

International Studies of Physical Education Using SOFIT: A Review

Nicole J. Smith¹, Thomas L. McKenzie², Amber J. Hammons³

¹Department of Kinesiology, California State University, Fresno, Fresno, USA

²School of Exercise & Nutritional Sciences, San Diego State University, San Diego, USA

³Department of Child and Family Science, California State University, Fresno, USA

Email: njsmith@csufresno.edu

How to cite this paper: Smith, N. J., McKenzie, T. L., & Hammons, A. J. (2019). International Studies of Physical Education Using SOFIT: A Review. *Advances in Physical Education*, 9, 53-74.
<https://doi.org/10.4236/ape.2019.91005>

Received: November 14, 2018

Accepted: January 7, 2019

Published: January 10, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Objective evaluations are essential to improving physical education (PE) policy and practice, and the System for Observing Fitness Instruction Time (SOFIT) is a valid and reliable tool designed to reach this end. This review assesses peer-reviewed studies that used SOFIT to describe preK-12 PE in international schools. Methods were informed by Preferred Reporting Items for Systematic Reviews (PRISMA) and articles were located by searching nine library databases and Google Scholar. A total of 739 records were located, 567 were screened, and 29 full-text articles were scrutinized. Data extraction was conducted to evaluate the characteristics of the 29 studies and to synthesize commonly reported SOFIT variables. The studies, conducted on 5 continents, included direct observations of 2703 lessons in 348 schools taught by more than 600 teachers in 10 different countries. There was substantial variability in study characteristics, how results were reported, and in study outcomes. All studies assessed physical activity (PA) and 90% (n = 26) assessed both PA and lesson context. More than two-thirds of the studies (69%; n = 20) assessed PA, lesson context, and teacher behavior. A common goal of the reviewed studies was to describe PE using SOFIT, however, researcher modifications to the established protocol and variability in how results were reported limited data syntheses and generalizations. As SOFIT is widely endorsed for assessing PE policies and practices, researchers could improve the generalizability of their study findings by adhering to the standard SOFIT protocol and by reporting results in a consistent manner.

Keywords

Evaluation, Exercise, Physical Activity, Physical Education, Schools

1. Introduction

1.1. The Importance of PreK-12 Physical Education (PE)

The World Health Organization (WHO) recommends that children and adolescents engage in at least 60 minutes of moderate to vigorous physical activity (MVPA) daily that includes muscle and bone strengthening activities at least three times per week (WHO, 2011). Unfortunately, more than 80% of adolescents do not meet the guidelines (WHO, 2011) and increasing physical activity (PA) among school-age children is a global priority (WHO, 2018a).

The consequences of physical inactivity are severe as sedentary living is associated with numerous health conditions. Physical inactivity is associated with increased risk for overweight and obesity and the consequences become apparent at a young age. The World Health Organization (WHO), for example, has indicated the prevalence of obesity worldwide has tripled since the onset of the obesity crisis in the 1970's and that millions of children worldwide are already overweight or obese by age five (WHO, 2018b).

There is global consensus that physical education (PE) is an essential program within preK to grade 12 (preK-12) schools, largely because of its potential to increase PA and play an important role in obesity prevention (UNESCO, 2015). Schools reach nearly all children and most countries have established recommendations for PE that recognize the importance of engaging students in health-enhancing MVPA during PE in order to develop student physical fitness and motor skills and to promote the engagement of lifetime PA (Hardman, 2014).

Although key stakeholders recognize that quality PE programs are a worthwhile public health investment, numerous barriers impact both the quantity and quality of PE, including limited schedules, inadequately trained teachers, lack of curricular resources, and insufficient equipment and facilities (McKenzie & Lounsbery, 2009). Assessing how PE is conducted is an important step in overcoming these barriers.

Global efforts to evaluate children's PA and the quality of PE and other school-based PA opportunities are currently underway (Hardman, 2014; Tremblay et al., 2016). The Active Healthy Kids Global Alliance, for example, recently published Report Cards on PA for international schools from 38 countries located on 6 continents (Tremblay et al., 2016). As well, in 2013 the United Nations Educational, Scientific and Cultural Organization (UNESCO) published the results of a worldwide survey of PE administered in 232 countries (Hardman, 2014). These efforts demonstrate a commitment to monitoring PE and improving its quality worldwide; experts acknowledge, however, that current data are limited, partly because objective assessment tools have not been widely adopted (Hardman, 2014; Tremblay et al., 2016).

1.2. The System for Observing Fitness Instruction Time (SOFIT)

The System for Observing Fitness Instruction Time (SOFIT) is a valid and reliable

instrument for objectively assessing PE programs (McKenzie, 2012; McKenzie, Sallis, & Nader, 1991a; McKenzie & Smith, 2017). SOFIT provides objective and contextually rich-data on the conduct of PE lessons and has been widely used. Observers are trained to use SOFIT via a standardized observation protocol that includes video segments for both instruction and assessment. Momentary time sampling methods (i.e., 10 seconds observe; 10 seconds record) are employed to simultaneously code student PA levels (i.e., lying down, sitting, standing, walking/moderate, vigorous), lesson context (i.e., how lesson time is being spent—management, knowledge, fitness, skill development, game play, free time), and teacher behavior (i.e., time spent promoting fitness, demonstrating fitness, instructing generally, managing, observing, or doing other tasks) or teacher interactions (i.e., instances of promoting “in-class” or “out-of-class” PA). Observers also record lesson start and end times, lesson location, target student gender, teacher gender, grade level, and the number of boys and girls engaged in the lesson.

SOFIT student activity codes have been validated using a variety of methods, including heart rate monitoring, accelerometry, and pedometry (McKenzie et al., 1991a; Ridgers, Stratton, & McKenzie, 2010; McNamee & van der Mars, 2005). The validity of the contextual and behavioral categories is also well-established, with studies consistently reporting significant relationships between student PA levels, how lesson time is allocated, and how teachers spend their time and interact with students (McKenzie, et al., 1991a; McKenzie, Sallis, & Nader, 1991b; McKenzie et al., 1995; McKenzie, Marshall, Sallis, & Conway, 2000; Smith, Monnat, & Lounsbery, 2015). A recent review of SOFIT studies conducted in the US found consistently high inter-observer agreement (i.e., reliabilities > 85%) (McKenzie & Smith, 2017).

1.3. Purpose

The current investigation reviews studies that used SOFIT to assess PE in preK-12 schools located outside of the US. Specifically, our objectives are to describe the characteristics of international SOFIT studies and to quantitatively synthesize results for the SOFIT main variables (i.e., student PA levels, lesson context, teacher behavior) and two other commonly reported variables--class size and lesson length.

1.4. Significance

SOFIT has been widely used to assess PE internationally, and this investigation complements a review of SOFIT studies published in the US between 1991-2016 (McKenzie & Smith, 2017). This review increases awareness about research findings from studies that have utilized SOFIT to describe PA, lesson contexts, and teacher promotion of PA in international settings. The findings have important implications for public health stakeholders, teacher preparation programs, and researchers. Foremost, the findings increase awareness about the

potential of PE to increase PA internationally. This is important because of the need to obtain objective evidence about opportunities for children and adolescents to accrue health-related PA. The SOFIT data specifically shed light on how teachers allocate lesson time and interact with students during PE. These factors have important implications for designing professional development for current and future teachers. Finally, this review identifies the strengths and limitations of existing international SOFIT studies and should lead to improving the data collection methods and the reporting of results in future studies. As well, because SOFIT has been recommended for surveillance (McKenzie & Smith, 2017; IOM, 2013), our data summaries for student activity, lesson context, teacher behavior, class size, and lesson length contribute to efforts to monitor PE globally (WHO, 2018a; UNESCO, 2015; Hardman, 2014).

2. Materials and Methods

2.1. Review Guidelines

Based on the recommendations of Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; see **Figure 1**), we completed a series of steps (Liberati, et al., 2009). First, we determined inclusion and exclusion criteria for potential studies and then conducted a comprehensive search. We removed duplicates from the resulting lists and then screened the remaining abstracts and records. We obtained full-texts of selected papers to confirm their eligibility for inclusion and then extracted relevant data from studies meeting inclusion criteria.

2.2. Inclusion Criteria

To be included in the review, studies had to: 1) use the standard SOFIT protocol; 2) describe PE lessons taught in typical preK-12 schools located outside of the US; and 3) be published in English in a peer-reviewed journal between 1991-2017. **Table 1** describes the 29 studies meeting these general criteria. Of these, 12 met three additional criteria in order to be included in a quantitative synthesis (**Table 2**). These were: a) include data from at least 30 typical PE lessons that were not influenced by an experiment or intervention; b) report mean scores and standard deviations for the main SOFIT variables; and c) provide evidence that the observational data were collected reliably throughout the study.

2.3. Search Terms and Information Sources

We searched nine databases for full-text, peer-reviewed research articles using the terms “physical education” OR “PE” AND “System for Observing Fitness Instruction Time” OR “SOFIT” AND “lesson context.” The databases were: 1) Academic Search Ultimate; 2) CINAHL Plus with Full Text (EBSCO); 3) Education Research Complete (EBSCO); 4) PsycINFO; 5) SPORTDiscus with full text (EBSCO); 6) Physical Education Index (ProQuest); 7) PubMed; 8) Science Direct (Elsevier); and 9) Web of Science. As well, we searched the reference lists of selected papers and used Google Scholar to locate additional relevant papers.

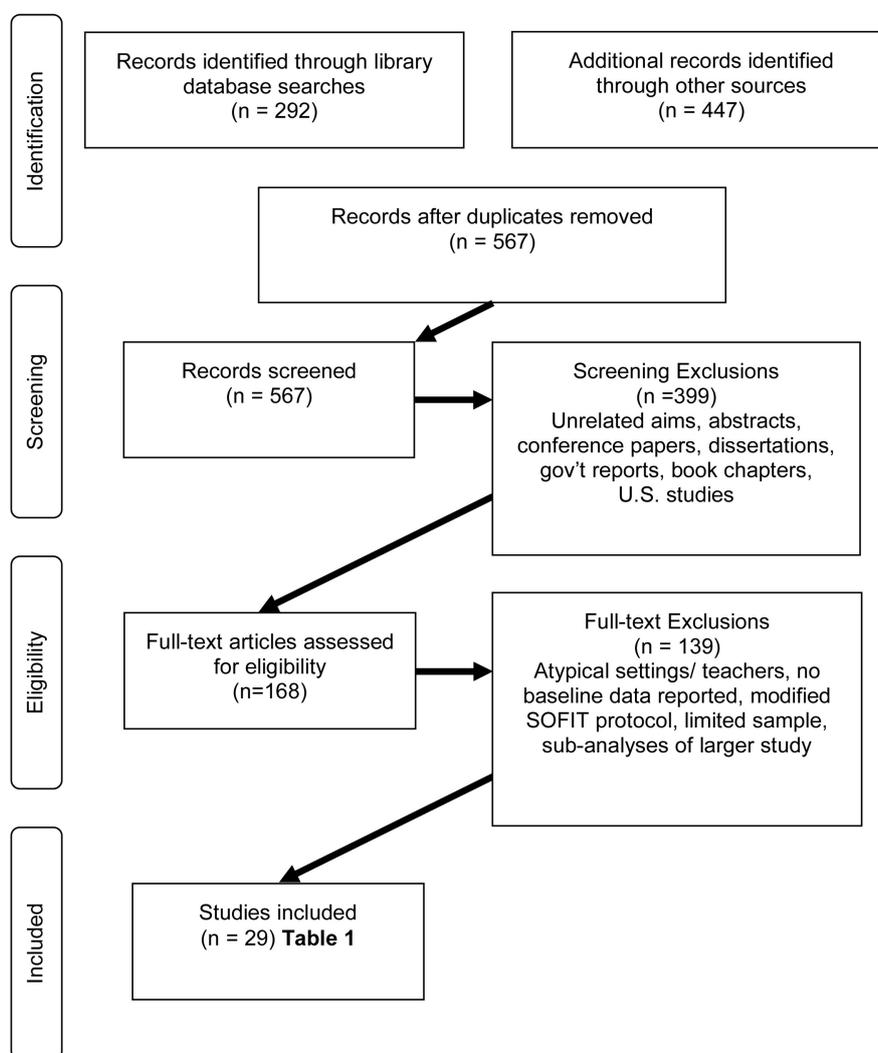


Figure 1. PRISMA flow diagram.

2.4. Data Extraction

All authors played a role in the process. The first author was responsible for initial data extraction with help from the third author and two student assistants. The first and third authors reviewed full-texts independently, and in the rare case of a disagreement, the second author arbitrated final decisions. The study characteristics extracted from the 29 papers that met initial inclusion criteria included: 1) author; 2) publication year [1991-2017]; 3) country; 4) study design [intervention/descriptive]; 5) study aims; 6) sample size [i.e., schools, lessons, teachers, classes]; 7) reliability [i.e., certification of observers prior to data collection and the maintenance of reliability throughout the study]; 8) main SOFIT categories [i.e., student PA levels, lesson context, teacher behavior, and teacher interaction]; and 9) analyses of other selected variables [e.g., student gender, teacher preparation, lesson location, PE dosage, energy expenditure, interaction between lesson context and MVPA, class size, and lesson length] (**Table 1**).

Table 1. Characteristics of Selected International SOFIT Studies (n = 29). (Note: Not all studies reported all characteristics).

First Author (Yr.) State/Region Study Design (D, I) ¹	Description/Design/Sample Schools (S); Lessons (L); Teachers (T); Classes (C)	Observer Reliabilities Identified ²	Main SOFIT Categories ³	Analysis ⁴	Other ⁵
		CR; FR	PA; LC; TB1; TB2	SG; TP; LL	PED; T; EE; I; CS
Preschools, (n = 2)					
Chow et al. (2015) Hong Kong (D)	Children's PA and associated variables Cross-sectional S = 4; L = 90; T = 25; C = 23	CR ⁶ ; FR	PA; LC; TB1	SG; LL	PED; T; EE; I; CS
Van Cauwenberghe (2012) Belgium (D)	PA levels and association with LC, and TB Cross-sectional S = 35; L = 35; T = 35; C = 35	CR ⁶ ; FR ⁷	PA; LC; TB1	SG; TP	PED ⁹ ; T; I; CS
Elementary Schools, (n = 16)					
Barnett et al. (2002) Australia (D)	How active are rural children in Australian PE? Non-experimental S = 18; L = 231	CR ⁶ ; FR	PA; LC	SG	T; I
Cardon et al. (2004) Belgium (D)	Physical activity levels in ES PE: Swimming vs non-swimming classes S = 16; L = 78; C = 39	CR ⁶ ; FR ⁷	PA	TP ⁸ ; LL	PED ⁹ ; T
Chow et al. (2008) Hong Kong (D)	Children's PA and environmental influences during PE S = 42; L = 368; T = 105; C = 126	CR ⁶ ; FR	PA; LC; TB1	TP ⁸ ; LL	PED ⁹ ; T; I; CS
da Cunha et al. (2016) Brazil (I)	Effect of educational program on children's energy expenditure during PE Randomized control trial S = 8; L = 79; T = 53; C = 48	FR ⁷	PA; LC; TB1	TP	T; EE
Da Costa et al. (2016) Brazil (D)	PA and LC in PE Cross-sectional sub-analyses S = 5; L = 12; T = 12; C = 12		PA; LC	LL	PED ⁹ ; T; I; CS
Gharib et al. (2015) Mexico (D)	Influence of LC and TB on PA Cross-sectional S = 20; L = 58; T = 58; C = 58		PA; LC; TB1	SG	PED; T; I
Hall-Lopez et al. (2017) Mexico (D)	MVPA during outdoor PE and recess Cross-sectional comparative analysis S = 23; L = 63; T = 63; C = 63	CR ⁶ ; FR	PA; LC		T
Jennings-Aburto et al. (2009) Mexico (D)	PA during the school day in public primary schools in Mexico City Cross-sectional S = 12; L = 26; T = 12; C = 12	CR ⁶ ; FR ⁶	PA; LC	SG	PED; T; I
Miller et al. (2016) Australia (I)	Efficacy of a game-centered approach on MVPA in PE Randomized control trial S = 1; L = 30; T = 4; C = 4	CR ⁶	PA	SG; TP	
Powell et al. (2016) United Kingdom (I)	Effectiveness of the SHARP Principles Model on MVPA Quasi-experimental, non-equivalent groups S = 2; L = 28; T = 15; C = 4	CR ⁶ ; FR ⁶	PA; LC; TB2	TP	CS

Continued

Safdie et al. (2013) Mexico (I)	Impact of school-based intervention on obesity risk factors in Mexican children Randomized control trial S = 27; L = 60; T = 38; C = 38		PA			PED ⁹
Sheehan (2015) Canada (D)	Assessment of MVPA during PE taught by PE specialist Cross-sectional S = 1; L = 54; T = 2; C = 14		PA; LC; TB1	SG; TP ⁸		I
Telford et al. (2016) Australia (I)	Four-year specialist-taught PE and PA: LOOK Cluster randomized control trial S = 29; L = 193; C = 68	CR ⁶	PA; LC; TB1	SG; TP		PED; T
Usher et al. (2016) Australia (D)	PA levels during PE Cross-sectional S = 10; L = 30; T = 10; C = 10	CR ⁶	PA; LC; TB2	TP ⁸		PED
van Beurden et al. (2003) Australia (I)	Evaluation of "Move it Groove it" Quasi-experimental S = 18; L = 465	CR ⁶ ; FR	PA; LC	SG; TP		T; I
Verstraete et al. (2007) Belgium (I)	Effectiveness of a two-year health-related PE intervention Quasi-experimental pretest-posttest design S = 16; L = 78; T = 16; C = 39	CR; FR ⁷	PA; LC	SG; TP ⁸		PED ⁹ ; I
Secondary Schools, (n = 10)						
Chow et al. (2009) Hong Kong (D)	PA and environmental influences during secondary PE Cross-sectional S = 30; L = 238; T = 65; C = 123	CR ⁶ ; FR	PA; LC; TB1	SG; TP ⁸ ; LL		PED; T; I; CS
Curtner-Smith et al. (1995) SW England (D)	PA during diverse activities in a health-related fitness PE program S = 5; L = 40; T = 20; C = 40	CR; FR ⁷	PA; LC; TB1	TP ⁸		T; CS
Curtner-Smith et al. (1996) SW England (D)	PE during summer term in one English town Cross sectional S = 5; L = 40; T = 20; C = 40	CR; FR ⁷	PA; LC; TB1	TP ⁸		T; CS
Dudley (2012) Australia (D)	Changes in PA, LC, and TB during PE Longitudinal cross-sectional descriptive S = 6; L = 132; C = 48	CR ⁶ ; FR	PA; LC; TB2			T; CS
Dudley (2012) Australia (D)	PA levels and movement skill instruction in PE Baseline cross-sectional S = 6; L = 81; C = 27	CR ⁶ ; FR	PA; LC; TB2	SG		T; I; CS
Fairclough & Stratton (2006) England (I)	Effects of a PE intervention to improve student PA Quasi-experimental S = 1; L = 12; T = 2; C = 2	CR ⁶ ; FR ⁷	PA; LC; TB1	SG; TP ⁸		T; EE; CS
Marques et al. (2017) Portugal (D)	Description of PE in one secondary school Cross-sectional S = 1; L = 30; T = 10	FR ⁷	PA; LC; TB1			PED ⁹ ; T
Mersh & Fairclough (2010) England (D)	PA, LC, and TB in one middle school Case study S = 1; L = 30; T = 2; C = 2	CR	PA; LC; TB1	SG; TP ⁸		PED; T; I

Continued

Smith et al. (2015) East of England (I)	PA of boys and girls during invasion games: Direct instruction vs. tactical games Quasi-experimental pre-test—post-test design S = 2; L = 48; T = 4; C = 4	CR ⁶ ; FR	PA; LC; TB1	SG; TP	T
Sutherland et al. (2016) Australia (D)	PA, LC, and TB in PE Cross-sectional descriptive S = 10; L = 100	CR ⁶ ; FR	PA; LC; TB2	SG; TP; LL	T; CS
Combined Schools, (n = 1)					
Santa Maria et al. (2010) Argentina (D)	Energy expenditure in PE in private elementary schools L = 55; C = 55	CR ⁶	PA; LC; TB1	SG	T; EE; CS

Notes: ¹Descriptive (D); Intervention (I); ²Certification Reliability (CR), Field Reliability (FR); ³Physical Activity (PA), Lesson Context (LC), Teacher Behavior (TB1), Teacher PA Promotion (TB2); ⁴Student Gender (SG), Teacher Preparation (TP), Lesson Location (LL); ⁵PE Dosage (PED); Lesson Length (T); Estimated Energy Expenditure (EE); Interaction between lesson context and physical activity (I), Class Size (CS), ⁶data not shown; ⁷video or audio recorded; ⁸All lessons taught by PE Specialists; ⁹Not measured objectively.

2.5. Quantitative Data Syntheses

We limited quantitative data syntheses to the main SOFIT variables (i.e., student PA levels, lesson context, teacher behavior) and two other commonly reported variables (class size and lesson length). Mean scores, standard deviation values, and sample size were extracted using an Excel tool. The range of mean scores was determined by sorting data from low to high values for each variable. Lower and upper values for the 95th confidence interval were estimated for MVPA%, a measure of PA intensity during lessons, using the formula:

$$\mu = M \pm t(s_M)$$

(Stangroom, 2018). Excel for Mac version 15.30 was used to compute the median, first and third quartiles, and interquartile range. **Figure 2** provides a forest plot to illustrate the average MVPA% for the studies as well as the lower and upper values for the 95th confidence interval. MVPA% from 0 - 100% is noted on the abscissa and the studies are listed in ascending order on the ordinate.

3. Results

3.1. Search Results

Figure 1 illustrates the number of records located, screened, and included in our report. We located a total of 739 records from 9 databases (n = 292) and other sources (n = 447). After removing duplicates (n = 172), we screened 567 records for eligibility and excluded 399 more for the reasons identified in **Figure 1**. We then evaluated 168 full-text articles and excluded 139 of them for reasons summarized in **Figure 1**, resulting in 29 studies that met inclusion criteria (**Table 1**).

3.2. Study Characteristics

Table 1 provides a detailed summary of the characteristics of the 29 studies

Table 2. Range of Study means, medians, and interquartile ranges for main SOFIT variables.

Variable	St ¹	L ²	Range of Means	Median	Q1	Q3	IQR
PHYSICAL ACTIVITY	12	1170					
Lying Down	12	1170	0.0 - 3.0	0.3	0.37	.74	.37
Sitting	12	1170	10.9 - 31.7	18.2	12.7	27.2	14.5
Standing	12	1170	16.3 - 65.8	37.4	29.4	42.2	12.8
Walking/moderate	12	1170	4.4 - 39.1	28.5	18.6	33.4	14.8
Vigorous	12	1170	9.0 - 23.8	18.8	13.4	21.5	8.1
MVPA	12	1170	20.9 - 58.2	41.9	37.8	50.5	12.7
LESSON CONTEXT	9	1050					
Management	9	1050	14.0 - 30.8	19.5	17.4	22.4	5.0
Knowledge	9	1050	7.1 - 26.3	15.2	12.4	17.7	5.3
Fitness Activity	9	1050	7.1 - 32.5	14.5	11.3	19.8	8.5
Skill Practice	9	1050	5.2 - 43.8	16.6	12.2	34.3	22.1
Game Play	9	1050	5.1 - 31.2	12.2	11.6	31.1	19.5
Other	9	1050	0.0 - 10	2.1	1.1	3.1	2.0
TEACHER BEHAVIOR	7	841					
Promotes fitness	7	841	0.0 - 21.2	9.3	.22	12.0	11.8
Demonstrates	6	751	0.0 - 13.0	5.8	1.1	9.8	8.7
General Instruction	7	841	6.7 - 69.2	54.7	50.5	64.0	13.5
Class Management	7	841	18.1 - 46.5	23.2	20.9	24.0	3.1
Observe	7	841	3.1 - 25.4	10.6	4.9	10.9	6.0
Other Tasks	6	801	0.1 - 0.7	0.3	0.2	.55	0.4
TEACHER INTERACTIONS	3	232					
IN Class	3	232	10.1 - 30.8	28.6	19.4	29.7	10.4
OUT of Class ³	1	100	-	-	-	-	-
None	3	232	68.9 - 89.7	71.2	70.1	80.5	10.4
CLASS SIZE	4	816	18.5 - 32.8	22.6	20.6	26.1	5.5
LESSON LENGTH	6	344	19.8 - 50.0	39.9	36.9	43.3	6.4

Notes: ¹Number of studies; ²Number of lessons observed; ³Promotion of out-of-class activity was reported in only one study (Sutherland et al., 2016; Mean intervals = 0.3%; SD = 0.8%).

meeting inclusion criteria. They included the direct observations of 2703 PE lessons that were taught by at least 603 teachers in more than 348 schools.

3.2.1. Setting & Participants

The studies were conducted in preschool ($n = 2$), elementary ($n = 16$), and secondary ($n = 10$) school settings and one included both elementary and secondary grade levels. Studies took place on five continents [Australia ($n = 8$), Europe ($n = 10$), South America ($n = 7$), Asia ($n = 3$), and North America ($n = 1$)]. They included 10 different countries/territories, with most studies taking place in Australia ($n = 8$), England ($n = 5$), and Mexico ($n = 4$).

3.2.2. Design & Main Variables Reported

Twenty studies (69%) were descriptive (D) and nine (31%) were part of an intervention (I). All 29 used the SOFIT PA codes and 26 (90%) also assessed lesson context. More than two-thirds ($n = 20$; 69%) described all three major categories--PA, lesson context, and teacher behavior. More studies used the original 6-category teacher behavior codes ($n = 15$; 52%) than the newer 3-category teacher interaction codes ($n = 5$; 17%).

3.2.3. Observer Reliability

Twenty-three studies (80%) described how data collectors were certified prior to starting data collection and 20 (69%) described the periodic assessment of observers (i.e., reliability) in the field during the data collection period. Studies consistently reported reliability scores met or exceeded the criteria standard ($\geq 85\%$ agreement; McKenzie, 2012) with inter observer agreements ranging between 80% - 90% for each main SOFIT variable (Mode = 85%) with between 84% - 100% for PA, 86% - 100% for lesson context, and 80% - 96% for teacher behavior.

3.2.4. Study Analyses & Other Variables Reported

Seventeen studies (59%) examined student gender, including 13 that compared boys and girls within the same lessons and four that investigated differences by class gender composition (i.e., boys-only, girls-only, and co-educational classes). Eight studies (28%) examined differences based on the preparation of teachers, mainly PE specialists vs classroom teachers. Ten studies (34%) described lessons taught only by PE specialists. Six studies (21%) investigated the location of lessons, with most comparing lessons taught indoors vs outdoors. Cardon et al. (2004), however, compared swimming and non-swimming lessons and Sutherland et al. (2016) compared lessons taught in rural and urban schools.

Twenty-three studies (79%) reported actual (i.e., observed) lesson length and 13 (45%) provided scheduled lesson length. PE dosage (i.e., lesson frequency \times lesson length) was reported anecdotally, but not objectively assessed. Thirteen studies (45%) reported the number of boys and girls present in class and 13 described student activity levels during the different lesson contexts. Only four studies (14%) reported estimated student energy expenditure rates (i.e., an overall measure of PA intensity).

3.2.5. Syntheses of Results Reported in the Studies

Table 2 presents the range of mean scores, medians, and interquartile ranges for

the SOFIT main variables that were identified in the 29 studies meeting the inclusion criteria for quantitative data syntheses (i.e., included at least 30 typical PE lessons not influenced by an intervention; reported mean scores and standard deviations; and provided evidence of observer reliability throughout the study). **Figure 2** provides a forest plot of mean MVPA% including the lower and upper values for the 95th confidence interval for 11 of the 12 studies included in the synthesis of MVPA%. **Table 2** and **Figure 2** indicate that there was substantial variability in the results both within and among the 29 studies. What follows is a description of syntheses for PA, lesson context, teacher behavior, teacher interactions, observed class size, and lesson length.

3.2.6. Physical Activity

Twelve of the 29 studies (41%) met the inclusion criteria for quantitative syntheses of PA. They included two preschool, three elementary, and six secondary school studies and a total of 1,170 lessons ($n = 125$ preschool; $n = 465$ elementary; $n = 580$ secondary) from 170 schools taught by more than 323 teachers (**Table 2**). Students typically spent most of lesson time standing (Median = 37.4%; IQR = 29.4% - 42.2%) or walking (Median = 28.5%; IQR = 18.6% - 33.4%) and little time being vigorous (Median = 18.8%; IQR = 13.4% - 21.5%). Study means for vigorous PA ranged between 9.0% (SD = 6.2%) and 23.8% (SD = 5.6%; **Table 2**). **Table 2** also shows that the means for MVPA% among studies ranged between 20.9% (SD = 21.0%) and 58.2% (SD = 5.7%), with the median MVPA% being 41.9% (IQR = 37.8% - 50.5%), which is 8.1% lower than the $\geq 50\%$ public health objective.

Figure 2 shows substantial variability in MVPA% (Walking/Moderate plus Vigorous) within and among the studies. **Figure 2** also shows that the mean MVPA% for 5 of the studies met or exceeded the public health objective of $\geq 50\%$

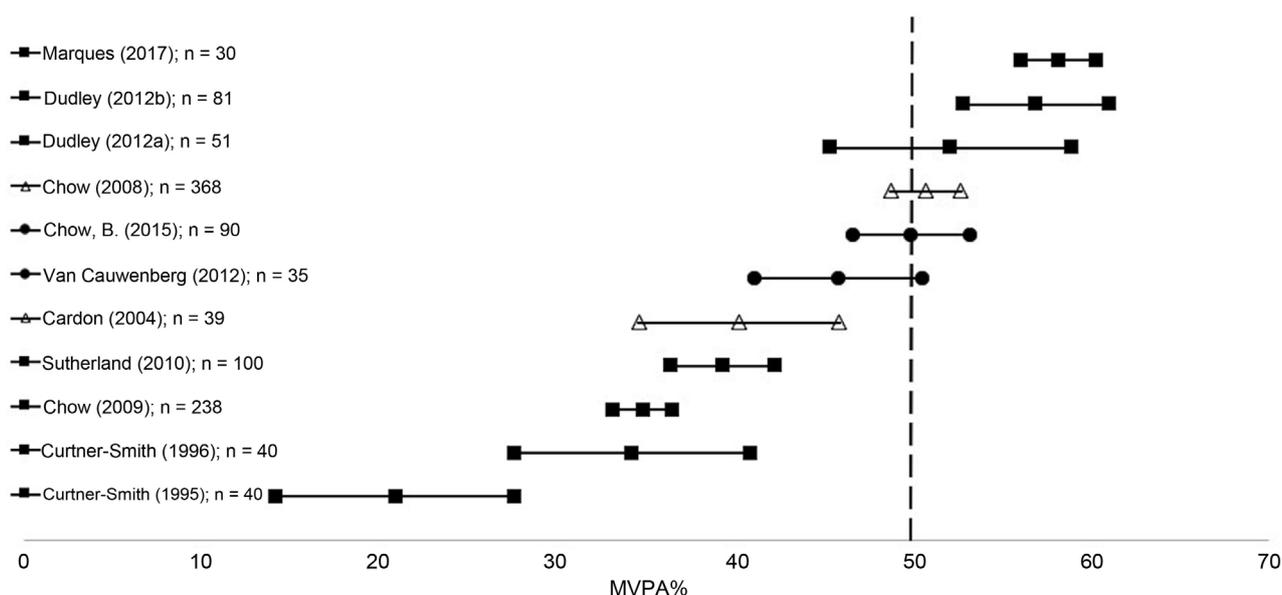


Figure 2. Forest plot of mean MVPA% in international SOFIT studies. ■ Secondary; ▲ Elementary; ● PreK.

MVPA. Mean MVPA% was above the median in the two preschool studies (45.8%, 49.9%), but below it in four secondary and one elementary school study. **Figure 2** also shows variability in MVPA% was particularly high in the secondary school studies, with MVPA% ranging between 20.9% and 58.2% (see **Table 2**).

Analyses for student gender were reported in 6 of the 12 studies that met the criterion for PA% syntheses (data not shown). Boys were typically observed being more physically active than girls, both when compared during coeducational lessons and when class gender composition (i.e., boys-only, girls-only lessons) was considered. For example, students in boys-only classes in Hong Kong secondary schools engaged in MVPA during 38.2% of lesson time compared to 31.8% for students in girls-only classes (Chow, McKenzie, & Louie, 2009). There was one exception, with Verstraete et al. (2007) reporting no gender differences MVPA% in their elementary school study in Belgium.

Other significant findings related to MVPA% were reported. For example, Van Cauwenberghe et al. (2011) reported students accumulated greater MVPA during lessons taught by early childhood specialists than non-specialists in Belgium preschools. Additionally, Sutherland et al. (2016) found higher MVPA% during Australian secondary school lessons taught by more experienced teachers and in those conducted in urban versus rural schools. Cardon et al. (2004) also reported that MVPA% increased during swimming lessons than in non-swimming lessons in elementary school PE in Belgium (Mean MVPA% = 52% vs. 40%).

3.2.7. Lesson Context

Nine studies (n = 31%) met the inclusion criteria for quantitative data syntheses for lesson context. These included 2 preschool, 2 elementary, and 5 secondary school studies for a total of 1,050 lessons (n = 125 preschool; n = 426 elementary; n = 500 secondary) in 150 schools taught by more than 304 teachers (**Table 2**). There was substantial variability both within and among studies in how teachers allocated time to the different lesson contexts (**Table 2**). Overall mean management time during lessons among studies ranged between 14.0% (SD = 9.4%) and 30.8% (SD = 13.2%), while time allocated for knowledge ranged between 7.1% (SD = 7.6%) and 26.3% (SD = 12.9%) and fitness activity time ranged between 7.1% (SD = 11.4%) and 32.5% (SD = 27.0%). The variability in lesson time allocation among studies was greatest for skill practice and game play, with skill practice time ranging between 5.2% (SD = 14.6%) and 43.8% (SD = 22.6%) of lessons and game play time ranging between from 5.1% (SD = 14.4%) and 46.6% (SD = 28.0%).

Noteworthy findings related to skill practice and game play were found for school level and country of origin. Skill practice was the most prevalent lesson context in the two preschool studies (Chow, McKenzie, & Louie, 2015; Van Cauwenberghe, Labarque, Gubbels, DeBourdeaudhuij, & Cardon, 2011) where it averaged 41.7% and 43.8% of lesson time. In comparison, game play was the most prevalent context in four of the five secondary school studies. On average

in these four studies, game play ranged between 12.1% and 46.6% of lesson time and skill practice occurred between 5.2% and 16.5% of lessons (data not shown). The exception was the Hong Kong secondary school study (Chow, et al., 2009) which reported students spent 36.5% of lesson time in skill practice and 12.1% of it in game play. Relative to country of origin, skill practice was the most prevalent context in all three Hong Kong studies (Chow, McKenzie, & Louie, 2008; Chow et al., 2009; Chow et al., 2015), regardless of school level (preschool, elementary, secondary) and game play was the most prevalent context in all three Australian secondary school studies (Dudley, Okely, Cotton, Pearson, & Caputi, 2012a; Dudley, Okely, Pearson, Cotton, & Caputi, 2012b; Sutherland, Campbell, Lubans et al., 2016).

Only six studies assessed MVPA% during different lesson contexts. Generally, lesson time allocated for fitness activities, skill practice, and game play was positively associated with MVPA%, and time for management and knowledge was negatively associated with it (Chow, et al., 2008; Chow, et al., 2009; Chow, et al., 2015; van Beurden, et al., 2003; Van Cauwenberghe, et al., 2011; Verstraete, 2007). The Verstraete et al. (2007) study found that involving teachers in a professional development intervention led to them being more efficient in allocating lesson time and this subsequently increased student MVPA%.

3.2.8. Teacher Behavior

Seven studies (n = 24%) met the inclusion criteria for a quantitative syntheses for teacher behavior. These included two preschool, one elementary, and four secondary school studies for a total of 841 lessons (n = 125 preschool, n = 368 elementary, and n = 348 secondary) from 122 schools taught by 280 teachers (Table 2). General instruction was most the most prevalent teacher behavior, and it occurred between 49.1% (SD = 12.8) and 69.2% (SD = 15.4) of the time in six of the seven studies (data not shown). In contrast, the same studies found teachers spent between 18.1% (SD = 11.4%) and 24.2% (SD = 20.7%) of lesson time in management and less than 13% of lesson time in fitness promotion (data not shown). The one exception was the Hong Kong preschool study (Chow, et al., 2015) where teachers were observed managing nearly half the time (Mean = 46.5%; SD = 21.5%) and spending little lesson time in general instruction (Mean = 6.7%; SD = 8.4%; data not shown).

3.2.9. Teacher Interactions

Five studies (17%) described teacher interactions, but only three secondary studies met the inclusion criteria for a quantitative synthesis. These included a total of 232 lessons from 22 schools taught by more than 48 teachers (Table 2). Teachers promoted student engagement in PA during PE between 10.1% (SD = 8.2%) and 30.8% (SD = 19.4%) of the 10 second observation intervals (Median = 28.6%; IQR = 19.4% - 29.7%; Table 2). Meanwhile, only Sutherland et al. (2016) found that that teachers promoted PA beyond the current lesson and they reported that it occurred rarely (Mean = 0.3% of intervals; SD = 0.8%; Table 2).

3.2.10. Observed Class Size

Thirteen studies (45%) reported observed class size, but only four (14%), met the criteria for inclusion in a quantitative synthesis. These included one preschool study ($n = 125$ lessons) and three secondary school studies ($n = 318$ lessons) for a total of 408 observed lessons in 44 schools taught by 130 teachers (Median = 22.6 students; IQR = 20.6 - 26.1; **Table 2**). The smallest classes observed were reported by Curtner-Smith, et al., 1995 in secondary schools in England (Mean = 18.5 students; SD = 6.0) and the largest were reported by Chow et al., 2009 in secondary schools in Hong Kong (Mean = 32.8 students; SD = 9.0).

3.2.11. Lesson Length

Lesson length was described in 23 studies (79%), but only six (21%) met the inclusion criteria for a quantitative synthesis. These included two preschool, one elementary, and three secondary school studies for a total of 344 total lessons ($n = 125$ preschool; $n = 39$ elementary; $n = 180$ secondary) in 75 schools and taught by more than 100 teachers (**Table 2**). Mean study lesson length ranged from 19.8 minutes (SD = 4.2) in four preschools in Hong Kong (Chow et al., 2015) to 43.8 minutes (SD = 11.8) in five secondary schools in England (Median = 39.9; IQR = 36.9 - 43.3; Curtner-Smith et al., 1995). Two Australian studies were not included in the quantitative syntheses because they did not report data means and standard deviations; nonetheless, lesson length in these cases ranged widely, between 19 - 110 minutes (data not shown; Dudley et al., 2012a; Dudley et al., 2012b).

Thirteen studies (45%) reported the number of PE minutes scheduled weekly, but only Chow et al. (2015) indicated students (preschool) had PE daily (between 25 - 30 minutes a day). The other 12 studies reported that students were typically scheduled to have PE lessons 1 - 2 days per week (Mode = 2 days per week) that they were between 20 - 120 minutes long (data not shown).

Actual observed lesson length was typically shorter than the scheduled lesson length because of student transitions to the instructional areas. Studies in Hong Kong elementary and secondary schools reported actual observed lessons were from 22% to 27% shorter than their scheduled lengths (Chow et al., 2008; Chow et al., 2009). In the Cardon et al. (2004) study, mean scheduled time was much longer for swimming lessons than regular lessons (83.0 min; SD = 22.0 min vs. 50.8 min; SD = 7.1 min; data not shown), however, lesson scheduled length was not significantly associated with the proportion of time that students were engaged in MVPA.

4. Discussion

Our purpose was to review SOFIT PE studies conducted in preK-12 schools outside the US. We located 739 records and systematically assessed 29 studies that were conducted in 10 different countries on 5 continents. Data for these studies were obtained via trained observers that used the same SOFIT instrument reliably to directly assess 2703 lessons that were taught by more than 603 teachers in

348 schools. Most of the 29 studies were conducted in elementary and secondary schools, but two involved preschools.

4.1. Study Characteristics

All 29 studies used SOFIT to describe PA, 90% described PA and lesson context, and 69% assessed PA, lesson context, and teacher behavior. Relative to teacher behavior, more studies assessed how teachers spent lesson time generally (i.e., teacher behavior categories, $n = 15$; 52%) rather than assessing teacher interactions related to promoting PA (teacher interaction, $n = 5$; 17%). Assessments of teachers promoting PA “in” and “out” of PE lessons are thus limited; as teacher promotion of PA is important, future studies should focus on it.

Although 90% of studies examined both PA and lesson contexts, only 13 (45%) assessed PA levels during the different contexts. Such an analysis requires entering data line-by-line data rather than entering lesson summary scores only. Entering data line-by-line is especially recommended for intervention studies because it will enable a more fine-tuned analysis of how changes in MVPA came about.

Synthesizing the results of studies was challenging because papers often did not always report specific information, such as for sample sizes (i.e., number of schools, teachers, and/or classes), field reliability tests, and standard deviations. Precision in sample size (i.e., number of schools, teachers, and classes) was lacking in numerous studies. Specifically, within the 29 studies where data were synthesized, one paper did not identify the number of schools, six did not identify the number of teachers, and five did not indicate the number of different classes observed. Additionally, it was not always clear if “lessons” and “classes” were distinct or if the terms were synonymous. Accurate and complete reporting of sample sizes is essential for understanding the scope of studies and should be reported consistently (e.g., how many schools were included, how many teachers, and how many distinct classes).

In some cases, the trustworthiness of the data was limited because observer reliabilities were not reported. Reliabilities were reported for 25 of the 29 studies, and the results consistently exceeded the established SOFIT protocol standard (i.e., >85% agreement). Not all studies reported detailed scores for certification and field tests, and subsequently 12 (41%) were excluded from quantitative syntheses because they did not provide sufficient evidence of data reliability throughout the study. Of these, four did not report any reliabilities, seven described reliabilities only during observer training, and one reported a low kappa value (i.e., $\kappa = 0.091$). A strength of SOFIT is that following a standardized protocol makes comparisons among studies possible, but this is only appropriate when the data are trustworthy (i.e., reliable).

Syntheses of lesson length and class size were limited, mainly because standard deviation scores and or reliabilities were not reported. Nearly 80% of studies ($n = 23$) described actual lesson length and 45% ($n = 13$) reported class size; however, only six and four studies, respectively, were synthesized, primarily

because standard deviation scores were not reported and/or it was not clear if observer reliabilities were maintained. Lesson length and class size have important implications for PE dosage and program quality, and it is important that this information be included in studies. Future reports should also include standard deviation scores and results of reliability assessments.

4.2. Lesson Characteristics

Only 12 out of 29 studies met the criteria for synthesis of PA% and fewer studies qualified for syntheses of other variables (i.e., lesson context, teacher behavior, teacher interactions, class size, and lesson length). Nonetheless, important findings emerged relative to the variability of study means scores and ranges of means among studies (**Table 2**). Results for PA were highly variable, but overall, students spent large amounts of lesson time being inactive. They spent more time standing (Median = 37.4%) compared to walking/moderate (Median = 28.5%) and engaging in vigorous PA (Median = 18.8%). Study means for vigorous PA% were between 9.0% - 23.8%. With lessons being infrequent and oftentimes of short duration, it appears little time was available for students to improve their physical fitness.

Only five of 29 studies met the public health goal of 50% MVPA. Further, there were differences by student gender, with boys accruing more MVPA than girls. This was found during both coeducational lessons and during boys-only and girls-only lessons. Teachers should strive to achieve the public health goal of 50% MVPA and provide more equitable PA opportunities for boys and girls.

An important finding was the large variability among studies in how time was allocated to the different lesson contexts (**Table 2**). Study means ranged widely; 41.5% for game play, 38.6% for skill practice, 25.4% for fitness, 16.8% for management, and 19.2% for knowledge (**Table 2**). These findings indicate lesson efficiency can be improved and suggest that assessing PA during different lesson contexts is important.

There was also variability in teacher behavior among the studies. Teachers spent more time in general instruction, rather than demonstrating and promoting fitness. In the four studies that assessed teacher interactions, teacher promotion of MVPA beyond the immediate lesson was observed rarely (during less than 1% of observation intervals).

The variability in PA, time spent in lesson contexts, and teacher behavior within and among studies illustrates that the conduct PE is substantially different and may depend on where a child lives and goes to school. Time allocations for different lesson contexts and teacher behaviors reflect both programmatic goals and teacher expertise. PE stakeholders can benefit from ongoing dialogue related to PE curricula and instructional methods with the aim of greater consistency within and among programs worldwide.

Class size and lesson length varied widely and these variables have important implications for program outcomes. [Chow et al. \(2008\)](#) counted an average of

33.6 students in Hong Kong lessons (range = 15 to 45), nearly twice as many as many as in the two studies by Curtner-Smith and colleagues in England in 1995 and 1996 and a third larger than the two secondary school studies in Australia reported by Dudley and colleagues in 2012. As well, PE was typically offered only two days per week with daily PE was identified only for the children in Hong Kong preschools. As well, the total minutes per lesson (e.g., 20 - 120 minutes) and per week varied widely. In many cases investigators reported that there were regional recommendations for PE time, but they also identified that school administrators were responsible for making site-based scheduling decisions for PE. Greater consistency in class size and lesson length at the school site level could ensure students have more equitable opportunities to become physically educated regardless of where they live.

4.3. Comparisons with the US Review

The findings of this investigation are similar to those reported in our review of SOFIT studies conducted in the US (McKenzie & Smith, 2017). For example, the challenges with synthesizing data were similar due important information being left out or reported inconsistently (i.e., observer reliabilities, sample sizes, and standard deviations). Nonetheless, there was similar variability in lesson characteristics in both the US studies and the current ones (e.g., how time was spent in lesson contexts).

A major difference between the US and international studies is the sample size, especially the number of lessons and schools observed. The 29 US studies included observations of 12,256 lessons, nearly five times the number of the 29 international studies. This difference is likely because SOFIT was used in randomized control-trials (e.g., SPARK, MSPAN, CATCH, TAAG) that were conducted in the US and sponsored by the National Institutes of Health (NIH).

5. Conclusion

The current description is limited to the assessment of the peer-reviewed reports of 29 different investigations that included direct observations of 2703 lessons using SOFIT in schools in 10 countries. Our syntheses of the main SOFIT variables were restricted to only the 12 studies that included at least 30 typical PE lessons that were not influenced by experiment or intervention, identified mean scores and standard deviations for main SOFIT variables, and provided evidence of observer reliability throughout the study. As the original study locations (e.g., county, city, school district, and schools) and the lessons themselves were not selected at random, our results may not accurately reflect the conduct of PE globally.

Nonetheless, the review has important implications for increasing awareness about the characteristics of preK-12 PE in international schools and for the conduct of future PE studies. Assessing PE is essential for improving its quality, and SOFIT has potential as a ground truthing tool that helps inform program-

matic and instructional improvement efforts. In order to realize this potential, however, there is need for additional observations of PE in preK-12 international schools and for greater consistency in study design and how results are reported.

To inform policy and best practices that could improve PE globally, it is important for future investigations using direct observation to establish observer reliability prior to the start of data collection and continue to assess it throughout the study. As well, the utility and generalizability of the results of these studies can be improved by reporting sample sizes, means, and standard deviations scores in a consistent manner. Improved generalizability could result from investigators adhering to the standard SOFIT protocol and using the observer training videos that available for no cost on YouTube. For larger studies, investigators should consider using the *iSOFIT*iOS application. This app is free and it has the potential to streamline data entry and reporting processes (e.g., it generates data graphs immediately and can export data files via email).

SOFIT provides objective data on student physical activity levels and how teachers allocate lesson time and behave during lessons. The resulting information can be used to assess how well these factors align with programmatic and instructional goals. PE goals may differ by country, state/province, school district, school, grade level, and even teacher. SOFIT was developed with the belief that PE should be conducted in a pleasant environment that provides students with ample amounts of MVPA in order for them to accrue health benefits while simultaneously becoming physically fit and motorically skilled. The instrument examines the potential of lessons relative to these goals; it does not assess opportunities for students to reach other relevant PE goals such as cognitive, social, and emotional outcomes.

Acknowledgements

We thank California State University Fresno students Jenna Aoki and Calixte Aholu for their assistance with data extraction.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

Authors' Contributions

The study was conceptualized by NS and TM. NS was responsible for all aspects of the process including the literature search, study selection, data extraction, data synthesis, and manuscript preparation. TM guided study conceptualization and methodology and made substantial contributions during the writing process. AH assisted with study selection, data extraction, and assessment of reliabilities. All authors read and reviewed the final version of the manuscript and agree with the order of presentation of the authors.

Funding

This research received no funding from agencies in the public, commercial, or not-for-profit sectors.

Availability of Data and Material

Supplementary data are available upon request.

References

- Barnett, L. M., van Beurden, E., Zask, A., Brooks, L. O., & Dietrich, U. C. (2002). How Active Are Rural Children in Australian Physical Education? *Journal of Science & Medicine in Sport*, *5*, 253-265. [https://doi.org/10.1016/S1440-2440\(02\)80011-1](https://doi.org/10.1016/S1440-2440(02)80011-1)
- Cardon, G., Verstraete, S., De Clercq, D., & De Bourdeaudhuij, I. (2004). Physical Activity Levels in Elementary-school Physical Education: A Comparison of Swimming and Nonswimming Classes. *Journal of Teaching in Physical Education*, *23*, 252-263. <https://doi.org/10.1123/jtpe.23.3.252>
- Chow, B. C., McKenzie, T. L., & Louie, L. (2008). Children's Physical Activity and Environmental Influences during Elementary School Physical Education. *Journal of Teaching in Physical Education*, *27*, 38-50. <https://doi.org/10.1123/jtpe.27.1.38>
- Chow, B. C., McKenzie, T. L., & Louie, L. (2009). Physical Activity and Environmental Influences during Secondary School Physical Education. *Journal of Teaching in Physical Education*, *28*, 21-37. <https://doi.org/10.1123/jtpe.28.1.21>
- Chow, B. C., McKenzie, T. L., & Louie, L. (2015). Children's Physical Activity and Associated Variables during Preschool Physical Education. *Advances in Physical Education*, *5*, 39-49. <https://doi.org/10.4236/ape.2015.51005>
- Curtner-Smith, M. D., Chen, W., & Kerr, I. G. (1995). Health-Related Fitness in Secondary School Physical Education: A Descriptive Analytic Study. *Educational Studies*, *21*, 55-66. <https://doi.org/10.1080/0305569950210105>
- Curtner-Smith, M. D., Kerr, I. G., & Clapp, A. J. (1996). The Impact of a National Curriculum Physical Education on the Teaching of Health-Related Fitness: A Case Study in One English Town. *European Journal of Physical Education*, *1*, 66-83. <https://doi.org/10.1080/1740898960010107>
- da Costa, B. G. G., da Silva, K. S., Hino, A. A. F., Leal, D. B., Gripa, L. T., & de Assis, M. A. A. (2016). Context and Intensity of Primary School Physical Education Classes in Brazilian Schoolchildren. *Mathews Journal of Pediatrics*, *1*, 003. http://www.mathewsopenaccess.com/PDF/Pediatrics/M_J_Pedi_1_1_003.pdf
- Da Cunha, C. T., Poblacion, A. P., Colugnati, F. A. B., Taddei, J. A. A. C., & Bracco, M. M. (2016). Effect of an Educational Program on Schoolchildren's Energy Expenditure during Physical Education Classes. *Medical Express (Sao Paulo, online)*, *3*, M160104. <https://doi.org/10.5935/MedicalExpress.2016.01.04>
- Dudley, D. A., Okely, A. D., Cotton, W. G., Pearson, P., & Caputi, P. (2012a). Physical Activity Levels and Movement Skill Instruction in Secondary School Physical Education. *Journal of Science & Medicine in Sport*, *15*, 231-237. <https://doi.org/10.1016/j.jsams.2011.10.005>
- Dudley, D. A., Okely, A. D., Pearson, P., Cotton, W. G., & Caputi, P. (2012b). Changes in Physical Activity Levels, Lesson Context, and Teacher Interaction during Physical Education in Culturally and Linguistically Diverse Australian Schools. *International Journal of Behavioral Nutrition & Physical Activity*, *9*, 114.

- <https://doi.org/10.1186/1479-5868-9-114>
- Fairclough, S. J., & Stratton, G. (2006). Effects of a Physical Education Intervention to Improve Student Activity Levels. *Physical Education & Sport Pedagogy, 11*, 29-44. <https://doi.org/10.1080/17408980500467613>
- Gharib, H., Galaviz, K. I., Lee, R. E., Safdie, M., Tolentino, L., Barquera, S., & Levesque, L. (2015). The Influence of Physical Education Lesson Context and Teacher Behaviour on Student Physical Activity in Mexico. *Retos, 28*, 160-164.
- Hall-Lopez, J. A., Ochoa-Martinez, P. Y., Burrueal, R. Z., Ortiz, L. R. M., & Bunuel, P. S. L. (2017). Moderate to Vigorous Physical Activity during Recess and Physical Education among Mexican Elementary School Students. *Retos, 31*, 137-139.
- Hardman, K. (2014). *World-Wide Survey of school Physical Education*. Paris: United Nations Educational, Scientific and Cultural Organization (UNESCO). <http://unesdoc.unesco.org/images/0022/002293/229335e.pdf>
- Institute of Medicine (IOM) (2013). *Educating the Student Body: Taking Physical Activity and Physical Education to School*. Washington DC: The National Academies Press.
- Jennings-Aburto, N., Nava, F., Bonvecchio, A., Safdie, M., Gonzalez-Casanova, I., Gust, T., & Rivera, J. (2009). Physical Activity during the School Day in Primary Schools in Mexico City. *Salud Pública México, 51*, 141-147. <https://doi.org/10.1590/S0036-36342009000200010>
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Loannidis, J. P. A. et al. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Health-Care Interventions: Explanation and Elaboration. *PLOS Medicine, 6*, e1000100. <https://doi.org/10.1371/journal.pmed.1000100>
- Marques, A., Ferro, N., Martins, J., & da Costa, F. C. (2017). The Performing of a Secondary Physical Education Department Committed to the Portuguese Physical Education National Curriculum. *Motricidade, 13*, 100-111. <http://www.scielo.mec.pt/pdf/mot/v13s1/v13s1a14.pdf>
- McKenzie, T. L. (2012). *SOFIT Protocol*. <https://activelivingresearch.org/sofit-system-observing-fitness-instruction-time>
- McKenzie, T. L., & Lounsbury, M. A. F. (2009). School Physical Education: The Pill Not Taken. *American Journal of Lifestyle Medicine, 3*, 219. <https://doi.org/10.1177/1559827609331562>
- McKenzie, T. L., & Smith, N. J. (2017). Studies of Physical Education in the United States using SOFIT: A Review. *Research Quarterly for Exercise & Sport, 88*, 492-502. <https://doi.org/10.1080/02701367.2017.1376028>
- McKenzie, T. L., Feldman, H., Woods, S. E., Romero, K. A., Dahlstrom, V., Stone, E. J. et al. (1995). Children's Activity Levels and Lesson Context during Third-Grade Physical Education. *Research Quarterly for Exercise & Sport, 66*, 184-193. <https://doi.org/10.1080/02701367.1995.10608832>
- McKenzie, T. L., Marshall, S. J., Sallis, J. F., & Conway, T. L. (2000). Student Activity Levels, Lesson Context, and Teacher Behavior during Middle School Physical Education. *Research Quarterly for Exercise & Sport, 71*, 249-259. <https://doi.org/10.1080/02701367.2000.10608905>
- McKenzie, T. L., Sallis, J. F., & Nader, P. R. (1991a). SOFIT: System for Observing Fitness Instruction Time. *Journal of Teaching in Physical Education, 11*, 195-205. <https://doi.org/10.1123/jtpe.11.2.195>
- McKenzie, T. L., Sallis, J. F., Nader, P. R., Patterson, T. L., Elder, J. P., Berry, C. C. et al. (1991b). BEACHES: An Observational System for Assessing Children's Eating and

- Physical Activity Behaviors and Associated Events. *Journal of Applied Behavior Analysis*, 24, 141-151. <https://doi.org/10.1901/jaba.1991.24-141>
- 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. <https://doi.org/10.1123/jtpe.24.3.282>
- Mersh, R., & Fairclough, S. J. (2010). Physical Activity, Lesson Context, and Teacher Behaviours within the Revised English National Curriculum for Physical Education: A Case Study of One School. *European Physical Education Review*, 16, 29-45. <https://doi.org/10.1177/1356336X10369199>
- Miller, A., Christensen, E., Eather, N., Gray, S., Sproule, J., Keay, J., & Lubans, D. (2016). Can Physical Education and Physical Activity Outcomes Be Developed Simultaneously Using a Game-Centered Approach? *European Physical Education Review*, 22, 113-133. <https://doi.org/10.1177/1356336X15594548>
- Powell, E., Woodfield, L. A., & Nevil, A. M. (2016). Increasing Physical Activity Levels in Primary School Physical Education: The SHARP Principles Model. *Preventive Medicine Reports*, 3, 7-13. <https://doi.org/10.1016/j.pmedr.2015.11.007>
- 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 & Health*, 7, 17-25. <https://doi.org/10.1123/jpah.7.1.17>
- Safdie, M., Jennings-Aburto, N., Levesque, L., Janssen, I., Campirano-Nunez, F., Lopez-Olmedo, N. et al. (2013). Impact of a School-Based Intervention Program on Obesity Risk Factors in Mexican Children. *SaludPublica de Mexico*, 55, S374-S387.
- Santa Maria, C. J., Laino, F. A., & Pintamalli, J. C. (2010). Energy Expenditure and Qualitative Aspects of Physical Education Lessons in Buenos Aires City, Argentina. *Physical Education & Sport*, 8, 9-20. <http://facta.junis.ni.ac.rs/pe/pe201001/pe201001-02.pdf>
- Sheehan, D. (2015). Instructor Behavior and Youth Physical Activity in Recreation Center Programs: The Role of Management in Improving Outcomes. *Journal of Park & Recreation Administration*, 33, 1-16. <https://js.sagamorepub.com/jpra/article/viewFile/6783/5353>
- Smith, L., Harvey, S., Savory, L., Fairclough, S., Kozub, S., & Kerr, C. (2015). Physical Activity Levels and Motivational Responses of Boys and Girls: A Comparison of Direct Instruction and Tactical Games Models of Games Teaching in Physical Education. *European Physical Education Review*, 21, 93-113. <https://doi.org/10.1177/1356336X14555293>
- Smith, N. J., Monnat, S. M., & Lounsbury, M. A. F. (2015). Physical Activity in Physical Education: Are Longer Lessons Better? *Journal of School Health*, 85, 141-148. <https://doi.org/10.1111/josh.12233>
- Stangroom, J. (2018). *Social Science Statistics. Single Sample Confidence Interval Calculator*. <https://www.socscistatistics.com/confidenceinterval/Default2.aspx>
- Sutherland, R., Campbell, E., Lubans, D. R., Morgan, P. J., Okely, A. D., Nathan, N. et al. (2016). Physical Education in Secondary Schools Located in Low-Income Communities: Physical Activity Levels, Lesson Context and Teacher Interaction. *Journal of Science & Medicine in Sport*, 19, 135-141. <https://doi.org/10.1016/j.jsams.2014.12.003>
- Telford, R. M., Olive, L. S., Cochrane, T., Davey, R., & Telford, R. D. (2016). Outcomes of a Four-Year Specialist Taught Physical Education Program on Physical Activity: A Cluster Randomized Controlled Trial, the Look Study. *International Journal of Behavioral Nutrition & Physical Activity*, 13, 64. <https://doi.org/10.1186/s12966-016-0388-4>
- Tremblay, M. S., Barnes, J. D., Gonzalez, S. A., Katzmarzyk, P. T., Onywera, V. O., Reilly, J. J. et al. (2016). Global Matrix 2.0: Report Card Grades on the Physical Activity of

- Children and Youth Comparing 38 Countries. *Journal of Physical Activity & Health*, 13, s343-s366. <https://doi.org/10.1123/jpah.2016-0594>
- United Nations Educational, Scientific and Cultural Organization (UNESCO) (2015). *Quality Physical Education: Guidelines for Policy-Makers*. Paris: UNESCO. <http://unesdoc.unesco.org/images/0023/002311/231101E.pdf>
- Usher, W., Keegan, R., & Edwards, A. (2016). Taking the “Physical” out of Physical Education. *Cogent Education*, 3, 1181025. <https://doi.org/10.1080/2331186X.2016.1181025>
- van Beurden, E., Barnett, L. M., Zask, A., Dietrich, U. C., Brooks, L. O., & Beard, J. (2003). Can We Skill and Activate Children through Primary School Physical Education Lessons? “Move It Groove It”—A Collaborative Health Promotion Intervention. *Preventive Medicine*, 36, 493-501. [https://doi.org/10.1016/S0091-7435\(02\)00044-0](https://doi.org/10.1016/S0091-7435(02)00044-0)
- Van Cauwenberghe, E., Labarque, V., Gubbels, J., DeBourdeaudhuij, I., & Cardon, G. (2011). Preschooler’s Physical Activity Levels and Associations with Lesson Context, Teacher’s Behavior, and Environment during Preschool Physical Education. *Early Childhood Research Quarterly*, 27, 221-230.
- Verstraete, S., Cardon, G., DeClercq, D., & DeBourdeaudhuij, I. (2007). Effectiveness of Two-Year Health-Related Physical Education Intervention in Elementary Schools. *Journal of Teaching Physical Education*, 26, 20-34. <https://doi.org/10.1123/jtpe.26.1.20>
- World Health Organization (WHO) (2011). *Global Recommendations on Physical Activity for Health: 5-17 Years Old*. <https://www.who.int/dietphysicalactivity/pa/en/>
- World Health Organization (WHO) (2018a). *Global Action Plan on Physical Activity 2018-2030: More Active People for a Healthier World* (p. 104). Geneva: World Health Organization. <http://apps.who.int/iris/bitstream/handle/10665/272722/9789241514187-eng.pdf>
- World Health Organization (WHO) (2018b). *Fact Sheets Obesity and Overweight*. <http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>

Abbreviations

CS	Class size
I	Interaction between LC and MVPA
LC	Lesson context
LL	Lesson location
PA	Physical activity
PE	Physical education
PreK-12	Preschool-Kindergarten-12 th grade
MVPA	Moderate to vigorous physical activity
SOFIT	System for Observing Fitness Instruction Time
T	Lesson length
TB1	Teacher behavior
TB2	Teacher interaction