TBM

ORIGINAL RESEARCH

Recess environment and curriculum intervention on children's physical activity: IPLAY

Claudio R. Nigg,[1] Eve Kutchman,[2] Katie Amato,[1] Christine A. Schaefer,[3] Guangxiang Zhang,[4] Md Mahabub Ul Anwar,[1] Peter Anthamatten,[5] Raymond C. Browning,[3] Lois Brink,[2] James Hill[2]

[1]Office of Public Health Studies, University of Hawai'i, Honolulu, HI, USA
[2]Anschutz Health and Wellness Center, University of Colorado, Aurora, CO, USA
[3]Department of Health and Exercise Science, Colorado State University, Fort Collins, CO, USA
[4]Biostatistics and Data Management Core, University of Hawai'i, Honolulu, HI, USA
[5]Department of Geography and Environmental Sciences, University of Colorado Denver, Denver, CO, USA

Correspondence to: Claudio R. Nigg, cnigg@hawaii.edu

Cite this as: *TBM* 2019;9:202–216
doi: 10.1093/tbm/iby015

Abstract

Understanding the impacts of the built environment on physical activity (PA) is essential to promoting children's PA. The purpose of this study was to investigate the effects of schoolyard renovations and a PA recess curriculum alone and in combination on children's PA. This was a 2 (learning landscape [LL] vs. non-LL) × 2 (curriculum intervention vs. no curriculum intervention) factorial design with random assignment to the curriculum intervention, and six elementary schools per condition. PA outcomes were assessed preprogram, mid-program, immediate postprogram, and one year postprogram. No meaningful intervention effects were found. Lack of an effect may be due to the brief dose of recess, the curriculum not being integrated within the schoolyard, the LL implementation occurring prior to the study, or the already high levels of PA. Potential avenues to promote PA include making recess longer, integrating recess into the school curricula, and developing recess PA curricula integrating schoolyards.

Keywords

Youth, Elementary school, Surroundings, Exercise, Intervention, Playgrounds

Implications

Practice: Recess as operationalized in these schools may not be enough of an intervention opportunity to promote physical activity (PA); additional opportunities to accumulate PA within the school day may be needed.

Policy: There are several potential avenues to increase the likelihood that recess could be a valuable part of PA promotion effort through making recess longer, integrating recess into the school curricula, and developing next-generation recess PA curricula that capitalizes on and explicitly incorporates the schoolyard environment.

Research: Researchers need to investigate the meaning of high-frequency PA data and what implication this has on current recommendations.

Promoting children's physical activity (PA) has become a primary public health goal in the USA [1]. At least 60 min of daily moderate- to vigorous-intensity PA (MVPA) is recommended for children [2, 3], which reduces body adiposity, increases aerobic fitness, reduces blood pressure, and improves bone mass, among other health benefits [3]. However, less than half of 6- to 11-year-old U.S. children meet these guidelines based on self-report [4]. Furthermore, MVPA prevalence decreases after childhood; therefore, PA should be promoted for children [4, 5].

Schools provide the opportunity for cost-effective and efficient delivery of PA interventions due to their reach, the time children spend in school, and the PA equipment present in schoolyards [6]. PA at school may be especially important for low-income, urban, minority children where PA opportunities and facilities are often limited [7]. Children's MVPA occurs at various times throughout the school day, including recess, physical education classes, lunch, and regular classroom time [8]. Due to academic demands, opportunities for school day PA via physical education classes may be reduced [9]; however, recess and afterschool programs provide students with access to school facilities and may promote increased PA among youth [10, 11]. Schoolyards, where recess most often occurs, offer safe and accessible PA opportunities for school-aged children regardless of ethnicity, income level, and sex [12].

Research on specific impacts of recess and school environments on children's PA behavior is limited but growing. Fjørtoft and colleagues [13], for instance, found that students engaged in the most intense activity in handball courts in Norway, which suggests that such facilities are important components of activity spaces in Norway. Previous work in Denver, Colorado, has shown that children's PA behavior is related both to the type of surface [14] and the density of features on the school grounds [15]. Small, inexpensive interventions that have changed the structure of the physical environment have shown significant positive differences, including painting the playgrounds [16] and providing game equipment [17]. Even increasing the number of balls available to youth showed a positive correlation in a cross-sectional study [18]. Systematic reviews have also been conducted to document the effects of recess interventions on children's PA in the USA and internationally. One review of interventions in primary school settings found that PA was significantly increased by intervention strategies that combined playground markings, court rotation, and movable equipment but was decreased by an intervention using active video games [19]. A review examining correlates of children's and adolescents' PA during school recess found overall facility provision, movable equipment, and perceived encouragement from parents, peers, and staff were all associated with higher PA levels [20]. One of the most effective strategies identified by a review of preschool and primary school interventions was the involvement of teachers as proponents of PA in a semi-structured recess environment, providing children the option of teacher-led activities [21].

Given the need to increase PA among children and the positive individual effects of recess interventions and environmental changes on PA, an intervention that combines the two approaches may result in greater recess PA than each individual approach. Therefore, the purpose of this study was to investigate the effects of schoolyard renovations and a PA curriculum delivered during recess alone and in combination on children's PA. It was hypothesized that schoolyard renovations completed 3 years prior, a PA curriculum delivered during recess, and the combination of the two would improve PA indicators compared with a no-treatment control. An additional exploratory hypothesis was put forth to examine whether the combined condition would outperform either condition in isolation.

METHODS

Study design and participants

The Intervention for PhysicaL Activity and Youth (IPLAY) was developed in collaboration with the University of Colorado at Denver College of Architecture and Planning, the Denver Public School (DPS) System, the Colorado State University Exercise Science Laboratory, and the University of Hawaii's Office of Public Health Studies. IPLAY was a 5-year study that employed a 2 (learning landscape [LL] vs. non-LL) × 2 (curriculum intervention vs. no curriculum intervention) factorial design (resulting in four groups: control, curriculum only, LL only, and LL-curriculum) with repeated measures (pre-program—Year 1 [Y1], mid-program—Year 2 [Y2], immediate postprogram—Year 3 [Y3], and one year postprogram—Year 4 [Y4]) with random assignment to the curriculum intervention. The LL intervention, completed 3 years prior to the curriculum intervention, had transformed schoolyards into attractive and safe multiuse playgrounds tailored to the needs and desires of the local community. The decision to use 3-year existing renovated playgrounds was based on pilot results showing no differences in children's PA between past years versus 3+ years ago renovated schoolyards [4], allowing more time for integrating renovations into school. The curriculum intervention combined the Sports, Play, and Recreation for Kids (SPARK), Active Recreation (AR), and Balance First curricula and was delivered for 8 weeks each fall and spring semester for 2 years. The intervention aimed at increasing recess PA, which may generalize to other parts of the day.

SPARK curriculum

SPARK is an evidence-based children's PA and health curricula [22]. The SPARK curriculum has been designed to provide PA opportunities, regardless of the child's experience or ability. The curriculum is easy to use and modify to fit recess setting constraints. Instructors can modify lessons by altering the intensity, duration, and difficulty level of the specific activities [22]. For the current study, local university students with previous experience working with children were recruited as SPARK Research Assistants (RAs) to serve as curriculum intervention instructors. All participating RAs attended a 2-day training workshop led by certified SPARK trainers. The same RAs implemented the SPARK intervention for the study duration maximizing delivery consistency. Due to the promising literature on recess and PA [19–21] and to avoid interfering with the class offerings, the SPARK curriculum was delivered for the current project during lunch recess with an average recess length of 18.95 min (SD = 4.42). All students attended recess (unless excused) and SPARK curriculum activities were available for all children to participate, but participation in SPARK was not mandatory. No specific efforts were made to adapt or integrate the SPARK curriculum to the specific schoolyards. Children were allowed to join at will, and on average, 25% of the school population excluding absentees participated in each session, about 19 children per session (range: 7%–88%).

LL initiative

The LL initiative transformed neglected Denver public elementary schoolyards into attractive and safe multiuse playgrounds tailored to local community needs and desires (Appendix Fig. A1). Prior to renovations, playgrounds within the DPS system averaged 50 years old. In 2000, 75 of the DPS elementary schools were identified as requiring moderate to extensive renovation or upgrades to meet adequacy standards. They consisted of hard play surfaces such as gravel or concrete, were devoid of plant life, and had limited play equipment. These schoolyards were considered "one-size-fits-all," with younger children often using age inappropriate equipment.

The LLs are developed through a hands-on, service-learning curriculum at the College of Architecture and Planning (at the University of Colorado, Denver), which enlists graduate students to develop master plans targeting existing and proposed uses, relationships between uses, programmatic requirements for uses, maintenance and safety issues, and preliminary cost estimates. Once the master plan is complete, the schools participating in the alliance program move into a design studio where graduate students synthesize the pieces of the master plan into a detailed site design. This process involves the students, teachers, parents, and communities through focus groups and community planning meetings. All work together to develop a comprehensive and detailed set of design development drawings. Landscape architects finalize the students' drawings in the form of construction documents.

The LL initiative transforms these rundown playgrounds into spaces accommodating a variety of activities to provide a high stimulus level for children of all ages with more opportunities for tactile, motor, and sensory experiences and by integrating learning with PA. The LLs provide participatory landscapes that support children's healthy development by encouraging outdoor play and learning, offering socialization tools, and improving opportunities for PA while reflecting the unique culture and history of the people, the school, and the neighborhood it serves. The distinctive elements of LL schoolyards include community gateways and gathering spaces, public art works, age appropriate play equipment, grass playing fields, colorful structured and unstructured asphalt games, custom shade structures, habitat areas, and nature play. The grass fields, often referred to as multipurpose fields, are improved through gravel removal, weed removal, new sod, grading, and irrigation. These fields provide places for a variety of children's activities including soccer, baseball, football, and tag. Hard surfaces for games are renovated to include basketball, tetherball courts, foursquare, hopscotch, and wall ball.

Traditional developmentally appropriate equipment that has improved accessibility and safety is provided at each site. Data has shown that children perceive more activity structures in renovated LLs compared with non-LLs [23]. To ensure the LLs had been integrated into the schools and the administration and teachers were familiar with the LLs, schools were selected that had renovated their schoolyards at least 3 years prior to the study.

Participants

Twenty-four urban DPS schools were recruited based on the willingness to cooperate with random assignment, implementing a curriculum intervention, and 4-year data collection [24]. Twelve randomly selected LLs were assigned a matched, control school (non-LL) on the basis of the school's size, ethnic population, and the percentage of students receiving free or reduced lunch. One school from each pair was randomly assigned to the curriculum condition and the other to the noncurriculum condition. To facilitate implementing this study, we enrolled schools in two waves. In the first project year, one half of the selected schools (three in each study condition; Wave 1) began the study, with the remaining 12 schools beginning the study the following year (Wave 2; Fig. 1). The study was approved by the Colorado Multiple Institutional Review Board (IRB), Colorado State University IRB, and the University of Hawaii IRB.

Measures

Intervention process data collection (Year 1, Wave 1 only)

Measurement RAs (different from the SPARK RAs) were asked to record PA-related information but were uninformed about the study purpose or what intervention condition the participating schools belonged to. Measurement RAs were assigned to specific schools for the duration of the project. Thus, measurement RAs were blinded to condition and used the SPARK session checklist [22] to assess 14

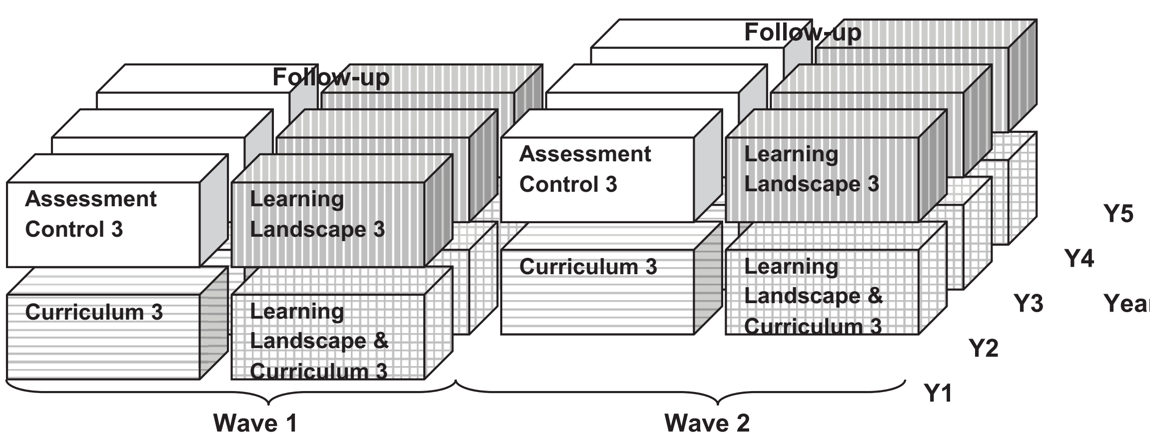

Fig 1 | The 4 (group) × 4 (time point) study design for two recruitment waves. The number of schools in each group is also identified.

items that addressed the intervention implementation quality for Y1 for Wave 1 schools.

RAs also completed log sheets for each intervention session to evaluate the delivery of intervention components by tracking the activity name and the number of participants per session. A session entry was recorded for each unique combination of grade and time frame at the school; a school with different grades on the playground at different times received one entry for each grade level. If multiple grades shared the same specific time frame for recess, however, the grades were grouped together because it was not possible to distinguish the number of participants from each grade. RAs reported whether implementation was executed as planned, if it was carried out according to the objectives and rules of the game with little or no modification to the existing curriculum—with the intent to encourage thoughtful lesson-planning—prior to recess. Zero participants were recorded when children chose not to participate in the activities or were absent from recess due to a school-related factor, such as field trips or the elimination of recess as a consequence of poor behavior.

Due to resource limitations and the positive results (noted below) the process evaluation efforts were discontinued after Year 1, Wave 1.

Outcome measures

Outcome measures were collected at baseline (Y1), annually during the 2-year intervention (Y2 and Y3), and one year after the completion of the intervention (Y4) using three complementary lines of evidence: direct observation, self-report surveys, and accelerometer measurements. Y4 measurements consisted of direct observation, self-report surveys, and accelerometry for Wave 1, whereas only direct observation was performed in Wave 2 due to time and resource limitations. Three lines of evidence were used to address the limitations of any one method, including social desirability effects from self-report surveys, the inability to measure every type of activity from accelerometry measurements, and the restriction of sampling moments in time from direct observation.

SOPLAY: System for Observing PhysicaL Activity in Youth

Direct observation of recess PA was measured using the System for Observing PhysicaL Activity in Youth (SOPLAY). Observers attended a 2-day training led by a certified SOPLAY instructor that included direct practice using modeling, videotaped segments, and field practice in the observational setting. All schoolyards were divided into 10 to 20 observation zones (depending on sizes and attributes) and were systematically observed, on average, 4.3 times (*SD* = 2.0) per recess (range: 1–12) on four nonconsecutive days from April through May. Using momentary time sampling, trained observers rated individual children's PA levels (sedentary, walking, or very active), separately for boys and girls (all students in all grades could potentially be observed). For inter-rater reliability, two observers simultaneously observed each zone for 25% of the total time measured (i.e., one day per school) and showed an inter-rater agreement of 77.89%, 91.87%, 90.55%, and 89.92% for Y1, Y2, Y3, and Y4, respectively. Throughout the study, all observers were blind to the IPLAY study curriculum conditions and hypotheses.

The observed sedentary, walking, and very active PA levels have been validated by heart rate monitors [25] and accelerometers [5], and they can be used to estimate an energy expenditure rate (EER). Total EER per zone (kcals/kg/min) was estimated by multiplying the number of children at each PA level by a constant (0.051kcal/kg/min for sedentary, 0.096kcal/kg/min for walking, and 0.144kcal/kg/min for very active) and summing those values [26]. In addition, a "per child" estimate of EER per zone was calculated by dividing total EER by the number of children observed in the zone, excluding records with no children in a zone. Total and per child EER (by sex) were averaged by zone for each day of observation, and then averaged across days; therefore, EER represents the average total and per child EER in each zone within a schoolyard. The intra-class correlation (*n* = 96; 24 schools, over 4 days) assessing consistency in total EER was 0.87 (95% CI 0.80 to 0.92).

Self-report survey

A subsample of four LLs and four non-LLs were randomly selected for survey administration, with two classes of fourth and two classes of fifth graders completing the survey per school (for a total of 32 classes). Each year, survey classroom teachers received a gift card with a value of up to $215 as an incentive. The incentives constituted $40 for personal use, $80 for classroom supplies, and up to $95 for classroom use, based on number of consented students. Trained research assistants and classroom teachers distributed consent forms followed by verbal and written reminders 2 weeks before survey administration. Over the course of the entire study, a total of 2,917 students were notified of the study. For the four intervention schools, 1,329 students were approached and 848 consented (64% return rate). The return rate from one intervention school was extremely low (6%), while the three other intervention schools ranged from 58% to 98%. For the control schools, a similar trend occurred: There were 1,588 students approached and 815 consented, yielding an overall return rate of 51%, where one school had a particularly low consent rate (5%) and the remaining three schools' return rates ranged from 50% to 82%. The low consent rates in two of the schools were due mainly to principal and staff turnover in those schools. Table 1 presents the return rate by year. Parental consent was obtained prior to data collection and informed student assent was given immediately before survey administration. A research assistant read the survey aloud to the children, while two others were available to answer questions.

Participants reported their height, weight, sex, and age in years. This information was used to calculate body mass index (BMI) for age. A study assessing the accuracy of height and weight of children aged 6 to 11 compared self-report data from surveys with objectively collected height and weight using stadiometers and scales from accelerometry data in the same IPLAY participant pool for baseline of the current study and found that by fifth grade, children are nearly as accurate as adults in self-reporting height and weight [27].

PA was measured using an adapted Godin Leisure-Time Exercise Questionnaire [28], which used modified response categories to allow participants to record how many days per week and minutes per day they perform strenuous, moderate, and mild PA. These PA intensity levels were defined, and several examples were given. Responses for days per week ranged from 0 to 7 days and responses for minutes per day ranged from 0 to 60+ min, reported in 10-min increments. Daily moderate and strenuous PA was combined to determine minutes per day spent in MVPA. The Leisure-Time Exercise Questionnaire was found to be significantly related to Caltrac accelerometer readings (r = .32), metabolic equivalents (METs; r = .36), treadmill exercise time (r = .57), percentage of body fat (r = −.43), and VO2max (r = .56) [29] and was significantly related to the PA stages of change across populations [30], including children [31]. Sallis et al. [32] reports good test–retest reliability (r = .81) and adequate validity (r = .39) when compared with kilo calories expended per day in a sample of 5th, 8th, and 11th graders.

Accelerometer PA assessment

During the spring of each year, we collected six consecutive days of accelerometer data on a cross-sectional subset of one class of first-, third-, and fifth-grade students in 8 of the 24 schools (three classes from each of 2 schools per condition). These were the same schools as participated in the self-report assessments, and the fifth-grade accelerometry class was one of the fifth-grade self-report classes. Data were collected during April and May of 2010–2013, after the conclusion of the spring intervention period. Over the course of the whole study period, a total of 1,367 students participated in the accelerometer data collection (due to an administrative error, recruitment data for accelerometry are no longer available). Each year, the same teacher's class was selected for accelerometer data collection. In cases where teachers were no longer employed at the school, study staff worked with principals to select a replacement teacher in the same grade. During the first year (Y1 Wave 1 schools only), we used the Actical accelerometer (Philips Respironics, Bend, OR), a lightweight (17 g), omni-directional,

Table 1 | Return rate of survey respondents over time

	Notified (n)	Consented (n)	Returned (n)	% Return rate
Control				
Year 1	297	112	112	38
Year 2	219	78	78	36
Year 3	200	124	124	62
Year 4[a]	95	77	77	81
Curriculum				
Year 1	121	100	100	83
Year 2	132	129	129	98
Year 3	199	162	162	81
Year 4[a]	102	80	80	78
LL				
Year 1	259	144	144	56
Year 2	210	122	122	58
Year 3	202	112	112	55
Year 4[a]	106	46	46	43
LL-Curriculum				
Year 1	274	80	80	29
Year 2	197	109	109	55
Year 3	209	158	158	76
Year 4[a]	95	30	30	32

LL learning landscape.

[a]Wave 1 only.

waterproof device that detects low frequency accelerations (0.5–2.0 Hz). It generates an analog voltage signal that is then filtered, amplified, and digitized by an A-to-D converter at 32 Hz. These digitized values are summed over the epoch and stored in the device. These stored values are proportional to the duration and magnitude of the movement [33]. Devices were calibrated by the manufacturer prior to use. During Y1, in the Wave 1 schools (4 schools), we collected data in 15-s epochs, the shortest available for the Actical device. After this baseline data collection, the benefits of a device that would collect and output unprocessed acceleration data so that we could select an appropriate epoch length (e.g., 1 s) became evident. Because of this, we began using the GENEActiv accelerometer device (Activinsights Limited, Cambridge shire, UK), a light-weight (16 g), wrist-worn, tri-axial, waterproof device that collects high frequency acceleration data up to 100 Hz. It has been validated for use among both children and adults [34]. Devices were calibrated by the manufacturer prior to use and data were collected at 30 Hz (a subset of data collected in three schools during the spring of 2011) and 75 Hz (all remaining data).

Data collection procedures

On the day the accelerometers were distributed, study staff explained the study and assented all children who had returned parental consent forms. Children then assembled in the hall where study staff assigned a device serial number to each child and attached the accelerometer device to the child's nondominant wrist using a semi-nonremovable hospital-type band (MedTech Wristbands, Orlando, FL). Children were instructed to go about their normal daily activities while wearing the device consecutively for the next 6 days. Study staff measured each child's height to the nearest 0.5 cm (standard tape measure) and weight to the nearest 0.2 kg (Health O Meter professional scale, Model 349KLX) while wearing shoes. BMI percentiles were calculated based on the Centers for Disease Control and Prevention (CDC) growth charts. Children were assigned as normal weight (NW, <85th percentile BMI-for-age score) or overweight/obese (OW/OB, ≥85th percentile BMI-for-age score).

Parents and teachers were given instructions about the devices and were asked to report abnormalities in activity during the data collection. Teachers provided school day schedules indicating when school started and ended and when children were at lunch, recess, and physical education classes. On the sixth day of data collection, researchers returned to the school to collect the devices. Each child received $10 when the accelerometer was put on and $10 when it was returned. An additional $30 was given to each of the parents and $25 to the teacher of the class.

Data processing. Actical data were downloaded using the Actical software (Version 2.12). A custom Matlab (Mathworks, Inc., v12.0, Natick, MA) program was created to process the accelerometer data and clean for nonwear (see below). Periods of 60 min or greater of zero count values (Actical device) or values below the laboratory-established nonwear threshold (<0.06 g s) were summed over the day to assess completeness of the data file.

All GENEA devices were downloaded using a USB 2.0 Charging Cradle and the GENEActiv software (Version 2.1). We created a Matlab program to read and filter the data file. We applied a low pass filter with a cutoff frequency of 15 Hz to the data to remove noise in the signal not representative of human movement. Once the data were filtered, we calculated a signal vector, the Euclidian Norm minus one (ENMO, see equation 1, where f = sampling frequency). This low pass ENMO value (LPENMO) was calculated on a per-second basis.

$$\mathrm{LPENMO} = \left(\sum_{i=1}^{f} \left|\left(\sqrt{x^2 + y^2 + z^2} \,/ -1\right)\right|\right) / (\mathrm{f}) \qquad (1)$$

After filtering the data, we cleaned the files to remove periods of nonwear. Using a custom Matlab program, we identified periods of 60 consecutive minutes of LPENMO values below 0.06 g s (laboratory established nonwear threshold). These periods of time were summed over each day to assess completeness of the data file.

Data for any day found to have less than 10 hr of wear time or at less than four valid days were considered invalid and removed from the accelerometer data files (n = 31). Custom time intervals were created to identify standard time periods throughout the day. These intervals include the full day (FD; 6 am–11 pm), school day (SD; school-specific start and end time), and lunch recess (LR; class-specific start and end time). The times used for SD and LR were determined from the class schedule completed by teachers. After identifying custom intervals, the Matlab program applied published Actical cut-points established using ROC curves [35] to the baseline Actical data and GENEActiv wrist cut-points derived using the same methodology as the Actical cut-points to the GENEActiv data. Cut-points were applied to determine the number of minutes and percent of time spent in sedentary (SED), light (LPA), moderate (MPA), vigorous (VPA), and moderate-vigorous PA (MVPA; sum of MPA and VPA) during each of the custom intervals. For the Acticals, the SED = 0–52, LPA = 53–387, MPA = 388–1210, VPA = 1211+, and MVPA = 388+. For the GENEActiv, the SED = 0–0.0935, LPA = 0.0936–0.1846, MPA = 0.1847–0.4531, VPA = 0.4532+, and MVPA = 0.1847+.

Statistical analysis

PA main outcomes were first summarized by means and standard deviations by intervention groups and study year. Three complementary lines of evidence used different models to evaluate the intervention effects. Statistical analyses were conducted using SAS® v 9.3 and a two-sided p-value <.05 was regarded as statistically significant. The p-values of pairwise post hoc comparisons of four intervention groups at a time point were corrected by Tukey–Kramer multiple comparison adjustments [36]. Main effects of the two treatment conditions were also explored.

SOPLAY outcomes

To evaluate the differences among four study intervention groups over four points in time, the aggregate school-level PA outcomes from SOPLAY data were fitted by linear mixed-effects repeated measure models. As the PA outcome measures from Y1 to Y4 from the same school were not independent, the year variable was treated as a random effects variable. The fixed effects variables included three terms: intervention and time main effects, and Intervention × Time interaction. The mixed-effects models could compare the differences in treatment effects at a given time point even if the Intervention × Time interaction was significant. If the interaction term was not significant, then the model was refitted after removing the interaction term.

Survey outcomes

For the survey, PA outcome data were fitted by linear mixed-effects models. As the children within an individual school might not be independent, the school cluster effect was taken into account. Therefore, the school was introduced as a random effect (intercept) in the model. The fixed effects included intervention and time main effects, and Intervention × Time interaction, and if significant, children's age and sex. Kenward–Roger's approximation was used to estimate denominator degrees of freedom [37].

Accelerometer outcomes

To confirm that output from the Actical device was comparable with that of the GENEActiv, data collected during a calibration study while wearing both monitors were analyzed (see [35] for a description). Briefly, 24 children participated in a variety of activities while wearing both monitors as well as a portable metabolic system (Oxycon Mobile, Yorba Linda, CA). Pearson product-moment correlation coefficient of 0.86 ($p < .05$) indicated a strong linear relationship between the two device outputs.

Descriptive statistics of accelerometer participants, reported as frequencies for sex, are summarized in Table 2. The dependent variables (i.e., full day minutes of MVPA and percent of time in school day and lunch recess in MVPA) were analyzed with a linear mixed-effects model with similar model specification as survey outcomes. Fixed effects included the year and intervention conditions, their two-way interaction, and when significant, children's age, sex, and BMI. School cluster effects were included as random effects (intercept). Kenward–Roger's approximation was used to estimate denominator degrees of freedom [36].

Power considerations

Sample size estimates were derived to obtain a sufficient number of schools and students to ensure adequate statistical power to detect differences between intervention arms in PA level, at the school level, and PA outcomes among children at the schools. The minimum detectable difference (MDD) was calculated based on the z-test for a comparison of means with adjustment for clustering in a group-randomized trial (GRT). To adjust the variance for a GRT, we included the design effect term in the calculations. The design effect (DE) gives the ratio of efficiency of the GRT to a randomized clinical trial and can be estimated as $\theta = (1 + (m - 1)\rho)$, where m represents the number of individuals in each community and ρ is the intraclass correlation (ICC). Postprogram measures were considered as the primary outcomes. We conservatively used an ICC of 0.1, a critical level of 0.05 and 0.01 (Bonferroni corrected for four–five comparisons), and a power of 80% to determine the MDD in means. The estimation of the variance and reasonable MDDs come from prior research. For the LL intervention, our pilot project results showed 67.29% (SD = 9.82%) compared with 61.47% (SD = 9.26%) of the children were either walking or very active in the built versus the not-built sites, respectively, resulting in a d of 0.61. Our pilot study using the Fun 5/SPARK AR curriculum resulted in an increase in observed MVPA from the baseline mean of 12.56% (SD = 6.99% across sites) to 29.68% (SD = 11.45%) during the after-school time for the pilot study [38]. In the first-year dissemination we found very similar results, increasing moderate and vigorous PA (MVPA) from 14% (SD = 11%) to 23% (SD = 9%) of the after-school program session of time spent in MVPA [10], resulting in a d of 1.82. We calculate that with a total of 6 schools per arm and 20–50 observation made per school during the postprogram period in question, we will be able to detect an effect size d of 0.7 in pairwise comparisons (e.g., LL+SPARK vs. control) with $\alpha = .05$ and a d of 0.8 with $\alpha = .01$. Note that we will be able to state definitely whether each intervention arm is different from the control condition, but will be only able to detect fairly large difference between LL+SPARK and the expected combined sum of LL and SPARK. For secondary analysis, with 2 schools per arm and an average of 69 students per school, the effect size d is 1.2–1.3 in pairwise comparisons (e.g., LL+SPARK vs. control) with $\alpha = .05$ when the school is the unit of comparison and d is 0.4 when student is the unit of comparison. There was sufficient power to test the hypotheses of interest.

Table 2 | Demographic characteristics of SOPLAY schools, survey participants, and accelerometer participants by treatment group

Item		Control	Curriculum	LL	LL-Curriculum
SOPLAY schools					
Total number of students (*n*)		3,491	2,831	3,009	2,339
% Free lunch		69.0	62.9	55.9	66.6
Ethnicity[a]	% Hispanic	64.3	54.2	63.9	53.2
	% White	25.4	40.0	27.7	32.0
Survey participants					
% Female (*n*)	Year 1	55.6 (108)	52.1 (94)	37.8 (136)	68.5 (54)
	Year 2	55.7 (79)	48.4 (126)	46.3 (121)	63.6 (107)
	Year 3	48.0 (123)	52.3 (153)	51.8 (112)	51.6 (156)
	Year 4[b]	50.0 (74)	50.0 (82)	58.7 (46)	46.5 (43)
Mean age (*SD*)	Year 1	10.3 (0.8)	10.2 (0.8)	10.3 (0.8)	10.1 (0.8)
	Year 2	10.2 (0.6)	10.2 (0.8)	10.2 (0.8)	10.3 (0.8)
	Year 3	10.2 (0.7)	10.2 (0.7)	10.1 (0.7)	10.3 (0.8)
	Year 4[b]	10.1 (0.8)	10.2 (0.8)	10.4 (0.6)	10.2 (0.7)
Mean BMI percentile (*SD*)	Year 1	65.8 (34.0)	65.1 (33.2)	57.4 (33.8)	70.6 (31.1)
	Year 2	57.6 (36.2)	57.3 (33.7)	40.5 (33.6)	61.1 (34.0)
	Year 3	61.8 (36.9)	62.4 (31.9)	55.5 (32.7)	66.6 (34.1)
	Year 4[b]	63.1 (37.9)	55.7 (34.6)	59.5 (36.4)	60.1 (38.6)
Ethnicity[a]					
% Hispanic	Year 1	76.5	34.5	48.9	85.7
	Year 2	76.6	48.4	37.1	80.4
	Year 3	75.9	37.5	54.1	79.7
	Year 4[b]	83.8	34.6	61.4	87.8
% White	Year 1	15.7	46.1	33.8	0.0
	Year 2	13.0	36.1	41.4	5.4
	Year 3	15.7	38.6	19.7	8.5
	Year 4[b]	16.2	60.3	22.7	4.9
Accelerometer participants					
% Female (*n*)	Year 1	46.1 (76)	47.8 (69)	42.9 (127)	60.3 (74)
	Year 2	45.0 (90)	51.8 (51)	45.1 (127)	55.4 (111)
	Year 3	49.5 (90)	52.1 (125)	42.5 (96)	50.0 (115)
	Year 4[b]	42.4 (50)	52.1 (60)	53.4 (65)	61.2 (41)
Mean BMI percentile (*SD*)	Year 1	69.7 (27.5)	61.1 (30.3)	54.6 (30.7)	65.8 (31.2)
	Year 2	67.0 (27.4)	61.3 (31.4)	48.8 (30.8)	65.7 (31.4)
	Year 3	66.6 (29.8)	59.2 (30.9)	56.9 (31.7)	77.4 (29.0)
	Year 4[b]	64.7 (32.9)	61.5 (31.1)	58.7 (29.4)	61 (33.4)

SOPLAY System for Observing PhysicaL Activity in Youth; *LL* learning landscape; *BMI* body mass index.

[a]Ethnicity ≠ 100, as non-White/non-Hispanic is not reported in the table.

[b]Wave 1 only.

RESULTS

Descriptive of demographics and SOPLAY, survey, and accelerometer

A summary of baseline demographic characteristics of SOPLAY and self-reported surveys is provided in Table 2. Control schools had highest student enrollment (3,491) followed by LL schools (3,009). More than half of the students received free lunch and more than half of the students were Hispanic (neither percentage receiving free lunch nor ethnicity were associated with the main PA outcomes [$p > .05$]). For the self-reported survey sample, students ranged in average age from 10.08 ($SD = 0.65$) to 10.39 ($SD = 0.61$) years old. The largest ethnic group represented was Hispanic, followed by White, and almost half were girls.

INTERVENTION OUTCOMES

Intervention process data (Year 1, Wave 1 only)

A total of 58 process evaluations (SPARK session checklists) were conducted over the fall and spring interventions for the first year of implementation. The total score on evaluations ranged from 3 to 14 points with mean of 11.6 (83.1%) with 12 indicators observed to be implemented 80% or more of the time. Indicators with percent observed were enthusiasm from instructor (98.3%), participant enjoyed the activities (96.6%), clear articulation of instructions

(94.8%), provision of a safe environment (93.1%), positive feedback from the instructor (93.1%), instructor was prepared with equipment (91.4%), adequate learner to equipment ratio (91.4%), session started within 5 min of recess release (89.7%), participants were active at least 50% of the time (87.9%), minimal management and transitions times (86.2%), instructor was prepared with lesson plan (82.8%), appropriate group sizes (82.8%), activity specific feedback (55.2%), and used existing play structures (20.7%).

There were a total of 1,497 session entries for the 12 intervention schools across 16 intervention weeks, with each school contributing between 114 and 215 entries for Year 1, Wave 1. These data were compiled from instructor lesson plans. A total of 1,364 activities were implemented during the first intervention year. Among the activities observed, 82.3% were modified forms of SPARK-AR curriculum activities and 17.7% were classified as non-SPARK activities. Counting all the observed sessions (including non-SPARK curriculum), 70.4% were classified as "implemented as planned," without significant changes to the prescribed curriculum or planned activity. This represents the preparedness of the instructor. The mean number of children participating per session was 19, ranging from 0 to 109 and children spent an average of 18 min on each session. The average recess, as recorded by the RAs, was 24 min; thus, 75% of the recess time was spent on the prescribed curriculum or planned activity.

SOPLAY

Roughly three quarters (75.5%) of students were observed in MVPA during the recess time, while 41.3% of students were observed in VPA. The linear mixed-effects models showed no significant Intervention × Time interaction effects for all four PA main outcomes, four $F(9, 60) < 1.0$, $p > .60$. There were generally no significant differences among the four intervention groups for all four PA main outcomes at a study year, Y1, four $F(3, 60) < 1.0$, $p > .47$; Y2, four $F(3, 60) < 1.27$, $p > .20$; Y3, EER, $F(3, 60) = 3.99$, $p = .012$; the other three $F(3, 60) < 1.13$, $p > .11$; Y4, four $F(3, 60) < 1.91$, $p > .14$. Although the EER level was significant among the four intervention groups in Y3, none of the six pairwise group comparisons was significant after Tukey–Kramer multiple comparison adjustment, for example, LL-Curriculum versus Control, difference = 4.65 kcal/kg/min, $t(df = 60) = 3.26$, adjusted $p = .11$.

After removing the Intervention × Time interaction term from the model, intervention main effects were significant for EER, $F(3, 20) = 4.11$, $p = .02$, and not for the other three main PA outcomes, three $F(3, 20) < 1.0$, $p > .49$. For the time main effect, compared with Y1, the study end point at Y4 was comparable for EER (the school-level PA volume indicator), $F(1, 69) = 1.85$, $p = .18$. However, the other three PA main outcomes (average PA level indicators of a school) all consistently had higher average activity levels at Y4 compared with Y1 within each of four groups: EER per child, $F(1, 69) = 10.30$, $p = .0020$; MVPA%, $F(1, 69) = 9.42$, $p = .0031$; and VPA%, $F(1, 69) = 7.78$, $p = .0069$. Neither main effect for LL or SPARK was significant for any SOPLAY PA outcomes across time points (all $p > .05$).

Survey (Fig. 2)

The survey data analysis comprised $n = 1,614$ students, including all four time points. The survey was multiple cross-sectional and assessed by the linear mixed-effects model, with children nested within school as random effect. The school cluster effects were estimated using ICC, and were minimal (0.0095). The Intervention × Time interaction effect was significant, $F(9, 1190) = 2.31$, $p = .014$. However, there were no significant differences among four intervention groups at any study year, that is, at Y1, $F(3, 13.3) = 2.62$, $p = .09$; at Y2, $F(3, 11.4) = .94$, $p = .45$; at Y3, $F(3, 8.54) = 2.61$, $p = .12$; at Y4, $F(3, 23.4) = 1.15$, $p = .35$.

When sex, age, and BMI were also included as fixed effects in the above model, males had 10.36 more MVPA minutes per day than females, $F(1, 1568) = 44.08$, $p < .0001$, but age was not significant, $F(1, 1567) = 3.57$, $p = .06$, nor was BMI, $F(1, 1188) = 0.55$, $p = .46$. Similar results were observed after adjusting for sex and age compared with the model, for example, Intervention × Time interaction effect, $F(9, 1294) = 2.71$, $p = .0039$; no significant differences among four intervention groups at any study year were observed, that is, Y1, $F(3, 12.1) = 2.12$, $p = .15$; Y2, $F(3, 9.93) = 1.11$, $p = .39$; Y3, $F(3, 7.74) = 2.50$, $p = .14$; and Y4, $F(3, 20.6) = 1.37$, $p = .28$. Neither main effect for LL or SPARK was significant for MVPA across time points (all $p > .05$).

Accelerometer

A total of 1,367 students participated in the accelerometer portion of the study with all three MVPA outcomes. The 4 years of accelerometer data were multiple cross-sectional and analyzed using the linear mixed-effects model with children nested within school as a random effect. We had a usable accelerometer data rate of 88.7% over the 4 years of assessment (including noncompliance and device failure [unable to separate]) with individual study year usable accelerometer data rates of 92.8%, 82.1%, 82.9%, and 87.5% for Y1, Y2, Y3, and Y4, respectively.

%MVPA (lunch and recess, Fig. 3A). There was a significant Intervention × Time interaction effect, $F(9, 1311) = 15.15$, $p < .0001$. The school cluster effects were estimated using ICC, which was 0.13. However, there were no significant differences among the four intervention groups at any study year, that is, Y1, $F(3, 4.96) = 2.76$, $p = .15$; Y2, $F(3, 4.72) = 2.01$, $p = .24$; Y3, $F(3, 4.37) = 1.77$, $p = .28$; and Y4, $F(3, 5.93) = 4.67$, $p = .05$.

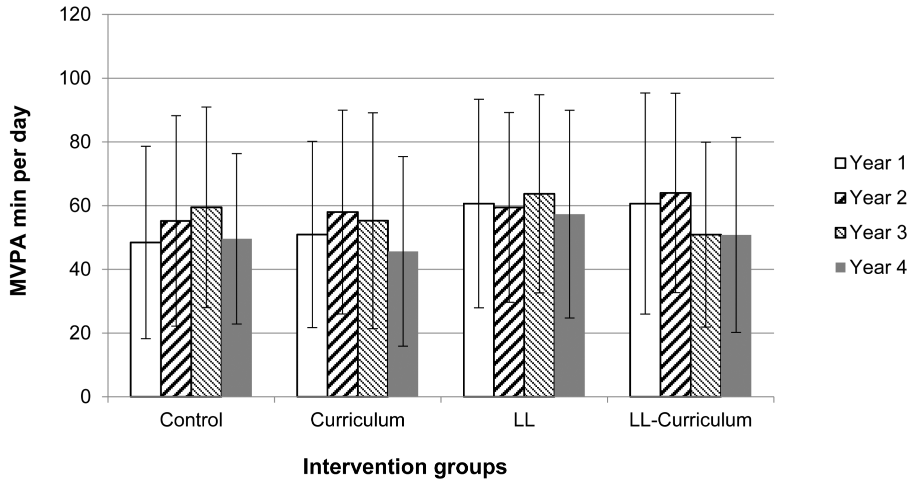

Fig 2 | Survey outcomes for mean (*SD*) of self-reported moderate- to vigorous-intensity physical activity (MVPA) minutes per day by intervention group and year. The linear mixed-effect model was used to assess the cross-sectional survey, with children nested within school as random effect. There was a significant Intervention × Time interaction effect ($p = .014$) but no significant differences among the groups at any specific study year. *LL* learning landscape.

The sex effect was significant (difference = 7.72%, higher for males vs. females) when age, sex, and BMI were included in the model, $F(1, 1339) = 102.06$, $p < .0001$, but neither the age nor the BMI effect was not significant, while similar results were observed, for example, Intervention × Time interaction effect, $F(9, 1314) = 16.30$, $p < .0001$, with no significant differences among the four intervention groups at any study year, that is, Y1, $F(3, 4.87) = 2.52$, $p = 0.17$; Y2, $F(3, 4.66) = 1.94$, $p = .25$; Y3, $F(3, 4.35) = 1.86$, $p = 0.27$; and Y4, $F(3, 5.74) = 4.67$, $p = .06$.

%MVPA (school day, Fig. 3B). The school cluster effects (ICC) was 0.03. There was a significant Intervention × Time interaction effect, $F(9, 886) = 12.86$, $p < .0001$. There were some significant differences among the four intervention groups at some study years, Y1, $F(3, 7.14) = 18.05$, $p = .0010$; Y2, $F(3, 6.40) = 1.91$, $p = .22$; Y3, $F(3, 5.56) = 6.13$, $p = .033$; and Y4, $F(3, 11.3) = 3.88$, $p = .040$. There were no significant differences across the four intervention groups at Y3 and Y4 after Tukey–Kramer multiple comparison adjustments, although at Y1 there were significant differences between some pairwise comparisons of intervention groups: LL-Curriculum was higher versus Control, difference = 3.83%, $t(df = 7.97) = 3.88$, adjusted $p = .011$; LL-Curriculum was higher versus Curriculum, difference = 7.25%, $t(df = 8.25) = 7.18$, adjusted $p < .0001$; and LL-Curriculum was higher versus LL, difference = 4.49%, $t(df = 6.17) = 5.01$, adjusted $p < .0001$.

Age was significant when age, sex, and BMI were also included in the model, $F(1, 1340) = 19.41$, $p < .0011$, and males had higher PA levels than females, difference = 1.75%, $F(1, 1340) = 72.95$, $p < .0001$, but BMI was not significant, $F(1, 1340) = 3.05$, $p = .08$, while similar results were still observed, for example, Intervention × Time interaction effect, $F(9, 1021) = 14.06$, $p < .0001$; intervention effects at Y1, $F(3, 6.80) = 17.58$, $p = .0014$; at Y2, $F(3, 6.16) = 1.60$, $p = .28$; at Y3, $F(3, 5.37) = 4.73$, $p = .06$; at Y4, $F(3, 10.20) = 3.99$, $p = .04$. Although there were no significant pairwise group comparisons at Y4, at Y1, there were significant differences between some pairwise comparisons of intervention groups: LL-Curriculum was higher versus Control, difference = 3.87%, $t(df = 7.44) = 3.76$, adjusted $p = .017$; LL-Curriculum was higher versus Curriculum, difference = 7.45%, $t(df = 7.73) = 7.08$, adjusted $p < .0001$; and LL-Curriculum was higher versus LL, difference = 4.70%, $t(df = 5.86) = 4.94$, adjusted $p = .0001$.

Minutes per day of MVPA (Fig. 3C). The school cluster effects (ICC) was 0.021. There was a significant Intervention × Time interaction effect, $F(9, 656) = 3.08$, $p = .0012$. However, there were no significant differences among the four intervention groups at any study year, Y1, $F(3, 6.76) = 4.01$, $p = .06$; at Y2, $F(3, 6.10) = 0.39$, $p = .77$; at Y3, $F(3, 5.28) = 3.12$, $p = .12$; and at Y4, $F(3, 11.30) = 3.01$, $p = .08$.

When age, sex, and BMI were also included in the model, sex differences were significant, that is, 15.22 min more for males per full day, $F(1, 1340) = 59.71$, $p < .0001$; as were age differences, $F(1, 1341) = 21.03$, $p < .0001$; but not BMI, $F(1, 1339) = 0.12$, $p = .73$; while similar results were still observed, for example, Intervention × Time interaction effect, $F(9, 664) = 3.40$, $p = .0004$, with no significant intervention effects at Y1, $F(3, 6.47) = 4.51$, $p = .05$; at Y2, $F(3, 5.87) = 0.47$, $p = .71$; at Y3, $F(3, 5.06) = 2.57$, $p = .17$; and at Y4, $F(3, 10.80) = 3.28$, $p = .06$.

Neither main effect for LL or SPARK was significant for any of the three accelerometry PA outcomes across time points (all $p > .05$).

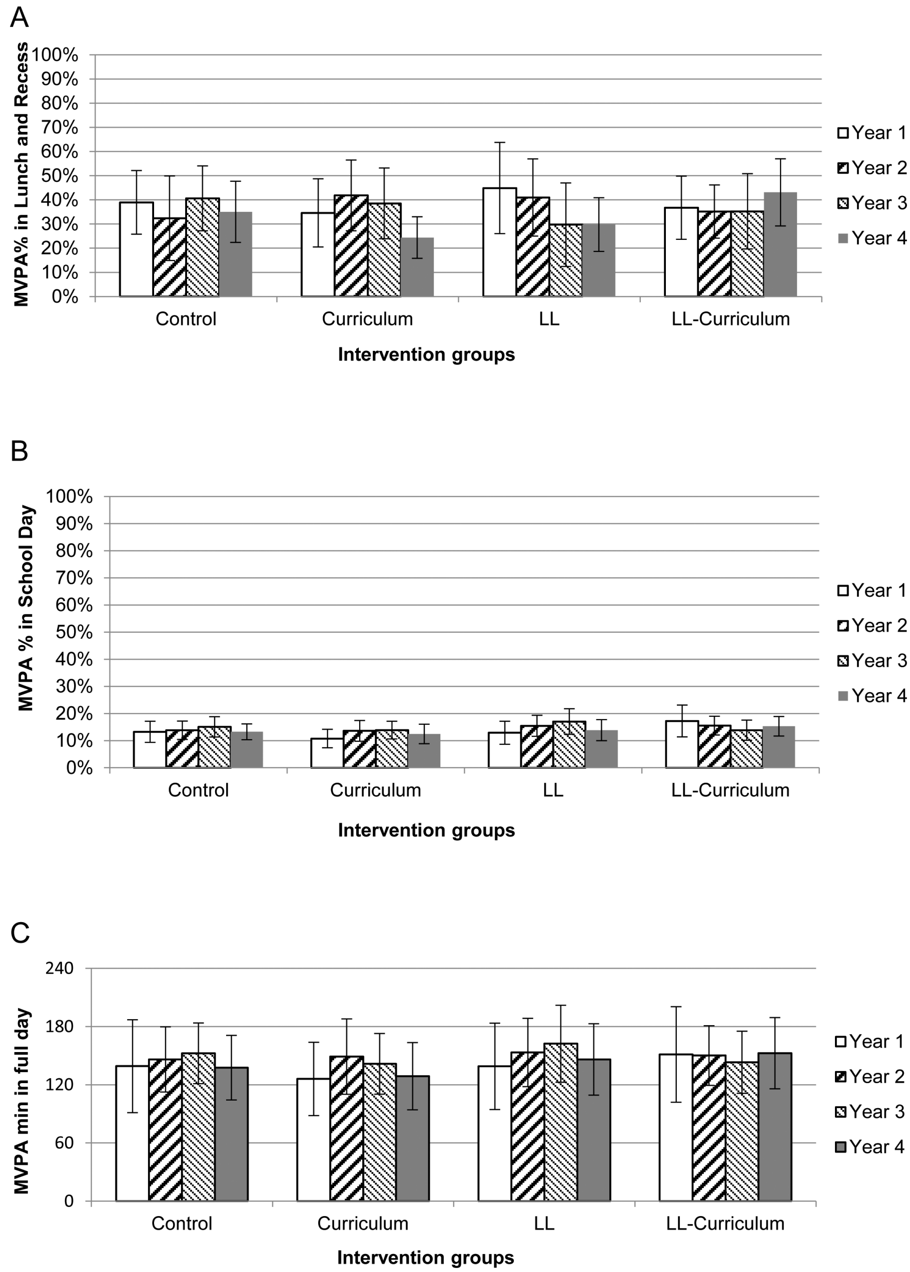

Fig 3 | Accelerometer outcomes for means (*SD*) of percentage of lunch and recess time in moderate- to vigorous-intensity physical activity (MVPA) per day (A), percentage of school day time in MVPA per day (B), and minutes per day in MVPA (C), by intervention group and year. The linear mixed-effects model was used to analyze accelerometer data with children nested within school as a random effect. Percentage of lunch and recess time in MVPA: Intervention × Time interaction, $p < .0001$. Percentage of school day time in MVPA: Intervention × Time interaction, $p < .0001$. Tukey–Kramer multiple comparison adjustments were used to assess pairwise comparisons of intervention groups for percentage of school day time in MVPA at Y1 with significant differences between some: LL (learning landscape)-Curriculum was higher versus Control, adjusted $p = .011$; LL-Curriculum was higher versus Curriculum, adjusted $p < .0001$; and LL-Curriculum was higher versus LL, adjusted $p < .0001$. Minutes per day of MVPA: Intervention × Time interaction, $p = .0012$.

DISCUSSION

This study investigated the effects of elementary schoolyard renovations and a PA curriculum delivered during recess, both alone and in combination, on children's PA. Results using different methods (i.e., observation, self-report survey, and accelerometry) and targeting different parts of the day (recess, outside the school day, and the entire day) all indicated a lack of an effect for any intervention condition compared with control. The lack of effects resulting from the IPLAY study reflects other intervention trials summarized in a recent meta-analysis which reported little to no effect of PA interventions on accumulation of MVPA [39]. The magnitude of effect across 14,326 participants was approximately 4 min of additional MVPA per day [39]. This evidence suggests that interventions must not only attempt to increase PA during already established free time, but also provide additional PA opportunities [40].

Our primary outcome data (direct observation, SOPLAY) do not allow for effectiveness analysis for children at risk (not meeting guidelines), and our other methods (survey and accelerometer) did not have enough statistical power for a meaningful analysis of at-risk children. Other studies have shown that although no effects of PA interventions were found overall, children with low levels of PA did have an increase in PA [11]. A study of the moderating influences of baseline PA levels on a recess intervention found that children with low baseline PA levels benefited more from the involvement of staff combined with equipment availability than children with higher baseline PA levels [41]. This suggests that the IPLAY study may have increased PA levels in the least active children.

Recess has been promoted as a solution to promote PA among children as school day PA via physical education classes may be reduced [9]. However, our study shows that recess by itself does not seem to be long enough (15–20 min) to produce change in children's PA. In addition, some available recess time was needed to manage the children to get ready for recess and to bring them back in from recess. This managing of children affected our ability (decreased the time available) to implement the curriculum during recess. If recess is meant to serve as a meaningful PA-promoting opportunity, it may need to be lengthened, or coupled with PA-promoting classroom activities such as math class summing steps taken during recess, language arts classes providing reading and writing assignments about PA to do during recess, and science classes teaching that energy balance can occur during recess.

SPARK has been implemented in various school settings [10, 11, 40]. However, to our knowledge, no studies have evaluated SPARK AR implemented during recess, alone or in combination with environmental modifications. Our results suggest that SPARK AR delivered during recess did not result in greater amounts of accumulated MVPA during the lunch recess period or other parts of the day. It is possible that individuals participating in the SPARK AR curriculum during recess simply replaced one type of activity with another activity of equally high intensity. Our sample overall spent approximately 38% of recess time engaged in MVPA, which aligns with guidelines suggested by Ridgers and colleagues [42] of 40%, providing possible evidence of activity substitution. Furthermore, the structured SPARK AR delivery period had concluded prior to accelerometer data collection. However, it may have been that once the structured SPARK AR curriculum implemented by trained study staff was completed, teachers, staff, and children did not continue to engage in these activities. Although this design enabled us to examine whether SPARK AR was adopted by the schools after the structured program delivery (i.e., program sustainability), this is a significant limitation in assessing the true effectiveness of SPARK AR. In the future, the intervention curriculum may need to be repeated throughout the year and measurements should be conducted while the SPARK AR curriculum is implemented to determine if it has an effect on MVPA.

The lack of effect of the schoolyard renovations is likely due to the fact that schools were selected that had renovated their schoolyards at least 3 years prior. This was counter to our expectation that students would learn to use the new environment over time. This expectation was indicated by our pilot results which showed no differences in PA between newly (past year) renovated schoolyards and schoolyards that were renovated three or more years ago [24]. For this reason, the study was designed to be longitudinal at the school level instead of the individual level. However, one of the possible mechanisms by which environment renovations promote PA is through a novelty effect. If children are present to see the schoolyard change, they may be more motivated to use the new and improved schoolyard structures. Due to the way that schools were selected, most students were not attending the school before the renovation occurred. Consequently, they perceived no change in the environment, which likely translated to a lack of change in motivation and behavior. This suggests that we may have an opportunity to capitalize on environment changes when they happen and use that time as a "teachable moment" to promote PA. Another possible implication is that environments could be developed that can be easily modified so that a playground is dynamic and thus children's PA would consistently be positively influenced.

We believe the lack of significant associations observed in our exploratory hypothesis comparing the combined intervention with the environment or curriculum intervention separately again may be due in part to short recess and lack of novelty effect. Another plausible explanation is that the curriculum is a standalone curriculum that was not integrated with the schoolyard environment, and therefore did not take advantage of the opportunities provided by the renovated schoolyards, which had more structures. For recess curriculum to increase the likelihood to promote PA, they should be adapted to incorporate the specific features provided by the schoolyard, which in theory would increase opportunities upon renovation. For example, tag games could incorporate the playground structure with using a specific item (standing on the bridge) being considered a safe "home base."

Another possible explanation is that the population we studied did not need an intervention. Data all indicated a high level of PA before the study began. Observation data indicated that over 70% engaged in MVPA during recess, self-report data indicated that about half of the participants met MVPA recommendations, and accelerometer data indicated that the vast majority were physically active for a significant portion of the day (>120 min). This suggests that PA levels have increased in the Denver area the last few years, potentially as a result of recent public health efforts, which needs to be confirmed through representative prevalence studies.

The differences between our methods of PA measurement are likely due to methodology (objective vs. self-report). Other studies that objectively measured samples of children reported an average of 134 min per day using accelerometer with a 1-min epoch among 9- to 12-year-old children [43] and approximately 86.1 min of daily MVPA using accelerometers collected at 2-s epochs among 8- to 10-year-old children [44]. Direct observation studies also demonstrate relatively large estimates of MVPA, including a study of preadolescent children that reported an average of 117 min of MVPA per day [45]. Brink et al. [24] observed that 65% of children engaged in MVPA on schoolyards and Anthamatten et al. [46] found that the rate was as high as 80% during lunch recess. These estimates are in stark contrast to the National Health and Nutrition Examination Survey (NHANES) estimates of daily MVPA which report that only 42% of children ages 6–11 are meeting the guideline of 60 min of MVPA per day [47] and Nigg et al. [10] who reported between 40 and50 min per day of self-reported leisure time MVPA across five cohorts (during 2004–2008) of elementary afterschool participants. The difference seems to be in the more granular data collection used in observation and accelerometry compared with self-report.

Our sample was very active according to the accelerometer data collected at very high frequencies (≤75 Hz). These are some of the first data in a free-living environment collecting PA data at a high resolution. Epidemiological trials using high-frequency accelerometer data and tracking chronic disease outcomes are needed to inform guidelines appropriate for this kind of data. These discrepant findings should raise questions about how we quantify and interpret MVPA data collected via accelerometry. Additional direct observation data should be collected to gain a better understanding of how and how much children move. A more in-depth examination of the metabolic consequences of short, sporadic bouts of movement also needs to be undertaken. Finally, these results suggest the need for high fidelity, direct observational data, as well as standardization of accelerometry data-processing techniques.

Although this study was a quasiexperimental design with blocked random assignment of schools to the curriculum condition, some limitations do need to be considered. The quality of SOPLAY data is dependent on the quality of the trained observers. In this study, the baseline inter-rater agreement (79.4%) was at the lower bound of acceptability (80%), whereas inter-rater agreement for all other years was excellent (>90%). This may have influenced the baseline data estimates. Furthermore, self-report survey data may be affected by social desirability and recall bias. Response rates for the survey data were less than 50% on average and so it is possible that there was selection bias. For one intervention and one control school, the response rates were extremely low (6% and 5%, respectively); how this influenced the results (as it occurred in both groups) is unclear. In addition, the survey and accelerometer samples were taken from a random selection of eight schools, limiting the ability to generalize results to other populations. The use of two different accelerometers (switching accelerometer after the first study assessment) may be viewed as a significant limitation. However, our analysis of the calibration data revealed a strong correlation in device output (R = .86). Finally, due to the lack of effect, there were no data collected from teachers about whether they would implement the curriculum without outside support. This is a practical limitation for sustainability of such efforts, as we do not have information whether teachers would go through training and implement a curriculum like this during the lunch recess. Future efforts are recommended to implement recess curriculum using school staff for practicality, that students are more familiar with school staff and for sustainability reasons.

These limitations notwithstanding, several lessons were learned for design, practical, and methodological issues which may inform future translational research in this area. This is in addition to the lessons of capitalizing on the novelty effect of playground renovations, integrating school playgrounds into the curriculum, and maximizing the short recess periods mentioned above. Due to the timing of our assessments, we did not collect data while SPARK AR was being implemented, but rather after the fact. A true measure of the SPARK AR effect should assess outcomes while the curriculum is being implemented. Further when principals changed, we assumed the school was still on board and more attention should have been focused on welcoming the new principal and bringing them up to speed. Finally, it is recommended to ensure adequate sample size to investigate the intervention effect on at-risk participants who have the most to benefit from these types of interventions.

Compliance with Ethical Standards

Primary Data: The authors have full control of all primary data and agree to allow the journal to review their data if requested. This manuscript represents results of original work that have not been published elsewhere (except as an abstract in conference proceedings).

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Approval: This article does not contain any studies with animals performed by any of the authors. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the institutional review board.

Informed Consent: Informed consent was obtained from all individual participants included in the study.

Acknowledgments: The authors would like to acknowledge the 24 schools that participated and the contributions of Alana D. Steffen, Sarah Lampe, Betsy Brunner, Lisamarie Bensman, Erin Fiene, Melanie Mainar, and Maura O'Neal. The authors acknowledge the input of Myra Smith in revising parts of the manuscript. This study was funded by the National Institute of Child Health and Human Development NICHD/NCI/NIDDK R01HD057229 and in part by the National Institute on Minority Health and Health Disparities NIMHD U54MD007584, G12MD007601, and the National Institute of General Medical Sciences NIGMS P20GM103466.

References

1. U.S. Department of Health and Human Services. *Healthy People 2020 objective target areas and page numbers.* Available at http://www.healthypeople.gov/2020/topicsobjectives2020/pdfs/HP2020objectives.pdf. Accessibility verified December 25, 2013.
2. Centers for Disease Control and Prevention. *How much physical activity do children need?* 2011. Available at http://www.cdc.gov/physicalactivity/everyone/guidelines/children.html. Accessibility verified December 25, 2013.
3. U.S. Department of Health and Human Services. *Physical activity guidelines for Americans.* Washington, DC: U.S. Department of Health and Human Services; 2008. Available at http://www.health.gov/paguidelines/. Accessibility verified December 25, 2013.
4. National Survey of Children's Health. *NSCH 2011/12. Data query from the Child and Adolescent Health Measurement Initiative, Data Resource Center for Child and Adolescent Health website.* Child & Adolescent Health Measurement Initiative; 2012. Available at http://childhealthdata.org/browse/survey/results?q=2465&r=1&g=453. Accessibility verified November 25, 2013.
5. McKenzie TL, Sallis JF, Armstrong CA. Association between direct observation and accelerometer measures of children's physical activity during physical education and recess. *Med Sci Sports Exerc.* 1994;26(5 suppl):S143.
6. Langille JL, Rodgers WM. Exploring the influence of a social ecological model on school-based physical activity. *Health Educ Behav.* 2010;37(6):879–894.
7. Umstattd Meyer MR, Sharkey JR, Patterson MS, Dean WR. Understanding contextual barriers, supports, and opportunities for physical activity among Mexican-origin children in Texas border colonias: a descriptive study. *bmc Public Health.* 2013;13:14.
8. Nettlefold L, McKay HA, Warburton DE, McGuire KA, Bredin SS, Naylor PJ. The challenge of low physical activity during the school day: at recess, lunch and in physical education. *Br J Sports Med.* 2011;45(10):813–819.
9. Slater SJ, Nicholson L, Chriqui J, Turner L, Chaloupka F. The impact of state laws and district policies on physical education and recess practices in a nationally representative sample of US public elementary schools. *Arch Pediatr Adolesc Med.* 2012;166(4):311–316.
10. Nigg C, Geller K, Adams P, Hamada M, Hwang P, Chung R. Successful dissemination of Fun 5—a physical activity and nutrition program for children. *Transl Behav Med.* 2012;2(3):276–285.
11. Iversen CS, Nigg C, Titchenal CA. The impact of an elementary after-school nutrition and physical activity program on children's fruit and vegetable intake, physical activity, and body mass index: fun 5. *Hawaii Med j.* 2011;70(7 suppl 1):37–41.
12. Escalante Y, García-Hermoso A, Backx K, Saavedra JM. Playground designs to increase physical activity levels during school recess: a systematic review. *Health Educ Behav.* 2014;41(2):138–144.
13. Fjørtoft I, Löfman O, Halvorsen Thorén K. Schoolyard physical activity in 14-year-old adolescents assessed by mobile GPS and heart rate monitoring analysed by GIS. *Scand J Public Health.* 2010;38(5 suppl):28–37.
14. Anthamatten P, Brink L, Kingston B, Kutchman E, Lampe S, Nigg C. An assessment of schoolyard features and behavior patterns in children's utilization and physical activity. *j Phys Act Health.* 2014;11(3):564–573.
15. Anthamatten P, Fiene E, Kutchman E, et al. A microgeographic analysis of physical activity behavior within elementary school grounds. *Am j Health Promot.* 2014;28(6):403–412.
16. Stratton G. Promoting children's physical activity in primary school: an intervention study using playground markings. *Ergonomics.* 2000;43(10):1538–1546.
17. Verstraete SJ, Cardon GM, De Clercq DL, De Bourdeaudhuij IM. Increasing children's physical activity levels during recess periods in elementary schools: the effects of providing game equipment. *Eur J Public Health.* 2006;16(4):415–419.
18. Zask A, van Beurden E, Barnett L, Brooks LO, Dietrich UC. Active school playgrounds-myth or reality? Results of the "move it groove it" project. *Prev Med.* 2001;33(5):402–408.
19. Parrish AM, Okely AD, Stanley RM, Ridgers ND. The effect of school recess interventions on physical activity: a systematic review. *Sports Med.* 2013;43(4):287–299.
20. Ridgers ND, Salmon J, Parrish AM, Stanley RM, Okely AD. Physical activity during school recess: a systematic review. *Am J Prev Med.* 2012;43(3):320–328.
21. Ickes MJ, Erwin H, Beighle A. Systematic review of recess interventions to increase physical activity. *j Phys Act Health.* 2013;10(6):910–926.
22. McKenzie TL, Sallis JF, Rosengard P. Beyond the stucco tower: design, development, and dissemination of the SPARK physical education programs. *Quest.* 2009;61(1):114–127.
23. Nigg CR, Steffen AD, Amato K, et al. Does perceived environment matter for children in the context of social cognitive variables? *Med Sci Sport Exerc.* 2012;44:s90.
24. Brink LA, Nigg CR, Lampe SM, Kingston BA, Mootz AL, van Vliet W. Influence of schoolyard renovations on children's physical activity: the Learning Landscapes Program. *Am J Public Health.* 2010;100(9):1672–1678.
25. Rowe PJ, Schuldheisz JM, van der Mars H. Measuring physical activity in physical education: validation of the SOFIT direct observation instrument for use with first to eighth grade students. *Pediatr Exerc Sci.* 1997;9(2):136–149.
26. McKenzie TL, Sallis JF, Nader PR, et al. BEACHES: an observational system for assessing children's eating and physical activity behaviors and associated events. *jJAppl Behav Anal.* 1991;24(1):141–151.
27. Beck J, Schaefer CA, Nace H, et al. Accuracy of self-reported height and weight in children aged 6 to 11 years. *Prev Chronic Dis.* 2012;9:E119.
28. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. *Can J Appl Sport Sci.* 1985;10(3):141–146.
29. Jacobs DR Jr, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Med Sci Sports Exerc.* 1993;25(1):81–91.
30. Lee RE, Nigg CR, DiClemente CC, Courneya KS. Validating motivational readiness for exercise behavior with adolescents. *Res Q Exerc Sport.* 2001;72(4):401–410.
31. Haas S, Nigg CR. Construct validation of the stages of change with strenuous, moderate, and mild physical activity and sedentary behaviour among children. *j Sci Med Sport.* 2009;12(5):586–591.
32. Sallis JF, Nader PR, Broyles SL, et al. Correlates of physical activity at home in Mexican-American and Anglo-American preschool children. *Health Psychol.* 1993;12(5):390–398.
33. Heil DP. Predicting activity energy expenditure using the Actical activity monitor. *Res q Exerc Sport.* 2006;77(1):64–80.
34. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. *Med Sci Sports Exerc.* 2011;43(6):1085–1093.
35. Schaefer CA, Nace H, Browning R. Establishing wrist-based cutpoints for the Actical accelerometer in elementary school-aged children. *J Phys Act Health.* 2014;11(3):604–613.
36. Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O. *SAS for Mixed Models.* 2nd ed. Cary, NC: SAS Institute Inc; 2006:17–56.
37. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics.* 1997;53(3):983–997.
38. Nigg CR, Battista J, Chang JA, Yamashita M, Chung RS. Physical activity outcomes of a pilot intervention using SPARK Active Recreation in elementary after school programs. *J Sport Exerc Psychol.* 2004;26:S144–S145.
39. Metcalf B, Henley W, Wilkin T. Effectiveness of intervention on physical activity of children: systematic review and meta-analysis of controlled trials with objectively measured outcomes (EarlyBird 54). *BMJ.* 2012;345:e5888.
40. Herrick H, Thompson H, Kinder J, Madsen KA. Use of SPARK to promote after-school physical activity. *J Sch Health.* 2012;82(10):457–461.
41. Saint-Maurice PF, Welk GJ, Russell DW, Huberty J. Moderating influences of baseline activity levels in school physical activity programming for children: the Ready for Recess project. *bmc Public Health.* 2014;14:103.
42. Ridgers ND, Stratton G, Fairclough SJ. Assessing physical activity during recess using accelerometry. *Prev Med.* 2005;41(1):102–107.
43. Nader PR, Bradley RH, Houts RM, McRitchie SL, O'Brien M. Moderate-to-vigorous physical activity from ages 9 to 15 years. *JAMA.* 2008;300(3):295–305.

44. Baquet G, Stratton G, Van Praagh E, Berthoin S. Improving physical activity assessment in prepubertal children with high-frequency accelerometry monitoring: a methodological issue. *Prev Med.* 2007;44(2):143–147.
45. Sleap M, Warburton P. Physical activity levels of 5-11-year-old children in England as determined by continuous observation. *Res q Exerc Sport.* 1992;63(3):238–245.
46. Anthamatten P, Brink L, Lampe S, Greenwood E, Kingston B, Nigg C. An assessment of schoolyard renovation strategies to encourage children's physical activity. *Int J Behav Nutr Phys Act.* 2011;8:27.
47. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008;40(1):181–188.

APPENDIX

Fig A1 View of playground at an elementary school before a learning landscape (LL) was installed in 2000 and after in 2003.