Research Quarterly for Exercise and Sport

ISSN: 0270-1367 (Print) 2168-3824 (Online) Journal homepage: http://www.tandfonline.com/loi/urqe20

Top 10 Research Questions Related to Assessing Physical Activity and Its Contexts Using Systematic Observation

Thomas L. McKenzie & Hans van der Mars

To cite this article: Thomas L. McKenzie & Hans van der Mars (2015) Top 10 Research Questions Related to Assessing Physical Activity and Its Contexts Using Systematic Observation, Research Quarterly for Exercise and Sport, 86:1, 13-29, DOI: 10.1080/02701367.2015.991264

To link to this article: http://dx.doi.org/10.1080/02701367.2015.991264

Published online: 09 Feb 2015.

Research Quarterly for Exercise and Sport, *86*, 13–29, 2015

ISSN 0270-1367 print/ISSN 2168-3824 online
DOI: 10.1080/02701367.2015.991264

Top 10 Research Questions Related to Assessing Physical Activity and Its Contexts Using Systematic Observation

Thomas L. McKenzie
San Diego State University

Hans van der Mars
Arizona State University

Numerous methods are available to assess physical activity (PA) but systematic observation (SO) excels in being able to provide contextually rich data on the setting in which the activity occurs. As SO is particularly useful for determining how activity is influenced by the immediate physical and social environments, its use is becoming more popular. Observation tools have the advantages of flexibility, high internal validity, low inference, and low participant burden, while their disadvantages include the need for careful observer training and recalibration, inaccessibility to certain environments, and potential participant reactivity. There is a need for both scientists and practitioners to have additional information on observation techniques and systems relative to making environmental and policy decisions about PA, and in this article, we describe concepts and identify questions related to using SO in researching PA behavior. We present 10 general questions in 3 sections, including those related to: (a) ensuring data accuracy through the selection of the most appropriate methodological protocols; (b) investigating PA in school settings, including physical education, recess, and other programs; and (c) investigating PA in community settings (e.g., parks, recreation centers, youth and adult sport programs) and homes.

Keywords: environment, exercise, observational research, research methods

Physical activity (PA) is a behavior that is an important contributor to the health of people of all ages (U.S. Department of Health and Human Services [USDHHS], 2008a, 2008b). It does not occur in a vacuum, but is place-based, and to fully understand it and especially to implement effective interventions, the context in which PA occurs must be considered. Systematic observation (SO) exceeds all other measures of PA in identifying the physical and social contexts in which it occurs. This article provides insight into using SO as an approach to collecting PA data and its associated variables and offers direction for future research in the area. Prior to presenting our Top 10 categories of questions, we provide a brief overview of why SO is considered a primary means of studying PA and identify key considerations for using it, including the selection of coding protocols, observer reliability, validity, and training, and how much and how often data need to be collected. We also highlight several misconceptions regarding the use of SO.

As a behavior, PA takes on many diverse forms of engaging in bodily movement that result in energy expenditure. It is inherently transitory and multidimensional in nature, and there is substantial intraperson and interperson variability relative to its frequency, intensity, duration, mode, and location. Nonetheless, PA can be observed directly, and with proper training, observers can deliver accurate data that are reliable and valid. SO is also adept at measuring sedentary behavior (e.g., Marshall & Merchant, 2013), which is now considered not just the opposite of PA but to have its own independent—but

Correspondence should be addressed to Thomas L. McKenzie, Professor Emeritus, 5127 Walsh Way, San Diego, CA 92115-1148. E-mail: tmckenzie@sdsu.edu

negative—influence on health (Owen, Healy, Howard, & Dunstan, 2012).

Advantages of Using Systematic Observation to Assess Physical Activity

Primary advantages of choosing SO over other methods of assessing PA (e.g., self-reports, accelerometers, pedometers, heart rate monitoring, doubly labeled water) are that it is a direct method and it allows for the simultaneous generation of information on both the physical environment and social environment. As a direct method, it can provide objective information that has strong internal (or face) validity (i.e., WYSIWYG—that is, “What You See Is What You Get”). In contrast, heart rate monitoring and doubly labeled water do not actually assess PA, pedometers and accelerometers do not function well in aquatic and strength-training environments, and self-reports assess only perceptions of PA engagement and often these are made distal to the location and occurrence of the PA.

Sallis (2009) has described PA as being “place-dependent” in that it always occurs in specific locations, each with uniquely built or natural physical and social characteristics. Factors such as access to facilities; presence of equipment and amenities; adult supervision in parks, schools, and homes; and level of structure/organization and fees for activity programs are all mediating variables for PA (e.g., McKenzie, Cohen, Sehgal, Williamson, & Golinelli, 2006; McKenzie, Marshall, Sallis, & Conway, 2000), and they can be observed. Thus, SO is an especially attractive data collection approach for researchers using cognitive-behavioral and ecological frameworks to advance the understanding of how environmental interventions impact PA (Sallis, Owen, & Fisher, 2008). The environment is a powerful determinant of PA (Sallis, Floyd, Rodriguez, & Saelens, 2012), and understanding it serves as the basis of key intervention strategies aimed at reducing the nation’s health burdens (Bauer, Briss, Goodman, & Bowman, 2014). SO can be used to investigate PA in mostly any setting (including aquatic environments), and there is little, if any, burden on the person(s) being assessed; and in some cases such as in open park environments, people may not realize their PA is being observed. As well, the data obtained using SO are usually in a format (e.g., frequency, minutes, walking) that is easily understood by practitioners, administrative personnel, and policymakers without much need for interpretation by researchers. Additionally, as addressed further in Question 4, advances in computer hardware and software technology now enable researchers to more readily enter, store, and analyze data, including via compact handheld devices such as tablets.

Limitations of Systematic Observation

SO is not without limitations. Observers may not always have access to all locations where people might be active (e.g., practice sessions of professional athletes, private clubs), and there is potential for observers to be biased (even unintentionally) and for participants to behave differently in the presence of an observer (i.e., reactivity). Time is required for observers to travel to and collect data in the target environment; thus, SO is often viewed as labor-intensive (e.g., Montoye, Kemper, Saris, & Washburn, 1996). Also, time is needed for ensuring observers are trained properly prior to data collection, and similar to mechanical and digital devices, observers need to be maintained and recalibrated throughout a study. These and other potential sources of observer error are highlighted later in this article.

SO often remains overlooked as a viable method for conducting research on PA behavior (e.g., Corder, Ekelund, Steele, Wareham, & Brage, 2008; Strath et al., 2013). A possible reason that SO is ignored is that researchers are often not familiar with the specific SO processes and protocols. Another reason is that many PA researchers are trained from a physiological perspective, and thus, energy expenditure is used as the central outcome variable rather than PA itself. Studying PA from socioecological or (eco) behavioral perspectives is fundamentally different from using a physiological approach, especially because there is an eye toward identifying the role of concurrent antecedent and consequential stimuli that either support or suppress PA (e.g., Cooper, Heron, & Heward, 2007). It is this detailed perspective that makes SO such an effective data collection method for intervention studies that include environmental and reinforcement components.

SECTION 1: METHODOLOGICAL QUESTIONS

As in all scientific endeavors, methodological data validity (or accuracy) is a central criterion that must be met for results to be credible. In the context of observing PA, validity refers to the ability of a tool to produce data that accurately represent the PA of the people observed and the context they are in. Questions 1 through 4 address different methodological aspects of using SO in this regard. All four questions focus on ensuring that the data collected provide an accurate representation of (a) the type, volume, and intensity of PA (and sedentary) behavior; and (b) the environmental contexts within which people are engaging in PA.

1. Which Observation Tactics Should Be Used for Assessing Physical Activity and Its Contexts?

SO is one of several methods available to measure PA; others include self-report questionnaires, interviews, pedometry, accelerometry, heart rate monitoring, Global Positioning Systems (GPS), and doubly labeled water, among others (Corder et al., 2008; Welk, 2002). Each method has its advantages and limitations (e.g., cost, ease of use, accuracy, focus). As an example, self-report questionnaires have often been the method of choice in large-scale population studies. Using self-reports with children, however, is problematic given that they are challenged with accurately understanding concepts and recalling the amount, duration, and type of activities. GPS data are limited in that they typically measure only location and time and distance traveled and they do not work well indoors. Doubly labeled water is useful for estimating energy expenditure but is expensive and provides little information on PA, and thus, is rarely used in field studies.

SO, which involves the observation and recording of observable events using standard procedures, typically uses one or more of the four basic observation tactics: event recording (ER), duration recording (DR), interval recording (IR), and momentary time sampling (MTS). ER and DR are preferred tactics as they can produce the most accurate depiction of PA behavior relative to its frequency, duration, and intensity. Researchers, however, are often interested in recording many different behaviors simultaneously (e.g., student PA, lesson context, teacher behavior), so IR and/or MTS would be more advantageous. Table 1 includes information on the basic features of each observation tactic.

ER provides a frequency count of behavior, and raw data are generally converted to rate per minute, percent of total, and sometimes as ratios. ER would be useful in determining events such as the number of people engaged in PA at various levels of intensity and in particular activity venues (e.g., gymnasiums, playgrounds). Relative to contextual variables, researchers can use ER to determine aspects such as amount and type of equipment available, number of activity venues that are accessible and usable, and the type of activities occurring.

DR, in contrast, provides information on the length of time a behavior lasted and should be employed when measuring (chains of) behaviors that are continuous and/or can last for extended periods of time. Typically, such data are expressed in minutes or percent of observed time that people engaged in PA at specific intensity levels (e.g., sedentary, moderate, vigorous).

IR allows observers to measure the (non)occurrence of PA during specified time intervals. Interval length, typically from 3 s to 10 s, depends on the complexity of the observation system (e.g., number of PA categories). Raw IR data are typically converted to percent of intervals; and unlike ER and DR, IR data can be used to estimate both frequency and duration. When using IR, investigators choose between "whole-interval" and "partial-interval" recording. During whole-interval recording, the behavior must be present for the entire length of the interval; on the other hand, even a fleeting occurrence of the behavior would be coded when using partial-interval recording. With partial-interval recording, observers choose between two cuing formats. The "record-only" format informs the observer to record the behavior of interest at a prerecorded "record" cue and is appropriate if three or fewer behaviors are to be charted. The "observe/record" cuing format paces observers to observe for a set time and allows them time to record the behaviors.

TABLE 1
Basic Features of Systematic Observation Tactics

Features	*Event Recording*	*Duration Recording*	*Interval Recording*	*Momentary Time Sampling*
Focus	Behavior/event frequency or occurrence	Behavior/event length	Occurrence of behavior *during* predetermined intervals	Occurrence of behavior *at the end* of predetermined intervals (which can be standard or variable)
Data units presented	Number; rate per unit of time (e.g., minutes); percent of total; ratio	Time (minutes, hours); percent of observed time	Percent of intervals; estimated frequency or duration	Percent of people observed; percent of events observed
Type of behavior targeted	Short-duration PA behavior (e.g., throws, tennis strokes, jump shot)	Continuous PA behavior (e.g., rope jumping, jogging, cycling, swimming)	PA behavior that is variable in frequency and duration (e.g., soccer, hockey game play)	PA behavior that is variable in frequency and duration (e.g., soccer, hockey game play)
Interval length	NA	NA	3 s–10 s	1 min to 60 min (possibly longer)
Standard variation(s)	NA	NA	a. Whole-interval recording b. Partial-interval recording	a. Sampling of individual's behavior b. Sampling of group behavior ("group time sampling")
Cuing format	NA	NA	a. "Record-only" format b. "Observe/record" format	NA

Note. PA = physical activity; NA = not applicable.

MTS is a recording tactic where the coding decision on a behavior occurs exactly on what is happening at the end of each interval. This tactic can be used to measure the occurrence of behavior of either individuals or groups, with the latter case called "group time sampling." Interval lengths with MTS generally range from 1 min to 60 min. While waiting for the end of an interval, observers can collect other relevant data (e.g., environmental conditions). When individuals are being observed, the raw data are usually converted to percent of intervals, and when groups are being observed, the number of people engaged in a particular behavior is converted to the percent of the total number of observed people across all samples taken.

There are numerous research questions that need to be answered relative to selecting the appropriate observation tactics. To be sure, IR and MTS are appropriate tactics for providing estimates of PA frequency and duration. Future research, however, could be focused on questions that can help researchers select the best observation tactic(s) based on the nature of the predominant PA of the participants. For example, which observation tactic (and cuing format) provides the most accurate data for PA that is naturally continuous (e.g., swimming, cross country, aerobic conditioning)? Also, which observation tactic (and cuing format) offers the most accurate data for PA that is naturally discrete and intermittent (e.g., archery, curling, bowling)? Replication of studies across the many types of PA (and the context in which they occur) is also needed relative to the establishment of data collection protocols specific to activity types and settings.

In the next methods-related question, we introduce SO tools designed specifically for assessing PA in various contexts. Although PA behavior categories have been validated on numerous occasions, additional validation studies are warranted that target coding conventions for specific situations. This may require behavioral researchers to team up with colleagues in exercise physiology to establish validity for agreed-upon coding conventions that can be more standardized and produce more accurate data.

2. What Are the Essential Components of an Observation System for Assessing Physical Activity and Its Contexts?

As a direct measure of behavior with high internal validity, SO is frequently used as a criterion measure for validating other data collection tools such as pedometers and accelerometers (e.g., Finn & Specker, 2000; McClain, Abraham, Brusseau, & Tudor-Locke, 2008). Validating PA-level categories is perhaps the most important step in the process of designing an SO tool, but it is an extensive and time-consuming process. An alternative is to choose a system with the PA categories already validated and then modify the contextual categories. The context category definitions can be adjusted with relative ease, while strong face validity is maintained using the original PA-level categories. Another advantage of doing this is that it enables consistency in data collection, thereby facilitating greater between-project comparisons and strengthening the potential for generality of methods and findings.

Several instruments using the same validated PA levels have been designed specifically for assessing PA in various contexts (e.g., physical education lessons, schools, parks, playgrounds, homes), and we highlight them in this article. They include: (a) System for Observing Fitness Instruction Time (SOFIT; McKenzie, Sallis, & Nader, 1991), (b) System for Observing Play and Leisure in Youth (SOPLAY; McKenzie et al., 2000), (c) System for Observing Play and Active Recreation in Communities (SOPARC; McKenzie et al., 2006), (d) System for Observing Children's Activity and Relationships during Play (SOCARP; Ridgers, Stratton, & McKenzie, 2010), and (e) Behaviors of Eating and Activity for Children's Health: Evaluation System (BEACHES; McKenzie, Sallis, Patterson, et al., 1991). Table 2 presents the essential features of each system. Full descriptions are beyond the scope of this article, but the coding protocols for the tools are available on the Active Living Research Web site (http://www.activelivingresearch.org).

The PA-level categories in the tools presented in Table 2 have all been validated during the course of several studies with typically developing individuals and using a variety of criterion measures such as heart rate, energy expenditure, pedometers, and accelerometers (e.g., Heath, Coleman, Lensegrav, & Fallon, 2006; McKenzie, Sallis, Patterson, et al., 1991; Pope, Coleman, Gonzalez, Barron, & Heath, 2002; Ridgers et al., 2010; Rowe, Schuldheisz, & van der Mars, 1997; Rowe, van der Mars, Schuldheisz, & Fox, 2004). The PA categories have also been validated for use with participants with handicapping conditions (e.g., Faison-Hodge & Porretta, 2004; Sit, Capio, Cerin, & McKenzie, 2013). This latter group may be especially prone to inactivity and consequently may be more susceptible to subsequent chronic diseases (Rimmer & Marques, 2012; Rimmer, Schiller, & Chen, 2012). The PA categories have also been validated for use in classroom settings (Honas et al., 2008). The contextual and instructor behavior variables that are typically used with these five related systems all have strong face validity and are defined in the system protocols and training videos.

The original SOFIT instrument included five levels of PA, including three sedentary body positions (i.e., lying down, sitting, standing), walking (moderate), and "vigorous." Based on subsequent validation work (e.g., Rowe et al., 2004), the three categories identifying body positions have been merged into a single category (i.e., sedentary) because of their similarity in energy expenditure, and they are used for coding convenience in the SOPLAY and SOPARC systems. Thus, given the extensive body of literature already available, researchers can be quite confident that these SO tools can provide a valid

TABLE 2
Sample Validated Systematic Observation Tools for Assessing Physical Activity and Its Contexts

Features	*SOFIT*	*SOPLAY*	*SOPARC*	*SOCARP*	*BEACHES*
Observation tactic	Momentary time sampling Interval recording	Momentary time sampling	Momentary time sampling	Momentary time sampling Interval recording	Momentary time sampling Interval recording
Typical coding format	10 s observe/10 s record	NA	NA	10 s observe/10 s record	15 s observe/15 s record
Main target	Individual students	All present in area	All present in area	Individual children	Individual children
Main location	Physical education lessons	School campus activity areas	Park/recreation settings	Playgrounds and other play areas	Home environments
Coding decision level	PA level	PA level	PA level	PA level	PA level
	Lesson context (e.g., management, fitness, skill development, game play)	Area context	Area context	Group size	Social context
	Instructor behavior	Activity type	Activity type	Activity type Interactions with peers	Physical context Food ingestion Media viewing
Demographic/ context data	Lesson content	School	Park/recreation area	Temperature	Child location
	Lesson location	Temperature	Temperature	Area contexts (accessible, usable, organized, supervised, equipped)	Presence of others
	Number of students	Time of day (before school; lunch/recess, after school)	Day and time of day	Area size	PA prompts and consequences
	Student gender	Area contexts (accessible, usable, organized, supervised, equipped)	Gender		Ingesting food
	Class gender composition		Age group		Viewing media
	Teacher gender		Race/ethnicity Area contexts (accessible, usable, organized, supervised, equipped) Area size		
First referenced	McKenzie, Sallis, & Nader (1991)	McKenzie, Marshall, Sallis, & Conway (2000)	McKenzie, Cohen, Sehgal, Williamson, & Golinelli (2006)	Ridgers, Stratton, & McKenzie (2010)	McKenzie, Sallis, Patterson, et al. (1991)

Note. SOFIT = System for Observing Fitness Instruction Time; SOPLAY = System for Observing Play and Leisure in Youth; SOPARC = System for Observing Play and Active Recreation in Communities; SOCARP = System for Observing Children's Activity and Relationships during Play; BEACHES = Behaviors of Eating and Activity for Children's Health: Evaluation System; NA = not applicable. The five systems summarized use the same five physical activity (PA)-level codes; these have been validated using numerous measures. Their protocols maybe downloaded for free from Active Living Research (http://activelivingresearch.org).

representation of PA levels. Further investigation, however, is warranted relative to the use of varying coding formats (e.g., interval length when using MTS and length of "observe/record" intervals with IR).

Compared with DR, which provides the most precise record of the amount of time a person spends at certain PA levels, timed-based recording tactics (i.e., IR and MTS) have the inherent risk of overestimating or underestimating the actual frequency and/or duration of PA. When using SO, the general rule is that more observation sessions and more samples per session are better. This, however, needs to be balanced according to temporal and financial costs. With more continuous activities (e.g., swimming, group exercise), data accuracy may be maintained even with MTS interval lengths that are well beyond the standard 20-s coding format. However, in activities that have many inherent breaks (e.g., tackle football), more samples (i.e., shorter intervals) per observation session are likely needed. Thus, across the wide spectrum of activities (i.e., sports, fitness, unstructured play), the following research question is pertinent: Given a particular activity, how far can the interval length be stretched without losing data accuracy? Answers to this question have both research and practical implications. Meanwhile, McNamee and van der Mars

(2005) indicated that interval lengths up 90 s in physical education could provide PA data at acceptable accuracy levels. Similarly, what are the optimal interval lengths in settings where activities may change (e.g., in physical education) or where participants have activity choices such as in recreational settings?

Researchers are often concerned about how many days are needed, how long observations are, and how many observations are needed to ensure accuracy, establish generalizability, and/or conduct power calculations for determining intervention effects. Answers to these questions are challenging and are based on multiple factors, such as the population to be studied, time of year (i.e., season), time of day, and weather conditions (e.g., rain, temperature). Frequent visits to the target environment prior to the start of data collection can aid in determining frequency and duration of observations. Nonetheless, there is some evidence for the minimum number of observation samples needed in studying physical education (e.g., Bailey et al., 1995) and neighborhood parks. Cohen et al. (2011), for example, studied parks in four U.S. regions and determined that weekly park use and user characteristics such as PA levels and age and race/ethnicity groupings could be predicted by observing four times a day for four days. Potential mediating factors for studies of physical education include lesson objectives, time in the unit of instruction, within- and across-lesson variation in content, type, amount of equipment, and facility size (e.g., Levin, McKenzie, Hussey, Kelder, & Lytle, 2001). In home settings, Klesges and colleagues (1984) recommended a minimum of four observation sessions to adequately estimate young children's PA patterns. Nonetheless, more studies assessing the generalizability of observational data are needed.

3. How Do Observer Training Protocols and Observer Experience Affect Potential Sources of Error?

Data credibility can be ensured only when observers have completed proper training and are monitored via periodic reliability checks throughout data collection. Prior to training, trainees should receive an observation protocol that includes all category definitions (with multiple examples for each), coding symbols, coding conventions, samples of coding sheets, steps for compiling raw data and calculating observer reliability scores, and answers to frequently asked questions. Training typically consists of the following:

1. *Orientation to SO and the specific tool to be used:* Trainees learn about the research project in general terms, but not the specific hypotheses. They learn about possible ethical issues, the need for objectivity, maintaining confidentiality, and observer etiquette.
2. *Memorizing behavior categories and accompanying coding symbols:* Trainees participate in directed video practice of all coding protocols (i.e., pacing and coding formats and conventions).
3. *Video assessments using "gold standard" coding records:* Trainees code video segments previously coded by a certified observer and make comparisons to the "gold standard."
4. *"Live" field-based practice of using parts of the observation system and the full observation system:* Trainers are available to answer questions, provide feedback, and determine potential gaps in observers' competence.
5. *Field-based observer reliability checks with a certified assessor:* This is a formal assessment of observers, and starting actual data collection is contingent on successfully meeting established interobserver agreement (IOA) levels.
6. *Reliability checks should be conducted throughout all phases of a study.* Retraining should occur after extended breaks (e.g., summer).

Sources of Observer Error

The goal of SO is to produce reliable and accurate records of PA and its contexts, but errors may occur as a consequence of "observer drift," "environmental complexity," "observer reactivity," and "observer bias and cheating." Drift reflects the tendency for observers to misinterpret category definitions and coding conventions over time. It can be avoided by conducting regular gold standard checks and using a rotation system for reliability observers. When identified, additional training (i.e., recalibration) should be provided immediately. Environmental complexity occurs when features of the setting make observing challenging, such as when many people are moving too fast, when many different things are occurring simultaneously, or when it is noisy. For example, with SOPLAY and SOPARC, complexity can be minimized by subdividing targeted activity areas into smaller sections and observing each subsection separately. The risk for observer reactivity is heightened when the primary observer is aware that an interobserver reliability check is being conducted. Though labor-intensive, one strategy is to conduct reliability assessments frequently. Observer bias is typically unintentional and it can be detected through reliability checks. Observer cheating (i.e., falsifying data) is rare and can minimized by (a) forwarding the data immediately after completing the observations, (b) having someone other than the observers calculate reliability percentages, and (c) conducting unannounced random reliability checks. Additionally, the advent of recording using modern electronic devices permits data to be automatically time- and location-stamped.

Determining and Reporting Observer Reliability

Reliability refers to the ability of observers to be consistent and is reflected in the level of agreement of data recorded (typically expressed in percent) by two trained, independent observers coding the same events at the same time. This is referred to as "interobserver" reliability or IOA. With videos, a single observer's scores from two separate coding sessions can be compared (i.e., "intraobserver" reliability). With tools like SOFIT, observer independence during IOA is established by employing two earphones connected to a device (e.g., iPhone) via a Y-adapter that allows both observers to simultaneously hear the "observe" and "record" cues from a prerecorded MP3 digital file. When using SOPLAY and SOPARC, observer independence is achieved by having the two observers positioned at least 15 feet apart from each other and one observer cueing the other when it is time to start an observation scan.

Observer reliability can be determined in numerous ways, including interval-by-interval (I-I) scores, intraclass correlation coefficients, and Kappa statistics. Kappa's advantage lies in the fact that it considers chance agreement, and it typically produces lower scores than provided by I-I. With more complex systems, such as those highlighted in this article, IOA percentages of 80% or better are deemed acceptable, while SO tools with fewer behavior categories should strive for IOA scores of at least 90% (Cooper et al., 2007; van der Mars, 1989). In studies, researchers should avoid reporting only summary scores and include reliabilities for all reported PA and contextual variables.

The six-step observer training process outlined in this article is the standard protocol, but there is no evidence to show that this is the best or most efficient means for training and maintaining competent observers. Future research should focus on how variations in the standard protocol affect the efficiency with which observers reach acceptable levels of observer reliability and produce accuracy against gold standards. For example, variations in the amount of time spent on video-based and live observation practice may affect the ability of observers to meet accepted criteria within a reasonable amount of time.

Related to the issues surrounding potential sources of observer error, researchers with an interest in measurement theory could consider employing generalizability (G) theory to determine the relative contribution of the various sources of observer error. Generalizability theory is a statistical theory that can be used to evaluate the dependability (i.e., reliability) of behavioral measurement (Shavelson & Webb, 1991). For example, the application of generalization theory might aid in establishing whether observer error is primarily a consequence of observer drift, environmental complexity, or observer reactivity. Moreover, observer experience could also be studied, as it is a potential mediating variable relative to potential sources of observer error. For example, are more experienced observers less likely to err in their observation through drift, reactivity, etc.?

4. What Roles Do Video and Technological Innovations Play in Advancing the Systematic Observation of Physical Activity?

With the emergence of digital technologies, the number of computer SO software tools has increased significantly (Castelliano, Perea, Alday, & Mendo, 2008). Early SO users were limited to paper-and-pencil approaches to data collection, but in the last few years the availability of compact hardware like tablets (e.g., iPads) has also given rise to the development of SO apps, such as iSOPARC, which is available for free through the iTunes App Store.

Great advances have also occurred in being able to create video records of events, including recording and storage becoming easier, more compact, and less expensive. The use of reel-to-reel video recording in the 1970s was considered a revolutionary development, and the subsequent VHS, Betamax, and VHS-C cassette videos were still analog-based. Today's digital recording devices (e.g., GoPro cameras) are much less obtrusive and allow recording events for extended periods of time. A related technological development is the ability to use remote recording of events that might be of interest to researchers studying facility usage in recreation centers and parks. Digital video samples can be recorded at user-defined intervals (e.g., mornings, afternoons) for set amounts of time (e.g., every 30 min). As well, with the emergence of high-capacity devices such as external hard drives and cloud-based storage, storing and backing up video records is no longer a barrier.

Key advantages of recent software technology include: (a) the ability to create "permanent records" of events, allowing for delayed observer coding as well as intraoberserver agreement and IOA checks; (b) faster compiling, summarizing, and storing of data; (c) being able to retrieve records with ease and conduct observations at any time, and if necessary recode sessions when observer reliability checks are subpar; and (d) ease of transferring raw data to more advanced statistical software packages for in-depth analysis. These advantages make these tools especially attractive for larger-scale (e.g., multisite) research projects in which SO is used.

Castelliano et al. (2008) identified several useful criteria enabling researchers to judge the quality and usability of digital software-based SO tools. These include: (a) user friendliness, such as software that permits users to adjust the tool to their needs instead of being locked into a "closed" system; (b) the option of being able to self-define target behaviors and contextual variables; (c) the use of automatic time stamping of observed events to indicate the time, duration, and location of the events; (d) the use of video-based observations data, which can be linked directly to the video record and allow for quick retrieval of events for review; and (e) basic analytical features allowing for descriptive data on the frequency and/or duration of events.

Yet, despite the ongoing technological advances, there are important methodological questions that deserve to be asked. As noted earlier, the central issue in all PA-focused research is the accuracy of data. In itself, the use of software programs and apps does not guarantee or improve data accuracy. A human must still observe and record the correct behavior and contextual codes. The aforementioned advantages lie beyond the actual act of observing and recording behavioral events. SO is less likely to be overlooked as a viable data collection method when researchers can demonstrate similar data accuracy between traditional (i.e., paper and pencil) and computerized data collection methods. Thus, a central question around the use of software programs and apps remains: How and to what extent does the use of technology-based data collection tools affect data accuracy when studying PA and related contextual variables?

Additional research questions relate to whether to observe "live" or to code from video recordings, and the answers depend on multiple factors. For example, gaining access to settings may be challenging given today's concerns for privacy, and school and recreation officials may be hesitant about allowing video recording. As well, participants are more likely to change their behavior when a camera is present than when an observer is there alone. Within the context of research on PA behavior, it is yet unknown whether participant reactivity is temporary or sustained over time and whether or not there are gender differences.

As noted earlier, observer training is essential to data accuracy, and there are implications when using advanced technology for data collection. Beyond the regular protocols, observers require training in all aspects of using both the hardware and software. It takes time to learn new technology, and it is essential that data collectors be able to swiftly and accurately use the correct keys for coding, make on-the-spot corrections, and save files and transfer them—all without losing valuable data. As technology continues to evolve, one question that needs answering is which observer training protocol is most efficient in preparing users to collect data with accuracy when using electronic tools such as tablets or iPads. Moreover, much is yet to be learned about systematic variation in how training modules are arranged (e.g., number and type of video-based examples) relative to observers being able to accurately discriminate among behavior and contextual categories. Answers would help establish the most efficient observer training protocols.

Relatedly, with the prospect of large-scale, multisite projects, training protocols may be delivered via the Internet, thus potentially reducing the costs associated with preparing observers. Whether data accuracy can be maintained when training is delivered remotely requires further research.

In summary, technology advancements offer a number of attractive features for SO researchers. Decisions on the "what" and "how much" technology to use should depend on how data can be expediently collected with the best possible accuracy and with consideration given to the research question, costs, available resources, and the rigor of observer training.

SECTION 2: SCHOOL-SETTING QUESTIONS

The significant role that schools play in providing and promoting PA and subsequently contributing to population health is well recognized (e.g., Institute of Medicine [IOM], 2013; Kahn et al., 2002; Pate et al., 2006; USDHHS, 2012), and the recent *Educating the Student Body: Taking Physical Activity and Physical Education to School* report (IOM, 2013) suggested that schools should provide children and adolescents with at least 50% of their recommended 60 min of daily moderate-to-vigorous PA (MVPA). The potential impact of federal-, state-, district-, school-, and program-level policies (or lack thereof) on PA is now beginning to be better understood (e.g., Amis, Wright, Dyson, Vardaman, & Ferry, 2012; Beets, Huberty, & Beighle, 2013; Kahan & McKenzie, in press; Lounsbery, McKenzie, Morrow, Monnat, & Holt, 2013), and there is evidence that the strength of a policy or mandate relative to its language and enforcement (i.e., degree of oversight, reporting requirements, and consequences), along with associated funding for its implementation, will affect its implementation adherence (e.g., Amis et al., 2012).

Meanwhile, most policy-related studies on school-based PA and physical education have depended on self-report surveys, often completed by respondents in state offices that are a far distance from the schools. Rarely do these studies employ what has come to be known as "ground-truthing," which includes directly observing actual PA levels of individuals (or groups) simultaneously with context-based policy and environmental variables. Acquiring such information would allow for much stronger evidence on the role and impact of policies, mandates, and laws.

Additionally, there has never been a comprehensive surveillance study of PA on school campuses across the United States or even within an individual state using direct observation; doing so would provide an increased understanding of the policies and practices related to PA in schools and would provide guidance for improved mandates. SO is particularly important in assessing PA in preschools (e.g., Pate, McIver, Dowda, Brown, & Addy, 2008) and primary schools where young children are unable to respond to questionnaires reliably. Although physical education is typically identified as the most salient opportunity for PA in schools, it is clear that physical education alone cannot provide sufficient MVPA minutes and that a more comprehensive or collaborative effort is needed for children and youth to meet national recommendations (USDHHS, 2012). This section therefore identifies research questions related to observing PA during physical education as well as during other on-campus programs, including recess and in before-school and after-school settings.

5. What Important Aspects Related to Physical Activity in Physical Education Can Be Studied Using Systematic Observation?

The important role that physical education plays in providing and promoting PA has long been recognized (IOM, 2013). This includes a stringent review of the available evidence established more than 15 years ago resulting in physical education being identified as one of only six interventions strongly recommended for increasing PA by the national *Task Force on Community Preventive Service* (Kahn et al., 2002). Sallis and colleagues (Sallis & McKenzie, 1991; Sallis, McKenzie, et al., 2012) have also argued that physical education is a cost-effective public health resource, and physical educators are uniquely well positioned to address inactivity. Physical education is: (a) part of the formalized school curriculum in all states, thereby reaching nearly all children; (b) the only venue where the least active children are likely to experience PA at higher intensities; and (c) the only formal requirement for PA engagement that many people have during their entire lives (discounting military service). In response to the American College of Sports Medicine promotion of PA through the theme of "exercise is medicine," McKenzie and Lounsbery (2009, 2014) have described physical education as "the pill not taken." They identified student PA during lessons as a major consideration in assessing physical education teacher and program effectiveness and expressed concern about the limited dosage of physical education (i.e., frequency and duration) as well as its content, palatability, and who delivers it (e.g., physical education specialists, classroom teachers). Many of these concerns can be addressed/examined using SO of physical education lessons.

Physical education dosage is calculated using lesson frequency and duration, and it is often expressed in physical education minutes per week. There is a dearth of objective data on how much physical education students actually receive. Although national professional recommendations exist (e.g., 150 min/week and 225 min/week for elementary and secondary schools, respectively), only about 40% of U.S. states have policies specifying time allocations for physical education (and not all of these have them for both the elementary and secondary school levels). The amount of time recommended for physical education in these states ($n = 19$) is only about 60% of that recommended by professionals (Kahan & McKenzie, in press), and it is well known that districts and schools frequently do not follow state physical education recommendations (e.g., Amis et al., 2012; Sanchez-Vaznaugh, O'Sullivan, & Egerter, 2013). School administrators often report their school's physical education schedule to district or state supervisors, but these reports may not accurately reflect the actual frequency and duration of lessons or opportunities for PA during those times. SO can verify and enhance self-reports by providing an accurate accounting of the number and proportion of lessons actually held and the proportion of scheduled time in which students engaged in MVPA during those lessons.

SO can also help answer relevant research questions other than dosage. SOFIT has been used to simultaneously assess PA, lesson context (i.e., how PE is delivered), and instructor behavior in numerous countries and with students from the preschool through university levels. As a result, it has been used frequently, including as a means to assess the effectiveness of small-scale and large-scale physical education interventions and their maintenance. SOFIT and other tools could also be used to explore research questions such as how PA in physical education may differ (a) during different lesson contexts (e.g., fitness, skill development, game play, knowledge); (b) during different sport and physical activities; (c) during lessons taught by male and female instructors; (d) across grade levels; (e) by boys and girls; (f) during coeducational and single-gender lessons; (g) by lesson location (e.g., indoor vs. outdoor lessons); and (h) by instructor certification or preparation (e.g., certified physical education teachers vs. classroom teachers; those with and without in-service development).

Another example of a key policy-related challenge to school physical education is the lack of state-level policies preventing physical education waivers/substitutions by individual districts and schools (National Association for Sport and Physical Education & American Heart Association, 2012). Common substitutions include Junior Reserve Officer Training Corps (JROTC), marching band, cheerleading, and interscholastic sports. Recently, SO was used to compare differences in PA, lesson contexts, and teacher promotion of PA in physical education and JROTC classes (Lounsbery, Holt, Monnat, Funk, & McKenzie, 2014). In addition to replicating this work, further investigations are warranted for using SO to study student PA contexts during other substituted programs such as marching band and interscholastic sports—both extracurricular programs that have fundamentally different goals and objectives from physical education.

Potential research questions related to using SO to assess instructional effectiveness in physical education also apply to other activity areas where instruction is involved. This includes before-school and after-school programs on campus, with young children in child care, and with adults during instruction in recreation and park settings. Discussion of additional research topics for these areas follows in subsequent sections.

6. What Important Aspects Related to Physical Activity During Recess and Classroom Breaks Can Be Studied Using Systematic Observation?

Schools vary substantially in their policies related to recess and classroom breaks as well as in the number, duration, and type of sessions that they provide (Lounsbery et al., 2013; Slater, Nicholson, Chriqui, Turner, & Chaloupka, 2012; Whitt-Glover, Porter, & Yancey, 2013). These "break from academics" sessions are important for various reasons other than providing PA (e.g., American Academy of Pediatrics, 2013), and they are much more common at the elementary school level than at the secondary school level. Allocated time for recess in elementary schools has dwindled in favor of time for "core academic subjects" since the enactment of No Child Left Behind (Center on Education Policy, 2008). Classroom activity breaks are scheduled much less frequently than recess, and to date, there is little objective information on their widespread occurrence or how much PA students accrue during these breaks.

Unlike physical education classes where students are required to engage in PA, participating actively during recess is usually voluntary and there is frequent debate over whether the activities during recess and activity breaks should be structured or unstructured. Nonetheless, research reviews indicate that when recess and activity breaks are held, they provide substantial opportunities for PA during the school day and they can increase children's overall daily activity (Beighle, 2012; Ward, 2011). Boys are typically more active than girls during recess and the session context matters, with PA levels usually being higher when loose equipment (e.g., balls, jumping ropes) is available (McKenzie, Crespo, Baquero, & Elder, 2010). Interventions targeting recess periods have included implementing active supervision, providing activity-enhancing equipment, activity cards, and playground markings (Beighle, 2012) with each resulting in higher PA levels. Many of the observational studies during recess have taken place in Australia and the United Kingdom, and there is a need to investigate PA baseline levels and the effects of interventions in other countries.

There is potential for using SO to explore research questions such as how PA during recess or activity breaks may be mediated by (a) children's gender, age/grade grouping, or specific handicapping condition (e.g., auditory or visual impairments, attention-deficit hyperactivity disorder); (b) session location (e.g., indoors vs. outdoors); (c) weather condition; (d) space size and different playground markings, type and/or amount of equipment, and structures; (e) type of activity (e.g., sport, recreational games); (f) structured vs. unstructured sessions; and (g) supervisor gender or certification/training. As well, there is ample opportunity to use SO to examine the social structures of leisure-time periods (e.g., group size, activity type [competitive vs. noncompetitive]) and children and supervisor interactions regarding both PA and social engagement (e.g., bullying) as is done using SOCARP (e.g., Ridgers, Carter, Stratton, & McKenzie, 2011; Ridgers et al., 2010).

Only a small number of studies have focused on the use and effects of classroom PA breaks, so additional research is warranted. SO, for example, could be used to examine the frequency of activity breaks and to assess classroom behavior such as levels of attentiveness and on-task behavior before and after recess breaks and/or in-class PA breaks (e.g., Mahar et al., 2006). In addition to these variables, it is particularly important to use SO on school campuses to assess student PA and how it is affected by state/district/school policies related to recess and classroom breaks. For example, a policy may exist indicating that recess is to be provided, while in reality, recess may be cancelled entirely (e.g., for academic testing, lack of suitable indoor space during inclement weather) or individual children are prohibited from participating in recess for academic or disciplinary reasons. Such data cannot be obtained without observing directly.

7. What Important Aspects Related to On-Campus Physical Activity Before and After School Can Be Studied Using Systematic Observation?

Most SO studies of PA in schools have been conducted in elementary schools and during physical education classes. Question 5 focused on physical education and Question 6 focused on recess and classroom activity breaks, which both occur primarily in elementary schools. With student participation in secondary school physical education being limited, it is important to examine opportunities for PA in other settings too. This can be accomplished by observing during: (a) *interscholastic activities* (programs providing competition with other schools); (b) *intramural activities* (programs providing competition for students within the same school); (c) *club activities* (both competitive and noncompetitive PA groups that meet regularly, such as a dance or aerobics club); and (d) *other activities* (sporadic or miscellaneous PA programs, such as a dance workshop). Unlike in physical education, where attendance and participation is required, student engagement in these extracurricular programs is voluntary. This leads to including important research questions beyond PA levels in such programs, such as how many students participate and for how often and how long.

Tools for objectively assessing PA during "open" leisure-time periods were not available until recently, when SOPLAY was validated (McKenzie et al., 2000) and used to assess an intervention in 24 middle schools (Sallis et al., 2003). As indicated in Table 2, SOPLAY uses MTS, and during the initial baseline study, observers assessed the PA of students in all activity areas ($N = 151$) before school, during lunch, and after school. Each student in an activity area was coded as being sedentary, walking, or vigorous using systematic scans, and the characteristics of the activity areas were simultaneously coded (i.e., as usable, accessible, supervised, equipped, and organized). Although the activity areas were nearly always usable, they were accessible only about 50% of the time and were rarely supervised or equipped for PA. More boys used the activity areas than girls and boys were more active when present. A greater proportion of students in attendance at a school were in the activity areas at lunchtime (i.e., 20%) than before school (4%) and after school (2%).

These data are from a single sample of observations taken in 24 schools, but they illustrate the vast amount of relevant information that can be obtained from SO and suggest the possibility that modifications to school policies (e.g., increased accessibility to activity areas) and the environment (e.g., increased provision of supervision, equipment, and supervised activities) might attract more students, especially girls, to activity programs and areas. A number of recent studies have also used SO to examine differences in PA provided by intramural versus interscholastic programs at middle schools (Bocarro et al., 2012; Bocarro, Kanters, Edwards, Casper, & McKenzie, 2014), the implications of shared use of school PA facilities (Kanters, Bocarro, et al., 2014), and the implications of staff behaviors on PA during after-school programs (Huberty, Beets, Beighle, & McKenzie, 2013). Nonetheless, the widespread contributions that various organized and unorganized extracurricular programs make to student PA on campuses remain relatively unexplored.

With few objective data available on school policies, grade levels, facility usage, participation rates and activity levels by gender and race/ethnicity, or sponsorship (i.e., whether a program is offered by the school or an outside organization), the extracurricular field is ripe for research using SO. For example, involvement in high school interscholastic sports has increased for 24 consecutive years, and in 2012 to 2013, more than 7.7 million students participated in high school interscholastic sports (National Federation of State High School Associations, 2013). To support this growth, PA facilities on high school campuses have grown increasingly expansive. These facilities are major investments for communities, yet little is known about their level of use across all 7 days of the week, who uses them, and how environmental (e.g., increased access, usability) and policy-related interventions might affect their use by students, staff and adults from surrounding neighborhoods. SO can help answer these questions.

SECTION 3: COMMUNITY-SETTING QUESTIONS

It is clear that children and adolescents are unable to accrue all their recommended daily 60 min of MVPA at school. Additionally, most adults are not enrolled in school and are likely, at best, to have only limited access to school PA facilities. Subsequently, other opportunities must be available if people are to realistically meet recommended PA levels. The significant role that nonschool settings such as parks, trails, and playgrounds play in providing and promoting PA is becoming better recognized (e.g., Bedimo-Rung, Mowen, & Cohen, 2005; Cohen et al., 2007; Mowen, 2010). Most of the information on PA in these settings is obtained using self-reports from adult participants or program administrators. Accelerometers, pedometers, and heart rate monitors can provide objective information on individuals, but they are limited when doing place-based research where different people come and go, often sporadically, and important environmental and contextual variables may change frequently.

This section identifies research questions related to observing PA in nonschool environments such as community parks, youth and adult sports, and home settings. Because environmental- and policy-based interventions affect not only specific individuals but the entire population, it is important that researchers consider applying recently developed SO scanning tactics such as those used in SOPLAY and SOPARC.

8. What Important Aspects Related to Physical Activity in Park and Recreation Settings Can Be Studied Using Systematic Observation?

Parks and other common areas in communities are important locations for population PA, and Mowen (2010) reported there is a need for more studies examining specific park features such as trails, sports fields, playgrounds, and support facilities, as well as the condition and design of those features with regard to area visitation and activity levels, especially among lower-income, racial and ethnic, youth, and rural populations. As well, there is a need to examine the role of park and recreation center management and administrative policies and practices, such as programming, staffing, supervision, and promotion efforts, in increasing park use and PA levels.

SO is well suited to assist in these studies. SOPARC, for example, was designed to assess PA, general area use, and contextual factors in community parks and recreation centers and has been validated and found to be a reliable indicator of park use (Cohen et al., 2011). It assesses area users' PA levels, gender, activity modes/types, and apparent age and race/ethnicity groupings while simultaneously providing contextual information on targeted activity areas (i.e., accessibility, usability, and levels of supervision and organization). SOPARC methodology analyzes park and other activity area use through MTS by counting the number and type of area users and coding their activities.

With multiple observations, SOPARC provides a valid measure of weekly park activity levels and use by gender, age and race/ethnicity grouping, and PA during seasons or across the entire year (Cohen et al., 2011). In addition to numerous cross-sectional studies (e.g., Cohen et al., 2013; Kaczynski, Stanis, Hastmann, & Besenyi, 2011), SOPARC has been used to assess a variety of interventions including the implementation of community fitness programs in

Recife, Brazil (Parra et al., 2010), modifying policies and programming (Cohen, Williamson, Sehgal, Marsh, & McKenzie, 2009), and adding a bike path (Cohen et al., 2008), fitness zones (Cohen, Marsh, Williamson, Golinelli, & McKenzie, 2012), and pocket parks (Cohen et al., 2014). SOPARC is also being used as the main tool in the first-ever national observational study of neighborhood parks in 25 randomly selected U.S. cities with a population of more than 100,000 (City Parks Alliance, 2013). Observations were made within each park ($n = 174$) during 4 days in the spring/summer of 2014 and will be repeated during the same time period in 2016.

An adaptation of SOPARC (System for Observing Physical Activity and Recreation in Natural Areas) that adds contextual activity support factors (e.g., horses, ATVs) is available for researchers interested in investigating PA in wilderness areas and campgrounds (Sasidharian & McKenzie, 2014). In addition to regular training, observers in wilderness areas need instruction on taking additional safety precautions, such as to travel in pairs; carry sufficient water, an emergency first-aid kit, two-way radio, cell phone, and whistle; dress appropriately (e.g., hat, hiking shoes); and protect against the elements (e.g., sun, bugs, adverse temperatures).

In addition to parks, out-of-school programs (e.g., Beets et al., 2013; Weaver, Beets, Webster, & Huberty, 2014), sport and fitness camps (e.g., Bullen, Reed, & Mayer, 1964; McKenzie, 1986), and active transportation in the community (Ryan & Lindsey, 2013) are understudied areas, and these are also ripe for exploration using SO. As well, park and recreation settings provide structured educational programs; thus, many of the research questions related to studying PA in physical education classes that were identified in Question 5 are also applicable in these environments.

9. What Important Aspects Related to Physical Activity in Youth and Adult Sport Settings Can Be Studied Using Systematic Observation?

Numerous observational tools are available for measuring behavior specifically in sports settings (e.g., Castelliano et al., 2008; Darst, Zakrajsek, & Mancini, 1989), but none focus directly on PA as a primary process and/or outcome variable. Meanwhile, research on adult PA levels during leisure-time sports is rare, but participation in organized sport has been shown to be important for the health of youths (Geidne, Quennerstedt, & Eriksson, 2013) and has shown to be associated with their higher PA levels (e.g., Pate, Trost, Levin, & Dowda, 2000; Pfeiffer et al., 2006). For example, 12- to 14-year-old boys and girls have been shown to acquire about 60% of their daily MVPA in sport settings (Katzmarzyk, Walker, & Malina, 2001). There is some additional evidence (based on accelerometer data) that children engage in MVPA for 33% to 46% of the time during sport practices and games and that there is a high percentage of inactivity time (Guagliano, Rosenkranz, & Kolt, 2013; Sacheck et al., 2011). Meanwhile, Leek and colleagues found that fewer than 25% of youth sport participants obtained the recommended 60 min of daily MVPA during practices (Leek et al., 2011).

Youth and adult sport coaching is essentially an unregulated endeavor, with most coaching done on a volunteer basis and with little oversight from regulatory agencies. Nonetheless, coach education programs and workshops are common throughout the nation, with most being sport-specific. How effective they are in changing coach/instructor behavior and whether the results improve learner/participant PA is essentially unknown. Meanwhile, there is a need to improve the instructional efficiency of sports practices to maximize time so participants can engage in more PA, develop skills, and become physically fit. Three examples of how this was investigated using SO follow. In an older study, McKenzie and Rushall (1974) used frequency counts to calculate laps per minute to assess the effectiveness of a self-management intervention strategy (i.e., program boards) with age-group swimmers. In a second study, Kanters, McKenzie, and colleagues (2014) used a modification of SOFIT (IR) to assess player PA levels and the practice context during 82 ice hockey practices of children aged 9 to 10 years. Of these practices, 39 operated under a traditional structure and 43 used a new instructional model. Overall, players spent 44% of practice time engaged in sedentary activities, 33% in moderate PA, and 23% in vigorous PA. Individual player MVPA did not differ significantly between the practice types, but the practices incorporating the new instructional model accommodated about 60% more players while having twice as many coaches, a lower player-to-coach ratio, a greater percentage of time in vigorous PA, and more time dedicated specifically to skill drills/activities. In the third example, SOFIT is currently being used in a randomized control trial to investigate the efficacy of coach education sessions on increasing MVPA and reducing inactivity during a girls' youth basketball program (Guagliano, Lonsdale, Kolt, & Rosenkranz, 2014).

Because sport practices, dance rehearsals, and physical education classes share many common instructional elements, important research questions related to SO in those environments are similar (see Question 5). In sport and dance settings, SO could be used to explore many factors, including whether PA might differ (a) by male and female participants, (b) during different sport and physical activities, (c) during sessions taught by male and female instructors, (d) across age and league levels, (e) during coeducational and single-gender sessions, (f) by session location (e.g., indoor vs. outdoor lessons), (g) by starters vs. substitutes, (h) by players of different positions, (i) by instructor/coach certification or preparation, and (j) during different session practice contexts (e.g., physical fitness, specific skill development, scrimmage, game play, knowledge).

10. What Important Aspects Related to Physical Activity in Home and Early Care and Education Settings Can Be Studied Using Systematic Observation?

Children develop PA habits early, and these habits are influenced first by family members in their home environment. As well, most young children spend a significant amount of time in early care and education (ECE) programs as well as with family child-care providers and other caregivers (e.g., relatives, friends, neighbors). Determining children's levels of PA and sedentary behavior in these settings as well as the physical and social environmental factors that influence them are important research questions that SO can help answer.

Few studies to date have used SO to simultaneously assess children's PA activity and associated conditions in the home, but doing so could shine a light on factors contributing to reduced PA and increased obesity in children. As noted earlier in this article, extended time spent in sedentary behavior is an independent risk factor for chronic disease. Little is known as yet about the impact of contextual variables in the home and ECE environment on children's PA and sedentary behavior. The richness of the data that can be obtained from observations is illustrated in a study of the PA of 6-year-old Mexican American children ($n = 139$) and associated contexts at home using BEACHES (McKenzie et al., 2008). The study was designed to assess whether selected physical and social environmental factors were associated with children's PA and sedentary behavior in an at-risk population. Overall, when observed, the children were primarily indoors (74% of the time), sedentary (74% of the time), and engaged in little vigorous PA (11% of time). Time spent being sedentary was associated with being indoors, parents being present, and time spent viewing media and ingesting food. It was clear that the PA prompts they received depended on their location, gender, and who was providing the prompts. Boys were prompted to be sedentary twice as often as girls (9% vs. 5% of total intervals), especially when indoors, where 85% of boys' activity-related prompts were to be sedentary.

Studies such as this one can contribute to understanding the potential impact of the home environment on the sedentary and activity behaviors of children. Intervention studies are also needed to determine whether changing specific factors can influence children's PA. Fortunately, SO can be used to assess both process and outcome variables during intervention studies. Observations of families need not be limited to the home setting itself, but with the advent of mobile devices, families can be followed wherever they go, including to the San Diego Zoo where observations were made to test the generalization of a family PA and nutrition intervention (Patterson et al., 1988). As well, SO studies can be undertaken to investigate how factors might impact the PA of the same children in different environments, such as home and school settings (e.g., McKenzie, Sallis, Nader, Broyles, & Nelson, 1992; Sit et al., 2013).

There are numerous observation systems for assessing children's PA at home (McKenzie, 1991), but the most commonly used and elaborate ones are the Observational System for Recording Physical Activity in Children-Home (McIver, Brown, Pfeiffer, Down, & Pare, 2009) and BEACHES. Both of these systems have the ability to simultaneously assess behavioral, social, and environmental factors that are potential influences on children's PA, and both can be used reliably in homes, child-care settings, and preschools. Both can assess potentially modifiable conditions (e.g., being indoors, presence of adults, PA prompting, viewing media, ingesting food) that have been shown to be associated with sedentary behavior, and thus, they would be helpful in answering questions about the impact and sustainability of interventions. Observations using these systems are typically combined with other measures such as a home checklist. For example, while using BEACHES, observers usually also complete an inventory of the presence of toys/apparatus that are likely to be used by the child to engage in PA. These include climbing apparatus (e.g., monkey bars), riding toys (e.g., tricycles), and active manipulatives (e.g., balls).

Most of the current SO studies in home settings target young children, but it would be possible to expand the systems to simultaneously assess the PA behaviors of adults who are present. Such an extension would be advantageous for researchers interesting in directly assessing the impact of parent modeling on children's PA.

CLOSING COMMENTS

The development of national PA recommendations (USDHHS, 2008a, 2012) and the National Physical Activity Plan (http://www.physicalactivityplan.org) are evidence of the importance of PA to society. This latter initiative reflects the multifaceted approach that is needed to engage all children, youth, and adults in ample amounts of health-enhancing PA, and no single institution, program, or venue can accomplish this alone. The important role that context plays relative to PA and sedentary behavior is now widely recognized and offers an expanded array of intervention research opportunities. In this regard, SO has a distinct advantage over other PA data collection tools in that information can be collected simultaneously on the volume and intensity of PA behavior and the numerous/diverse physical and social contextual variables that potentially suppress or facilitate PA.

In this article, we introduced concepts and questions related to using SO in researching PA behavior.

We presented general questions in three sections, including those related to: (a) ensuring data accuracy through the selection of the most appropriate methodological protocols; (b) investigating PA throughout the day in school settings, including physical education, recess, and other programs; and (c) investigating PA in community settings (e.g., parks, recreation centers, youth and adult sport programs) and home settings.

There is already an emerging body of literature for a number of the primary questions we posed. However, as Sidman (1960) reminded scientists more than five decades ago, the first rule of science is to replicate previous research. In the natural sciences, this is a commonly accepted principle. It is perhaps even more important in the study of PA behavior given (a) the wide variation in environmental and social conditions (e.g., climate, weather, urban vs. rural location, differential access, and opportunity as a consequence of economic disparities) that affect PA, and (b) the many types of interventions available. In the context of the medical sciences, Murad and Montori (2013) highlighted the limits of using single studies as a definitive basis for clinical practice. That is, confidence in the generality of interventions, methods, and findings can only be strengthened if/when substantial bodies of evidence can be reviewed systematically. Evidence-based "best available knowledge" is always evolving, and we encourage researchers to consider including SO as a data collection method in their future studies of PA.

REFERENCES

American Academy of Pediatrics. (2013). The crucial role of recess in school. *Pediatrics*, *131*, 183–188.

Amis, J. M., Wright, P. M., Dyson, B., Vardaman, J. M., & Ferry, H. (2012). Implementing childhood obesity policy in a new educational environment: The cases of Mississippi and Tennessee. *American Journal of Public Health*, *102*, 1406–1413.

Bailey, R. C., Olson, J., Pepper, S. L., Porszasz, J., Barstow, T. J., & Cooper, D. M. (1995). The level and tempo of children's physical activities: An observational study. *Medicine & Science in Sports & Exercise*, *27*, 1033–1041.

Bauer, U. E., Briss, P. A., Goodman, R. A., & Bowman, B. A. (2014). Prevention of chronic disease in the 21st century: Elimination of the leading preventable causes of premature death and disability in the USA. *The Lancet*, *384*, 45–52.

Bedimo-Rung, A. L., Mowen, A. J., & Cohen, D. A. (2005). The significance of parks to physical activity and public health: A conceptual model. *American Journal of Preventive Medicine*, *28*(Suppl. 2), 159–168.

Beets, M. W., Huberty, J., & Beighle, A. (2013). Systematic observation of physical activity in afterschool programs: Preliminary findings from Movin' Afterschool intervention. *Journal of Physical Activity and Health*, *10*, 974–981.

Beighle, A. (2012). *Increasing physical activity through recess: A research brief*. Princeton, NJ: Active Living Research. Retrieved from http://www.activelivingresearch.org/increasing-physical-activity-through-recess

Bocarro, J. N., Kanters, M. A., Cernin, E., Floyd, M. F., Casper, J. M., Suau, L. J., & McKenzie, T. L. (2012). School sport policy and school-based physical activity environments and their association with observed physical activity in middle school children. *Health & Place*, *18*, 31–38. doi:101016/j.healthplace.2011.08.007

Bocarro, J. N., Kanters, M. A., Edwards, M., Casper, J., & McKenzie, T. L. (2014). Prioritizing school intramural and interscholastic programs based on observed physical activity. *American Journal of Health Promotion*, *28*(Suppl. 3), S65–S71. doi:10.4278/ajhp.130430-QUAN-205

Bullen, B. A., Reed, R. B., & Mayer, J. (1964). Physical activity of obese and non-obese adolescent girls. *American Journal of Clinical Nutrition*, *14*, 211–223.

Castelliano, J., Perea, A., Alday, L., & Mendo, A. H. (2008). The measuring and observation tool in sports. *Behavior Research Methods*, *80*, 898–905.

Center on Education Policy. (2008). *NCLB Year 5: Instructional time in elementary schools: A closer look at changes for specific subjects* (A report in the series *From the Capital to the Classroom: Year 5 of the No Child Left Behind Act*). Washington, DC: Author. Retrieved from http://www.cep-dc.org/displayDocument.cfm?DocumentID=309

City Parks Alliance. (2013, December 2). *City Parks Alliance to participate in first national study of urban parks and physical activity* (Press release). Washington, DC: Author. Retrieved from http://www.cityparksalliance.org/storage/documents/Press_Releases/CPA_Press_Release–December_2_2013.pdf

Cohen, D. A., Lapham, S., Evenson, K. R., Williamson, S., Golinelli, D., Ward, P., . . . McKenzie, T. L. (2013). Use of neighborhood parks: Does socioeconomic status matter? A four city study. *Public Health*, *127*, 325–332. doi:10.1016/j.puhe.2013.01.003

Cohen, D. A., Marsh, T., Williamson, S., Golinelli, D., & McKenzie, T. L. (2012). Impact and cost-effectiveness of family Fitness Zones: A natural experiment in urban public parks. *Health & Place*, *18*, 39–45. doi:10.1016/j.healthplace.2011.09.008

Cohen, D. A., Marsh, T., Williamson, S., Han, B., Derose, K. P., Golinelli, D., & McKenzie, T. L. (2014). The potential for pocket parks to increase physical activity. *American Journal of Health Promotion*, *28*(Suppl. 3), S19–S26. doi:10.4278/ajhp.130430-QUAN-213

Cohen, D. A., McKenzie, T. L., Sehgal, A., Williamson, S., Golinelli, D., & Lurie, N. (2007). Contribution of public parks to physical activity. *American Journal of Public Health*, *97*, 509–514.

Cohen, D., Sehgal, A., Williamson, S., Golinelli, D., McKenzie, T. L., Capone-Newton, P., & Lurie, N. (2008). Impact of a new bicycle path on physical activity. *Preventive Medicine*, *46*, 80–81.

Cohen, D. A., Setodji, C., Evenson, K. R., Ward, P., Hillier, A., Lapham, S., . . . Marsh, T. (2011). How much observation is enough? Refining the administration of SOPARC. *Journal of Physical Activity and Health*, *8*, 1117–1123.

Cohen, D. A., Williamson, S., Sehgal, A., Marsh, T., & McKenzie, T. L. (2009). Effects of park improvements on park use and physical activity: Policy and programming implications. *American Journal of Preventive Medicine*, *37*, 475–480. doi:10.1016/j.amepre.2009.07.017

Cooper, J. O., Heron, T. E., & Heward, W. E. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Corder, K., Ekelund, U., Steele, R. M., Wareham, N. J., & Brage, S. (2008). Assessment of physical activity in youth. *Journal of Applied Physiology*, *105*, 977–987.

Darst, P., Zakrajsek, D., & Mancini, V. (Eds.). (1989). *Analyzing physical education and sport instruction* (2nd ed.). Champaign, IL: Human Kinetics.

Faison-Hodge, J., & Porretta, D. L. (2004). Physical activity levels of students with mental retardation and students without disabilities. *Adapted Physical Activity Quarterly*, *21*, 139–152.

Finn, K. J., & Specker, B. (2000). Comparison of Actiwatch activity monitor and Children's Activity Rating Scale in children. *Medicine & Science in Sports & Exercise*, *32*, 1794–1797.

Geidne, S., Quennerstedt, M., & Eriksson, C. (2013). The youth sports club as a health promoting setting: An integrative review of research. *Scandinavian Journal of Public Health*, *41*, 269–283.

Guagliano, J. M., Lonsdale, C., Kolt, G. S., & Rosenkranz, R. R. (2014). Increasing girls' physical activity levels during an organised youth sport basketball program: A randomised controlled trial protocol. *BMC Public Health*, *14*, 383. doi:10.1186/1471-2458-14-383

Guagliano, J. M., Rosenkranz, R. R., & Kolt, G. S. (2013). Girls' physical activity levels during organized sports in Australia. *Medicine & Science in Sports & Exercise*, *45*, 116–122.

Heath, E. M., Coleman, K. J., Lensegrav, T., & Fallon, J. A. (2006). Using momentary time sampling to estimate minutes of physical activity in physical education: Validation of scores for the System for Observing Fitness Instruction Time. *Research Quarterly for Exercise and Sport*, *77*, 142–146.

Honas, J. J., Washburn, R. A., Smith, B. K., Greene, J. L., Cook-Wiens, G., & Donnelly, J. E. (2008). The System for Observing Fitness Instruction Time (SOFIT) as a measure of energy expenditure during classroom-based physical activity. *Pediatric Exercise Science*, *20*, 439–445.

Huberty, J. L., Beets, M. W., Beighle, A., & McKenzie, T. L. (2013). Association of staff behaviors and afterschool program features to physical activity: Findings from Movin' Afterschool. *Journal of Physical Activity and Health*, *10*, 423–429.

Institute of Medicine. (2013). *Educating the student body: Taking physical activity and physical education to school*. Washington, DC: National Academies Press.

Kaczynski, A. T., Stanis, S., Hastmann, T., & Besenyi, G. (2011). Variations in observed park physical activity intensity level by gender, race, and age: Individual and joint effects. *Journal of Physical Activity and Health*, *8*(Suppl. 2), S151–S160.

Kahan, D., & McKenzie, T. L. (in press). The potential and reality of physical education in controlling overweight and obesity. *American Journal of Public Health*, *105*(4). doi:10.2105/AJPH.2014.302355

Kahn, E. B., Ramsey, L. T., Brownson, R. C., Heath, G. W., Howze, E. H., Powell, K. E., ... Corso, P. (2002). The effectiveness of interventions to increase physical activity: A systematic review. *American Journal of Preventive Medicine*, *22*(Suppl. 4), 73–107.

Kanters, M. A., Bocarro, J. N., Filardo, M., Edwards, M., McKenzie, T. L., & Floyd, M. (2014). Shared use of school facilities with community organizations and afterschool physical activity program participation: A cost–benefit assessment. *Journal of School Health*, *84*, 302–309.

Kanters, M. A., McKenzie, T. L., Edwards, M., Bocarro, J., Mahar, M., & Hodge, C. (2014, March). *Youth sport practice model gets more kids active with more time practicing skills* (Research brief). Retrieved from http://activelivingresearch.org/sites/default/files/Kanters_Brief_PhysicalActivityDuringSportPractice.pdf

Katzmarzyk, P. T., Walker, P., & Malina, R. M. (2001). A time-motion study of organized youth sports. *Journal of Human Movement Studies*, *40*, 325–334.

Klesges, R. C., Coates, T. J., Moldenhauer-Klesges, L. M., Holzer, B., Gustavson, J., & Barnes, B. (1984). The FATS: An observational system for assessing physical activity in children and associated parent behavior. *Behavioral Assessment*, *6*, 333–345.

Leek, D., Carlson, J. A., Cain, K. L., Henrichon, S., Rosenberg, D., Patrick, K., & Sallis, J. F. (2011). Physical activity during youth sports practices. *Archives of Pediatrics and Adolescent Medicine*, *165*, 294–299.

Levin, S., McKenzie, T. L., Hussey, J. R., Kelder, S., & Lytle, L. (2001). Variability of physical activity in physical education lessons across elementary school grades. *Measurement in Physical Education and Exercise Science*, *5*, 207–218.

Lounsbery, M. A., Holt, K. A., Monnat, S. A., Funk, B., & McKenzie, T. L. (2014). JROTC as a substitute for PE: Really? *Research Quarterly for Exercise and Sport*, *85*, 413–419. doi:10.1080/02701367.2014.930408

Lounsbery, M. A., McKenzie, T. L., Morrow, J. R., Monnat, S., & Holt, K. (2013). District and school physical education policies: Implications for physical education and recess time. *Annals of Behavioral Medicine*, *45*(Suppl. 1), S131–S141. doi:10.1007/s12160-012-9427-9

Mahar, M. T., Murphy, S. K., Rowe, D. A., Golden, J., Shields, A. T., & Raedeke, T. D. (2006). Effects of a classroom-based program on physical activity and on-task behavior. *Medicine & Science in Sports & Exercise*, *38*, 2086–2094.

Marshall, S. J., & Merchant, G. (2013). Advancing the science of sedentary behavior measurement. *American Journal of Preventive Medicine*, *44*, 190–191.

McClain, J. J., Abraham, T. L., Brusseau, T. A., & Tudor-Locke, C. (2008). Epoch length and accelerometer outputs in children: Comparison to direct observation. *Medicine & Science in Sports & Exercise*, *40*, 2080–2087.

McIver, K. L., Brown, W. H., Pfeiffer, K. A., Down, M., & Pare, R. R. (2009). Assessing children's physical activity in their homes: The Observational System for Recording Physical Activity in Children-Home. *Journal of Applied Behavior Analysis*, *42*, 1–16.

McKenzie, T. L. (1986). A behaviorally-oriented residential camping program for obese children and adolescents. *Education & Treatment of Children*, *9*(1), 38–46.

McKenzie, T. L. (1991). Observational measures of children's physical activity. *Journal of School Health*, *61*, 224–227.

McKenzie, T. L., Baquero, B., Crespo, N., Arredondo, E., Campbell, N., & Elder, J. P. (2008). Environmental correlates of physical activity in Mexican-American children at home. *Journal of Physical Activity and Health*, *5*, 579–591.

McKenzie, T. L., Cohen, D. A., Sehgal, A., Williamson, S., & Golinelli, D. (2006). System for Observing Play and Recreation in Communities (SOPARC): Reliability and feasibility measures. *Journal of Physical Activity and Health*, *1*, S203–S217.

McKenzie, T. L., Crespo, N. C., Baquero, B., & Elder, J. P. (2010). Leisure-time physical activity in elementary schools: Analysis of contextual conditions. *Journal of School Health*, *80*, 470–477. doi:10.1111/j.1746-1561.2010.00530.x

McKenzie, T. L., & Lounsbery, M. A. F. (2009). School physical education: The pill not taken. *American Journal of Lifestyle Medicine*, *3*, 219–225. doi:10.1177/1559827609331562

McKenzie, T. L., & Lounsbery, M. A. F. (2014). The pill not taken: Revisiting physical education teacher effectiveness in a public health context. *Research Quarterly for Exercise and Sport*, *85*, 287–292. doi:10.1080/02701367.2014.931203

McKenzie, T. L., Marshall, S. J., Sallis, J. F., & Conway, T. L. (2000). Leisure-time physical activity in school environments: An observational study using SOPLAY. *Preventive Medicine*, *30*, 70–77.

McKenzie, T. L., & Rushall, B. S. (1974). Effects of self-recording on attendance and performance in a competitive swimming environment. *Journal of Applied Behavior Analysis*, *7*, 199–206.

McKenzie, T. L., Sallis, J. F., & Nader, P. R. (1991). SOFIT: System for Observing Fitness Instruction Time. *Journal of Teaching in Physical Education*, *11*, 195–205.

McKenzie, T. L., Sallis, J. F., Nader, P. R., Broyles, S. L., & Nelson, J. A. (1992). Anglo- and Mexican-American preschoolers at home and at recess: Activity patterns and environmental influences. *Journal of Developmental and Behavioral Pediatrics*, *13*, 173–180.

McKenzie, T. L., Sallis, J. F., Patterson, T. L., Elder, J. P., Berry, C. C., Rupp, J. W., ... Nader, P. R. (1991). BEACHES: An observational system for assessing children's eating and physical activity behaviors and associated events. *Journal of Applied Behavior Analysis*, *24*, 141–151.

McNamee, J., & van der Mars, H. (2005). Accuracy of momentary time sampling: A comparison of varying interval lengths using SOFIT. *Journal of Teaching in Physical Education*, *24*, 282–292.

Montoye, H. J., Kemper, H., Saris, W., & Washburn, R. A. (1996). *Measuring physical activity and energy expenditure*. Champaign, IL: Human Kinetics.

Mowen, A. (2010). *Parks, playgrounds and Active Living: A research synthesis*. Princeton, NJ: Active Living Research. Retrieved from http://www.activelivingresearch.org/parks-playgrounds-and-active-living

Murad, M. H., & Montori, V. M. (2013). Synthesizing evidence: Shifting the focus from individual studies to the body of evidence. *Journal of the American Medical Association*, *309*, 2217–2218. doi:10.1001/jama.2013.5616

National Association for Sport and Physical Education & American Heart Association. (2012). *Shape of the nation report: Status of physical education in the USA*. Reston, VA: American Alliance for Health, Physical Education, Recreation and Dance.

National Federation of State High School Associations. (2013). *Annual report-2013*. Retrieved from http://www.nfhs.org/media/885658/2013-nfhs-annual-report.pdf

Owen, N., Healy, G. N., Howard, B., & Dunstan, D. W. (2012). *Too much sitting: Health risks of sedentary behavior and opportunities to change*. Washington DC: President's Council on Fitness, Sports and Nutrition.

Parra, D. C., McKenzie, T. L., Ribeiro, I., Akira, C., Dreisinger, M., Coniglio, K., … Simoes, E. (2010). Assessing physical activity in public parks in Brazil using systematic observation. *American Journal of Public Health*, *100*, 1420–1426.

Pate, R. R., Davis, M. G., Robinson, T. N., Stone, E. J., McKenzie, T. L., & Young, J. C. (2006). Promoting physical activity in children and youth: A leadership role for schools (AHA Scientific Statement). *Circulation*, *114*, 1214–1224.

Pate, R. R., McIver, K., Dowda, M., Brown, W. H., & Addy, C. (2008). Directly observed physical activity levels in preschool children. *Journal of School Health*, *78*, 438–444.

Pate, R. R., Trost, S., Levin, S., & Dowda, M. (2000). Sports participation and health-related behaviors among US youth. *Archives of Pediatrics and Adolescent Medicine*, *154*, 904–911. doi:10.1001/archpedi.154.9.904

Patterson, T. L., Sallis, J. F., Nader, P. R., Rupp, J. W., McKenzie, T. L., Roppe, B., & Bartok, P. (1988). Direct observation of physical activity and dietary behaviors in a structured environment: Effects of a family-based health promotion program. *Journal of Behavioral Medicine*, *11*, 447–458.

Pfeiffer, K. A., Dowda, M., Dishman, R. K., McIver, K. L., Sirard, J. R., Ward, D. S., & Pate, R. R. (2006). Sport participation and physical activity in adolescent females across a four-year period. *Journal of Adolescent Health*, *39*, 523–529.

Pope, R. P., Coleman, K. J., Gonzalez, E. C., Barron, F., & Heath, E. M. (2002). Validity of a revised System for Observing Fitness Instruction Time (SOFIT). *Pediatric Exercise Science*, *14*, 135–146.

Ridgers, N. D., Carter, L. M., Stratton, G., & McKenzie, T. L. (2011). Examining children's physical activity and play behaviors during school playtime over time. *Health Education Research*, *26*, 586–595.

Ridgers, N. D., Stratton, G., & McKenzie, T. L. (2010). Reliability and validity of the System for Observing Children's Activity and Relationships during Play (SOCARP). *Journal of Physical Activity and Health*, *7*, 17–25.

Rimmer, J. H., & Marques, A. C. (2012). Physical activity for people with disabilities. *The Lancet*, *380*, 193–195. doi:10.1016/S0140-6736(12)61028-9

Rimmer, J. H., Schiller, W., & Chen, M. D. (2012). Effects of disability-associated low energy expenditure deconditioning syndrome. *Exercise Sport Sciences Reviews*, *40*, 22–29.

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

Rowe, P. J., van der Mars, H., Schuldheisz, J. M., & Fox, S. (2004). Measuring physical activity in physical education: Validating SOFIT for use with high school students. *Journal of Teaching in Physical Education*, *23*, 235–251.

Ryan, S., & Lindsey, G. (2013). *Counting bicyclists and pedestrians to inform transportation planning: A research brief*. Princeton, NJ: Active Living Research. Retrieved from http://www.activelivingresearch.org/counting-bicyclists-and-pedestrians-inform-transportation-planning

Sacheck, J. M., Nelson, T. F., Ficker, L., Kafka, T., Kuder, J., & Economos, C. D. (2011). Physical activity during soccer and its contribution to physical activity recommendations in normal weight and overweight children. *Pediatric Exercise Science*, *23*, 281–292.

Sallis, J. F. (2009). Measuring physical activity environments: A brief history. *American Journal of Preventive Medicine*, *36*(Suppl. 4), S86–S92. doi:10.1016/j.amepre.2009.01.002

Sallis, J. F., Floyd, M. F., Rodriguez, D. A., & Saelens, B. E. (2012). Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation*, *125*, 729–737. doi:10.1161/CIRCULATIONAHA.110.969022

Sallis, J. F., & McKenzie, T. L. (1991). Physical education's role in public health. *Research Quarterly for Exercise and Sport*, *62*, 124–137.

Sallis, J. F., McKenzie, T. L., Beets, M. W., Beighle, A., Erwin, H., & Lee, S. (2012). Physical education's role in public health: Steps forward and backward over 20 years and HOPE for the future. *Research Quarterly for Exercise and Sport*, *83*, 125–135. doi:10.1080/02701367.2012.10599842

Sallis, J. F., McKenzie, T. L., Conway, T. L., Elder, J. P., Prochaska, J. J., Brown, M., … Alcaraz, J. E. (2003). Environmental interventions for eating and physical activity: A randomized controlled trial in middle schools. *American Journal of Preventive Medicine*, *24*, 209–217.

Sallis, J. F., Owen, N., & Fisher, E. B. (2008). Ecological models of health behavior. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), *Health behavior and health education: Theory, research, and practice* (4th ed., pp. 465–486). San Francisco, CA: Jossey-Bass.

Sanchez-Vaznaugh, E. V., O'Sullivan, M., & Egerter, S. (2013). *When school districts fail to comply with state physical education laws, the fitness of California's children lags* (Policy brief). Retrieved from http://activelivingresearch.org/sites/default/files/PolicyBrief_PEPolicyCali.pdf

Sasidharian, V., & McKenzie, T. L. (2014). SOPARNA: *System for Observing Physical Activity and Recreation in Natural Areas: Description and procedures manual*. San Diego, CA: San Diego State University. Retrieved from http://activelivingresearch.org/node/13208

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Sidman, M. (1960). *Tactics of scientific research*. New York, NY: Basic Books.

Sit, C. H. P., Capio, C. M., Cerin, E., & McKenzie, T. L. (2013). Assessment of measures of physical activity of children with cerebral palsy at home and school: A pilot study. *Journal of Child and Adolescent Behavior*, *1*, 112. doi:10.41172/jcalb.100112

Slater, S. J., Nicholson, L., Chriqui, J., Turner, L., & Chaloupka, F. (2012). The impact of state laws and district policies on physical education and recess practices in a nationally representative sample of US public elementary schools. *Archives of Pediatrics and Adolescent Medicine*, *166*, 311–316. doi:10.1001/archpediatrics.2011.1133

Strath, S. J., Kaminsky, L. A., Ainsworth, B. E., Ekelund, U., Freedson, P. S., Gary, R. A., … Swartz, A. M. (2013). Guide to assessment of physical activity: Clinical and research applications: A scientific statement from the American Heart Association. *Circulation*, *128*, 2259–2279.

U.S. Department of Health and Human Services. (2008a). *Physical Activity Guidelines Advisory Committee report*. Washington, DC: Author. Retrieved from http://www.health.gov/PAGuidelines/report/pdf/committeereport.pdf

U.S. Department of Health and Human Services. (2008b). *Physical Activity Guidelines for Americans*. Washington, DC: Author. Retrieved from http://www.health.gov/PAGuidelines

U.S. Department of Health and Human Services. (2012). *Physical Activity Guidelines for Americans midcourse report: Strategies to increase physical activity among youth*. Washington, DC: Author. Retrieved from http://www.health.gov/paguidelines/midcourse/pag-mid-course-report-final.pdf

van der Mars, H. (1989). Observer reliability: Issues and procedures. In P. W. Darst, D. Zakrajsek, & V. H. Mancini (Eds.), *Analyzing physical education and sport instruction* (2nd ed., pp. 54–80). Champaign, IL: Human Kinetics.

Ward, D. (2011). *School policies on physical education and physical activity: A research synthesis*. Princeton, NJ: Active Living Research. Retrieved from http://www.activelivingresearch.org/school-policies-physical-education-and-physical-activity

Weaver, R. G., Beets, M. W., Webster, C., & Huberty, J. (2014). System for Observing Staff Promotion of Activity and Nutrition (SOSPAN). *Journal of Physical Activity and Health*, *11*, 173–185.

Welk, G. (Ed.). (2002). *Physical activity assessments for health-related research*. Champaign, IL: Human Kinetics.

Whitt-Glover, M., Porter, A., & Yancey, T. (2013). *Do short physical activity breaks in classrooms work? A research brief*. Princeton, NJ: Active Living Research. Retrieved from http://www.activelivingresearch.org/do-short-physical-activity-breaks-classrooms-work