+, 270.1 [SE=27.0] minutes/week; YAP, 273.1 [SE=13.5] minutes/week). Both the upper and lower limit of the 95% CI for YAP scores (95% CI=259.6, 286.6) were within 10% of GT3X+ MVPA (10% interval=243.1, 297.1) (Figure 1).

Predicted $MVPA_{minutes/week}$ based on weekend items were not significantly correlated with $MVPA_{minutes/week}$ recorded by the GT3X+ (r=0.16, p=0.31) (Appendix Figure 1C). The LOA ranged from −70.2% to 131.6% of average GT3X+ minutes of MVPA and the MAPE was 32.5%. Predicted (185.8 [SE=14.1] minutes) and measured (203.5 [SE=20.4] minutes) $MVPA_{minutes/week}$ during the weekend differed by −17.8 minutes (approximately −9 minutes/day) and reached agreement when the equivalence interval was set at 20% (162.8–244.3 $MVPA_{minutes/week}$ measured by the GT3X+) (Figure 1).

Predicted $SB_{minutes/week}$ were significantly correlated with the GT3X+ data (r=0.32, p=0.02) (Appendix Figure 1D). The LOA ranged from −39.6% to 29.2% of average GT3X+ minutes of SB and the MAPE was 12.2%. The predicted $SB_{minutes/week}$ (1,372.0 [SE=97.6] minutes/week or 22.9 [SE=1.6] hours/week) differed from GT3X+ $SB_{minutes/week}$ (1,447.0 [SE=268.2] minutes/week or 24.1 [SE=4.5] hours/week) by −75.6 minutes (−1.2 hours/week or −15 minutes/weekday) and the associated 95% CI was within the 10% equivalence zone (Figure 2).

DISCUSSION

This study examined the utility of a streamlined calibration method for re-scaling the YAP to match objective data from wrist-worn ActiGraph data. The value of the present calibration is that the YAP estimates are designed to correspond with wrist-worn ActiGraph data that are used within segments of the FLASHE sample. The YAP and the proposed calibration algorithms can be implemented in other national or community projects where it is unfeasible to collect objective data to assess youth PA patterns. Overall, the estimated time spent in MVPA from the YAP were statistically equivalent to the device-measured estimates for both in-school and out-of-school PA (using the 10%–15% equivalence zone criteria), but agreement was lower for weekend MVPA. These findings were expected given that activity at school and out of school represent a more familiar and routinized context when compared with weekend days.[30]

The indices of validity of the YAP can be interpreted relative to other comparable self-report tools as well as against previous validation work done with the YAP. The proportion of variability in GT3X+ explained by the YAP algorithms was low, but predicted estimates were consistent with other self-report instruments in the PA domain (with r ranging from

0.30 to 0.50).[7] The designs and outcomes in other studies on self-report tools vary substantially, and as such, the performance of the YAP can be most effectively interpreted when comparing the present results to those of a similarly structured self-report. For example, the International Physical Activity Questionnaire for Adolescents (IPAQ-A) is a self-report measure that also uses context-related cues (i.e., school and out-of-school items) to ask about time spent in MVPA.[31] The IPAQ-A was evaluated in 248 European adolescents and the study[31] found correlations of 0.15 and 0.27 between minutes of moderate and vigorous PA, respectively, obtained from the IPAQ and an ActiGraph accelerometer. The IPAQ-A was found to overestimate PA measured by the ActiGraph by approximately 38 minutes/day for moderate PA and 25 minutes/day for vigorous PA (i.e., 63 minutes/day of MVPA).[31] Overall, the authors found better correlation indices for the calibrated YAP in the present study. The YAP–ActiGraph correlations ranged from 0.16 to 0.52 and the predicted minutes of MVPA obtained from the YAP differed from ActiGraph MVPA estimates by approximately 11 minutes/week (or approximately 2 minutes/day). This value is substantially lower than the 63 minutes/day found with the IPAQ-A or an equivalent 441 minutes/week difference from ActiGraph estimates for MVPA. The present study was also designed to extend previous work with the YAP by testing calibration with different monitors and methods. For example, a key difference is that the calibration was done by combining school and out-of-school items into an overall indicator prior to calibration rather than calibrating individual items as in the authors' original work.

In general, the findings from the present study were consistent with the original development work on the YAP.[19] For example, the observed MAPE values for school PA (31.5%), out-of-school PA (29.0%), weekend PA (32.5%), and SB (12.5%) were relatively similar to the previously developed YAP algorithms (school, 28.8%; out of school, 37.8%; weekend, 45.5%; SB, 17.5%; P Saint-Maurice & G Welk, Iowa State University, unpublished observations, 2015). However, it is important to note that the magnitude of individual-level error reported here is similar to that observed in the authors' prior work as well as with other self-report instruments[32] and with most calibration equations developed for accelerometry-based activity monitors.[33] The main difference in findings from this and prior validation studies of the YAP[19] was in the measurement properties associated with estimates of SB. In this study, the composite sedentary score based on YAP items was not a significant predictor of measured SB. However, the authors' past work demonstrated that this section had the strongest predictive coefficients when calibrated against the SenseWear Armband.[19]

Inherent differences between the GT3X+ and SenseWear Armband may partially explain some of the incongruences between the present calibration and the previous validation work done with the YAP. However, this is difficult to assess given the challenges associated with harmonizing different objective measures of PA.[34,35] For example, in this study, Crouter and colleagues'[23] cut points that were specifically developed for wrist output in youth were used. Surprisingly, with this criterion measure, gender was not a significant predictor of objective estimates of PA. This finding was consistent with another study in this supplement examining PA patterns in a sample of more than 400 youth.[36] Previous research using hip-worn accelerometers and pedometers has consistently shown boys to be more active than girls so it is possible that the present findings may be influenced by the unique nature of the wrist-worn monitors. The decision to process the data using the Crouter et al.[23] cut points

was supported by a parallel study conducted as part of the FLASHE evaluation. In this pilot study, the authors found that Crouter and colleagues' method had the best agreement (when compared with other wrist-based methods) with the SenseWear Armband. Estimates of SB and MVPA differed on average by 2 and 35 minutes/day, respectively. Another paper in this supplement[36] also revealed more-realistic PA patterns with the method of Crouter et al. than with the GGIR method available to process raw ActiGraph data at the wrist. Additional research is still needed to determine the most appropriate ways to process wrist-worn data, but the main point is that the YAP estimates as described in this study are linked to the unique characteristics of both the monitor (i.e., GT3X+) and data processing methods (i.e., Crouter and colleagues' method) used in the calibration.

Limitations

The availability of calibrated estimates of MVPA and SB within FLASHE presents a number of advantages for future research, but it is important to keep in mind that these values are also rough estimates. Additional limitations of the study relate to the lack of control over the PA measurement protocol that was administered from a distance. Participants were provided with detailed instructions on how to use the accelerometer for 7 days and complete the YAP on the following day so that data would be temporally matched. However, some participants completed the YAP weeks later after the completion of the 7-day protocol with the GT3X+. The authors were able to minimize this issue by excluding participants that had more than a 2-day gap between completing the accelerometer protocol and filling out the YAP. A final limitation of the present study was the inability to control for inherent error in the measures. PA is a behavior that varies naturally from day to day[37,38] and from season to season[39] within an individual. This variability represents random error that can lead to the attenuation of statistical measures of effect (e.g., correlation and regression coefficients, ORs).[40,41] Studies have demonstrated the attenuation of relative risk estimates from self-report measures when used in epidemiology studies.[42,43] The incorporation of measurement error models prior to calibration of the YAP would likely help to reduce error and improve precision for both individual- and group-level estimates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. Prog Cardiovasc Dis. 2015; 57(4):387–395. https://doi.org/10.1016/j.pcad.2014.10.005. [PubMed: 25446555]

2. Corder K, Ekelund U, Steele RM, Wareham NJ, Brage S. Assessment of physical activity in youth. J Appl Physiol. 2008; 105(3):977–987. https://doi.org/10.1152/japplphysiol.00094.2008. [PubMed: 18635884]

3. Strath SJ, Kaminsky LA, Ainsworth BE, et al. Guide to the assessment of physical activity: Clinical and research applications: a scientific statement from the American Heart Association. Circulation. 2013; 128(20):2259–2279. https://doi.org/10.1161/01.cir.0000435708.67487.da. [PubMed: 24126387]

4. Pettee Gabriel KK, Morrow JR Jr, Woolsey AL. Framework for physical activity as a complex and multidimensional behavior. J Phys Act Health. 2012; 9(Suppl 1):S11–18. https://doi.org/10.1123/jpah.9.s1.s11. [PubMed: 22287443]

5. Troiano RP, Pettee Gabriel KK, Welk GJ, Owen N, Sternfeld B. Reported physical activity and sedentary behavior: why do you ask? J Phys Act Health. 2012; 9(Suppl 1):S68–75. https://doi.org/10.1123/jpah.9.s1.s68. [PubMed: 22287450]

6. Haskell WL. Physical activity by self-report: a brief history and future issues. J Phys Act Health. 2012; 9(Suppl 1):S5–10. https://doi.org/10.1123/jpah.9.s1.s5. [PubMed: 22287448]

7. Chinapaw MJ, Mokkink LB, van Poppel MN, van Mechelen W, Terwee CB. Physical activity questionnaires for youth: a systematic review of measurement properties. Sports Med. 2010; 40(7): 539–563. https://doi.org/10.2165/11530770-000000000-00000. [PubMed: 20545380]

8. Nusser SM, Beyler NK, Welk GJ, Carriquiry AL, Fuller WA, King BM. Modeling errors in physical activity recall data. J Phys Act Health. 2012; 9(Suppl 1):S56–67. https://doi.org/10.1123/jpah.9.s1.s56. [PubMed: 22287449]

9. Biddle SJ, Gorely T, Pearson N, Bull FC. An assessment of self-reported physical activity instruments in young people for population surveillance: Project ALPHA. Int J Behav Nutr Phys Act. 2011; 8:1. https://doi.org/10.1186/1479-5868-8-1. [PubMed: 21194492]

10. Welk GJ, Corbin CB, Dale D. Measurement issues in the assessment of physical activity in children. Res Q Exerc Sport. 2000; 71(2 Suppl):S59–73. https://doi.org/10.1080/02701367.2000.11082788. [PubMed: 10925827]

11. Baranowski T, Domel SB. A cognitive model of children's reporting of food intake. Am J Clin Nutr. 1994; 59(1 Suppl):212S–217S. [PubMed: 8279427]

12. Kang, M., Rowe, DA. Issues and challenges in sedentary behavior measurement; Meas Phys Educ Exerc Sci. 2015. p. 105-115.https://doi.org/10.1080/1091367X.2015.1055566

13. Atkin AJ, Gorely T, Clemes SA, et al. Methods of Measurement in epidemiology: sedentary Behaviour. Int J Epidemiol. 2012; 41(5):1460–1471. https://doi.org/10.1093/ije/dys118. [PubMed: 23045206]

14. Bowles HR. Measurement of active and sedentary behaviors: closing the gaps in self-report methods. J Phys Act Health. 2012; 9(Suppl 1):S1–4. https://doi.org/10.1123/jpah.9.s1.s1. [PubMed: 22287442]

15. Tooze JA, Troiano RP, Carroll RJ, Moshfegh AJ, Freedman LS. A measurement error model for physical activity level as measured by a questionnaire with application to the 1999–2006 NHANES questionnaire. Am J Epidemiol. 2013; 177(11):1199–1208. https://doi.org/10.1093/aje/kws379. [PubMed: 23595007]

16. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc. 2008; 40(1):181–188. https://doi.org/10.1249/mss.0b013e31815a51b3. [PubMed: 18091006]

17. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S.: adults compliance with the Physical Activity Guidelines for Americans. Am J Prev Med. 2011; 40(4):454–461. https://doi.org/10.1016/j.amepre.2010.12.016. [PubMed: 21406280]

18. Saint-Maurice PF, Welk GJ. Web-based assessments of physical activity in youth: considerations for design and scale calibration. J Med Internet Res. 2014; 16(12):e269. https://doi.org/10.2196/jmir.3626. [PubMed: 25448192]

19. Saint-Maurice PF, Welk GJ. Validity and Calibration of the Youth Activity Profile. PLoS One. 2015; 10(12):e0143949. https://doi.org/10.1371/journal.pone.0143949. [PubMed: 26630346]

20. Saint-Maurice PF, Welk GJ, Beyler NK, Bartee RT, Heelan K. Calibration of self-report tools for physical activity research: The Physical Activity Questionnaire (PAQ). BMC Public Health. 2014; 14(461):1–9. https://doi.org/10.1186/1471-2458-14-461. [PubMed: 24383435]

21. Nebeling L, Oh A, Hennessy E, et al. The Family Life, Activity, Sun, Health, and Eating Study: Development and participant characteristics. Am J Prev Med. 2017 In press. [THIS THEME].

22. Oh A, Davis T, Dwyer L, et al. The Family Life, Activity, Sun, Health and Eating (FLASHE) study: Study design. Am J Prev Med. 2017 In press. [THIS THEME].

23. Crouter SE, Flynn JI, Bassett DR Jr. Estimating physical activity in youth using a wrist accelerometer. Med Sci Sports Exerc. 2015; 47(5):944–951. https://doi.org/10.1249/MSS.0000000000000502. [PubMed: 25207928]

24. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc. 2011; 43(2):357–364. https://doi.org/10.1249/MSS.0b013e3181ed61a3. [PubMed: 20581716]

25. Harrell JS, McMurray RG, Baggett CD, Pennell ML, Pearce PF, Bangdiwala SI. Energy costs of physical activities in children and adolescents. Med Sci Sports Exerc. 2005; 37(2):329–336. https://doi.org/10.1249/01.MSS.0000153115.33762.3F. [PubMed: 15692331]

26. Ridley K, Olds TS. Assigning energy costs to activities in children: A review and synthesis. Med Sci Sports Exerc. 2008; 40(8):1439–1446. https://doi.org/10.1249/MSS.0b013e31817279ef. [PubMed: 18614948]

27. Esliger DW, Copeland JL, Barnes JD, Tremblay MS. Standardizing and optimizing the use of accelerometer data for free-living physical activity monitoring. J Phys Act Health. 2005; 3:366–383. https://doi.org/10.1123/jpah.2.3.366.

28. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1(8476):307–310. https://doi.org/10.1016/S0140-6736(86)90837-8. [PubMed: 2868172]

29. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm. 1987; 15(6):657–680. https://doi.org/10.1007/BF01068419. [PubMed: 3450848]

30. Robin J, Wynn J, Moscovitch M. The spatial scaffold: The effects of spatial context on memory for events. J Exp Psychol Learn Mem Cogn. 2016; 42(2):308–315. https://doi.org/10.1037/xlm0000167. [PubMed: 26375782]

31. Hagstromer M, Bergman P, De Bourdeaudhuij I, et al. Concurrent validity of a modified version of the International Physical Activity Questionnaire (IPAQ-A) in European adolescents: The HELENA Study. Int J Obes (Lond). 2008; 32(Suppl 5):S42–48. https://doi.org/10.1038/ijo.2008.182.

32. Helmerhorst HJ, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. Int J Behav Nutr Phys Act. 2012; 9:103. https://doi.org/10.1186/1479-5868-9-103. [PubMed: 22938557]

33. Bassett DR Jr, Rowlands A, Trost SG. Calibration and validation of wearable monitors. Med Sci Sports Exerc. 2012; 44(1 Suppl 1):S32–38. https://doi.org/10.1249/MSS.0b013e3182399cf7. [PubMed: 22157772]

34. Brazendale K, Beets MW, Bornstein DB, et al. Equating accelerometer estimates among youth: The Rosetta Stone 2. J Sci Med Sport. 2016; 19(3):242–249. https://doi.org/10.1016/j.jsams.2015.02.006. [PubMed: 25747468]

35. Bornstein DB, Beets MW, Byun W, et al. Equating accelerometer estimates of moderate-to-vigorous physical activity: in search of the Rosetta Stone. J Sci Med Sport. 2011; 14(5):404–410. https://doi.org/10.1016/j.jsams.2011.03.013. [PubMed: 21524938]

36. Kim Y, Hibbing P, Saint-Maurice PF, et al. Surveillance of youth physical activity and sedentary behavior with wrist-worn accelerometers. Am J Prev Med. 2017 In press. [THIS THEME].

37. Levin S, Jacobs DR Jr, Ainsworth BE, Richardson MT, Leon AS. Intra-individual variation and estimates of usual physical activity. Ann Epidemiol. 1999; 9(8):481–488. https://doi.org/10.1016/S1047-2797(99)00022-8. [PubMed: 10549881]

38. Matthews CE, Ainsworth BE, Thompson RW, Bassett DR Jr. Sources of variance in daily physical activity levels as measured by an accelerometer. Med Sci Sports Exerc. 2002; 34(8):1376–1381. https://doi.org/10.1097/00005768-200208000-00021. [PubMed: 12165695]

39. Matthews CE, Freedson PS, Hebert JR, et al. Seasonal variation in household, occupational, and leisure time physical activity: longitudinal analyses from the seasonal variation of blood cholesterol study. Am J Epidemiol. 2001; 153(2):172–183. https://doi.org/10.1093/aje/153.2.172. [PubMed: 11159163]

40. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: Prospective observational studies corrected for the regression dilution bias. Lancet. 1990; 335(8692):765–774. https://doi.org/10.1016/0140-6736(90)90878-9. [PubMed: 1969518]

41. Beaton GH, Milner J, Corey P, et al. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Am J Clin Nutr. 1979; 32(12):2546–2559. [PubMed: 506977]

42. Dodd KW, Guenther PM, Freedman LS, et al. Statistical methods for estimating usual intake of nutrients and foods: A review of the theory. J Am Diet Assoc. 2006; 106(10):1640–1650. https://doi.org/10.1016/j.jada.2006.07.011. [PubMed: 17000197]

43. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. Am J Epidemiol. 2003; 158(1):14–21. discussion 22–16. https://doi.org/10.1093/aje/kwg091. [PubMed: 12835281]
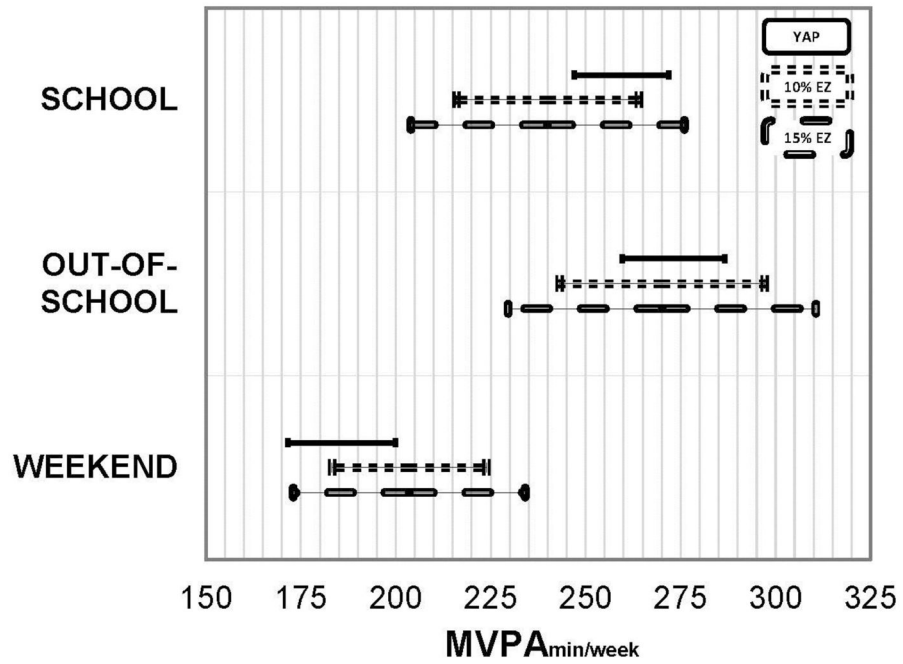
Figure 1.
Group-level agreement for school (n=51), out-of-school (n=51), and weekend (n=39) activity using a 10% and 20% equivalence zones.

Notes: The solid black line bar represents the CI for predicted YAP MVPA values while the dashed and the rectangular lines represent the 10% and 20% equivalence zones based on GT3X+ MVPA values, respectively.

YAP, Youth Activity Profile; MVPA, moderate-to-vigorous physical activity; EZ, equivalence zone
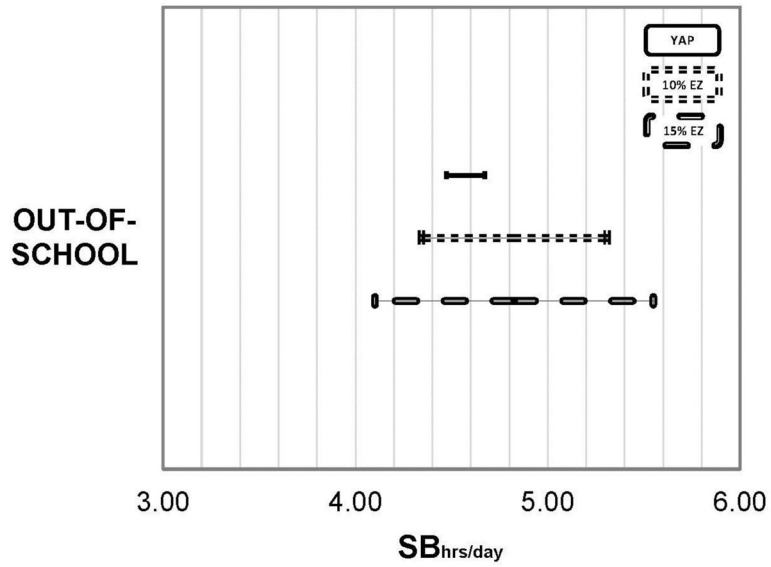
Figure 2.
Group-level agreement for out-of-school sedentary behavior (SB) (n=50) using a 10% and 20% equivalence zones.

Notes: The solid black line bar represents the CI for predicted YAP MVPA values while the dashed and the rectangular lines represent the 10% and 20% equivalence zones based on GT3X+ measured sedentary time, respectively.

YAP, Youth Activity Profile; MVPA, moderate-to-vigorous physical activity; SB, sedentary behavior

Table 1

Demographic Information, Stratified by Gender, Segment of the Week, and School Grade Level

Demographics	School		Out-of-School		Weekend	
	Middle school (n=31)	High school (n=48)	Middle school (n=30)	High school (n=50)	Middle school (n=19)	High school (n=38)
Boys						
Age	12.9 ±0.9	15.9 ±1.1	13.0 ±0.9	15.8 ±1.1	12.9 ±1.1	15.8 ±1.3
BMI	21.7 ±4.7	23.3 ±4.1	21.6 ±4.6	23.3 ±4.1	22.5 ±5.0	23.6 ±4.2
Height (cm)	161.7 ±12.6	176.5 ±8.7	162.1 ±12.8	175.5 ±8.6	159.5 ±14.2	175.9 ±8.5
Weight (kg)	57.4 ±16.4	72.8 ±15.4	57.4 ±16.7	72.2 ±15.9	58.1 ±17.5	73.5 ±16.7
$MVPA_{min/day}$	60.7 ±26.2	46.8 ±23.3	65.4 ±28.1	48.1 ±19.2	104.1 ±52.4	101.0 ±46.6
$MVPA_{\%/day}$	12.7 ±5.5	9.8 ±4.6	13.7 ±5.8	10.1 ±4.0	11.2 ±5.7	11.0 ±4.9
$SB_{min/day}$	NA	NA	277.3 ±55.9	298.9 ±52.9	NA	NA
$SB_{\%/day}$	NA	NA	58.1 ±10.6	62.7 ±8.9	NA	NA
CPM_{day}	1,662.8 ±687.1	1,275.8 ±490.5	1,686.6 ±642.4	1,291.6 ±462.6	1,423.7 ±640.5	1,350.6 ±481.7
Wear time (days)	4.0 ±0.2	3.9 ±0.3	3.9 ±0.3	3.9 ±0.3	1.9 ±0.2	1.9 ±0.3
Wear time (min/day)	477.5 ±32.8	475.1 ±39.0	477.1 ±31.2	475.8 ±41.3	924.9 ±37.7	916.9 ±35.8
Period duration	478.7 ±30.1	477.7 ±38.2	482.3 ±30.5	476.4 ±64.1	960.0 ±0.0	960.0 ±0.0
Girls	(n=33)	(n=57)	(n=30)	(n=60)	(n=27)	(n=46)
Age	13.2 ±0.9	15.5 ±1.1	13.3 ±0.9	15.5 ±1.1	13.2 ±1.0	15.5 ±1.0
BMI	20.8 ±4.8	22.6 ±4.1	20.6 ±4.9	23.1 ±5.0	20.5 ±4.1	22.7 ±4.1
Height (cm)	158.1 ±9.4	164.8 ±6.4	157.9 ±9.5	163.8 ±7.2	158.0 ±9.5	163.3 ±7.3
Weight (kg)	52.4 ±15.3	61.3 ±12.2	52.1 ±16.2	62.1 ±14.6	51.7 ±13.3	60.9 ±12.8
$MVPA_{min/day}$	60.5 ±25.3	50.6 ±22.4	67.9 ±29.4	56.3 ±26.9	115.3 ±51.3	98.7 ±46.5
$MVPA_{\%/day}$	12.8 ±5.2	10.5 ±4.7	14.1 ±6.0	12.0 ±5.5	12.4 ±5.4	10.9 ±5.1
$SB_{min/day}$	NA	NA	260.9 ±51.1	285.4 ±66.0	NA	NA
$SB_{\%/day}$	NA	NA	54.7 ±10.5	60.7 ±10.5	NA	NA
CPM_{day}	1,635.1 ±564.9	1,299.5 ±519.8	1,665.8 ±555.8	1,446.2 ±607.0	1,484.6 ±543.4	1,310.6 ±497.1
Wear time (days)	3.9 ±0.3	3.9 ±0.3	3.9 ±0.3	3.8 ±0.4	1.9 ±0.4	2.0 ±0.2
Wear time (min/day)	471.1 ±39.9	487.5 ±79.5	478.6 ±40.0	469.5 ±54.0	921.2 ±50.2	908.2 ±40.1
Period duration	474.2 ±36.1	491.8 ±89.3	486.3 ±37.6	476.3 ±49.7	960.0 ±0.0	960.0 ±0.0

CPM, counts per minute; MVPA, moderate-to-vigorous physical activity; SB, sedentary behavior; NA, not applicable

Table 2

Regression Coefficients for School, Out-of-School, Weekend, and Sedentary Activity Outcomes

Model	School (n=118)	Out-of- school (n=119)	Weekend (n=91)	Sedentary (n=116)
Intercept	**13.18±0.90***	**12.38±0.76***	**10.87±0.64***	**58.35±1.25***
Age	**−1.70±0.63***	**−0.98±0.39***	**−0.65±0.32***	0.56±0.85
YAP	**1.24±0.55***	**1.17±0.58***	**1.30±0.55***	3.44±2.15
Age x YAP	−0.35±0.44	0.14±0.31	0.45±0.27	1.07±1.33

Note: Boldface indicates statistical significance (*p<0.05).

β-weights were rounded to 2 decimal places in this table however the full β-weight estimate (with 4 decimal places) can be found in Appendix 2.

YAP, Youth Activity Profile