

Calibration of the Online Youth Activity Profile Assessment for School-Based Applications

Gregory J. Welk

Iowa State University

Pedro F. Saint-Maurice

National Cancer Institute

Philip M. Dixon

Iowa State University

Paul R. Hibbing

Children's Mercy Kansas City

Yang Bai

The University of Utah

Gabriella M. McLoughlin

Washington University

Michael Pereira da Silva

Federal University of Rio Grande

A balance between the feasibility and validity of measures is an important consideration for physical activity (PA) research—particularly in school-based research with youth. The present study extends previously tested calibration methods to develop and test new equations for an online version of the youth activity profile (YAP) tool, a self-report tool designed for school applications. Data were collected across different regions and seasons to develop more robust, generalizable equations. The study involved a total of 717 youth from 33 schools (374 elementary [ages 9–11 years], 224 middle [ages 11–14 years], and 119 high school [ages 14–18 years]) in two different states in the United States. Participants wore a Sensewear monitor for a full week and then completed the online YAP at school to report PA and sedentary behaviors in school and at home. Accelerometer data were processed using an R-based segmentation program to compute PA and sedentary behavior levels. Quantile regression models were used with half of the sample to develop item-specific YAP calibration equations, and these were cross validated with the remaining half of the sample. Computed values of mean absolute percentage error ranged from 15 to 25% with slightly lower error observed for the middle school sample. The new equations had improved precision compared with the previous versions when tested on the same sample. The online version of the YAP provides an efficient and effective way to capture school level estimates of PA and sedentary behaviors in youth.

Keywords: children, physical activity, sedentary behavior

Accurate assessments of physical activity (PA) and sedentary behavior (SB) are essential for answering many important public health research questions (Dunton et al., 2019; Hidding, Altenburg, Mokkink, Terwee, & Chinapaw, 2017). Emphasis in the field has been on refining approaches with monitor-based methods (Troiano, 2005; Troiano, McClain, Brychta, & Chen, 2014), but the challenges of employing monitors in research are well documented

(Lynch et al., 2019; Pedišić & Bauman, 2015). Parallel efforts are needed to enhance the accuracy and utility of more practical and cost-effective report-based instruments (Spruijt-Metz et al., 2018; Welk, Beyler, Kim, & Matthews, 2017).

A promising, but underutilized, harmonization strategy is the use of calibration methods that can rescale report-based tools so that estimates equate with those from monitor-based tools (Welk, 2019). Calibration and measurement error models have been widely used in the diet assessment literature and have similar utility for PA assessment (Beyler, Fuller, Nusser, & Welk, 2015; Nusser et al., 2012; Welk, Beyler, et al., 2017). In addition to being easier to administer, an advantage of calibrated report-based tools is that it enables context of the underlying behaviors to be captured—a key limitation of monitor-based methods (Bowles, 2012; Engelen et al., 2015; Troiano, Gabriel, Welk, Owen, & Sternfeld, 2012). Calibration methods offer promise for all segments of the population but may offer additional advantages for addressing challenges with assessing PA and SB in youth (Dunton et al., 2019; Ekellund, Tomkinson, & Armstrong, 2011; Lynch et al., 2019; Spruijt-Metz et al., 2018).

Numerous studies have reported on the applications of report-based measures for capturing youth PA and SB (Hidding et al., 2017; Spruijt-Metz et al., 2018; Welk, Corbin, & Dale, 2000; Welk, Morrow, & Saint-Maurice, 2017); however, instruments vary considerably in their design as well as in the demand on cognitive

© 2021 The Authors. Published by Human Kinetics, Inc. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, CC BY-NC-ND 4.0, which permits the copy and redistribution in any medium or format, provided it is not used for commercial purposes, no modifications are made, appropriate credit is given, and a link to the license is provided. See <http://creativecommons.org/licenses/by-nc-nd/4.0>. This license does not cover any third-party material that may appear with permission in the article. For commercial use, permission should be requested from Human Kinetics, Inc., through the Copyright Clearance Center (<http://www.copyright.com>).

Welk and Dixon are with Iowa State University, Ames, IA, USA. Saint-Maurice is with the National Cancer Institute, Rockville, MD, USA. Hibbing is with Children's Mercy Kansas City, Knoxville, TN, USA. Bai is with the University of Utah, Salt Lake City, UT, USA. McLoughlin is with Washington University in St. Louis, St. Louis, MO, USA. Pereira da Silva is with the Federal University of Rio Grande, Rio Grande, Rio Grande do Sul, Brazil. Welk (gwelk@iastate.edu) is corresponding author.

processes that influence recall. A recent review emphasized the importance of instrument design, contextual cues, and format for capturing more accurate information on youth behaviors (Saint-Maurice, Sousa, Welk, Matthews, & Berrigan, 2020). However, calibration equations are likely still needed to adjust for bias and known sources of error. A promising tool that was intentionally developed to enable calibration with monitor-based methods is the youth activity profile (YAP[®]; Saint-Maurice, Welk, Beyler, Barte, & Heelan, 2014).

The YAP was designed to facilitate evaluation of youth PA and SB for school-based applications, since this is a primary setting to reach and influence youth. Items are each framed to provide spatial and contextual cues, since this has been shown to improve accuracy of recall (Baranowski & Domel, 1994; Robin & Moscovitch, 2014). Students' reported behaviors in distinct segments of the day were then calibrated using regression-based methods to estimate time spent in PA and SB (Saint-Maurice & Welk, 2015). Our preliminary calibration of the YAP showed that PA and SB estimates from the YAP were within 10–25% from monitor-based estimates.

Subsequent research employed alternative analytic procedures to calibrate the YAP for use within the national FLASHE (Family Life, Activity, Sun, Health, and Eating) project led by the National Cancer Institute (Saint-Maurice et al., 2017). Another study recalibrated the YAP for related school-based applications in the United Kingdom (Fairclough et al., 2019). These studies and related work with other instruments (Saint-Maurice et al., 2017) have supported the utility of the YAP calibration methodology.

To facilitate broader adoption, a web-based version of the YAP was recently developed (www.youthactivityprofile.org).¹ The utility and feasibility of this online version has been demonstrated through its use within the FitnessGram NFL Play60 FitnessGram Partnership project, a participatory research initiative involving over 1,100 schools (Welk, Bai, Saint-Maurice, Allums-Featherston, & Candelaria, 2016). The online version provides clear advantages for school-based applications; however, formalized tests are needed to evaluate accuracy when the YAP is administered in school settings using the online version. Physical activity behaviors (and associated reporting on the YAP) also vary based on a number of factors including age, culture, geography, rurality, and season. The original study was conducted on a limited sample of youth from a rural midwestern U.S. state; therefore, there is a need to calibrate and test the online version of the YAP in a more diverse sample. Through an National Institutes of Health-funded study (NIH 5R21CA188641-02), we collected new data with the online version of the YAP using a larger and more distributed sample of youth from two different states and across different seasons. The present study is designed to develop and evaluate new calibration equations to enhance group-level estimation of PA and SB with this online version of the YAP. Consistent with best practices in PA assessment research (Welk et al., 2019), the utility of the updated algorithms are directly compared with the original versions to evaluate potential advantages of the online version for future research and practice applications.

Methods

The present study employs the same basic calibration process as in the original study (Saint-Maurice & Welk, 2015) and also uses the Sensewear armband (SWA) monitor as a reference/comparison device since it provides robust estimates of PA and SB in

youth (Arvidsson, Slinde, & Hulthén, 2009; Calabro, Stewart, & Welk, 2013; Calabro, Welk, & Eisenmann, 2009), while also providing a direct indicator of wear time. The key feature of the published calibration methodology (Saint-Maurice & Welk, 2014) is that youth are asked to wear the SWA for a full week and then complete the YAP as a recall-based assessment to report their behavior over the same time periods. By temporally matching the data from the YAP to the SWA, it is possible to develop generalized prediction equations that produce group-level estimates of PA and SB. A brief summary of the YAP is provided below followed by the specific procedures for the study.

Description of the YAP Assessment and Software Tool

The YAP is a self-administered, 7-day recall instrument that can be easily completed by youth as part of normal school-based assessment. The assessment is completed online using a web-based content management system using password protected conventions. Students are guided through the assessment using built-in instructions and prompts that encourage them to specifically report their behaviors over the previous week.

The specific assessment consists of 15 easily interpretable items divided into three thematic sections: (a) in-school PA, (b) out-of-school PA, and (c) sedentary habits. Items in the “in-school” section capture participation in PA during five specific segments (transportation to/from school, PA during PE, lunch, and recess). Items in the “out-of-school” section are subdivided into PA on weekdays (before school, after school, and during the evening) and PA on weekend days (Saturday and Sunday). Items on SB, ask about time spent watching TV, playing videogames, using computer, using cell phone, and an overall sedentary time item (see Table 1 for a breakdown of contexts assessed in the YAP).

A unique advantage of the YAP calibration methodology (Saint-Maurice & Welk, 2015) is that the generated equations predict the percentage of time spent in PA for each item. This allows the prediction equations to estimate moderate-to-vigorous physical activity (MVPA) for distinct portions of the day and to accommodate different segment durations (e.g., different amounts of PE or recess). For example, if schools have more scheduled PE or longer recess breaks, the estimates would yield proportionately larger estimates for predicted minutes of PA. The segmentation into school and home components also helps school leaders and researchers evaluate the contributions of school-based PA and total PA behaviors.

Design and Sample

The present study was designed to facilitate the creation of more robust and generalizable prediction equations for the online version of the YAP. Data were collected from two different states (Iowa and Texas) to capture differences in culture, weather, and geography. The Texas sample was primary urban/suburban (metro Dallas area), while the Iowa sample was primarily rural/small town. Efforts were made to recruit schools with varying demographics and to ensure similar sample sizes from elementary (Grades 4–5, ages 9–11 years); middle (Grades 6–8, ages 11–14 years); and high school (Grades 9–12, ages 14–18 years). We recruited a total of 33 schools (22 in Iowa and 11 in Texas). In Iowa, data were collected from six elementary, eight middle, and eight high schools. In

Table 1 Summary of Item Labels and Question Content for the YAP Segments

Segment/item	Question text
In-school	
PA to school	How many days did you walk or bike to school? (If you can't remember, try to estimate)
PE	During physical education, how often were you running and moving as part of the planned games or activities? (<i>If you didn't have PE, choose "I didn't have physical education"</i>)
Recess	During recess, how often were you playing sports, walking, running, or playing active games? (<i>If you didn't have a break at school, choose "I didn't have recess"</i>)
Lunch ^a	During lunch break, how often were you moving around, walking or playing? (<i>If you didn't have a break at school, choose "I didn't have lunch breaks"</i>)
PA from school	How many days did you walk or bike from school? (<i>If you can't remember, try to estimate</i>)
Out-of-school	
Before school	How many days before school (6:00–8:00 am) did you do some form of physical activity for at least 10 min? (<i>This includes activity at home NOT walking or biking to school</i>)
After school	How many days after school (between 3:00 and 6:00 pm) did you do some form of physical activity for at least 10 min? (This can include playing with your friends/family, team practices, or classes involving physical activity but <i>NOT walking or biking home from school</i>)
Evening	How many school evenings (6:00 and 10:00 pm) did you do some form of physical activity for at least 10 min? (This can include playing with your friends/family, team practices or classes involving physical activity but <i>NOT walking or biking home from school</i>)
Weekend	
Saturday	How much physical activity did you do last Saturday? (<i>This could be for exercise, work/chores, family outings, sports, dance, or play. If you don't remember, try to estimate</i>)
Sunday	How much physical activity did you do last Sunday? (<i>This could be for exercise, work/chores, family outings, sports, dance, or play. If you don't remember, try to estimate</i>)
Sedentary items	
Computer time	How much time did you spend using computers outside of school time? (<i>This doesn't include homework time but includes time on Facebook as well as time spent surfing the Internet, instant messaging, playing online video games or computer games</i>)
TV time	How much time did you spend watching TV outside of school time? (<i>This includes time spent watching movies or sports but NOT time spent playing video games</i>)
Video games	How much time did you spend playing video games outside of school time? (<i>This includes games on Nintendo DS, wii, Xbox, PlayStation, iTouch, iPad, or games on your phone</i>)
Cell phone	How much time did you spend using your cell phone after school? (<i>This includes time spent talking or texting</i>)
Overall sedentary	Which of the following best describes your typical sedentary habits at home? (<i>Try to think about a typical week and not just last week</i>)

Note. YAP = youth activity profile; PA = physical activity; PE = physical education.

^aLunch was not included in the final calibration and was replaced by an item to ask about classroom PA integration.

Texas, data were collected from three elementary, six middle, and two high schools. Students were individually recruited by sending out information packets with consent/assent forms to parents from intact classrooms in these schools. In accordance with the institutional review board approved protocol, parents provided informed consent for their child to participate and children completed an assent form. In total, we enrolled 575 students from Iowa and 472 students from Texas (total $N = 1,047$).

Data Collection Procedures

The data collection took place over a 2-year period in both Iowa and Texas to capture data across seasons and weather conditions. Specific efforts were made to balance the order and timing of data collection in elementary, middle, and high schools so that age-related differences in PA levels were not further confounded by seasonality or weather. Enrolled students from the targeted classrooms were met by members of the research team at the beginning of a school day. They received further information about the project and provided their assent to participate. The students were

asked to wear a SWA monitor for the next 7 days, including sleep but excluding swimming activities, showering, or other activities not appropriate for monitoring. Students were provided with instructions on wear and use of the monitor along with a log used to report times when the monitors were removed. While the SWA cannot be considered to be a true criterion measure, prior work has supported the validity (relative to doubly labeled water and other devices) for estimation of energy expenditure in youth (Calabro et al., 2013; Lopez, Brønd, Andersen, Dencker, & Arvidsson, 2018).

At the end of the 7-day monitoring protocol, research team members returned to the classroom to collect the monitors and to administer the online version of the YAP to obtain students' reports of PA and SB over the previous 7 days. Standardized instructions were provided to explain the nature of the assessment and to ensure that students had appropriate understanding of PA, the need to report based on the last 7 days, and the importance of completing the items as carefully as possible. Students were provided with access to the online system and were provided with individual login information to enable project IDs used in the monitor initialization to match the report-based data from the YAP.

Data Processing Procedures

Data processing of the YAP and the SWA followed our established methodology (Saint-Maurice & Welk, 2015; Saint-Maurice, Welk, Barteel, & Heelan, 2017). Deidentified student responses on the 15 YAP items were downloaded through the web-based YAP system and saved for subsequent processing. To enable calibration, the monitor-based data needed to be temporally linked to the corresponding YAP segments for each student. The SWA software (InnerView software, version 6.1) was used to create individual export files that were coded with project IDs to facilitate the merging process. The data segmentation process was conducted using customized R programs that incorporated schedule information for each class and school (e.g., school start time, PE periods, recess periods, etc.) as well as student responses for each item. The process generated a master data set with all participant individual records and a segmentation code that identified the segment of the day and week. Individual minutes recorded from the monitor throughout the week were coded for each segment (e.g., recess, lunch, etc.) using standard metabolic equivalents (SB: METs ≤ 2.0 ; MVPA: METs ≥ 4.0), based on unique youth MET criteria (Saint-Maurice, Kim, Welk, & Gaesser, 2016). The total number of minutes in either sedentary or MVPA were then aggregated to compute total minutes of SB or MVPA per segment per day, as well as the relative time spent at each behavior (e.g., percentage of time spent in MVPA at recess). The PA data were further aggregated into three primary day segments (in-school, out-of-school, and weekend), while the sedentary segment was aggregated to capture weekdays outside of school time.

As the calibration process is predicated on accurate monitor-based data, the individual records were carefully screened for compliance (separately for each segment) to assure quality and representativeness of the data. Accelerometer records were first screened for nonwear time which is objectively reported from the SWA output, since the sensors can detect when it is not on the arm. Participants with less than 70% of wear time for a specific time segment were excluded from the analytical data set. In addition, participants with less than three separate records in the week (for a given segment) were also excluded, except for the PE, Saturday, and Sunday segments (which had a minimum of one record for the week). Participants with incomplete YAP scores were also removed and not considered for further analyses. The final sample included 717 participants (68% of the initial sample).

Statistical Analysis

The statistical analyses involved separate calibration and cross-validation phases. The retained sample was randomly split into separate calibration ($n = 359$; 50%) and cross-validation data sets ($n = 358$; 50%) to provide an appropriate evaluation of the developed models. All analyses were conducted using SAS (version 9.4; Cary, NC), and we used $p < .05$ to define statistical significance.

In the calibration phase, we used quantile regression (Koenker, 2019) to predict the accelerometer median percentage of time spent in MVPA with YAP crude scores; grade level (elementary, middle, and high school); and sex (boys and girls) as predictors. The models specified a non-zero intercept and additive linear effects for the other terms.² Confidence intervals for the regression parameters were estimated by a Markov Chain Marginal Bootstrap (He & Hu, 2002) using 10,000 bootstrap samples.

Consistent with the original study, models were fit individually for each of the specific PA items. The models did not converge for the item capturing lunchtime due to insufficient variability in

lunchtime MVPA so this was dropped from the predictions. Segment-specific % MVPA predictions were converted into minutes of MVPA per week by multiplying by the associated length of the interval and the frequency of the segment in a week. These estimates were then aggregated into minutes of MVPA for three distinct day segments (in-school, out-of-school, and weekend) as shown in Table 1. We calibrated percentage of time spent in SB by predicting median percentage of time in SB generated from the accelerometer, using the aggregated YAP SB score, grade level, and sex as predictors. In the cross-validation phase, we used standard measurement agreement methods to examine the accuracy of the equations when used to compute school-level estimates of average time spent in MVPA and SB. The primary focus of the YAP calibration process is to provide a reasonably accurate group-level (i.e., grade level and/or school) estimate of time spent in MVPA. Hence, in the cross-validation phase, we examined if school estimates from the YAP were directly compared with parallel school estimates from the SWA.

Consistent with past work, we used mean absolute percentage error (MAPE) values, which capture the overall degree of accuracy of the estimates, and equivalence testing concepts (Dixon et al., 2018), which determine whether two measurement methods are significantly equivalent (within a specified zone). The MAPE values provide information about the precision of predictions for an individual school. The equivalence bounds provide information about the accuracy of predictions averaged over the population of schools. The equivalence test essentially evaluates whether the difference between two methods, δ , is within prespecified equivalence bounds (l , u). Because there are no standard equivalence bounds for PA measurement, we report the smallest equivalence bounds for which the corresponding equivalence test will have a p value $< .05$. These were computed as the smallest value, m , for which the equivalence region ($-m$ and m) completely includes the 90% confidence interval for the mean difference. This is equivalent to the 2 one-sided test approach with a 5% alpha.

Results

The demographics of the participants in the calibration and cross-validation phase of the evaluation are summarized in Table 2. There was an even distribution of participants from Iowa and Texas and balanced gender and age distribution in the sample.

The calibration process with the sample of 359 participants generated a series of item-specific quantile regression equations. The regression coefficients equations for estimating the percentage of time spent in MVPA/SB for each of the designated time periods in the YAP are provided in Table 3. The coefficients for YAP were positive for each activity segment and ranged from 0.92% (before school) to 5.9% (recess). For interpretation purposes, a one unit increase in YAP score on transport to school item would reflect a 2.5% higher estimate for the percentage of time in MVPA during transportation to school. With the exception of the physical education (PE) and before school items, the YAP items were significant ($p < .05$) predictors of monitor-based estimates of percentage of time spent in MVPA and SB, after adjusting for gender and grade level.

Cross-validation results are provided in Table 4 for the combined sample as well as segmented by age and gender. (Results were restricted to samples that had 10 or more students with available data.) The minimum sizes of 95% equivalence regions range from 7.2 min/day (for in-school PA) to 13.4 min/day (for out-

Table 2 Samples and Descriptive Statistics for the Calibration and Cross Validation

Site	Calibration sample (<i>n</i> = 359)		Validation sample (<i>n</i> = 358)		Total (<i>n</i> = 717)
	Iowa (<i>n</i> = 185)	Texas (<i>n</i> = 174)	Iowa (<i>n</i> = 182)	Texas (<i>n</i> = 176)	
Age, mean (<i>SD</i>)	11.3 (1.8)	11.1 (1.5)	11.3 (1.8)	11.2 (1.5)	11.2 (1.7)
Boys, <i>n</i> (%)	89 (48.1)	88 (50.6)	84 (46.2)	79 (44.9)	340 (47.4)
Girls, <i>n</i> (%)	96 (51.9)	86 (49.4)	98 (53.9)	97 (55.1)	377 (52.6)
Elementary, <i>n</i> (%)	84 (45.4)	102 (58.6)	87 (47.8)	101 (57.4)	374 (52.2)
Middle, <i>n</i> (%)	69 (37.3)	46 (26.4)	61 (33.5)	48 (27.3)	224 (31.2)
High school, <i>n</i> (%)	32 (17.3)	26 (14.9)	34 (18.7)	27 (15.3)	119 (16.6)
In-school %MVPA, ^a mean (<i>SD</i>)					
Transportation to school	11.1 (14.8)	10.4 (14.7)	9.1 (14.6)	13.4 (20.7)	11.1 (16.6)
Physical education	41.9 (33.7)	26.8 (28.3)	44.4 (34.9)	29.2 (28.7)	34.2 (31.8)
Recess	45.9 (30.4)	45.4 (30.5)	46.0 (29.8)	47.0 (31.2)	46.1 (30.3)
Lunch	2.2 (5.8)	5.9 (10.8)	3.0 (8.6)	5.4 (9.3)	4.2 (9.0)
Transportation to home	19.4 (17.7)	22.2 (23.3)	18.9 (19.8)	25.2 (21.7)	21.7 (21.1)
Out-of-school %MVPA, ^a mean (<i>SD</i>)					
Before school	10.6 (13.5)	8.2 (11.7)	9.1 (13.1)	10.1 (13.9)	9.5 (13.1)
After school	18.2 (13.9)	16.7 (16.5)	17.1 (15.1)	16.8 (15.7)	17.1 (15.4)
Evening	11.5 (9.8)	10.8 (10.5)	12.1 (10.2)	12.1 (11.1)	11.6 (10.4)
Weekend %MVPA, ^a mean (<i>SD</i>)					
Saturday	13.1 (12.9)	13.4 (13.1)	11.3 (11.3)	13.0 (11.8)	12.7 (12.3)
Sunday	12.2 (11.4)	9.9 (10.8)	10.4 (9.2)	9.5 (9.9)	10.6 (10.4)
Nonschool %SB, ^b mean (<i>SD</i>)	64.4 (19.5)	56.2 (22.9)	61.9 (19.5)	58.0 (21.6)	60.3 (21.1)

Note. MVPA = moderate-to-vigorous physical activity; SB = sedentary behavior.

^aPercentage of time spent in MVPA in 10-min bouts at each of the individual day segments. For example, on average, participants spent 34.2% of a physical education session in MVPA. ^bPercentage of time spent in sedentary time during out-of-school time.

of-school PA) when averaged over all schools and from 5.6 to 25.2 min/day when data are subset by grade level or gender. An alternative way to report the significance of these equivalence tests is relative to a prespecified equivalence bound. For in-school PA, an equivalence test using a prespecified bound of 10 min/day would be significant with $p < .05$, but one with a bound of 5 min/day would not ($p > .05$).

Table 4 also shows MAPE values for both the new equations, as well as comparison values based on the application of the original equations (Saint-Maurice & Welk, 2015). The MAPE values for new algorithms were consistently lower than those produced with the original equations, but patterns were fairly similar. Averaged over schools, the YAP-based predictions of MVPA were slightly lower than the SWA measurements for the combined data as well as for the age group and gender subsets. YAP predictions of sedentary time are frequently slightly larger than the SWA measurements, again on average. Overall, the MAPE values tended to be lower in the middle and elementary schools, but there was some variability across the time segments and age (see Figure 1).

The agreement in estimated MVPA and reported MAPE values was also computed separately at for each school to provide an indication of how the YAP would perform for school-specific applications (see Figure 2). The plot captures the average YAP and SWA data for the schools with at least 10 participants in the cross-validation data set (34% of the schools) to ensure generalizability of the estimates. The YAP estimates generally corresponded in magnitude and direction with the patterns produced by the SWA. However, the relationships exhibited some degree of random error

as there were inconsistent patterns across schools. For example, School 4 demonstrated good agreement between for the in-school items and poorer for out-of-school; this was the converse for School 7. These plots reflect mean bias for individual schools, and the inherent variability reflects the nature of the error that is captured with the more robust indicator of MAPE. The overall school-level MAPE values were 23%, 21%, 17%, and 8.4% for in-school PA, out-of-school PA, weekend PA, and sedentary time, respectively.

A summary stacked bar plot based on the cross-validation data shows the direct comparison of the accumulated MVPA estimates for the eight segments that collectively capture a typical school day for elementary aged youth (see Figure 3). The plot captures the overall degree of correspondence that would be expected when compiling data across schools to estimate youth PA behaviors.

Discussion

The purpose of this study was to develop and test updated calibration equations for the online version of the YAP. The study followed methods and approaches that were similar to our original study (Saint-Maurice & Welk, 2015) but also included a number of additional features to enhance the generalizability of the calibration equations. The present study used a larger and more diverse sample of students from schools in two different states to capture differences due to region and culture. It included both urban and rural schools to capture differences due to environmental and socio-demographic factors. It also utilized an intentionally extended data collection protocol that captured data across two full years (to

Table 3 Quantile Regression Coefficients and 95% CIs to Predict Median %MVPA for Individual Items in the Online YAP

Individual YAP Items	Intercept	YAP	Grade level	Sex
In-school (%MVPA)				
Transportation to school	3.33 [−1.74, 8.41]	2.50* [1.16, 3.84]	−1.67 [−4.1, 0.80]	0.00 [−3.7, 3.7]
PE	23.95 [−5.6, 53.6]	1.87 [−5.8, 9.5]	9.84* [1.5, 18.2]	−15.7* [−31.1, −0.2]
Recess	45.28* [24.1, 66.5]	5.90* [0.33, 11.5]	NA	−28.2* [−38.5, −17.8]
Lunch				
Transportation to home	12.50* [6.8, 18.2]	5.54* [3.3, 7.8]	−0.67 [−4.6, 3.3]	−5.17 [−10.9, 0.6]
Out-of-school (%MVPA)				
Before school	7.11* [3.1, 11.1]	0.92 [−0.1, 1.9]	−1.33 [−3.2, 0.6]	−4.44* [−7.7, 1.2]
After school	9.23* [4.4, 14.0]	3.30* [2.0, 4.6]	−2.83* [−5.1, −0.5]	−4.39* [−8.0, 0.7]
Evening	7.53* [4.2, 10.8]	1.91* [1.0, 2.8]	−2.41* [−3.9, −0.9]	−3.49* [−6.1, −0.9]
Weekend (%MVPA)				
Saturday	7.35* [3.8, 10.8]	2.63* [1.4, 3.8]	−4.09* [−5.6, −2.5]	−3.26* [−5.8, −0.7]
Sunday	7.77* [4.2, 11.3]	2.53* [1.3, 3.7]	−3.23* [−4.8, −1.6]	−3.85* [−6.4, −1.3]
Out-of-school (%SB)				
	41.58* [35.7, 47.4]	5.14* [1.6, 8.2]	12.01* [9.3, 14.7]	10.82* [6.5, 15.2]

Note. Regression coefficients were obtained to predict percentage of time spent in MVPA or percentage of time spent in sedentary generated from the accelerometer. The following coding was used in the prediction equations to estimate percentage of time in MVPA or SB–YAP: All items are scored from 0 to 4; grade level: 0 = *elementary school* (≤5th grade), 1 = *middle school* (6th–8th grades), and 2 = *high school* (≥9th grade); sex: 0 = *male* and 1 = *female*. The estimates of MVPA are obtained by multiplying the %MVPA value by the specific duration of time available for that segment. This may vary by application and setting. Note that calibration of the recess item was limited to elementary school participants. PE = physical education; CI = confidence interval; YAP = youth activity profile; NA = not applicable; MVPA = moderate-to-vigorous physical activity; SB = sedentary behavior.

* $p < .05$.

address variability due to weather and other factors). Finally, the study used a more naturalistic design with youth completing the online assessment with tablets or in school computer labs using a basic set of instructions and little support. These features add natural variability into the system, but lead to equations that are more generalizable for field-based research and professional applications.

The results support the utility of the online version and provide documentation of the bias and group level error in the estimations. The overall error (i.e., bias) averaged across the different segments was similar to previously reported outcomes with the print version (Saint-Maurice & Welk, 2015), but some differences were evident. For example, the bias for the estimate of PA in school was approximately 7 min/day compared with the previous value of 16 min/day. The bias for the out-of-school PA estimates was about 13 min/day compared with 4 min/day in the previous study. In interpreting these values, it is important to consider that these values reflect averaged errors across the whole sample and capitalize to some extent on the cancellation of errors. The equations may overestimate for some individuals and underestimate others so the small differences in averaged error are not particularly relevant. The MAPE values provide a better indicator of error since it reflects the expected level of precision when used in practice for a specific school.

The MAPE values for the four segments ranged from 10 to 23%, and these values are consistently better than those from the original study which ranged from 25 to 30%. Because error is population specific, we also tested the relative accuracy of the original and new algorithms in the same sample. The comparisons in Table 4 documented that the overall MAPE values for the new MVPA equations were lower than values resulting from the application of original equations to the same data (in-school: 9% lower, out-of-school: 24% lower, and weekend: 37% lower). The MAPE for the sedentary estimate was 12% higher, but when averaged, the MAPEs for the new equations are 15% lower than the previous values.

Comparisons by age revealed that the accuracy was somewhat stronger in middle school students than for elementary or high school students (see Table 4). This is perhaps due to a more structured school schedule for these students compared with those in lower or higher grades. Activity patterns in young children are intermittent and are therefore more difficult to capture (Engelen et al., 2015; Koning et al., 2018; Patnode et al., 2011), but an issue for older adolescents may be the larger variability in student activity profiles (Mayorga-Vega & Viciano, 2014; Troiano et al., 2008). We combined the data across all ages to enable age to be included in the model; however, age-specific models may provide advantages for further improving precision. For the in-

Table 4 Cross-Validation of Updated YAP Algorithms for Predicting Time Spent in MVPA (In-School, Out-of-School, and Weekend Segments) and Sedentary Time for the Combined Sample

Sample	Segment	SWA	YAP	ZOE	MAPE	Original ^a MAPE
Combined	In-school PA	29.6	22.6	7.2	23.2	25.6
	Out-of-school PA	54.8	41.9	13.4	21.0	27.8
	Weekend PA	88.1	76.6	12.4	17.1	27.1
	Sedentary time	105.9	108.6	3.0	8.4	7.5
Elementary school	In-school PA	31.6	24.7	7.2	21.1	25.4
	Out-of-school PA	68.9	52.3	17.2	22.5	16.0
	Weekend PA	107.1	96.9	11.3	16.4	21.1
	Sedentary time	95.7	98.7	3.7	9.9	9.2
Middle school	In-school PA	22.9	15.7	8.1	27.8	35.1
	Out-of-school PA	37.2	31.1	7.1	14.0	39.3
	Weekend PA	65.1	57.2	8.7	13.5	27.8
	Sedentary time	123.8	124.5	1.4	6.1	5.1
High school	In-school PA	33.8	26.8	7.8	22.8	11.7
	Out-of-school PA	39.0	26.9	14.1	27.2	46.1
	Weekend PA	65.9	44.0	25.2	24.9	44.0
	Sedentary time	106.3	113.3	—	6.6	5.3
Boys	In-school PA	36.5	31.2	5.6	15.5	17.5
	Out-of-school PA	68.3	53.3	15.6	23.3	23.4
	Weekend PA	106.1	86.7	21.0	20.0	15.6
	Sedentary time	104.0	99.8	4.7	8.3	8.2
Girls	In-school PA	26.8	18.5	8.9	40.5	45.8
	Out-of-school PA	48.0	37.4	11.6	21.3	40.1
	Weekend PA	75.6	62.9	13.9	16.1	30.7
	Sedentary time	112.1	121.8	10.3	10.8	8.2

Note. SWA = SenseWear armband; YAP = youth activity profile; ZOE = zone of equivalence; MAPE = mean absolute percentage error; PA = physical activity.

^aCalculations of MAPE from the original equations are provided for comparison.

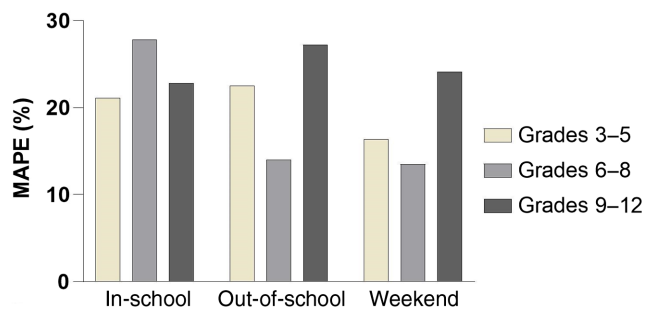


Figure 1 — Variability in MAPE values across time segments and age for PA items. MAPE = mean absolute percentage error; PA = physical activity.

school segment, MAPE was considerably lower for boys than for the girls (15% vs. 40%, respectively); however, no other discernible patterns were evident when comparing boys with girls. We included gender as a variable within the single model, but further research is warranted to determine whether gender-specific models would have advantages.

A number of other report-based tools have been developed to capture youth PA behaviors and different approaches have been used to create summarized outcomes. Methods that rely on youth to

report specific periods of time or to estimate intensity methods have proven to be challenging for youth to complete (Telford, Salmon, Jolley, & Crawford, 2004). The use of standard MET thresholds is also complicated by age-related differences in metabolism (Saint-Maurice, Kim, Welk, & Gaesser, 2016) as well as by the “cut point bias” due to thresholds in MET tables (Olds, Gomersall, Olds, & Ridley, 2019). There are fewer online tools available for youth, but the Multimedia Activity Recall for Children and Adolescents tool has been developed similarly to the YAP to enable youth to easily report behaviors (Foley, Maddison, Olds, & Ridley, 2012; Ridley, Olds, & Hill, 2006). The Multimedia Activity Recall for Children and Adolescents tool has good utility, but a relevant advantage of the YAP is that it relies on a calibration process to produce estimates of MVPA rather than directly using time values based on student reports or estimates. Because the YAP calibrations are item specific, it is possible to capture PA for different segments of the day. In addition, as models estimate the percentage of time spent, the estimates can also account for differences in durations of key segments such as PE or recess that may vary due to policy or programming efforts.

The current equations have reasonable accuracy, but presently only incorporate age and gender as predictors. Research has documented that PA patterns vary widely across a number of social, environmental, cultural, and regional dimensions (James-Burdumy et al., 2016; Singh, Kogan, & van Dyck, 2008), so further refinement may be needed to improve the generalizability of the

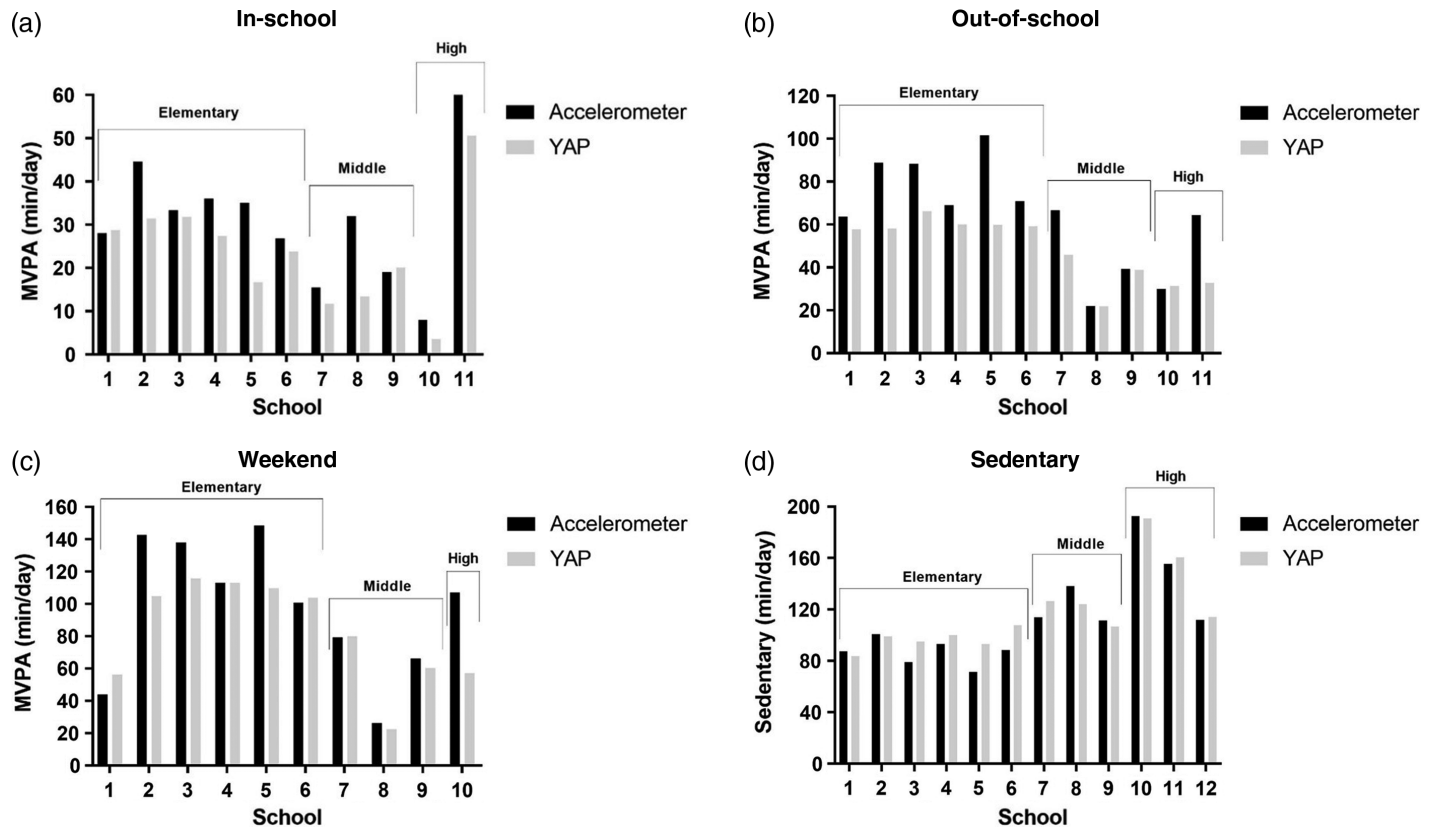


Figure 2 — Accelerometer and predicted (YAP) MVPA, stratified by school for in-school (a), out-of-school (b), weekend (c), and sedentary time (d). YAP = youth activity profile; MVPA = moderate-to-vigorous physical activity.

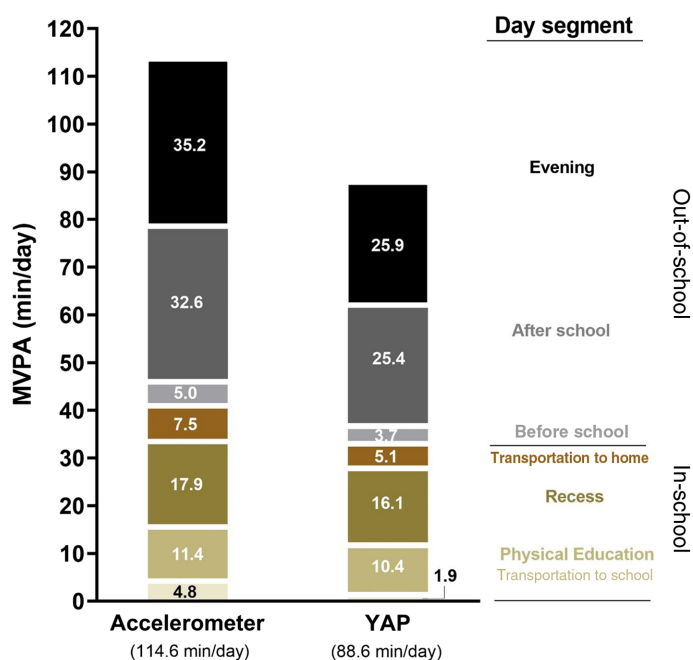


Figure 3 — Comparison of time allocations (in minutes per day) in various segments of the YAP relative to the estimates from the Sensewear Armband. YAP = youth activity profile; MVPA = moderate-to-vigorous physical activity.

predictions. We observed differences in agreement when equations developed in Iowa were applied to data from Texas (results not shown), and this is why a generalized approach was ultimately used. Therefore, it may be necessary to also consider geographic and sociodemographic variables in future research.

Results are also influenced by the calibration methodology, the choice of criterion, and other design-related factors in this type of research. In the present study, we utilized the item-specific calibration approach developed for the original YAP (Saint-Maurice et al., 2017). This approach offers more precision over the approach used in the FLASHE project (Saint-Maurice et al., 2017; Welk, Saint-Maurice, et al., 2017), which calibrated the YAP using longer “day segments” (e.g., in-school and out-of-school). The resulting MAPE values in FLASHE were considerably higher (in-school PA: 31.5%, out-of-school PA: 29.0%, and weekend PA: 32.5%), but that study also differed from the present one, since results were anchored to data from wrist-worn ActiGraph monitors (Pensacola, FL). There are documented limitations with the accuracy of available wrist-based methods (Ellingson et al., 2017; Kim et al., 2017; Lynch et al., 2019), so the results in FLASHE may also be influenced by this difference.

Our evaluation of the predictive accuracy of the new models focused on the characteristics of school averages because the focus of the YAP is on providing school-level assessments rather than attempting to provide predictions for specific individuals. While our cross-validation sample used an independent (hold out) sample, the same schools were used in the calibration and cross-validation samples. Therefore, the accuracy at the school level may be slightly overestimated. An alternative cross-validation strategy would be to

randomly partition schools into those in the calibration data set and those in the cross-validation data set. This strategy was used in the U.K. calibration study (Fairclough et al., 2019) but was not useful for our study because of the extreme range of number of individuals per school, ranging from 2 to 97.

Our estimates of error were similar to those in the U.K. calibration study (Fairclough et al., 2019), which reported statistical equivalence zones ranging from 15 to 20%. The present study and the study conducted in the United Kingdom both used the Sensewear armband; therefore, it is possible that results are more similar due to the use of the same comparison/criterion measure. We selected the SWA for our original work (and this study) because research has documented the accuracy of the SWA for predicting energy expenditure in youth sample (Calabro et al., 2013; Lee, Kim, Bai, Gaesser, & Welk, 2016; Lopez et al., 2018). It is also reliable, easy to use, and provides clear indications of wear time to facilitate the compliance check. These features provide advantages for the calibration process, but since the SWA is no longer available for purchase, it will be important to eventually test how the estimates compare with other criterion measurements. The calibration of the YAP is linked to (and predicated on) the accuracy and precision of the criterion method, but with open-source methods evolving in the future, it can potentially be recalibrated as the field builds consensus on methods for processing monitor-based methods (Keadle, Lyden, Strath, Staudenmayer, & Freedson, 2019; Rowlands, 2018).

An important and encouraging finding from the present analyses is the correspondence in overall school level estimates. As shown in Figure 2, the estimated values from the YAP matched the observed age and gender profiles obtained from the SWA to a reasonable extent. Figure 3 also documents good overall agreement on an item-specific level for a typical school day. Collectively, these results support the use of the YAP as an outcome measure in school-based studies where the emphasis is placed on examining change over time or differences between groups. The YAP is currently being used in several large school-based studies to provide population estimates of PA and SB. For example, in the School Wellness Integration Targeting Child Health® project (Chen et al., 2018), the YAP is used in formative ways to guide school programming and in summative ways to evaluate change. A more recent study from this project (McLoughlin et al., 2019) documented that youth who tracked their behaviors more regularly had larger gains in YAP scores than youth who were less engaged. These applications document the utility and the sensitivity to change for group level reporting. However, the high MAPE values observed for individual estimation observed in this study and the U.K. study by Fairclough et al. (2019) demonstrate that additional research is needed to have sufficient precision to enable estimation at the individual level.

Although the results support the utility of the online version of the YAP, the equations need additional refinement. The equations are based on a relatively small sample and only model variability by age and gender, so additional work is needed to evaluate the utility in segments of the population and across region and season. In considering these needs, it is important to consider the similar challenges in training machine learning algorithms to model data from monitor-based methods. Algorithms can only be improved when they are provided with more data. Thus, with continued refinement, the calibration of the YAP offers a cost-effective method to enhance understanding of youth PA and SB patterns, leading to potential use in surveillance research and filling a much-needed gap in the literature (Dunton et al., 2019). The incorporation

of robust measurement error models prior to calibration would be particularly important to provide the precision for surveillance applications and this work is in progress.

Overall, the results demonstrate reasonable group level estimates of PA and SB, but it is important to understand the potential limitations of the regression-based models. The assessment is administered and collected at the individual level, but it is unrealistic to try to accurately predict individual behaviors with these simple items. Validity is also a relative term, and there is still considerable error in the estimates—even at the group level. Therefore, it is important to understand the intended application and purpose of the tool. In the current form, the value of the YAP is that it provides a feasible way to capture this data on groups of children using an efficient and educationally sound method for school applications. The intended application is for capturing group level estimates for research and practice applications. While it is common to apply methods to estimate PA at the individual level, many research applications only necessitate an accurate group level estimate.

The refined calibration of the online version of the YAP provides new opportunities to advance school-based research and public health surveillance since it can be used to capture accurate group-level estimates of PA and SB from large numbers of youth in an efficient and cost-effective manner. We envision that researchers can use the YAP to assess group-level outcomes of intervention studies, detect differences among groups within ongoing studies, and as a means to capture patterns and trends for surveillance applications. For teachers and schools, we envision the YAP as a means to gather data on classes at different time points across the school year; such data would provide impetus for programmatic and policy-related decisions related to student health and wellness through data-driven decision-making processes (Marsh & Farrell, 2015).

Acknowledgment

The study was funded by a grant from the National Institutes of Health (R21CA188641).

Notes

1. The YAP is a copyrighted online tool but can be accessed for free by teachers through a collaboration with Active Schools. Customized versions can be developed for research or other applications. Visit www.youthactivityprofile.org or email the project team at yap@iastate.edu.
2. We considered alternative models without intercepts and versions that would rescale values to produce 0 min of MVPA for YAP scores of 0, but the selected model yielded the smallest MAPE scores across all responses in the calibration data set.

References

- Arvidsson, D., Slinde, F., & Hulthén, L. (2009). Free-living energy expenditure in children using multi-sensor activity monitors. *Clinical Nutrition*, 28(3), 305–312. PubMed ID: 19345453 doi:10.1016/j.clnu.2009.03.006
- Baranowski, T., & Domel, S.B. (1994). A cognitive model of children's reporting of food intake. *The American Journal of Clinical Nutrition*, 59(1), 212S–217S. doi:10.1093/ajcn/59.1.212S
- Beyler, N., Fuller, W., Nusser, S., & Welk, G. (2015). Predicting objective physical activity from self-report surveys: A model validation study

- using estimated generalized least-squares regression. *Journal of Applied Statistics*, 42(3), 555–565. doi:[10.1080/02664763.2014.978271](https://doi.org/10.1080/02664763.2014.978271)
- Bowles, H.R. (2012). Measurement of active and sedentary behaviors: Closing the gaps in self-report methods. *Journal of Physical Activity and Health*, 9(Suppl. 1), S1–S4. doi:[10.1123/jpah.9.s1.s1](https://doi.org/10.1123/jpah.9.s1.s1)
- Calabro, M.A., Stewart, J., & Welk, G.J. (2013). Validation of pattern-recognition monitors in children using doubly labeled water. *Medicine & Science in Sports & Exercise*, 45(7), 1313–1322. PubMed ID: [23299766](https://pubmed.ncbi.nlm.nih.gov/23299766/) doi:[10.1249/mss.0b013e31828579c3](https://doi.org/10.1249/mss.0b013e31828579c3)
- Calabro, M.A., Welk, G.J., & Eisenmann, J. (2009). Validation of the sensewear pro armband algorithms in children. *Medicine & Science in Sports & Exercise*, 41(9), 1714–1720. PubMed ID: [19657300](https://pubmed.ncbi.nlm.nih.gov/19657300/) doi:[10.1249/mss.0b013e3181a071cf](https://doi.org/10.1249/mss.0b013e3181a071cf)
- Chen, S., Dziewaltowski, D.A., Rosenkranz, R.R., Lanningham-Foster, L., Vazou, S., Gentile, D.A., ... Welk, G.J. (2018). Feasibility study of the SWITCH implementation process for enhancing school wellness. *BMC Public Health*, 18(1), 1119. PubMed ID: [30217186](https://pubmed.ncbi.nlm.nih.gov/30217186/) doi:[10.1186/s12889-018-6024-2](https://doi.org/10.1186/s12889-018-6024-2)
- Dixon, P.M., Saint-Maurice, P.F., Kim, Y., Hibbing, P., Bai, Y., & Welk, G.J. (2018). A primer on the use of equivalence testing for evaluating measurement agreement. *Medicine & Science in Sports & Exercise*, 50(4), 837–845. PubMed ID: [29135817](https://pubmed.ncbi.nlm.nih.gov/29135817/) doi:[10.1249/MSS.0000000000001481](https://doi.org/10.1249/MSS.0000000000001481)
- Dunton, G.F., Berrigan, D., Young, D.R., Pfeiffer, K.A., Lee, S.M., Slater, S.J., & Pate, R.R. (2019). Strategies to improve physical activity surveillance among youth in the United States. *The Journal of Pediatrics*, 210, 226–231. PubMed ID: [31078282](https://pubmed.ncbi.nlm.nih.gov/31078282/) doi:[10.1016/j.jpeds.2019.03.022](https://doi.org/10.1016/j.jpeds.2019.03.022)
- Ekelund, U., Tomkinson, G., & Armstrong, N. (2011). What proportion of youth are physically active? Measurement issues, levels and recent time trends. *British Journal of Sports Medicine*, 45(11), 859–865. PubMed ID: [21836170](https://pubmed.ncbi.nlm.nih.gov/21836170/) doi:[10.1136/bjsports-2011-090190](https://doi.org/10.1136/bjsports-2011-090190)
- Ellingson, L.D., Hibbing, P.R., Kim, Y., Frey-Law, L.A., Saint-Maurice, P.F., & Welk, G.J. (2017). Lab-based validation of different data processing methods for wrist-worn ActiGraph accelerometers in young adults. *Physiological Measurement*, 38(6), 1045. PubMed ID: [28481750](https://pubmed.ncbi.nlm.nih.gov/28481750/) doi:[10.1088/1361-6579/aa6d00](https://doi.org/10.1088/1361-6579/aa6d00)
- Engelen, L., Bundy, A., Lau, J., Naughton, G., Wyver, S., Bauman, A., & Baur, L. (2015). Understanding patterns of young children's physical activity after school—it's all about context: A cross-sectional study. *Journal of Physical Activity and Health*, 12(3), 335–339. PubMed ID: [24828415](https://pubmed.ncbi.nlm.nih.gov/24828415/) doi:[10.1123/jpah.2013-0153](https://doi.org/10.1123/jpah.2013-0153)
- Fairclough, S.J., Christian, D.L., Saint-Maurice, P.F., Hibbing, P.R., Noonan, R.J., Welk, G.J., ... Boddy, L.M. (2019). Calibration and validation of the Youth Activity Profile as a physical activity and sedentary behaviour surveillance tool for English youth. *International Journal of Environmental Research and Public Health*, 16(19), 3711. doi:[10.3390/ijerph16193711](https://doi.org/10.3390/ijerph16193711)
- Foley, L., Maddison, R., Olds, T., & Ridley, K. (2012). Self-report use-of-time tools for the assessment of physical activity and sedentary behaviour in young people: Systematic review. *Obesity Reviews*, 13(8), 711–722. PubMed ID: [22429291](https://pubmed.ncbi.nlm.nih.gov/22429291/) doi:[10.1111/j.1467-789X.2012.00993.x](https://doi.org/10.1111/j.1467-789X.2012.00993.x)
- He, X., & Hu, F. (2002). Markov chain marginal bootstrap. *Journal of the American Statistical Association*, 97(459), 783–795 doi:[10.1198/016214502388618591](https://doi.org/10.1198/016214502388618591)
- Hidding, L.M., Altenburg, T.M., Mokkink, L.B., Terwee, C.B., & Chinapaw, M.J.M. (2017). Systematic review of childhood sedentary behavior questionnaires: What do we know and what is next? *Sports Medicine*, 47(4), 677–699. PubMed ID: [27577686](https://pubmed.ncbi.nlm.nih.gov/27577686/) doi:[10.1007/s40279-016-0610-1](https://doi.org/10.1007/s40279-016-0610-1)
- James-Burdumy, S., Beyler, N., Borradaile, K., Bleeker, M., Maccarone, A., & Fortson, J. (2016). The impact of playworks on students' physical activity by race/ethnicity: Findings from a randomized controlled trial. *Journal of Physical Activity and Health*, 13(3), 275–280. PubMed ID: [26284816](https://pubmed.ncbi.nlm.nih.gov/26284816/) doi:[10.1123/jpah.2014-0607](https://doi.org/10.1123/jpah.2014-0607)
- Keadle, S.K., Lyden, K.A., Strath, S.J., Staudenmayer, J.W., & Freedson, P.S. (2019). A framework to evaluate devices that assess physical behavior. *Exercise and Sport Sciences Reviews*, 47(4), 206–214. PubMed ID: [31524786](https://pubmed.ncbi.nlm.nih.gov/31524786/) doi:[10.1249/JES.0000000000000206](https://doi.org/10.1249/JES.0000000000000206)
- Kim, Y., Hibbing, P., Saint-Maurice, P.F., Ellingson, L.D., Hennessy, E., Wolff-Hughes, D.L., ... Welk, G.J. (2017). Surveillance of youth physical activity and sedentary behavior with wrist accelerometry. *American Journal of Preventive Medicine*, 52(6), 872–879. PubMed ID: [28526364](https://pubmed.ncbi.nlm.nih.gov/28526364/) doi:[10.1016/j.amepre.2017.01.012](https://doi.org/10.1016/j.amepre.2017.01.012)
- Koenker, R. (2019). Quantile regression in R: A vignette. Retrieved from <https://cran.r-project.org/web/packages/quantreg/vignettes/rq.pdf>
- Koning, M., de Jong, A., de Jong, E., Visscher, T.L.S., Seidell, J.C., & Renders, C.M. (2018). Agreement between parent and child report of physical activity, sedentary and dietary behaviours in 9–12-year-old children and associations with children's weight status. *BMC Psychology*, 6(1), 14. PubMed ID: [29631618](https://pubmed.ncbi.nlm.nih.gov/29631618/) doi:[10.1186/s40359-018-0227-2](https://doi.org/10.1186/s40359-018-0227-2)
- Lee, J.-M., Kim, Y., Bai, Y., Gaesser, G.A., & Welk, G.J. (2016). Validation of the SenseWear mini armband in children during semi-structure activity settings. *Journal of Science and Medicine in Sport*, 19(1), 41–45. PubMed ID: [25459233](https://pubmed.ncbi.nlm.nih.gov/25459233/) doi:[10.1016/j.jsams.2014.10.004](https://doi.org/10.1016/j.jsams.2014.10.004)
- Lopez, G.A., Brønd, J.C., Andersen, L.B., Dencker, M., & Arvidsson, D. (2018). Validation of SenseWear Armband in children, adolescents, and adults. *Scandinavian Journal of Medicine & Science in Sports*, 28(2), 487–495. PubMed ID: [28543847](https://pubmed.ncbi.nlm.nih.gov/28543847/) doi:[10.1111/sms.12920](https://doi.org/10.1111/sms.12920)
- Lynch, B.A., Kaufman, T.K., Rajjo, T.I., Mohammed, K., Kumar, S., Murad, M.H., ... Levine, J.A. (2019). Accuracy of accelerometers for measuring physical activity and levels of sedentary behavior in children: A systematic review. *Journal of Primary Care & Community Health*, 10, 1–8. doi:[10.1177/2150132719874252](https://doi.org/10.1177/2150132719874252)
- Marsh, J.A., & Farrell, C.C. (2015). How leaders can support teachers with data-driven decision making. *Educational Management Administration & Leadership*, 43(2), 269–289. doi:[10.1177/1741143214537229](https://doi.org/10.1177/1741143214537229)
- Mayorga-Vega, D., & Viciano, J. (2014). Adolescents' physical activity in physical education, school recess, and extra-curricular sport by motivational profiles. *Perceptual and Motor Skills*, 118(3), 663–679. PubMed ID: [25068739](https://pubmed.ncbi.nlm.nih.gov/25068739/) doi:[10.2466/06.10.PMS.118k26w0](https://doi.org/10.2466/06.10.PMS.118k26w0)
- McLoughlin, G.M., Rosenkranz, R.R., Lee, J.A., Wolff, M.M., Chen, S., Dziewaltowski, D.A., ... Welk, G.J. (2019). The importance of self-monitoring for behavior change in youth: Findings from the SWITCH® school wellness feasibility study. *International Journal of Environmental Research and Public Health*, 16(20), 3806. doi:[10.3390/ijerph16203806](https://doi.org/10.3390/ijerph16203806)
- Nusser, S.M., Beyler, N.K., Welk, G.J., Carriquiry, A.L., Fuller, W.A., & King, B.M.N. (2012). Modeling errors in physical activity recall data. *Journal of Physical Activity and Health*, 9(Suppl. 1), S56–S67. doi:[10.1123/jpah.9.s1.s56](https://doi.org/10.1123/jpah.9.s1.s56)
- Olds, T.S., Gomersall, S.R., Olds, S.T., & Ridley, K. (2019). A source of systematic bias in self-reported physical activity: The cutpoint bias hypothesis. *Journal of Science and Medicine in Sport*, 22(8), 924–928. PubMed ID: [30956062](https://pubmed.ncbi.nlm.nih.gov/30956062/) doi:[10.1016/j.jsams.2019.03.006](https://doi.org/10.1016/j.jsams.2019.03.006)
- Patnode, C.D., Lytle, L.A., Erickson, D.J., Sirard, J.R., Barr-Anderson, D.J., & Story, M. (2011). Physical activity and sedentary activity patterns among children and adolescents: A latent class analysis

- approach. *Journal of Physical Activity and Health*, 8(4), 457–467. PubMed ID: 21597117 doi:10.1123/jpah.8.4.457
- Pedišić, Ž., & Bauman, A. (2015). Accelerometer-based measures in physical activity surveillance: Current practices and issues. *British Journal of Sports Medicine*, 49(4), 219–223. PubMed ID: 25370153 doi:10.1136/bjsports-2013-093407
- Ridley, K., Olds, T.S., & Hill, A. (2006). The multimedia activity recall for children and adolescents (MARCA): Development and evaluation. *International Journal of Behavioral Nutrition and Physical Activity*, 3(1), 10. doi:10.1186/1479-5868-3-10
- Robin, J., & Moscovitch, M. (2014). The effects of spatial contextual familiarity on remembered scenes, episodic memories, and imagined future events. *Journal of Experimental Psychology: Learning Memory and Cognition*, 40(2), 459–475. doi:10.1037/a0034886
- Rowlands, A.V. (2018). Moving forward with accelerometer-assessed physical activity: Two strategies to ensure meaningful, interpretable, and comparable measures. *Pediatric Exercise Science*, 30(4), 450–456. PubMed ID: 30304982 doi:10.1123/pes.2018-0201
- Saint-Maurice, P.F., Kim, Y., Hibbing, P., Oh, A.Y., Perna, F.M., & Welk, G.J. (2017). Calibration and validation of the youth activity profile: The FLASHE study. *American Journal of Preventive Medicine*, 52(6), 880–887. PubMed ID: 28526365 doi:10.1016/j.amepre.2016.12.010
- Saint-Maurice, P.F., Kim, Y., Welk, G.J., & Gaesser, G.A. (2016). Kids are not little adults: What MET threshold captures sedentary behavior in children? *European Journal of Applied Physiology*, 116(1), 29–38. PubMed ID: 26271677 doi:10.1007/s00421-015-3238-1
- Saint-Maurice, P.F., Sousa, S., Welk, G.J., Matthews, C.E., & Berrigan, D. (2020). Report-based measures of physical activity: Features, considerations, and resources. In S.J. Fairclough, T.A. Brusseau, & D.R. Lubans (Eds.), *Handbook of youth physical activity*. Routledge: London, UK.
- Saint-Maurice, P.F., & Welk, G.J. (2014). Web-based assessments of physical activity in youth: Considerations for design and scale calibration. *Journal of Medical Internet Research*, 16(12), e269–269. doi:10.2196/jmir.3626
- Saint-Maurice, P.F., & Welk, G.J. (2015). Validity and calibration of the youth activity profile. *PLoS One*, 10(12), e0143949. doi:10.1371/journal.pone.0143949
- Saint-Maurice, P.F., Welk, G.J., Bartee, R.T., & Heelan, K. (2017). Calibration of context-specific survey items to assess youth physical activity behaviour. *Journal of Sports Sciences*, 35(9), 866. PubMed ID: 27326748 doi:10.1080/02640414.2016.1194526
- Saint-Maurice, P.F., Welk, G.J., Beyler, N.K., Bartee, R.T., & Heelan, K.A. (2014). Calibration of self-report tools for physical activity research: The physical activity questionnaire (PAQ). *BMC Public Health*, 14(1), 461. doi:10.1186/1471-2458-14-461
- Singh, G.K., Kogan, M.D., & van Dyck, P.C. (2008). A multilevel analysis of state and regional disparities in childhood and adolescent obesity in the United States. *Journal of Community Health*, 33(2), 90–102. PubMed ID: 18049885 doi:10.1007/s10900-007-9071-7
- Spruijt-Metz, D., Wen, C.K.F., Bell, B.M., Intille, S., Huang, J.S., & Baranowski, T. (2018). Advances and controversies in diet and physical activity measurement in youth. *American Journal of Preventive Medicine*, 55(4), e81–e91. PubMed ID: 30135037 doi:10.1016/j.amepre.2018.06.012
- Telford, A., Salmon, J., Jolley, D., & Crawford, D. (2004). Reliability and validity of physical activity questionnaires for children: The children's leisure activities study survey (CLASS). *Pediatric Exercise Science*, 16(1), 64–78. doi:10.1123/pes.16.1.64
- Troiano, R.P. (2005). A timely meeting: Objective measurement of physical activity. *Medicine & Science in Sports & Exercise*, 37(11), S487–S489. doi:10.1249/01.mss.0000185473.32846.c3
- Troiano, R.P., Berrigan, D., Dodd, K.W., Mâsse, L.C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise*, 40(1), 181–188. PubMed ID: 18091006 doi:10.1249/mss.0b013e31815a51b3
- Troiano, R.P., Gabriel, K.K.P., Welk, G.J., Owen, N., & Sternfeld, B. (2012). Reported physical activity and sedentary behavior: Why do you ask? *Journal of Physical Activity and Health*, 9(Suppl. 1), S68–S75. doi:10.1123/jpah.9.s1.s68
- Troiano, R.P., McClain, J.J., Brychta, R.J., & Chen, K.Y. (2014). Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine*, 48(13), 1019–1023. PubMed ID: 24782483 doi:10.1136/bjsports-2014-093546
- Welk, G.J. (2019). Harmonizing monitor- and report-based estimates of physical activity through calibration. *Kinesiology Review*, 8(1), 16–24. doi:10.1123/kr.2018-0064
- Welk, G.J., Bai, Y., Lee, J.-M., Godino, J.O.B., Saint-Maurice, P.F., & Carr, L. (2019). Standardizing analytic methods and reporting in activity monitor validation studies. *Medicine & Science in Sports & Exercise*, 51(8), 1767–1780. PubMed ID: 30913159 doi:10.1249/MSS.0000000000001966
- Welk, G.J., Bai, Y., Saint-Maurice, P.F., Allums-Featherston, K., & Candelaria, N. (2016). Design and evaluation of the NFL PLAY 60 FITNESSGRAM partnership project. *Research Quarterly for Exercise and Sport*, 87(1), 1. PubMed ID: 26889580 doi:10.1080/02701367.2015.1127126
- Welk, G.J., Beyler, N.K., Kim, Y., & Matthews, C.E. (2017). Calibration of self-report measures of physical activity and sedentary behavior. *Medicine & Science in Sports & Exercise*, 49(7), 1473–1481. PubMed ID: 28240704 doi:10.1249/MSS.0000000000001237
- Welk, G.J., Corbin, C.B., & Dale, D. (2000). Measurement issues in the assessment of physical activity in children. *Research Quarterly for Exercise and Sport*, 71(Suppl. 2), 59–73. doi:10.1080/02701367.2000.11082788
- Welk, G.J., Morrow, J., & Saint-Maurice, P.F. (2017). *Measures registry user guide: Individual physical activity*. Washington, DC: National Collaborative on Childhood Obesity Research. Retrieved from http://nccor.org/tools-mruserguides/wp-content/uploads/2017/NCCOR_MR_User_Guide_Individual_PA-FINAL.pdf
- Welk, G.J., Saint-Maurice, P.F., Kim, Y., Ellingson, L.D., Hibbing, P., Wolff-Hughes, D.L., & Perna, F.M. (2017). Understanding and interpreting error in physical activity data: Insights from the FLASHE study. *American Journal of Preventive Medicine*, 52(6), 836–838. PubMed ID: 28526359 doi:10.1016/j.amepre.2017.03.001