Children at non-evidence-based/standard physical activity instruction programs engaged in higher levels of moderate-to-vigorous physical activity and showed greater improvements in fitness levels over time.

6

Evidence-based fitness promotion in an afterschool setting: Implementation fidelity and its policy implications

Jean M. Thaw, Manuela Villa, David Reitman, Christian DeLucia, Vanessa Gonzalez, K. Lori Hanson

REGULAR PHYSICAL ACTIVITY (PA) in children and adolescents is associated with many health and psychosocial benefits, including improved cardiovascular fitness, psychological well-being, and a decreased risk of obesity.[1] Unfortunately, PA levels among youth remain below recommended levels, with sharply decreasing PA levels occurring as children age through adolescence, especially among minority youth.[2]

Support for this study was provided by the Active Living Research Program of the Robert Wood Johnson Foundation (2010–2011) Grant 67304 (Jean Thaw, principal investigator) and by The Children's Trust. The authors would like to thank Dr. Paul Rosengard (SPARK) and Dr. Thom McKenzie (SPARK) for their helpful comments on this chapter.

Numerous barriers to implementing school-based PA promotion programs have been identified, including fewer school resources and increased academic demands that have reduced or eliminated the time previously allocated for physical education and recess.[3] Although school-based interventions are being implemented with some success, school-based PA currently contributes less than half of the total daily activity among children and adolescents, with almost half of the daily activity occurring after school hours.[4] Still, many children remain sedentary when school is not in session.[5]

OST programs are emerging as vital to PA promotion. Estimates suggest that as many as 8.4 million youth in the United States participate in some form of OST program, and public support is increasing for the development and funding of OST programs in public schools.[6] OST programs offer as much as one-third of the sixty minutes of daily moderate-to-vigorous PA (MVPA) recommended by the U.S. Centers for Disease Control and Prevention (CDC).[7] OST programs vary widely in terms of the types of PAs offered, staff training and background in PA, availability of indoor and outdoor spaces, and resources for PA promotion (for example, portable or fixed PA equipment).

Because research on best practices for PA in OST programs is in its early stages, it is difficult to discern which aspects of OST programs are most needed and effective for PA promotion.[8] OST programs that incorporate evidence-based approaches have positive impacts on attitudes about self-efficacy and self-esteem, prosocial behaviors, and school performance, whereas programs that do not feature these evidence-based elements are generally unsuccessful in improving any outcome (see the meta-analysis conducted by Durlak et al.).[9] Evidence-based PA curricula have resulted in significant improvements in fitness and sports skills (and, as a side effect, academic achievement) when implemented during school hours.[10] However, studies of evidence-based PA programs during afterschool hours have shown mixed results.[11]

Implementation fidelity in out-of-school time programs

One possible explanation for why some OST programs work and others do not might be variability in implementation fidelity.[12] Translating efficacious interventions into real-world settings and maintaining them are complicated, requiring a lasting commitment in terms of time, effort, and funding. Gauging the impact of these programs requires an understanding of "what" and "how well" the intervention was delivered. Overall, findings indicate that, in multiple research areas and target populations, higher levels of implementation are often associated with better outcomes.[13]

Unfortunately, evaluation of implementation quality is currently absent in the majority of PA instruction research in both schools and afterschool programs, although limited data do suggest improvements over "usual practice."[14] Nigg and colleagues recently evaluated the dissemination and sustainability of a PA and nutrition curriculum for children in state-run afterschool programs in Hawaii over a four-year period.[15] The program included an evidence-based PA curriculum called Sports, Play, and Active Recreation for Kids Active Recreation (SPARK AR) and a nutrition component. Implementation was defined as the correspondence between program execution in the afterschool program and research protocol, using the SPARK Lesson Quality Checklist published in the SPARK manual. Results indicated that proper program implementation rose over the four years, from 67 percent in year 1 to over 80 percent in years 3 and 4. Although there was no comparison group, the study suggests that higher levels of implementation, with ongoing training and support, are associated with increases in PA levels.

Local out-of-school policies and physical activity

To help children become more active, many national, state, and local organizations have developed policies to promote PA within the

OST setting.[16] Policy characteristics range from having a written policy in place, to ensuring that staff receive a sufficient amount of training for promoting PA, to monitoring and evaluating the PA and fitness of the children who attend OST programs.[17] In Miami-Dade County, FL, such policy determinations are often driven by one of the largest local funders of OST programs, The Children's Trust (Trust). The Trust supports over two hundred afterschool sites that serve more than 20,000 school-aged children, predominantly in vulnerable, at-risk communities. Pilot data obtained by our team revealed that 78 percent of these OST participants (80 percent of whom were African American or Hispanic) did not reach recommended age- and gender-specific fitness standards.[18] Over the past five years, the Trust has supported PA in OST programs by investing in activity-focused, evidence-based curricula (EBC) such as SPARK, mandating thirty minutes of daily PA in all Trust-funded OST programs, and incorporating a measure of cardiovascular fitness into their participant outcome assessments.[19]

As noted above, little is known about how the adoption of evidence-based PA curricula affects the level of PA during OST programs, creating a situation where funders and OST providers are investing resources in programs with unknown impact on PA levels.[20] The literature on OST health and PA promotion is limited and focuses primarily on small-scale initiatives implemented in highly controlled, research-driven settings.[21] Moreover, there are no clear guidelines concerning what constitutes reasonable expected gains for the kinds of PA interventions currently offered in these OST settings. Thus, the present study sought to obtain information about children's fitness and cardiovascular health when exposed to an evidence-based PA curriculum and compare these results to those found in non-curriculum-based programs. The primary study hypothesis was that exposure to evidence-based PA instruction (that is, SPARK), when implemented with fidelity, would result in higher PA levels and improved health outcomes than those found in non-evidence-based PA instruction.

Method

The present study sought to evaluate children's level of PA and fitness during participation in OST programs using a naturalistic, quasi-experimental design.

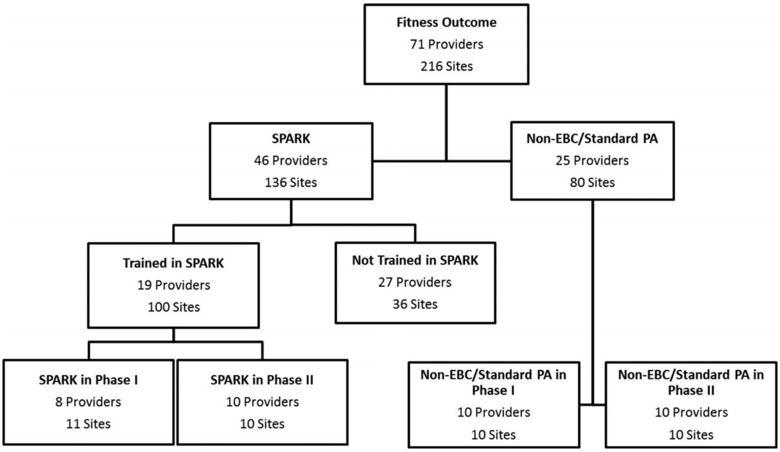
Study design

The study evaluated the impact of an evidence-based curriculum on PA and obesity-related health outcomes through a three-wave, quasi-experimental observation design in which OST programs were considered either intervention (evidence-based PA instruction) or comparison (non-EBC/standard PA) programs. The study was conducted in two phases to accommodate the evaluation of a larger number of sites and due to delayed approval for the study to be conducted at the OST programs located within the local public school system. Phase I occurred during the spring (February through June) and Phase II occurred during the fall (August through December), with each phase lasting four and one-half months. Individual- and site/program-level data were collected in three waves: at the beginning/baseline, middle, and end of each phase. At each wave, individual-level (that is, child's height, weight, body composition, and fitness level) and site/program-level (that is, direct observation of PA level and implementation quality of PA instruction) data were collected. The study was approved by the Nova Southeastern University Institutional Review Board and the Miami-Dade County Public School Research Review Committee, and participants provided consent or parental consent in the case of the students.

Site selection

Figure 6.1 outlines the site selection process. At the initiation of the project, the Trust funded 120 provider agencies, each of which offered an OST program at one or more sites (locations; range = 1–27 sites per provider). Participating sites were selected from the pool of OST provider agencies funded by the Trust. All 120 OST providers were required to include thirty minutes of PA daily,

Figure 6.1. Site selection process



although only a subset of provider agencies ($N = 71$; 216 sites) were eligible for study participation because they were contractually required to measure and report physical fitness as a participant outcome.

Of seventy-one eligible provider agencies, forty-six providers had a contractual agreement with the Trust to implement the evidence-based SPARK curriculum. However, only nineteen providers (41.3 percent) had received any type of SPARK training and were thus qualified to be part of the site selection process. Another twenty-five provider agencies offered non-EBC/standard PA instruction. Sites were recruited from a random sample of eligible provider agencies that had adopted either SPARK ($N = 21$ sites) or non-EBC/standard PA instruction ($N = 20$ sites) for their PA component.

Participants

A total of 659 children (315 boys and 344 girls) between the ages of 6 and 17 were recruited for participation in the study. A summary of the demographic characteristics of the sample is provided in Table 6.1. Participants in the SPARK and non-EBC/standard PA instruction programs differed in terms of age, ethnicity, yearly household income, and program attendance. Specifically,

Table 6.1. Baseline participant characteristics as a percentage or mean (standard deviations in parentheses) of the SPARK and non-EBC/standard physical activity groups

Characteristic	*SPARK (n = 320)*	*Standard PA (n = 339)*	*Overall (N = 659)*
Child age	8.73 (1.95)$_a$*	9.09 (2.12)$_b$	8.91 (2.04)
Child gender			
Male	48.7	46.2	47.4
Female	51.3	53.8	52.6
Child race/ethnicity			
Hispanic	63.0$_a$**	46.9$_b$	54.7
Non-Hispanic White	1.7	3.1	2.4
Non-Hispanic Black	35.0	48.1	41.8
Other	0.3	1.8	0.1
Child education			
Elementary	88.4	87.0	87.7
Middle school	11.3	12.0	11.7
High school	0.0	0.9	0.5
Income*			
Less than $25,000	47.9	38.6	43.2
$25,000 to $50,000	29.8	27.9	29.1
> $50,000	9.2	19.3	14.3
Body composition			
Baseline BMI	19.69 (3.81)	20.23 (4.62)	19.97 (4.25)
BMI > 85th percentile	50.6	47.0	48.8
% body fat (BIA)	23.75 (8.51)	24.49 (9.45)	24.14 (9.00)
Fitness			
Baseline PACER laps	13.65 (6.81)	14.51 (9.60)	14.10 (8.39)
% of children in the HFZ at baseline	43.5	49.1	46.6
OST program attendance			
Days of attendance (per week)	4.88 (0.47)	4.85 (0.54)	4.87 (0.51)
Participation in program for > 1 year	55.5$_a$	63.2$_b$*	59.5
Extracurricular activities			
Participation in organized sports/other PA	22.9	28.7	25.9
PA opportunities during free time (not OST)	82.2	81.2	81.7
Frequency of free time PA (not OST) in days	3.76 (2.18)	3.91 (2.04)	3.84 (2.11)

Note: Non-EBC = non-evidence-based curriculum; BMI = body mass index; BIA = bioelectrical impedance analysis; PACER = Progressive Aerobic Cardiovascular Endurance Run; HFZ = healthy fitness zone based on criterion-based norms for children ≥10 years old; OST = out-of-school time; PA = physical activity.

*$p < .05$; **$p < .01$. Percentages or means with differing subscripts within rows are significantly different based on one-way ANOVAs for continuous variables (child age) and chi-square analyses for categorical variables (child ethnicity, OST program attendance, income). Although not conveyed in the table, a higher proportion of children in the SPARK programs lived in homes with an estimated yearly income of under $35,000 (65.1 percent and 51.5 percent for SPARK and non-EBC/standard PA instruction, respectively).

participants in the SPARK programs were younger than participants in the non-EBC/standard PA instruction programs [$F(1,657) = 5.30$, $p = .02$] and had attended these programs for a shorter period of time [$\chi^2(1) = 4.05, p = .04$]. In addition, there were more Hispanic children in the SPARK programs than in the non-EBC/standard PA programs [$\chi^2(1) = 16.73, p = .00$.], and a higher proportion of children in the SPARK programs lived in homes with an estimated yearly income of under $35,000 (65.1 percent and 51.5 percent for SPARK and non-EBC/standard PA instruction, respectively). There were no differences in the number of children enrolled in extracurricular activities outside of the OST program or in PA opportunities outside of the OST program across the two types of PA instruction.

Measures

Body mass index (BMI). Children's height and weight were measured using standard procedures.[22] Heavy clothes, shoes, and socks were removed before weighing and measuring in a private setting. Height was measured to the nearest 0.1 cm using a portable stadiometer (Seca 214 Portable Stadiometer), and weight was measured to the nearest 0.1 kg with a high-precision electronic digital scale (BC-533; Tanita, Tokyo, Japan). Body mass index (BMI) was calculated as weight in kilograms divided by height in meters, squared (kg/m^2). Ages were calculated from the birthdates of the children and the date on which the measurements were made. The gender of each child was also recorded. Using standard CDC growth charts, each child's BMI percentile (and associated BMI *z*-score) for age and sex was calculated.[23] Consistent with the CDC growth charts, children were classified as overweight if their BMI was between the 85th and the 94th percentiles and obese if their BMI was at or above the 95th percentile.

Body composition. Percentage of body fat was measured using bioelectrical impedance analysis (BIA) with a high-accuracy body composition monitor (embedded in the electronic digital scale; BC-533; Tanita, Tokyo, Japan) that uses the resistance and reactance of electrical flow through the body to estimate percentage of body fat.

Physical fitness. Fitness was measured by the Progressive Aerobic Cardiovascular Endurance Run (PACER).[24] The PACER, sometimes called the "Beep Test," is a multistage shuttle run that progressively increases in difficulty by decreasing the time allotted to cover a 20 m distance. It is a valid measure of endurance as well as an acceptable predictor of aerobic capacity (maximal rate of oxygen consumption; VO_2 max) in children and adolescents.[25] During the PACER, students run back and forth between parallel lines placed 20 m apart, at a specified pace, which increases each minute. By increasing the energy demand each minute, the PACER offers a close approximation of the graded exercise test of VO_2 max (that is, aerobic fitness). The number of completed laps is the student's score in the PACER. Generally, higher scores indicate a higher level of fitness. Criterion-based standards are published by the Cooper Institute as part of FITNESSGRAM for children ages ten years and above.[26] Age- and gender-specific norms determine how PACER performance relates to classification in the "healthy fitness zone" (HFZ). The HFZ indicates the level of fitness associated with a low risk for future health problems.

Physical activity. Activity data were obtained through direct observation using the System for Observing Fitness Instruction Time (SOFIT).[27] The SOFIT provides simultaneous recordings of individual PA levels, the lesson context in which they occurred, and teacher behavior. SOFIT uses codes to estimate energy expenditure associated with PA, and detailed procedures for using the system are published elsewhere.[28] Lesson context refers to how lesson content is being allocated at the observation moment and includes options for class management, knowledge, physical fitness, skill development, game play, and free play. Teacher behavior classifies teacher interaction during the observation moment in terms of verbal encouragement of PA, motor skills, or fitness, and is coded as in-class promotion of PA, out-of-class promotion of PA (for example, verbal suggestions of PA at home), or neither. Briefly, the

PA levels of four randomly selected students, the lesson context, and teacher behavior were coded every twenty seconds throughout entire lessons, resulting in twelve observation intervals for each student. The five PA codes (lying, sitting, standing, walking, and vigorous) have been calibrated using heart rate monitoring and validated using pedometry and accelerometry.[29] Intervals in which either walking or vigorous activity was coded were summed to indicate MVPA. In addition, the total number of children actively participating in class was recorded.

Observation schedule. The quantity and quality of PA were observed on three scheduled days (waves) at each site (total of 122 observation days) during the two phases of the study. Observation days occurred at the beginning, middle, and end of each phase, and each site was observed at these times during the phase in which they were studied (twenty-one sites were observed during Phase I and twenty sites were observed during Phase II).

Observer training, assessment, and recalibration. Two full-time staff members and eight doctoral-level graduate students conducted all observations. Initial training included classroom lectures, videotape/lab assessment, and field practice. During training, the observers became certified by reaching an interobserver agreement (IOA) criterion of 80 percent on all variables on precoded "gold-standard" videotaped lessons.

Reliability assessment. Field-based interobserver reliabilities were conducted throughout the study during 9 of the 122 observations (7.4 percent). Equipped with a y-adapter and two earphone jacks, one senior certified staff assessor and one graduate-level certified assessor independently recorded the same students while being paced by the same Mp3 player. Percent of IOA was calculated overall and for each variable. Over the course of the study, the mean interobserver agreement was above the 80 percent recommendation: 91.01 percent for the overall measure, 86.40 percent for student activity, 95.78 percent for lesson context, and 90.86 percent for teacher behavior. The intraclass correlation for independent observers was 0.96 for MVPA minutes.

SPARK implementation fidelity/quality of PA instruction. Adherence to the SPARK curriculum and overall quality of PA instruction in the comparison condition were assessed using a modified version of the SPARK Lesson Quality Checklist provided in the SPARK manual (Lesson Quality Checklist-Revised [LQC-R], available upon request).[30] The original 21-item dichotomous (Yes/No) scale provided in the SPARK curriculum binder was designed to serve as a self-, peer-, or evaluative assessment of adherence to the SPARK instructional methodology. The checklist was modified for this study by adding content (for example, minimum lesson length, child enjoyment, and PA promotion outside of class) and changing the scoring scheme to a Likert-scale format for substantive items. The resulting measure consists of twenty-three items and three subscales (management, design, and instruction) intended to assess instructional effectiveness. Scores can range from 23 to 92 for the overall scale, from 11 to 44 for the 11-item design scale, from 3 to 12 for the 3-item management scale, and from 9 to 36 on the 9-item instruction scale. For the purposes of this study, only the total score was utilized in our analysis, though means were computed for the subscales. Coefficient alpha for the full scale in the current sample was 0.80. Importantly, although the original measure was intended to assess SPARK implementation fidelity, the modified measure focuses on instructional practices found in high-quality fitness programs without being specifically limited to SPARK. Modifying the measure in this way permitted an assessment of non-SPARK program adherence to “best practices.” Scores on the LQC-R are presented in two ways: as raw mean scores or as a percentage of total quality points available (that is, 92), which is what we refer to as “instructional competence” (see Table 6.3 for details).

Statistical analyses. In general, results from three classes of analyses are presented below. First, at the participant level, mixed-model MANCOVAs were used to examine change over time in continuous outcomes (that is, BMI, BIA, PACER laps) as a function of instruction type (that is, SPARK versus non-EBC/standard PA) and assessment wave, with child age, gender, and ethnicity included as covariates. A three-way mixed-model MANOVA with two within-subjects factors (wave with three levels, observation interval with twelve levels) and one between-subjects factor (SPARK versus non-EBC/standard PA instruction group) was used to examine group differences and change over time in PA (MVPA as observed with the SOFIT) at the site level. In such analyses, interactions between instruction type and assessment wave indicate that trends over time on the outcome vary significantly by instruction type. Main effects of time (in the absence of such interactions) suggest significant change over time that does not vary by instruction type. In addition, several group contrasts (for example, SPARK versus non-EBC/standard PA instruction) were analyzed using *t*-test or ANOVAs, which are equivalent when there is a single numerator degree of freedom. Finally, when outcomes were categorical (for example, classification into a certain activity class), the results of chi-square tests are reported.

Many of the longitudinal analyses discussed below were conducted using MANOVA models, which model appropriately the repeated observations nested within individual children (for example, change in PACER scores over the three assessment waves). It is worth noting that an additional level of nesting occurs in our data in that children are nested within various programs. Although this higher-order nesting was not explicitly modeled in the MANOVA framework, we repeated primary analyses (that is, models based on BMI, BAI, and PACER data) by specifying three-level random effect regression models using the mixed procedure in SAS (version 9.4).[31] Primary results—regarding the presence/absence of SPARK effects and general trends over time—were consistent across both modeling frameworks, so the results from the MANOVA models are presented below.

Exploratory analyses. For the purposes of the present study, SPARK programs were initially defined contractually—that is, the provider agency had a contractual agreement with the Trust to implement the evidence-based SPARK curriculum. However, SPARK programs can also be defined by the level of training in the SPARK curriculum or by the quality of PA instruction itself, as measured by the LQC-R. Therefore, two subsequent classifications were derived for exploratory purposes.

The first alternative classification used level of training (one aspect of fidelity) to operationally define SPARK sites as belonging to one of two groups: SPARK-DT (direct training) or SPARK-IT (indirect training). Under this scheme, SPARK-DT sites were so deemed only if they reported having had SPARK-trained instructors deliver the program on at least two of the three PA sessions observed during the study. To be SPARK-trained, OST staff attended formal SPARK workshops led by certified SPARK trainers. The one- to two-day six-hour workshops focus on SPARK philosophy and implementation. SPARK-IT sites were defined as those with staff who had been trained during a brief one-hour conference symposium sponsored by the Trust and/or by a SPARK-trained colleague instead of the SPARK-certified trainers. This classification defined eight SPARK-DT sites and twelve SPARK-IT sites. Mixed-model MANCOVAs were used to examine change over time in continuous outcomes as a function of training level (that is, SPARK-DT versus SPARK-IT versus non-EBC/standard PA) and assessment wave, with child age, gender, and ethnicity included as covariates. A three-way mixed-model MANOVA with two within-subjects factors and one between-subjects factor (training level; SPARK-DT versus SPARK-IT versus non-EBC/standard PA) was used to examine group differences and change over time in PA. Chi-square tests were used when outcomes were categorical.

A second classification of SPARK sites involved a process-based examination of the quality of PA instruction using the LQC-R. For this analysis, the sites' contractually defined PA instruction classification was disregarded, and all sites were classified according to their standing in the Quality of Physical Activity Instruction subscale continuum of the LQC-R. Using this classification for all forty-one sites in the study, twelve sites scoring at or above the 75th percentile for overall quality of instruction were compared to nine sites scoring at or below the 25th percentile. Mixed-model MANCOVAs were used to examine change over time in continuous outcomes as a function of quality of instruction (that is, 75th percentile versus 25th percentile) and assessment wave, with child age, gender, and ethnicity included as covariates. A three-way mixed-model MANOVA with two within-subjects factors (wave with three levels, interval with twelve levels) and one between-subjects factor (quality of instruction; 75th percentile versus 25th percentile group) was used to examine group differences and change over time in PA. Categorical outcomes were assessed using chi-square tests.

Results

Results from the participant-level, program-level, and exploratory analyses are presented below.

Body mass index. There were no significant differences in BMI between children in SPARK and non-EBC/standard PA instruction at baseline, $F(1,652) = 2.69, p = .10$. At baseline, 48.8 percent (50.6 percent SPARK, 47.0 percent non-EBC/standard PA) of the study participants were at or above the 85th percentile for BMI, indicating that they were classified as either overweight or obese (see Table 6.1). At Time 3, 45.1 percent (47.0 percent SPARK, 43.3 percent non-EBC/standard PA) were at or above the 85th percentile for BMI. Change in BMI over time was not significant [$F(2,449) = 2.22, p = .11$] for the overall sample or by the type of PA instruction [$F(2,449) = 0.11, p = .89$] when controlling for child age, gender, and ethnicity.

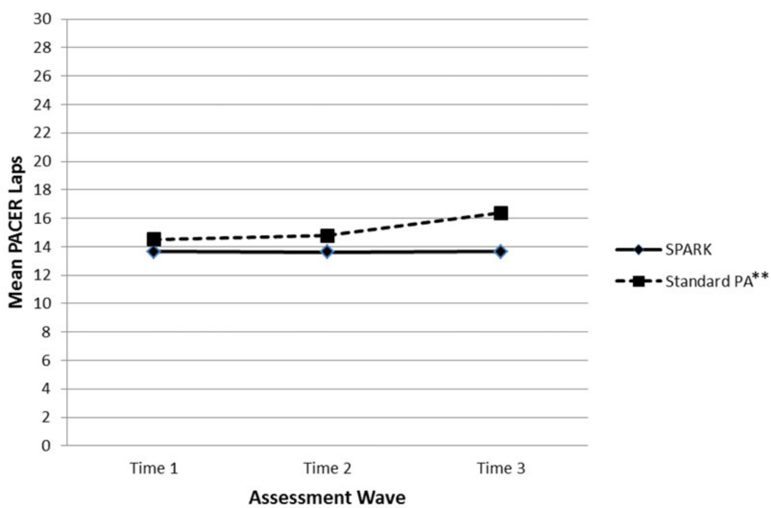
Body composition. Similar to BMI, there were no significant differences in BIA between the two groups at baseline, $F(1,648) = 1.11, p = .29$. When controlling for child age, gender, and ethnicity, BIA varied as function of the type of PA instruction over time, $F(2,437) = 3.20, p = .04$. Although body fat percentage increased between wave 1 and wave 3 for both groups, BIA increased slightly more for SPARK participants (2.09 percent) than for non-EBC/standard PA participants (1.51 percent). Although statistically significant, this difference of 0.58 percent is small and of modest clinical significance.

Physical fitness. Children's fitness levels were assessed with the PACER across the three waves. At baseline, the number of PACER laps run was similar across both types of PA instruction, $F(1,566) =$

1.57, $p = .22$. However, children in the non-EBC/standard PA condition achieved higher PACER scores at Time 3, $F(1,498) = 10.94$, $p < .01$. Given the sensitivity of PACER scores to age and gender as well as the unequal distribution of Hispanic children across the two types of programs, changes in fitness levels over time were assessed by comparing children's performance across the three assessment waves while controlling for child gender, age, and ethnicity. Figure 6.2 displays the means for PACER scores across the three waves according to each type of PA instruction. PACER performance improved over time for all participants, $F(2,364) = 3.53$, $p = .01$. Children in the non-EBC/standard PA condition improved more than those in the SPARK condition, $F(2,364) = 4.72$, $p = .01$.

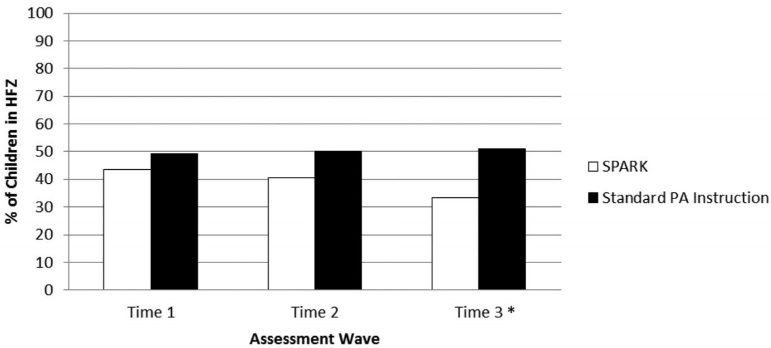
The percentage of children classified as being in the HFZ was also identified for each assessment wave. As mentioned previously,

Figure 6.2. Mean number of laps completed on the Progressive Aerobic Cardiovascular Endurance Run by SPARK and non-EBC/standard physical activity instruction groups across three assessment waves



Note: PACER = Progressive Aerobic Cardiovascular Endurance Run; PA = physical activity. All participants ran more laps over time, and participants in the Standard PA condition improved more than those in the SPARK condition, based on a mixed-model MANCOVA with child age, gender, and ethnicity included as covariates.
**$p < .01$.

Figure 6.3. Percent of children in the Healthy Fitness Zone in the SPARK and non-EBC/standard physical activity instruction groups across three assessment waves



Note: HFZ = healthy fitness zone; PA = physical activity. There was no difference in HFZ classification at Time 1 or Time 2. Programs using non-EBC/standard PA instruction had a higher number of children placing within the HFZ at Time 3 than did the SPARK group, based on chi-square analysis.
$^{*}p < .05$.

age- and gender-specific norms determine how PACER performance relates to classification in the HFZ. The HFZ indicates the level of fitness associated with a low risk for future health problems if this level of fitness is maintained. Figure 6.3 illustrates the distribution of children in the HFZ across both types of PA instruction. Although there was no difference in HFZ classification at baseline, programs using non-EBC/standard PA instruction had a higher number of children scoring within the HFZ at the third assessment wave [$\chi^2(1) = 5.43, p = .02$] than did the SPARK group.

Physical activity. Level of PA was assessed at each of the three waves for each participating site. A total of 122 observations were conducted on a total of 488 children (52.6 percent male and 47.4 percent female), and the percentage of intervals spent in each of the SOFIT categories (that is, PA level, lesson context, and teacher interaction) was computed at each wave and overall. Lesson length did not differ according to the type of PA instruction [$F(1,486) = 1.89, p = .17$], averaging thirty-one minutes for both conditions.

Table 6.2 presents the percentage of intervals children spent at various activity levels, in different lesson contexts, and

Table 6.2. Proportion of lesson time for participant activity, lesson context, and teacher behavior variables during SPARK and non-EBC/standard physical activity instruction averaged across three assessment waves

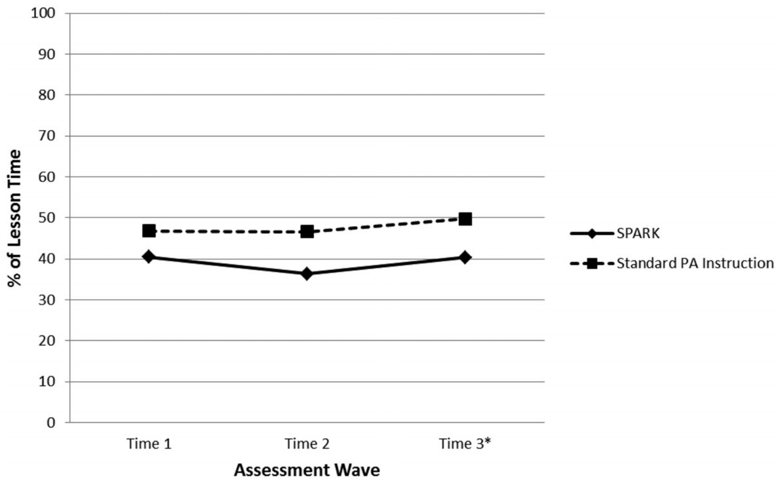
Category	*SPARK*	*Non-EBC/standard PA*
Activity intensity		
Lying down (%)	0.30	0.39
Sitting (%)	7.08	9.64
Standing (%)	53.57_a^{**}	41.49_b
Walking/moderate PA (%)	23.94	29.45
Running/vigorous PA (%)	15.11	19.03
MVPA (%)	39.05_a	48.48_b^{*}
Lesson context		
Management (%)	27.01	21.75
Knowledge (%)	4.93	3.39
Fitness (%)	7.69	10.24
Skill building (%)	8.20	13.38
Game (%)	49.47	38.91
Other (%)	2.74	12.32
Teacher behavior		
Promotes in-class MVPA (%)	29.60	30.87
Promotes out-of-class MVPA (%)	0.13	0.07
No promotion of MVPA (%)	70.26	69.06

Note: PA = physical activity; MVPA = moderate-to-vigorous physical activity (walking/moderate PA + running/vigorous PA).
$^{*}p < .05$; $^{**}p < .01$. Percentages with differing subscripts within rows are significantly different based on a three-way mixed-model MANOVA with two within-subjects factors (wave with three levels, interval with twelve levels) and one between-subjects factor (SPARK versus non-EBC/standard physical activity instruction group).

accompanied by different teacher behaviors averaged across the three assessment waves. SPARK and non-EBC/standard PA instruction sites did not differ in the amount of time spent within the various lesson contexts or in the instructors' promotion of in-class and out-of-class MVPA.

In terms of PA, Figure 6.4 indicates the percentage of intervals in which MVPA was observed at each wave. Children in the non-EBC/standard PA programs engaged in more MVPA at wave 3 and overall [$t(39) = 2.39, p = .02, d = .75$ and $t(39) = 2.61, p = .01, d = .82$, respectively]. Children in the SPARK programs spent more time "standing" when compared to those in the non-EBC/standard PA programs, $F(1,39) = 10.88, p < .01$.

Figure 6.4. Proportion of lesson time spent in moderate-to-vigorous physical activity during SPARK and non-EBC/standard physical activity instruction across three assessment waves



Note: PA = physical activity. Children in the non-EBC/standard PA instruction programs engaged in more MVPA at wave 3 and overall based on comparison *t*-tests.
$^{*}p < .05$.

SPARK implementation fidelity/Quality of PA instruction. Table 6.3 provides descriptive statistics for the LQC-R. Total quality of instruction scale scores ranged from 50.67 to 73.33 for SPARK sites and from 41.33 to 78.33 for the non-EBC/standard PA instruction sites. Overall lesson quality and quality of instruction (Instruction subscale) diminished over time but not as a function of type of PA instruction [$F(2,37) = 3.27, p = .05$ and $F(2,37) = 4.51, p = .02$, respectively]. Consideration of the group means (averaged across the three time periods) revealed that there was only a modest degree of improvement in "instructional competence" in SPARK [$M = 62.85$ (5.87) or 68.3 percent of the score maximum, 92] versus non-EBC/standard PA programs [$M = 60.25$ (7.36) or 65.4 percent of the score maximum, 92].

Discussion

The present quasi-experimental observation sought to evaluate the impact of evidence-based PA instruction when implemented by

Table 6.3. Program characteristics as a percentage or mean (standard deviations in parentheses) of the SPARK and non-EBC/standard physical activity instruction

Characteristic	*SPARK (n = 21)*	*Non-EBC/standard PA (n = 20)*	*Overall (N = 41)*
Staff qualifications/training			
At least one on-site staff member with PE degree (%)	23.8 (*n* = 5)	25.4 (*n* = 5)	25.0 (*n* = 10)
Instructor with PE degree at three assessment waves (%)	19.0 (*n* = 4)	15.0 (*n* = 3)	17.0 (*n* = 7)
Years of experience in PA instruction for observed staff	6.70 (6.84)	7.09 (7.97)	6.89 (7.32)
Instructor with some type of SPARK training at three assessment waves (%)	52.4 (*n* = 11)	N/A	N/A
Instructor was SPARK/Trust-trained at all three assessments (%)	23.8 (*n* = 5)	N/A	N/A
SPARK implementation			
Length of SPARK program implementation (years)	1.88 (1.39)	N/A	N/A
SPARK implementation (days/week)	3.67 (1.49)	N/A	N/A
Use of SPARK curriculum on a daily basis (%)	57.1 (*n* = 12)	N/A	N/A
Quality of PA instruction (score range)			
Design (11–44)	31.06 (2.81)	30.14 (4.00)	30.61 (3.43)
Management (3–12)	7.96 (1.31)	8.15 (1.71)	8.06 (1.50)
Instruction (9–36)*	23.82 (3.15)	21.95 (3.06)	22.91 (3.21)
Total quality of instruction (23–92)*	62.85 (5.87)	60.25 (7.36)	61.58 (6.69)
Total quality of instruction score range	50.67–73.33	41.33–78.33	
Instructional competence (%)	68.31 (*n* = 21)	65.49 (*n* = 20)	

Note: No differences were found between the two instruction groups. PA = physical activity; Trust = The Children's Trust; PE = physical education; N/A = not applicable.

*$p < .05$. Total quality of instruction scores and instruction subscale scores diminished over time, but not as a function of type of PA instruction, based on mixed-model MANOVA.

OST programs in a "real-world" urban setting. The study was initiated with the intention of informing funder policy decisions regarding PA instruction practices. While it is generally accepted that evidence-based approaches yield higher levels of PA when implemented by researchers under controlled conditions, findings are inconsistent when evidence-based PA instruction is implemented in the field, under presumably less controlled conditions.[32]

Consistent with national data, approximately half of the study participants were overweight or obese and exhibited fitness levels that placed them at risk for obesity-related complications at the study's inception, underscoring the need to promote PA in this community.[33] Children were engaged in MVPA during approximately 39 percent of the observation intervals in SPARK programs compared to 48 percent of the observation intervals in non-EBC/standard PA programs. However, while the non-EBC/standard PA programs here seemed to outperform the SPARK programs, both produced rates of MVPA that are consistent with previous research showing that PA-focused EBCs produce rates of MVPA ranging from almost one-third of the lesson time to almost half of the lesson time.[34] However, because the time allocated for PA sessions was limited in the current study ($M = 31$ minutes), the resulting proportion of that time in MVPA (approximately 12 minutes/day in SPARK programs and 15 minutes/day in non-EBC/standard PA programs) amounts to only 20–25 percent of the recommended daily sixty minutes of MVPA.[35] These findings are interesting on several levels. First, the data suggest that the non-EBC programs were very successful in producing MVPA (as they produced rates of MVPA that are actually consistent with better-performing EBCs). Second, SPARK effectiveness varied greatly in a nonexperimental setting, as discussed below. Third, to help children achieve the recommended sixty minutes of MVPA in the OST setting, changes will be needed in both the time allocated for fitness activity and the nature of those activities.

Although the OST programs provided opportunities for children to engage in PA, most of that time was spent in light-intensity activities such as standing rather than the MVPA recommended by numerous authorities on children's health and fitness, especially when a PA curriculum was used.[36] Our findings suggested that children at the non-EBC/standard PA sites spent more time engaged in MVPA at the third wave and overall than those who were at sites using the SPARK curriculum. Fitness levels seemed to follow these differences in MVPA, with children in the non-EBC/standard PA groups achieving higher levels of fitness than those in the SPARK group. It is interesting to note that the actual percentage of children in the HFZ, that is, the level of fitness associated with a low risk of future health problems, remained fairly constant for children at the non-EBC/standard PA sites while the proportion of children considered "fit" at the SPARK sites decreased over time (see Figure 6.3). To be classified as being in the HFZ, children must run a certain number of PACER laps, a number that increases with age (according to gender). Slight increases in PACER laps run will maintain HFZ status whereas flat trajectories of laps run (as seen in Figure 6.2) will leave more and more children outside of the HFZ.

Findings from the present study should be interpreted with caution and in the context of the "real-world" implementation of an EBC such as SPARK. Few PA instructors had a physical education degree and less than 25 percent had received formal training in the SPARK curriculum. In addition, results indicated that the SPARK curriculum was implemented with relatively low levels of fidelity (see Table 6.3). Indeed, SPARK programs achieved only a slightly higher level of instructional competence (68 percent of total quality points) than non-SPARK programs (65 percent). According to the executive director of SPARK (Paul Rosengard, personal communication, May 21, 2014), SPARK trainers and administrative staff consider implementation fidelity of less than 80 percent to be low and SPARK-certified programs would be expected to achieve levels of instructional competence even higher for best results. Although the SPARK programs in the current study had sufficient equipment for SPARK activities, they had only minimal training and none requested follow-up support (to our or the funder's knowledge).

As noted previously, only recently has any study included a measure of fidelity when evaluating the effectiveness of PA instruction in OST programs.[37] In their four-year SPARK dissemination study, Nigg and colleagues found a 67 percent rate of implementation fidelity in the first year of their study, prior to any consistent implementation support, a level that is strikingly similar to our estimate.[38] Over the course of their four-year study, they increased sustainability efforts to include PA equipment for all programs (updated when needed), booster sessions/mini-trainings each semester, time management planners, and frequent communication with programs as part of a quality improvement feedback loop. The authors also identified a variety of intangible factors that they believed were essential for improved implementation, such as building local capacity for SPARK training, developing strategies to minimize staff effort, and including strong provider partners as champions to the effort. Even with these rather substantial efforts to improve implementation fidelity, it took nearly three years for the project to achieve and sustain implementation fidelity rates of 80 percent. In the last two years of their study, higher levels of implementation fidelity resulting from ongoing training and support were associated with increases in PA levels.

Taking a broader view of training, research shows that the traditional approach to implementation, consisting of one-time in-service or orientation training followed by informal monitoring of staff progress, rarely produces lasting improvements in staff skill, implementation fidelity, or program quality.[39] By contrast, training approaches that incorporate coaching, where group sessions are replaced or supplemented with ongoing individual modeling, have received support.[40] Improvements in program quality are achieved when continuous, ongoing efforts at quality improvement are employed.[41] In the present study, we found that SPARK implementation in the OST setting lacked many of the elements that might have contributed to its effectiveness in other settings, such as the necessary level of training intensity and support typically found in controlled research settings and as recommended by SPARK program authors.[42]

One possible limitation of the current study was the quasi-experimental design: providers were not randomly assigned to SPARK versus non-EBC/standard PA instruction conditions. Instead, because this was a naturalistic study, providers "self-selected" into these two conditions. All were contractually required to offer thirty minutes of daily PA, but a subset offered non-EBC/standard PA instruction rather than SPARK (see Figure 6.1). Table 6.1 provides pretest data for youth as a function of instruction type. These data suggest that differences between instruction types were small in magnitude, especially for the primary physical fitness outcomes. In other words, these data suggest that large systematic differences between groups were not detected (at least on these measured characteristics). Nonetheless, analyses of the primary participant fitness outcomes included statistical control for age, gender, and ethnicity, given their association with the outcomes and instruction type. A second limitation concerns the possible effects of the seasons (fall, spring) when the two phases occurred. To identify possible seasonal differences, data were first compared according to phase and no differences were found. Therefore, the assessment waves were combined across the two phases.

In any study of this sort, what one can say about the effectiveness of an intervention is always constrained by concerns about the level of implementation fidelity. Low implementation fidelity reduces the likelihood of finding intervention effects and leaves investigators and public policy advocates wondering about the best way to allocate scarce resources.[43] In the context of the present study, without strong implementation fidelity it is hard to assess whether the resources expended to facilitate dissemination of evidence-based programs such as SPARK (and other EBC fitness-related programs) are worthwhile or cost effective (see Yates, 1994).[44] Studies on strategies to improve implementation quality of PA-focused curricula recommend continuous and consistent support, simple focused lessons with readily available alternatives, and individual assistance/modeling.[45] This ongoing support costs more than the traditional one-time training approach. Funders and OST providers are at the point where they must decide if they are willing to invest more in resources for continuous quality support.[46]

Regardless of implementation fidelity, it is important to note that the results reveal that these OST programs enabled participants to achieve MVPA at levels close to the recommended 50 percent of instruction time despite reaching only 20–25 percent of the daily recommended MVPA.[47] In fact, the Trust's requirement for OST programs to include daily PA programming, EBC or not, might be responsible for the observation that fitness levels, as measured by PACER performance, improved over time for all participants in the current study. However, these findings also reveal "room for improvement" when it comes to providing PA instruction in these settings and set the stage for research that might examine efforts to improve dissemination and implementation practices, such as enhancing support and training, and generally exploring how public funders of afterschool programs can ensure the greatest cost benefit to taxpayers and participants in funded programs. Future studies might seek to examine the qualitative factors that negatively affect curriculum implementation fidelity, with the goal of establishing community partnerships that could foster improved PA instruction, as recent studies have attempted with some success.[48] Most importantly, future studies might also seek to evaluate the impact of EBCs and instructional practices in the context of an intervention study where level of staff training and level of support are varied to determine the effects on implementation fidelity and fitness levels over a longer period of time.

Notes

1. U.S. Department of Health and Human Services. (2008). *Physical activity guidelines for Americans, 2008*. Washington, DC: U.S. Department of Health and Human Services, CDC, National Center for Chronic Disease Prevention and Health Promotion.

2. Pate, R. R., Freedson, P. S., Sallis, J. F., Taylor, W. C., Sirard, J., Trost, S. G., & Dowda, M. (2002). Compliance with physical activity guidelines: Prevalence in a population of children and youth. *Annals of Epidemiology*, *12*(5), 303–308. doi:10.1016/S1047-2797(01)00263-0; Ogden, C. L., Carroll, M. D., Kit, B. K., & Flegal, K. M. (2012). Prevalence of obesity and trends in body mass

index among US children and adolescents, 1999–2010. *JAMA: The Journal of the American Medical Association*, *307*(5), 483–490. doi:10.1001/jama.2012.40; Tudor-Locke, C., Johnson, W. D., & Katzmarzyk, P. T. (2010). Accelerometer-determined steps per day in US children and youth. *Medicine & Science in Sports & Exercise*, *42*(12), 2244–2250. doi:10.1249/MSS.0b013e3181e32d7f

3. Tudor-Locke, C., Lee, S. M., Morgan, C. F., Beighle, A., & Pangrazi, R. P. (2006). Children's pedometer-determined physical activity during the segmented school day. *Medicine & Science in Sports & Exercise*, *38*(10), 1732–1738. doi:10.1249/01.mss.0000230212.55119.98

4. Afterschool Alliance. (2009). *America after 3 pm: Key findings.* Retrieved from http://www.afterschoolalliance.org/documents/AA3PM_Key_Findings_2009.pdf.

5. Arundell, L., Ridgers, N. D., Veitch, J., Salmon, J., Hinkley, T., & Timperio, A. (2013). 5-year changes in afterschool physical activity and sedentary behavior. *American Journal of Preventive Medicine*, *44*(6), 605–611. doi:10.1016/j.amepre.2013.01.029; Beets, M. W., Huberty, J., Beighle, A., & Healthy Afterschool Program Network. (2012). Physical activity of children attending afterschool programs: Research- and practice-based implications. *American Journal of Preventive Medicine*, *42*(2), 180–184. doi:10.1016/jamapre.2011.10.007

6. Afterschool Alliance. (2009).

7. Trost, S. G., Rosenkranz, R. R., & Dzewaltowski, D. (2008). Physical activity levels among children attending after-school programs. *Medicine & Science in Sports & Exercise*, *40*(4), 622–629. doi:10.1249/MSS.0b013e318161eaa5; U.S. Department of Health and Human Services. (2008); Strong, W. B., Malina, R. M., Blimkie, C. J., Daniels, S. R., Dishman, R. K., Gutin, B., . . . Trudeau, F. (2005). Evidence based physical activity for school-age youth. *The Journal of Pediatrics*, *146*(6), 732–737. doi:10.1016/j.jpeds.2005.01.055

8. Strong et al. (2005); Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, *45*(3–4), 294–309. doi:10.1007/s10464-010-9300-6

9. Durlak et al. (2010).

10. McKenzie, T. L., Nader, P. R., Strikmiller, P. K., Yang, M., Stone, E. J., Perry, C. L., . . . Kelder, S. H. (1996). School physical education: Effect of the Child and Adolescent Trial for Cardiovascular Health. *Preventive Medicine*, *25*(4), 423–431. doi:10.1006/pmed.1996.0074; McKenzie T. L., Sallis, J. F., Prochaska, J. J., Conway, T. L., Marshall, S., & Rosengard, P. (2004). Evaluation of a two-year middle-school physical education intervention: M-SPAN. *Medicine & Science in Sports & Exercise*, *36*, 1382–1388. doi:10.1249/01.MSS.0000135792.20358.4D; McKenzie T. L., Sallis J. F., & Rosengard, P. (2009). Beyond the stucco tower: Design, development, and dissemination of the SPARK physical education programs. *American Academy of Kinesiology and Physical Education*, *61*(1), 114–127. doi:10.1080/00336297.2009.10483606; Sallis, J. F., McKenzie, T. L., Alcaraz, J. E., Kolody, B., Faucette, N., Hovell, M. F. (1997). The effects of a 2-year

physical education program (SPARK) on physical activity and fitness in elementary school students. *American Journal of Public Health*, *87*(8), 1328–1334. doi:10.2105/AJPH.87.8.1328

11. Trost et al. (2008); Beets, M. W., Beighler, A., Erwin, H., & Huberty, J. (2009). After-school program impact on physical activity and fitness: A meta-analysis. *American Journal of Preventive Medicine*, *36*(6), 527–537. doi:10.1016/j.amepre.2009.01.033; Dzewaltowski, D. A., Rosenkranz, R. R., Geller, K. S., Coleman, K. J., Welk, G. J., Hastmann, T. J., & Milliken, G. A. (2010). HOP'N after-school project: An obesity prevention randomized controlled trial. *International Journal of Behavioral Nutrition and Physical Activity*, 7(1), 90–101. doi:10.1186/1479-5868-7-90; Kelder, S., Hoelscher, D. M., Barroso, C. S., Walker, J. L., Cribb, P., & Hu, S. (2005). The CATCH kids club: A pilot after-school study for improving elementary students' nutrition and physical activity. *Public Health Nutrition*, *8*(2), 133–140. doi:10.1079/PHN2004678; Sharpe, E. K., Forrester, S., & Mandigo, J. (2011). Engaging community providers to create more active after-school environments: Results from the Ontario CATCH Kids Club Implementation Project. *Journal of Physical Activity and Health*, *8*(Suppl. 1), S26–S31; Slusser, W. M., Zharif, M. Z., Erausquin, J. T., Kinsler, J. J., Collin, D., & Prelip, M. L. (2013). Improving overweight among at-risk minority youth: Results of a pilot intervention in after-school programs. *Journal of Health Care for the Poor and Underserved*, *24*(2A), 12–24. doi:10.1353/hpu.2013.0111; Nigg, C. R., Geller, K., Adams, P., Hamada, M., Hwang, P., & Chung, R. (2012). Successful dissemination of Fun 5: A physical activity and nutrition program. *Translational Behavioral Medicine*, *2*(3), 276–285. doi:10.1007/s13142-012-0120-0

12. Beets et al. (2009); Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45. doi:10.1016/S0272-7358(97)00043-3

13. Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., . . . Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*, *23*(5), 443–451. doi:10.1037/0278-6133.23.5.443; Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3–4), 327–350. doi:10.1007/s10464-008-9165-0

14. Beets et al. (2009); 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*(2), 125–135. doi:10.1080/02701367.2012.10599842; Sallis et al. (2012).

15. Nigg et al. (2012).

16. Wiecha, J. L., Hall, G., Gannett, E., & Roth B. (2012). Development of healthy eating and physical activity quality standards for out-of-school time

programs. *Childhood Obesity*, *8*(6), 572–576. doi:10.1089/chi.2012.0030; Beets, M. W., Wallner, M., & Beighle, A. (2010). Defining standards and policies for promoting physical activity in afterschool programs. *Journal of School Health*, *80*(8), 411–417. doi:10.1111/j.1746-1561.2010.00521.x

17. Wiecha et al. (2012); Beets et al. (2010).

18. Thaw, J. M., & Reitman, D. (2010). *Fitness outcome report: Progressive Aerobic Cardiovascular Endurance Run (PACER) pilot year summary—school year 2008–2010* (Unpublished raw data). Miami, FL: Project RISE.

19. McKenzie et al. (2004); Sallis et al. (1997).

20. Beets, M. W., Huberty, J., Beighle, A., Moore, J. B., Webster, C., Ajja, R., & Weaver, G. (2012). Impact of policy environment characteristics on physical activity and sedentary behaviors of children attending afterschool programs. *Health Education & Behavior*, *40*(3), 296–304. doi:10.1177/1090198112459051

21. Sharpe et al. (2011).

22. Centers for Disease Control and Prevention (CDC). (2010). *National Health and Nutrition Examination Survey (NHANES): Anthropometry procedures manual*. Hyattsville, MD: National Center for Health Statistics.

23. Kuczmarski, R. J., Ogden, C. L., Guo, S. S., Grummer-Strawn, L. M., Flegal, K. M., Mei, Z., . . . Johnson, C. L. (2002). 2000 CDC growth charts for the United States: Methods and development. *Vital and Health Statistics: Data from the National Health Survey*, *11*(246), 1–190.

24. Mahar, M. T., & Rowe, D. A. (2008). Practical guidelines for valid and reliable youth fitness testing. *Measurement in Physical Education and Exercise Science*, *12*(3), 126–145. doi:10.1080/10913670802216106

25. Mahar & Rowe. (2008); Mahar, M. T., Welk, G. J., Rowe, D. A., Crotts, D. J., & McIver, K. L. (2006). Development and validation of a regression model to estimate VO_2 peak from PACER 20-m shuttle run performance. *Journal of Physical Activity and Health*, *3*(Suppl. 2), S34–S46. doi:10.1080/10913670802216106

26. Mahar & Rowe. (2008); Mahar et al. (2006); Meredith, M. D. & Welk, G. J. (Eds.). (2004). *FITNESSGRAM/ACTIVITYGRAM*: *Test administration manual* (3rd ed.). Champaign, IL: Human Kinetics.

27. McKenzie T. L., Sallis J. F., & Nader P. R. (1992). SOFIT: System for observing fitness instruction time. *The Journal of Teaching in Physical Education*, *11*, 195–205.

28. McKenzie et al. (1996); McKenzie et al. (1992).

29. McKenzie et al. (1992); Rowe, P. J., Schuldheisz, J. M., & Van der Mars, H. (1997). Validation of SOFIT for measuring physical activity of first- to eighth-grade students. *Pediatric Exercise Science*, *9*, 136–149; McKenzie, T. L., Sallis, J. F., & Armstrong, C. A. (1994). Association between direct observation and accelerometer measures of children's physical activity during physical education and recess. *Medicine & Science in Sports & Exercise*, *26*(Suppl. 5), S143. doi:10.1249/00005768-199405001-00805

30. McKenzie et al. (2009); Hedeker, D. & Gibbons, R. D. (2006). *Longitudinal data analysis*. New York, NY: Wiley.

31. McKenzie et al. (2009); Hedeker & Gibbons. (2006).

32. Trost et al. (2008); Beets et al. (2009); Dzewaltowski et al. (2010); Kelder et al. (2005); Sharpe et al. (2011).

33. Ogden et al. (2012).

34. Trost et al. (2008); Sharpe et al. (2011).

35. U.S. Department of Health and Human Services. (2008); Strong et al. (2005).

36. Strong et al. (2005).

37. Sallis et al. (2012).

38. Nigg et al. (2012).

39. Sheldon, J., Arbreton, A., Hopkins, L., & Grossman, J. B. (2010). Investing in success: Key strategies for building quality in after-school programs. *American Journal of Community Psychology*, *45*(3–4), 394–404. doi:10.1007/s10464-010-9296-y

40. Sheldon et al. (2010).

41. Nigg et al. (2012); Sheldon et al. (2010).

42. Sharpe et al. (2011); Nigg et al. (2012).

43. Beets et al. (2009).

44. Yates, B. T. (1994). Toward the incorporation of costs, cost-effectiveness analysis, and cost-benefit analysis into clinical research. *Journal of Consulting and Clinical Psychology*, *62*(4), 729–736. doi:10.1037/0022-006X.62.4.729

45. Kelder et al. (2005); Nigg et al. (2012); Sheldon et al. (2010).

46. Beets et al. (2012).

47. National Association for Sport and Physical Education. (2009a). *Appropriate instructional practice guidelines for elementary school physical education* (3rd ed.). Reston, VA: Author. Retrieved from http://www.cahperd.org/cms-assets/documents/ToolKit/NASPE_ApprroPrac/5287-207931.elementaryappropra c.pdf; National Association for Sport and Physical Education. (2009b). *Appropriate instructional practice guidelines for middle school physical education* (3rd ed.). Reston, VA: Author. Retrieved from http://www.cahperd.org/cms-assets/doc uments/ToolKit/NASPE_ApprroPrac/5289-666992.msappropracpdf; U.S. Department of Health and Human Services. (2008); Strong et al. (2005).

48. Bellg et al. (2004); Sheldon et al. (2010).

JEAN M. THAW *is codirector of Project RISE and part-time core faculty at the Center for Psychological Studies of Nova Southeastern University.*

MANUELA VILLA *is an adjunct professor at Nova Southeastern University.*

DAVID REITMAN *is codirector of Project RISE and an associate professor at the Center for Psychological Studies of Nova Southeastern University.*

CHRISTIAN DELUCIA *is an associate professor at the Center for Psychological Studies of Nova Southeastern University.*

VANESSA GONZALEZ *is a psychometrician at the American Welding Society.*

K. LORI HANSON *is the chief of research evaluation and strategic planning at The Children's Trust in Miami-Dade County, FL.*