



Research article

Development and piloting of the Movement Observations during Play (MOP) measure

Nila Joshi^{1,*}, Son Truong¹, Janet Loebach², Daniel Rainham^{1,3}, Becky Feicht¹ and Michelle R. Stone¹

¹ School of Health and Human Performance, Dalhousie University, Nova Scotia, Canada

² Cornell College of Human Ecology, Cornell University, New York, United States of America

³ Centre for Planetary Health and Sustainable Care, Nova Scotia, Canada

* **Correspondence:** Email: Nila.Joshi@dal.ca; Tel: +9024941167.

Abstract: *Background:* Fundamental movement skills (FMS) underpin children’s physical activity and healthy development, yet few observational tools capture the specific movements children demonstrate during child-led outdoor play, including outdoor loose parts play (OLPP). *Purpose:* Our purpose of this study was to develop and pilot Movement Observations during Play (MOP), an observational measure designed to document school-aged children’s FMS during play and examine its preliminary inter-rater reliability (IRR). *Methods:* MOP was developed through a review of relevant literature and adaptation of the Observational System for Recording Physical Activity in Children-Environmental (OSRAC-E). OSRAC-E movement categories were expanded to include 14 additional distinct movements, and 15 FMS were added based on the literature, stakeholder consultation, and iterative reliability rounds. MOP was then used in field-based observations across four outdoor play sites in Nova Scotia, Canada with children between the ages of 3–12 years. Two observers independently recorded children’s movements occurring within 15-second play events, and IRR was assessed across sites. *Results:* MOP captured 46 unique FMS across locomotor, object control, and stability categories. During reliability rounds, 40 of 46 skills (86.9%) were observed, suggesting the measure captures a broad range of movements expressed during outdoor play. Agreement for recording at least one of the same movements within a play event was high across sites ($\kappa = 0.93$ – 0.98 ; overall $\kappa = 0.96$). Observers also consistently recorded two of the same movements within a 15-second play event. Agreement for three or more matched movements was lower and more variable. Additional analyses showed strong consistency for commonly observed movements such as walking, standing, running, carrying, and holding, with

agreement improving across sites over time. *Conclusions:* MOP shows promise as an observational measure for documenting a wide range of FMS during child-led play, including OLPP, and for examining how children express movement in these play contexts.

Keywords: children; fundamental movement skills; loose parts; outdoor play; observations

Abbreviations: FMS: fundamental movement skills; PA: physical activity; OLPP: outdoor loose parts play; BM: behavioural mapping; PLEY: physical literacy in early years; MOP: movement observations during play

1. Introduction

1.1. Fundamental movement skills (FMS) in the context of outdoor play

Outdoor play offers endless opportunities for movement, helping children develop fundamental movement skills (FMS), which are associated with increasing physical activity (PA) and supporting overall physical health in childhood [1–3]. FMS are the foundation for more advanced movements required for PA, sports, games, and exploration [4]. Drawing from the work of Goodway and colleagues [4], FMS include locomotor skills, object control/ball/manipulative skills, and balance/stability skills. Locomotor skills are defined as movements where the body travels through space between two distinct points and include movements such as walking, running, and jumping. Object control skills involve movements where the body interacts with objects, manipulating them and using their movements to guide the object between two spatial points. Balance/stability skills are centered around limb-based motions and include movements such as balancing, stretching, and bending. FMS develops through maturation, instruction, habitual activities such as walking, and physical activities like outdoor play and sports [4]. These types of movements are important for fostering PA participation.

1.2. Association between outdoor loose parts play (OLPP) and children's FMS

Adding loose parts (LP) to children's play environments is hypothesized to enhance the diversity of play, and foster inventiveness, creativity, and discovery [5]. LP are characterized as “natural or manufactured materials with no specific set of directions that can be used alone or combined with other materials, moved, carried, combined, redesigned, lined up, and taken apart and put back together in multiple ways and used for play” [6]. In the context of play, LP can offer a wide range of opportunities, enabling children to interact with materials in diverse and imaginative ways that suit their interests, needs, preferences, or motivations [5]. Children can perceive and use LP in ways that suit them and their play needs, contributing to diverse use of FMS.

Several studies, predominantly comprising quasi-experimental and observational research, have explored the value of outdoor loose parts play (OLPP) to children's health and development, particularly their movement skill development [5,7–13]. Findings demonstrate that children move in diverse ways and use a combination of movements when engaging in OLPP, leading to the development of more intricate movement patterns [7]. In a study examining the impact of an OLPP intervention on

preschoolers' physical literacy (Physical Literacy in the Early Years (PLEY) project) [14], the repetition of movements enabled children to practice FMS, learn from mistakes, and build confidence, which motivated them to engage in more diverse movements [7], supporting their physical literacy [8]. While the findings demonstrate the value of OLPP in supporting children's FMS development and broader concepts like physical literacy, they are based on educators' perceptions of children's movements during OLPP. The authors note how typical quantitative assessments of children's FMS (e.g., Test of Gross Motor Development-3) [15] may not be appropriate for examining the role of OLPP in children's FMS development, as they do not include many of the movements that children engage in during this type of play (e.g., pulling, pushing, carrying, and lifting) [8].

1.3. Observational approaches to understanding outdoor play

Observational methods have been used to provide valuable insights into outdoor play [5,9,10,16–18] and have the potential to enhance our understanding of FMS, especially during OLPP. One approach to systematic observation is behavioral mapping (BM), which involves direct observation and recording of behaviors occurring in a particular setting at a specific time [19]. BM enables researchers to gather a comprehensive record of behaviors exhibited in the space being used and identify patterns and trends of particular behaviors [20]. Measures such as the Tool for Observing Play Outdoors (TOPO) [21] have been used with the BM approach, and have enabled researchers to collect quantitative behavioural and environmental data, enhancing our understanding of the dynamics and patterns of play in these settings.

The TOPO tool is an observational tool developed to capture children's play behaviors in outdoor settings [21]. The TOPO tool has well defined, distinct outdoor play typologies, and physical play is a focus of the measure [21]. The significance of examining physical play lies in the extent of PA involved and the variety of movements children engage in. This category of play behavior includes four subtypes of physical play: fine motor, gross motor, vestibular, and rough and tumble play. These subtypes capture and examine physical play in a more nuanced manner, specifically how children use large muscle groups (e.g., gross motor) and smaller muscle groups (e.g., fine motor), along with their vestibular systems (e.g., balancing on one foot or climbing), during play activities.

While the TOPO tool is valuable for understanding the contribution of OLPP to broader movements, it does not provide a high level of specificity of FMS that children demonstrate during OLPP. For example, as outlined in Loebach and Cox's paper [21], gross motor skills include activities such as climbing, running, throwing, catching, lifting, or carrying. The TOPO's play typology would not be able to identify these specific FMS, nor the FMS category they fell into (e.g., locomotor and object control); rather, FMS would be amalgamated into the category of gross motor skills. Identifying specific FMS and categories of FMS that children engage in during OLPP ensures a more profound understanding and appreciation of OLPP's contribution to children's FMS development, including the movement pattern complexity that OLPP appears to foster.

Our purpose of this paper is to describe the development process of Movement Observations during Play (MOP), an observational measure that can be used to document the FMS that school-aged children engage in during OLPP. We also report preliminary field-based inter-rater reliability for MOP using a BM protocol and explore how piloting MOP within BM can inform refinements to future observational procedures.

2. Methods

In the following section the development process of an observational measure of children's FMS during outdoor play and outline the procedure for piloting it in the field. In Phase 1, we outline the process of identifying observational tools for recording children's FMS within play contexts through a literature review. In Phase 2, we provide an overview of how FMS were identified for the observational measure, using literature and data reviews. In Phase 3, we describe the selected sites for piloting the observational measure, the designated observers, and the preliminary reliability rounds conducted to assess initial inter-rater reliability.

2.1. Phase one: Literature review

A literature review of observational tools used to measure children's FMS was conducted to develop an observational measure tailored to observing the specific movements (FMS) children engage in during outdoor play, including OLPP. Consultations were conducted with a university librarian (SM), and medical subject heading (MeSH) terms were identified to assist with the literature review. MeSH terms are keywords that help identify articles relevant to a particular research area. MeSH terms such as "fundamental movement skills", "outdoor play", and "motor skills" were used when conducting this review. Tools that assess children's FMS through observation were used to inform which movements are typically included in these tools. This served as a reference point to ensure that the observational measure captured a comprehensive range of FMS. This review also helped determine which tools are most appropriate for use in outdoor play and OLPP contexts, and how they can be modified to better reflect the unique characteristics of these contexts.

2.2. Phase two: Establishing FMS in outdoor play

In Phase two, we focused on establishing the specific FMS typically seen during outdoor play, including during OLPP, and integrating them within the adapted observational measure. Building on the initial literature search, an additional literature review was conducted, specifically focused on FMS exhibited by children during outdoor play, with a particular emphasis on movements associated with OLPP. To guide the identification and classification of movements, we adopted the FMS framework defined by Goodway and colleagues [4]. According to this framework, FMS are categorized into three domains: locomotor skills (e.g., running, jumping, skipping), object control skills (e.g., throwing, pulling, kicking), and stability skills (e.g., balancing, squatting, kneeling) [4]. These domains and definitions provided consistency throughout the review and were used to inform the selection of relevant movements and the structure of the adapted observational measure. The selected FMS framework reflects a widely accepted categorization of movement, making it an appropriate way to understand FMS in children. Additionally, this framework includes stability skills as a distinct category alongside locomotor and object control skills, enabling a more comprehensive understanding of the full range of movements children use during OLPP.

The review was conducted between January 2021-September 2021 using the following databases: PubMed, Google Scholar, EBSCOhost, and CINAHL. Search terms included variations and combinations of the following keywords: fundamental movement skills, outdoor play, outdoor loose parts play, physical literacy, movement, locomotor, object control, stability, gross motor, and fine

motor. These terms were selected to capture broad and specific conceptualizations of FMS in the context of outdoor play and OLPP. Various approaches used in previous studies to examine OLPP and describe children's movements were identified and used to inform the development of the MOP by helping to identify and categorize the types of movements children typically exhibit during OLPP [9–12,18].

In addition to the literature review, qualitative focus group data collected from the PLEY project [7,8,12,13], which centered on early childhood educators' perceptions of OLPP, were examined. Within focus groups, educators discussed the OLPP intervention and its influence on children's physical development, specifically their movements, including FMS used within play activities. A content-focused review of the transcripts was conducted by the author, with attention to educators' descriptions of children's movement patterns, including specific FMS used during OLPP. The author systematically extracted and documented references to different types of movements children were observed using during OLPP. To complement this, videos captured during a PLEY knowledge dissemination event were also reviewed by the author [22]. These videos were analyzed through close observational review, with the goal of identifying additional movements exhibited by children that may not have been mentioned in the focus groups. This analysis provided a list of FMS that were commonly observed during OLPP.

2.2.1. Development process and evidence supporting content of the Movement Observations during Play (MOP) measure

Throughout the development of MOP, several steps were taken to strengthen the measure's content and feasibility for field use. These steps were adapted from Loebach and Cox (2020), which described the development of the TOPO observational tool since both tools examine outdoor play behaviors [21]. In developing MOP, the authors consulted with Drs. Loebach and Cox to gain deeper insight into their approach, particularly during the early stages of tool development. This consultation helped ensure that the MOP development process aligned with the methodology used in creating TOPO. The development approach primarily focused on incorporating two key strategies: Synthesizing evidence through triangulation and peer review or debriefing of the data and research process.

Triangulation, a method of cross-verifying findings using multiple methods, data sources, or perspectives, helps strengthen confidence in the content and its definitions [23]. Triangulation was used through various means to develop this observational measure. This included conducting literature reviews to identify the appropriate tool for adaptation and to identify additional FMS observed during OLPP. Additionally, it involved cross-referencing the PLEY data and gathering input from outdoor play stakeholders (researchers, educators, community partners) on MOP.

Peer review and debriefing of the data and research process were conducted to improve clarity of definitions and coding rules of MOP. Throughout the development stages, ongoing discussions with outdoor play stakeholders were maintained, fostering a collaborative environment for sharing perspectives and collective expertise. This iterative process facilitated the identification of any missing FMS and enables the inclusion of additional descriptions to enhance the definition of FMS. Together, these steps strengthened the overall quality of MOP.

2.3. Phase three: Piloting the usability of the Movement Observations during Play (MOP) measure

The final phase focused on piloting MOP and collecting preliminary inter-rater reliability data of the observational measure in outdoor environments to ensure it could adequately capture the FMS that children engage in during outdoor play, including OLPP. Reliability testing was limited to an initial assessment of inter-rater reliability, examining agreement between two independent observers during real-time field observations. Intra-rater reliability, which examines the consistency of the same observer coding the same events at different points in time (e.g., using video recordings) was not assessed.

2.3.1. Observational approach: Behavioral Mapping (BM)

We used a BM approach with the observational measure. BM is an observational approach where behaviors observed can be recorded at a particular place during a particular time [19]. BM observations can be place-based or person-based, depending on the research question [20]. Place-based observations focus on the environment, examining how different settings influence behavior [20]. In contrast, person-based observations concentrate on individual behaviors and movements, tracking how specific people engage with their surroundings [20]. We utilized a place-based approach to pilot the use of MOP and how it can be used in different outdoor settings with diverse environmental and playground features. Members of the research team received BM training from Dr. Janet Loebach and Dr. Adina Cox, who provided guidance on the use of BM approaches in outdoor play settings. This training informed the team's use of place-based observation, scan sampling, and basemap-based recording within this study.

2.3.2. Sampling structure and unit of analysis

Observations were collected using a 15-second scan of children's play activities. The unit of analysis for inter rater reliability testing was the play event (i.e., each 15-second scan), which was independently coded by two observers in real time. Observations were event-based rather than participant-based, meaning scans captured whichever child(ren) were engaged in play within the observers' field of view at the time of the scan. Because scanning was event-based, it is possible that the same child was observed across scans within a session or across sessions. Child identifiers were not recorded during scanning (particularly at the public park site), which prevented confirming the total number of unique children represented in the full dataset and quantifying the extent of repeated observations.

2.3.3. Observer training and practice

Before preliminary inter-rater reliability testing, observers completed training and practice sessions to support the consistent use of MOP in real-time outdoor play settings. This phase focused on testing and refining the proposed observational measure prior to field-based reliability rounds. To enhance the observation and coding of children's movements (FMS), the research team engaged in coding practice using outdoor play videos provided by Dr. Loebach and Dr. Cox as part of the BM training module.

During the training sessions, the lead author presented 15-second videos of children playing with LP materials. Each video was viewed twice. After viewing each video, coders independently

completed the Google Form by inputting the movements they observed. Once all coders submitted their responses, the lead author displayed the results to compare coding decisions across observers. When codes did not match, coders were asked to explain their coding decision. These discussions were used to clarify movement definitions, refine coding decision rules, and determine whether additional movements needed to be added to the measure. No study data were collected during these practice sessions; rather, the purpose was to familiarize coders with MOP and refine the movement definitions prior to reliability testing.

Online coding practice continued until general consistency was observed among coders.

Following these practice sessions, coders transitioned to using the tablets that would be used during field observations. These sessions enabled coders to become familiar with using the tablets, applying the BM protocol in real time, and identifying the location of observed behaviors on the site basemaps. Once coders were comfortable using the tablets and identifying behaviors on the basemaps, the team proceeded with the preliminary reliability rounds.

2.3.4. Preliminary reliability rounds

Preliminary inter-rater reliability of MOP was assessed through multiple reliability rounds at each of the four study sites, each consisting of three one-hour sessions conducted over three consecutive days. Reliability rounds occurred sequentially, starting with Site A, followed by Sites B, C, and D. In each reliability round, two (2) observers watched the same child in the outdoor play space for a span of 15-second and then independently used MOP to code the FMS (up to five separate movements) demonstrated during the play event. Both observers then collectively moved through the observation period by simultaneously observing one child for 15 seconds and, immediately afterward, taking up to 1 minute to code the FMS observed during that child's play event. Once coding was complete, they moved on to another child and repeated this process over the full 1-hour observation period. A one-hour reliability round was conducted per day for three days at each study site. The lead researchers served as an observer in all sessions, ensuring consistency and continuity across all rounds. The second observer remained consistent across sites with the exception of two instances where they were unavailable; in these cases, another trained member of the research team substituted to ensure the reliability round could proceed as planned. In total, 12 hours of reliability rounds were conducted. A detailed description of the observational measure can be found in Appendix A. Data collection for this project was approved by Dalhousie University's Research Ethics Board (REB #2021-5783).

2.4. Site descriptions

In total, four sites were selected to conduct the reliability rounds. These sites were chosen based on their participation in the overarching study this paper is part of, and their suitability for piloting and evaluating the observational measure in diverse outdoor play environments. While all sites supported unstructured outdoor play for children aged 3 to 12, they varied in key characteristics such as setting (urban vs. suburban), type (public park, school, or childcare center), and the design of the outdoor space (e.g., inclusion of fixed equipment, natural elements, or loose parts). This range enabled the observational measure to be piloted across outdoor play contexts.

2.4.1. Site A: Urban park

The first site where the observers conducted their reliability rounds was in a large urban green space in the city of Halifax, Nova Scotia, Canada. This also included a fixed playground structure, an outdoor pool, a baseball field, and walking paths. Children who were visiting the park during the time of the observations were observed for the reliability rounds.

2.4.2. Site B: Urban private school

Site B was a private school in Halifax, Nova Scotia, with students from Pre-primary to Grade 12. The school includes two distinct outdoor play spaces. The first space is dedicated to pre-primary and primary school children (ages 3 to 5 years) and features a fixed playground structure placed on a gravel surface. This play area also includes trees and bushes, providing natural elements within the environment. The second play space is newly renovated and incorporates natural elements such as trees, bushes, a large pavement area, and a human-made hill. A fence encloses the play space. This play space serves children from Grades 1–6 (ages 6 to 12 years).

2.4.3. Site C: Suburban childcare center

This site was the outdoor play yard of an after-school program which serves students from pre-primary to grade 5 (ages 4 to 11 years). The site has a fixed playground, a large green space, a soccer field and a baseball field. The children engage in diverse outdoor experiences, exploring the woods, participating in gardening activities, and using LP. Additionally, the site includes a spacious open area featuring trees, bushes, and rocks, primarily used by school-aged children (ages 5–11 years).

2.4.4. Site D: Urban childcare center

Site D was the outdoor play yard of an after-school program serving children between the ages of 5–12 years and affiliated with the Boys and Girls Club, a prominent charitable organization dedicated to providing a safe and supportive environment for children and youth. The site has outdoor play areas, including a renovated basketball court where the children actively participate in structured activities such as basketball. Additionally, a large parking lot serves as a play area, and a small section of green space with garden boxes offers another space for children to play.

2.5. Data analysis

2.5.1. Analysis of inter rater reliability data

Data were organized and analyzed using SPSS Version 28 [24]. MOP the recording of up to five different movements within a single 15-second play observation. For recording purposes, these were entered into five movement fields (Movement 1 to Movement 5). These fields served only as slots for documenting multiple movements observed and were not intended to represent the sequence in which movements occurred. Because observations involved a 15-second scan, followed by a recording period (up to 1 minute), the order in which movements were documented was not treated as meaningful.

Observers may have recorded the same movements in different sequences depending on their recall of the play event. Therefore, during analysis, movements were matched across observers based solely on whether the same FMS was identified within the same 15-second observation scan, regardless of the order in which it was recorded. For example, if Observer 1 recorded “running” as Movement 1 and Observer 2 recorded “running” as Movement 4 during the same scan, this was treated as agreement for that movement. Each recorded movement within the corresponding 15-second scan was therefore coded as either “agree” or “not agree.” For instance, if Observer 1 recorded striking, pulling, running, lifting, and digging, and Observer 2 recorded walking, running, digging, lifting, and pulling, then 4 of the 5 movements were considered to agree. This approach was used to determine whether the same movement was identified during the same observational scan, rather than whether it was recorded in the same order.

Kappa statistics were used to assess the preliminary inter-rater reliability of MOP by examining the extent to which two observers identified the same FMS within the same 15-second observation. For each observation (scan), movements recorded by the two observers were matched to determine whether the same FMS had been identified, regardless of the movement field in which it was entered. These matched data were then used to calculate kappa coefficients, which estimate the level of agreement between observers beyond chance. A kappa value nearing 1 indicates higher agreement and greater reliability, whereas a value nearing 0 indicates lower agreement and less reliability [25]. Agreement was interpreted as follows: 0.00–0.20, no agreement; 0.21–0.39, minimal agreement; 0.40–0.59, weak agreement; 0.60–0.79, moderate agreement; 0.80–0.90, strong agreement; and above 0.90, almost perfect agreement [25].

2.5.2. Descriptive analyses of observer agreement

Additional analyses were conducted to support interpretation of the kappa findings. These included calculation of percent agreement between raters for commonly observed movements, as well as examination of the number of shared recordings of FMS, or no FMS, identified by both observers within each play event. These supplementary analyses were used to provide a more descriptive understanding of observer agreement across scans and to contextualize the kappa coefficients, particularly for movement categories that occurred more frequently.

3. Results

In this section, we present findings from the development and preliminary piloting of the MOP measure. It begins with results from the literature reviews that informed the measure’s design, followed by findings from the reliability rounds. These include the number of movements recorded using MOP, inter-rater reliability results, percent agreement among observers, and the number of shared FMS or instances of no FMS observed by both observers during each play event. Additionally, the most commonly observed movements are reported. Together, these findings provide insight into the initial piloting of MOP and its inter-rater reliability by illustrating how many movements were captured and offering guidance on applying the MOP observational protocol effectively in the field.

3.1. Development of the observational measure: Selection of a tool for modification

A total of nine observational tools were identified and considered in the development of MOP. These tools were assessed based on specific inclusion criteria: (1) The ability to observe and record FMS in the context of outdoor play, (2) inclusion of specific movements (e.g., pulling, jumping) rather than prompted or structured movements/tasks (e.g., sliding across a line), and (3) alignment with movements typically observed during outdoor play and OLPP that reflect recognized categories of FMS (locomotor, object control, and stability). To ensure that the observational measure was comprehensive, consultations were held with a research team member (HC) with expertise in children's physical literacy, PA, and OLPP. During these consultations, they reflected on the identified tools, their recording methods for FMS, and how these methods fit the context of outdoor play and OLPP. After several consultations, the Observational System for Recording Physical Activity-Elementary School (OSRAC-E) tool [26] was selected for its specificity in recording movement types during outdoor play. Developed as an extension of the Observational System for Recording Physical Activity-Preschool (OSRAC-P) [27], OSRAC-E collects PA data in elementary school environments. OSRAC-P, in turn, was created by integrating and revising two observational systems: the Children's Activity Rating Scale (CARS) and CASPER-II [27,28]. OSRAC-P records a child's activity level (e.g., fast, moderate, and stationary) and type of activity (e.g., running, sitting, walking, and riding) in preschool settings while capturing social factors (i.e., initiator of activity, group composition, and prompts for PA) and nonsocial environmental circumstances (i.e., child location and indoor and outdoor contexts) associated with the activity. Although there are several components to the OSRAC-P and the OSRAC-E, the only component selected and adapted from the tool for the development of MOP was activity type.

The OSRAC-E tool originally included 17 movement categories [26]. Upon reviewing the detailed definitions provided for each of these movements, the researcher identified additional distinct movements embedded within these definitions. For example, movements such as kneeling, squatting, and sitting were extracted from broader categories of sit/squat and treated as individual movement types. Fourteen of these movements were extracted and categorized as individual movement types, and an additional 15 (step up, step down, sliding, rolling a ball or object, striking, dribbling with hands, dribbling with feet, lifting, holding, carrying, digging, stirring, shaking, balancing, and shaking the body) were included from the literature review, consultations with stakeholders, and reliability rounds enabling a more detailed and comprehensive observational measure. The MOP now includes 46 movement types that are organized into the three categories of FMS as outlined by Goodway and colleagues [4]: Locomotor, object control, and stability, resulting in 18 movements in the locomotor category, 15 movements in the object control category, and 13 movements in the stability category. The movements added to MOP were primarily within the object control category. These additions included specific actions related to interacting with LP, such as digging or holding objects. A few movements were added to the stability and locomotor categories (balancing, shaking (body), step up and step down). The final measure, MOP, can be found in Appendix A.

MOP was developed for use in place-based, outdoor settings to capture children's natural movement during unstructured play. MOP is designed to capture up to five FMS during each observation of children's outdoor play, including OLPP. The decision to allow the recording of up to five different movements provides flexibility, considering the dynamic nature of outdoor play and the potential for multiple movements to occur during a single observation (scan). The decision was made

to use a 15-second scan during the outdoor play activities in alignment with the TOPO protocol [21]. In practice rounds, no more than five FMS were observed in a 15-second scan. Notably, there is no hierarchical order (e.g., the time and sequence of the observed FMS), which provides flexibility to the observer; the FMS observed during a 15-second scan do not need to be recorded in the sequence in which they occur. The observer simply records all FMS that they observed during the child's 15-second of play. The decision not to include a hierarchical order in the measure's design is to enable observers to simply watch the child's play unfold and then take time afterward to record those FMS they saw the child using within their play, rather than trying to record the sequence of FMS that are being used in real-time (which could compromise the quality of the observation).

3.2. Pilot testing of MOP

3.2.1. Overview of movements observed

A total of 40/46 (86.9%) of the FMS included within MOP were recorded during the reliability rounds. The most common FMS observed across all sites were walking, carrying, and holding. Observers were extremely consistent in the number of FMS recorded during each 15-second scan (difference between observers <1%); 87% of observations had a second movement recorded, 50% had a third movement recorded, and 17% had four movements recorded; only 5% of observations included five movements. The option of "Other" was available to observers when a movement was observed that was not included in MOP. During the reliability rounds, the "Other" category was selected only in two scans.

3.2.2. Descriptive data by site

Table 1 presents the descriptive data of the observations collected during the reliability rounds. While age and gender were not analyzed as part of the inter-rater reliability analysis, they were included to provide contextual information about the observed sample. In total, 250 play events were observed. The number of play events involving males and females was relatively similar at Sites A and B. However, at Sites C and D, males were observed in a significantly higher number of play events. A chi-square test was conducted to examine the distribution of gender and grade level across the four sites. Results revealed significant differences in gender and grade distributions across sites ($\chi^2(9) = 438.71$, $p < 0.001$), indicating that the representation of gender and grade levels varied by location. Reliability rounds were conducted sequentially, beginning with Site A in April 2022, followed by Site B (April 2022), Site C (May 2022), and Site D (May 2022).

Table 1. Descriptive data by site.

	Site A (April 2022)		Site B (April 2022)		Site C (May 2022)		Site D (May 2022)	
	# of play events	% of play events	# of play events	% of play events	# of play events	% of play events	# of play events	% of play events
Total play events	44		70		66		70	
Gender								
Male	22	50.0	36	51.4	39	59.1	55	78.6
Female	21	47.7	34	48.5	27	40.9	15	21.4
Null/Incomplete data	1	2.3	0	0	0	0	0	0
Grade								
Pre-primary-Primary	n/a	n/a	66	94.3	9	13.6	9	12.9
Grades 1–3	n/a	n/a	4	5.7	42	63.6	61	87.1
Grades 4–6	n/a	n/a	0	0	15	22.7	0	0

Note: N/A indicates data not available.

3.2.3. Inter-rater reliability of MOP

Table 2 provides an overview of the preliminary inter rater reliability of the observational measure, specifically, the initial agreement among observers in capturing up to five FMS during each 15-second observation of children’s outdoor play behavior. The movement entry fields (Recorded 1 of the same movements–Recorded 5 of the same movements) represent the recording slots used to document multiple movements observed during a single scan and do not reflect the order in which movements occurred. Results are presented for each site (Sites A–D) and for the total inter rater reliability (IRR). Total reliability round observations ranged from a low of 44 (Site A) to a high of 70 (Site B).

Table 2. Inter-rater reliability (Cohen’s kappa statistic) by site and number of movements recorded.

	Recorded 1 of the same movements	Recorded 2 of the same movements	Recorded 3 of the same movements	Recorded 4 of the same movements	Recorded 5 of the same movements
Site A	0.95	0.62	0.31	0.13	/
Site B	0.98	0.57	0.28	0.09	0.15
Site C	0.93	0.90	0.43	0.40	0.19
Site D	0.96	0.97	0.98	0.79	0.76
Total IRR	0.96	0.78	0.52	0.37	0.49

Note: Using Kappa statistics, the results show the level of agreement between the two observers. A value between 0 and 0.2 signifies no agreement, 0.21 to 0.39 signifies minimal agreement, 0.40 to 0.59 signifies weak agreement, 0.60 to 0.79 signifies moderate agreement, 0.80 to 0.90 signifies strong agreement, and above 0.90 signifies almost perfect agreement [29].

The preliminary reliability results indicate a moderate to high level of agreement among observers in capturing one or two of the same FMS within a short (15-second) play observation (Total IRR = 0.96, 0.78, respectively), and declining levels of agreement thereafter (i.e., Recorded 3–5 of the same movements; total IRR = 0.52, 0.37, 0.49, respectively). While the inter-rater reliability of the measure varied across the four sites (reliability rounds were conducted first in Site A, followed by Sites B, C, and D), the level of agreement among observers generally increased over time (that is, IRR agreement was generally higher at the last two sites observed).

To complement the Kappa statistics shown in Table 2, Table 3 presents the percent agreement for the most frequently observed movements. While Cohen’s kappa accounts for agreement expected by

chance, percent agreement provides a descriptive indication of how often observers coded the same movement during observations [25]. Together, these metrics offer a more complete understanding of observer consistency [25]. Agreement was high across commonly coded movements, including walking (84.9%), standing (83.3%), holding (80.1%), running (95.6%), and carrying (93.2%). Agreement was also strong for balancing (91.6%) and very high for sitting (98.4%), squatting (96.8%), and jumping (99.2%).

Table 3. Percent agreement between raters for commonly observed movements.

Movements	Frequency Observed Among Raters	Percent Agreement
Walking	41.8%–45.8%	84.9%
Standing	30.3%–36.7%	83.3%
Holding	20.4%–23.5%	80.1%
Running	21.9%–22.3%	95.6%
Carrying	15.9%–17.9%	93.2%
Balancing	12.7%–14.7%	91.6%
Sitting	12.7%–13.5%	98.4%
Squatting	9.6%	96.8%
Jumping	9.6%	99.2%

To further describe agreement between observers, we examined the number of shared MOP movement entries recorded within each 15-second play event across the four reliability sites. During each play event, observers could record up to five movements; therefore, agreement was summarized based on how many of the five possible movement entries were shared between observers. A shared entry was defined as either both observers recording the same movement within the same play event, regardless of the order in which it was entered, or both observers indicating that no additional movement was observed. Therefore, a score of zero shared entries indicates that the observers did not record any of the same movements, or shared absence of movement, within that play event. Overall, agreement increased across the reliability rounds conducted from April to May 2022. At Site A, observers recorded three to four shared entries in 70.5% of play events, increasing to 74.6% at Site B. Agreement improved further at Sites C and D, with observers recording four or more shared entries in 87.7% of play events at Site C and up to five shared entries in 87.1% of play events at Site D. Full agreement across all five movement entries also increased over time, from 11 of 44 play events at Site A to 61 of 70 play events at Site D. These findings suggest that observer agreement strengthened across the reliability rounds, with observers becoming increasingly consistent in identifying the same movements, as well as agreeing when no additional movement was observed during a play event.

4. Discussion

Our primary aim of this study was twofold. First, we sought to develop a preliminary observational measure, MOP, for recording various FMS as children engage in outdoor play, including OLPP, in response to the lack of observational tools designed for this context. Our second aim of the study was to pilot MOP within the field and assess the preliminary inter-rater reliability of the observational measure.

4.1. Development of MOP

The development of MOP involved several stages, including reviewing observational tools to identify those that could be adapted, while incorporating additional FMS identified through the literature review, stakeholder consultations, and data review processes. This resulted in the identification of 46 unique FMS across three movement categories. Although the OSRAC-E tool served as the foundation for MOP, the process of adapting and extending it was critical to the tool's contribution. Specifically, broader OSRAC-E movement categories were unpacked to identify 14 distinct embedded movements, and a further 15 movement types were incorporated based on the literature review, stakeholder consultations, and reliability rounds. These additions enabled MOP to provide a more detailed and contextually relevant way of observing movement during outdoor play and OLPP than would have been possible using the original broad OSRAC-E categories alone. This was important because it enabled MOP to capture a wider range of movement expressions that were likely to occur in child-led outdoor play, thereby strengthening its usefulness for research and practice.

4.2. Pilot testing of MOP

4.2.1. Most FMS within MOP were observed

In the reliability rounds, we found that 40 of 46 FMS from MOP were observed, indicating that the incorporated movements were relevant and adequately captured the range of children's FMS. This suggests that our literature and data review processes were thorough and effective in identifying the types of movements children engage in during outdoor play. While outdoor play is highly dynamic and influenced by environmental factors that can afford a wide variety of movements, our results of this study suggest that MOP was well-suited to capture the FMS most commonly demonstrated within our specific outdoor play contexts. Although other movements, such as pivoting, twisting, or dodging, have been noted in different studies and environments [30], the high use of the FMS identified in MOP, along with minimal reliance on the "Other" category, indicates that the observational measure effectively reflected the movements observed during outdoor play within our contexts. These results help validate the process we used to develop a comprehensive observational measure.

4.2.2. The number of observations per site increased over time

The number of observations per site varied. This variation was largely due to the research team becoming more accustomed to the measure over time. As a result, fewer observations were made at the first site, while the highest number was recorded at the last site. Although the observers adhered to the 15-second observational scan protocol, they took additional time to record and discuss the observed movements to ensure consistency. As a result, during this round, two additional movements were added to MOP: step up and step down. This indicates that spending more time enabled additional movements exhibited during outdoor play to be identified, which were initially overlooked during the practice sessions. Allowing additional time to record and discuss the movements during this round demonstrated a deliberate effort by the field team to ensure consistent observation and recording of FMS among all observers. This process helped clarify any ambiguities arising from observing outdoor play and aligned the observers' understanding of the different FMS. Identifying two additional

movements refined and expanded MOP, making it more comprehensive and capable of capturing a broader range of FMS.

4.2.3. Observers consistently recorded up to two of the same movements, and inter-rater reliability increased over time

Overall, there was substantial inter-rater agreement among observers in recording at least one movement within each play observation across the four sites, and most sites show moderate to almost perfect agreement for recording at least two movements. In contrast, the agreement among the raters for recording three to five of the same movements varied, ranging from weak to almost perfect. Site A showed a decline in reliability from recording one of the same movements to recording four of the same movements, with a substantial drop from 0.950 in recording one of the same movements to a very low 0.133 in recording four of the same movements. Site B showed similar trends, with almost perfect agreement for recording one of the same movements to weak agreement for recording two of the same movements and minimal to no agreement for recording three of the same movements, and recording four and five of the same movements, respectively. Site C demonstrated consistently higher Kappa values for recording one of the same movements to recording five of the same movements compared to Sites A and B, signaling a more uniform and consistent agreement among observers. This improvement could be due to having additional time to practice, as observers had already used the observational measure at two other sites. Finally, Site D displayed higher kappa values across recording one of the same movements to recording five of the same movements. This suggests a greater consensus among observers at Site D compared to the other sites and greater agreement through continued practice over time.

In the development of the TOPO tool, the researchers observed a similar increase in reliability over time [21]. During the second round of reliability testing, the researchers made refinements to address any coding discrepancies [21]. The overall preliminary inter-rater reliability throughout the study from Site A to Site D suggests an upward trend in agreement levels. However, this pattern should be interpreted cautiously because reliability rounds were conducted sequentially across sites. As a result, improvements in inter rater reliability may reflect increased observer familiarity and practice with MOP over time, but they may also reflect confounding differences between sites, including play-space layout, available play materials and loose parts, environmental characteristics, and the age groups or play behaviors of the children present during observation sessions. This pattern may reflect the potential influence of ongoing use and practice on reliability, emphasizing the importance of familiarity and experience in enhancing observer agreement. However, because observer practice and site-level differences could not be disentangled in this study, this interpretation should be considered a possible explanation rather than a definitive conclusion. Moreover, including regular feedback during practice sessions and opportunities to discuss and align our understanding of the movements we were observing and recording and having several practice sessions likely contributed to this improvement. The consistent improvement in these preliminary inter-rater reliability results across sites demonstrates the value of training and repeated use of the observational measure. The findings also suggest that initial discrepancies can be mitigated over time, leading to more accurate and consistent data collection. Similarly, it could also reflect the adaptability of MOP to different settings and contexts; however, further testing is needed to confirm its robustness and versatility.

4.2.4. Observers appeared most consistent when identifying at least one shared movement

The almost perfect agreement for recording one of the same movements suggests that observers were able to reliably capture at least one of the same movements during the short (15-second) play observation; the level of agreement between observers weakened as the number of recorded movements increased. High reliability for recording one of the same movements may suggest that observers were well-acquainted with the 46 different FMS included in MOP. This familiarity may have supported their ability to identify and select the observed movement during the 15-second scan, which highlights the importance of practice and familiarity with MOP before its application in the field. The progression in inter-rater reliability for recording two of the same movements may also suggest that over time and with practice, observers were able to recall at least two movements during each observational scan. Similarly, during a 15-second scan, the number of movements the observer can observe and the observer's natural limitations in recalling and recording these movements come into play. The results indicate that despite these constraints, both observers could reliably identify and recall at least two movements performed during the play observation. This finding supports the assumption that during short, discrete moments of play (15-second), there is a high likelihood that two trained observers will detect and record at least two of the same movements. However, as the complexity of play increases, often characterized by rapid or overlapping movements, the likelihood of full agreement between observers may diminish. There may have been lower reliability in recording three to five of the same movements, where movements occurred simultaneously or overlapped, making it challenging for observers to isolate and identify specific movements. Therefore, it is essential to ensure that observers are familiar with all 46 types of movements in MOP and can identify these movements, especially in combination with other movements. Similarly, using additional methods, such as video recording or including an open-text within MOP to briefly narrate the observed play during the 15-second scan, may help capture movements missed during the 15-second scan.

4.2.5. Observers were consistent in recording typical FMS used during outdoor play

In addition to the kappa statistics, percent agreement was examined for the most frequently observed movements to provide a descriptive indication of observer consistency. High percent agreement was observed for commonly coded movements, including walking, standing, running, carrying, and holding, as well as for stability-based movements such as balancing, sitting, squatting, and jumping. Agreement was particularly high for several movements (e.g., jumping and sitting), suggesting that observers were consistently able to identify clearly defined and visually distinct movements during the short observations. These findings complement the kappa results by demonstrating that while agreement for recording three to five movements was lower, observers were nonetheless highly consistent in identifying the most frequently occurring movements during play.

4.2.6. Most play events involved one or two movements, despite the capacity to record up to five

The findings suggest that when observing outdoor play, particularly with LP, up to two movements were typically recorded. This trend is also reflected in the data, where observations involving the recording of one movement or two movements were most common. Additionally, the analysis of play events showed that over time, there was an increasing consensus regarding the number

of movements observed (or not observed) across all five categories during each play event. This indicates that both observers consistently recorded up to two movements while consistently not recording three to five movements.

Together, these results suggest that observers were generally consistent in identifying common movements, but that capturing multiple movements within a brief observation period remained more challenging. Despite these challenges, the decision to enable the recording of up to five movements per 15-second observation was appropriate, as children's play is often complex and fluid, involving transitions between movements or multiple movements occurring at once. Moreover, environmental factors such as the design of the outdoor play space or the presence of loose parts can influence the types of movements children use, as noted by Lim³¹. Allowing up to five entries per observational scan enables observers to capture this dynamic range of movements, given the environment they are in.

4.2.7. Loose parts may increase movement complexity and observational challenge

A discrepancy was noted in the number of observational scans between observers at Sites B and C. Unlike the other two sites, these locations included a diverse range of LP, such as balls and skipping ropes, integrated into their play spaces. This suggests that the presence of LP may involve a greater variety of movements within a single play event, which could make observational scans more challenging. The complexity of movement involved in LP, such as the greater variety of movements or the use of different movement skill sequences, can lead to differences in how observers interpret and record the type of movement. Therefore, when observing a setting with LP, it is important to provide ample time for recording the movements observed, and ideally the observer should be in a location where they can visibly observe the child engaging with the LP to ensure the accuracy of the recorded movements.

4.3. *Implications of the Movement Observations during Play measure*

The development of MOP is significant as it aims to provide a better understanding of how outdoor play spaces, specifically those incorporating LP, may support the development of FMS in children. This observational measure represents an innovative approach to studying FMS within the context of outdoor play and OLPP. Our findings from this study provide preliminary support for the inter-rater reliability and feasibility of MOP across outdoor play settings and over time. However, because formal criterion and construct validation were not conducted, these findings should be interpreted as an initial step in the development of the measure rather than evidence of full measure validation. As with any observational measure, its effectiveness in practice is strengthened through repeated use and observer familiarity. Regular training and practice with the MOP measure may improve observer accuracy and consistency, especially when used in dynamic, fast-paced play environments.

Examining the FMS children use during outdoor play has the potential to offer valuable insights into how to support and enhance their physical development. Understanding these skills in the context of outdoor play, including OLPP, may help researchers and practitioners better understand how environmental affordances can facilitate physical growth and encourage more PA participation. Prior research indicates that OLPP can significantly enrich children's development by providing opportunities for movement [7,8,13]. Environments featuring elements like climbing structures, LP,

and open spaces enable children to engage in many FMS, such as balancing, jumping, throwing, and climbing. These dynamic settings may provide opportunities for children to practice and refine motor skills, build confidence, and increase motivation to engage in physical activities, which may contribute to their overall physical literacy. However future validation work is needed before broader claims can be made about the use of MOP in intervention design, practice, or outdoor play space planning.

5. Conclusions

In conclusion, this study provides an observational measure, MOP, tailored to capture FMS in school-aged children during outdoor play and OLPP. We outlined the comprehensive process of developing the measure, informed by a thorough literature review and data analysis, to include movements that are typically used during outdoor play and OLPP. The results of the rounds demonstrate preliminary support for the inter-rater reliability of MOP, with moderate to high agreement for recording up to two movements within a short (15-second) observational scan of children's outdoor play and notable improvements in reliability with continued practice of the measure. Additional agreement analyses strengthened this interpretation by showing high observer consistency for commonly observed FMS, including movements typically seen during outdoor play. These analyses also suggested that most play events involved one or two recorded movements, indicating that MOP was able to capture the movements most frequently expressed during play while enabling the documentation of additional movements during a single play event. By including 46 unique FMS in MOP, many specific to OLPP, this comprehensive measure enables diverse FMS that emerge during children's outdoor play and OLPP to be documented and explored. Together, these findings support MOP as a promising observational measure and highlight the importance of observer training and practice in achieving consistent outcomes. When using MOP, the duration of the observation period may impact the detail and accuracy of the recorded behaviors. It is important to consider additional methods that could help capture this, such as video recording or field notes. Overall, this research contributes valuable insights into how children's FMS can be observed and understood during outdoor play and OLPP. It also offers a movement-specific observational measure to support future research and interventions aimed at promoting children's movement in outdoor play settings.

The strengths and limitations of this study

One of the strengths of the MOP measure is that it provides a detailed breakdown of play behaviors into specific FMS exhibited during outdoor play, including OLPP. Although initially developed in the context of OLPP, this observational measure can be used to understand movements during outdoor play, given the similar nature of these types of activities. The high level of specificity of MOP enhances our understanding of movements during this type of outdoor play, enabling the opportunity to identify FMS that emerge, and providing a means to explore their contribution to children's health.

MOP also offers versatility in usage. For example, it can function as a standalone direct observational measure or be integrated with other observational approaches like BM. It caters to researchers aiming to understand children's FMS during outdoor play or OLPP and practitioners working with children, providing insight into how outdoor play spaces and the affordances within these influence FMS. Furthermore, while MOP was designed with outdoor environments in mind, its comprehensive list of 46 FMS means it could easily be adapted for indoor play environments. For

instance, it could be utilized to observe and record movements facilitated in indoor settings such as gymnasiums utilizing LP, play centers, or activity spaces. Data collected from MOP may provide insight into the types of FMS occurring in a specific play space, informing the design of indoor and outdoor environments. The data collected from MOP can also inform the curriculum of programs aimed at children, guiding the selection of activities and equipment that promote movements.

A key strength of MOP is that it builds upon a published and validated observational measure, the OSRAC-E, which has been widely used to observe children's play in elementary school settings [26]. MOP extends the utility of OSRAC-E by categorizing movements into FMS categories, locomotor, object control, and stability, and by adding additional movements that are commonly observed in outdoor play and OLPP. While no formal validation procedure (e.g., criterion or construct validation) was conducted for MOP, several steps were taken to ensure that the movements included were representative of those commonly used during outdoor play and OLPP. The development of MOP followed a comprehensive process modeled after the tool development approach described by Loebach and colleagues [21]. While we acknowledge that a formal validation study was beyond the scope of this paper, we view this as an important next step and have identified it as a limitation and future direction for further development and evaluation of MOP. Moreover, researchers should focus on establishing these comprehensive psychometric properties, echoing the validation pathways utilized by established play-based learning and development assessment batteries [31].

A limitation of MOP is its inability to capture the sequence of movements, as there is no hierarchical structure for recording them. While this lack of hierarchy enables flexibility in observation, it also means we cannot see how different movements are used together or in sequence. To address this issue in future studies, incorporating video or voice recordings, or written descriptions of each play event, could help reveal the context and sequence of movements. Understanding the sequence in which children use different movements could offer deeper insights into their physical development.

Similarly, MOP does not provide details on the proficiency of the observed movements. While understanding movement proficiency is crucial for assessing how well children perform skills, most tools that evaluate FMS do so in a structured manner, such as asking the child to run or skip [32]. Although this structured assessment is important, it can sometimes feel unnatural to the child. In contrast, an observational measure that examines various FMS during outdoor play, including play with LP, addresses this gap in the literature. This enables us to observe FMS in a natural setting, providing insights into how children use these skills in their everyday play.

A key limitation of this study is that only preliminary inter-rater reliability was examined. While this provides insight into the consistency between observers, intra-rater reliability, the ability of a single observer to consistently code the same events across time, was not evaluated. Thus, researchers could incorporate video recordings of observational sessions to enable intra-rater assessments and further validate the stability of the observational measure over time.

In addition, formal criterion and construct validation were not conducted in this study. MOP was developed through literature review, expert consultation, pilot testing, observer training, and inter-rater reliability testing; however, it was not formally compared against an established criterion measure, nor were construct validity hypotheses tested. As a result, the findings should be interpreted as evidence of preliminary inter-rater reliability and feasibility rather than full validation of the measure. As such, researchers should examine criterion-related validity by comparing MOP with other observational or standardized movement measures and construct validity by testing whether MOP produces expected patterns across play environments, age groups, or intervention conditions.

Because the lead researchers participated as an observer in all inter-rater reliability sessions, there is potential for expectancy effects and systematic bias in how movement was observed and recorded over time. This is a common limitation in direct observation research [33,34], particularly during tool development when a primary developer is closely involved in implementation. To mitigate this risk, we used a standardized observation protocol, explicit operational definitions, and paired independent coding during each scan, followed by structured discrepancy discussions aimed at clarifying decision rules rather than aligning judgments to a preferred outcome. That said, there is some degree of observer-related bias that may remain. Researchers should include inter-rater reliability testing led by observers who were not involved in tool development and incorporate video-based coding to assess intra-rater reliability and further reduce the influence of expectancy effects.

Children were also not individually identified during observations. This approach was consistent with the place-based BM approach and enabled children's movements to be observed in a naturalistic and minimally intrusive way. However, this also means that we could not determine whether some children were observed more frequently than others or assess the potential influence of repeated observations of the same children. As a result, the data should be interpreted at the observation level rather than the individual-child level. Researchers could consider approaches that enable individual-level tracking while maintaining ethical and practical considerations in naturalistic play settings.

The inter-rater reliability findings should also be interpreted cautiously because improvements over time may have been influenced by observer practice and differences between sites. As observers became more familiar with MOP through repeated use, coding consistency may have improved. Furthermore, each site differed in layout, play-space features, available LP, and the age group or complexity of children's play behaviors, which may have affected the ease or difficulty of coding. Therefore, changes in inter-rater reliability across sessions may not reflect observer learning alone. Researchers should examine MOP inter-rater reliability across a wider range of sites and play contexts to better understand how setting characteristics influence coding consistency.

Finally, utilizing MOP to capture FMS observed during play is resource intensive. Given the time that needs to be allocated to training, practice, and using this observational measure, it could require a lot of time and personnel, in terms of training and data collection. Another limitation is the subjectivity that comes with observational approaches. Observations inherently carry subjectivity, where the interpretation of the observed behavior is based on the observer's personal experiences and perceptions. Although efforts have been made to minimize this bias through detailed descriptions of each FMS, the subjective nature of observations remains a consideration.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

We would like to thank PLEY School members, Julie Campbell, Jane Cawley, Kaitlyn Mooney, and Brayden Thomas for their invaluable assistance with the development and testing of the observational measure. We also extend our gratitude to Dr. Hilary Caldwell for her expert consultation on the observational measure, and Shelley McKibbin for her support in database searches (literature review). Our sincere thanks go to Drs. Janet Loebach and Adina Cox for their training and the sharing of their expertise.

in behavioural mapping, which greatly enhanced the methodology of this study. Funding Support: This project was funded through a Research Nova Scotia Establishment Grant, 2019–2022, Grant Number 2187.

Authors' contributions

N.J. conceived the study, collected, organized, and analyzed the data, interpreted the findings, and drafted the manuscript. S.T., J.L., D.R., B.F., and M.S. provided feedback on the study design, analysis, interpretation of the findings, and critical revision of the manuscript. All authors read and approved the final manuscript.

Conflict of interest

The authors declare no conflicts of interest in this paper.

References

1. Coe DP (2020) Means of optimizing physical activity in the preschool environment. *Am J Lifestyle Med* 14: 16–23. <https://doi.org/10.1177/1559827618818419>
2. Jones D, Innerd A, Giles EL, et al. (2020) Association between fundamental motor skills and physical activity in the early years: A systematic review and meta-analysis. *J Sport Health Sci* 9: 542–552. <https://doi.org/10.1016/j.jshs.2020.03.001>
3. Loprinzi PD, Davis RE, Fu YC (2015) Early motor skill competence as a mediator of child and adult physical activity. *Prev Med Rep* 2: 833–838. <https://doi.org/10.1016/j.pmedr.2015.09.015>
4. Goodway JD, Ozmun JC, Gallahue DL (2021) *Understanding Motor Development: Infants, Children, Adolescents, Adults*, 8 Eds, Burlington, MA: Jones & Bartlett Learning.
5. Flannigan C, Dietze B (2018) Children, outdoor play, and loose parts. *J Child Stud* 42: 53–60. <https://doi.org/10.18357/jcs.v42i4.18103>
6. Lee EY, de Lannoy L, Li L, et al. (2022) Play, Learn, and Teach Outdoors—Network (PLaTO-Net): terminology, taxonomy, and ontology. *Int J Behav Nutr Phys Act* 19: 66. <https://doi.org/10.1186/s12966-022-01294-0>
7. Branje K, Stevens D, Hobson H, et al. (2021) Impact of an outdoor loose parts intervention on Nova Scotia preschoolers' fundamental movement skills: a multi-methods randomized controlled trial. *AIMS Public Health* 9: 194–215. <https://doi.org/10.3934/publichealth.2022015>
8. Caldwell HAT, Spencer RA, Joshi N, et al. (2023) Impact of an outdoor loose parts play intervention on Nova Scotian preschoolers' physical literacy: a mixed-methods randomized controlled trial. *BMC Public Health* 23: 1126. <https://doi.org/10.1186/s12889-023-16030-x>
9. Maxwell LE, Mitchell MR, Evans GW (2008) Effects of play equipment and loose parts on preschool children's outdoor play behavior: An observational study and design intervention. *Child Youth Environ* 18: 36–63. <https://doi.org/10.1353/cye.2008.0025>
10. Naish C, McCormack GR, Blackstaffe A, et al. (2023) An observational study on play and physical activity associated with a recreational facility-led park-based “loose parts” play intervention during the COVID-19 pandemic. *Children* 10: 1049. <https://doi.org/10.3390/children10061049>

11. Olsen H, Smith B (2017) Sandboxes, loose parts, and playground equipment: a descriptive exploration of outdoor play environments. *Early Child Dev Care* 187: 1055–1068. <https://doi.org/10.1080/03004430.2017.1282928>
12. Spencer RA, Joshi N, Branje K, et al. (2019) Educator perceptions on the benefits and challenges of loose parts play in the outdoor environments of childcare centres. *AIMS Public Health* 6: 461–476. <https://doi.org/10.3934/publichealth.2019.4.461>
13. Spencer RA, Joshi N, Branje K, et al. (2021) Early childhood educator perceptions of risky play in an outdoor loose parts intervention. *AIMS Public Health* 8: 213–228. <https://doi.org/10.3934/publichealth.2021017>
14. Houser NE, Cawley J, Kolen AM, et al. (2019) A loose parts randomized controlled trial to promote active outdoor play in preschool-aged children: Physical Literacy in the Early Years (PLEY) Project. *Methods Protoc* 2: 27. <https://doi.org/10.3390/mps2020027>
15. Ulrich DA (2017) Introduction to the special section: Evaluation of the psychometric properties of the TGMD-3. *J Mot Learn Dev* 5: 1–4. <https://doi.org/10.1123/jmld.2017-0020>
16. Boulton P, Thomas A (2022) How does play in the outdoors afford opportunities for schema development in young children? *Int J Play* 11: 184–201. <https://doi.org/10.1080/21594937.2022.2069348>
17. Loebach J, Cox A (2022) Playing in “The Backyard”: Environmental features and conditions of a natural playspace which support diverse outdoor play activities among younger children. *Int J Environ Res Public Health* 19: 12661. <https://doi.org/10.3390/ijerph191912661>
18. Sutton MJ (2011) In the hand and mind: The intersection of loose parts and imagination in evocative settings for young children. *Child Youth Environ* 21: 408–424. <https://doi.org/10.1353/cye.2011.0000>
19. Cosco NG, Moore RC, Islam MZ (2010) Behavior mapping: A method for linking preschool physical activity and outdoor design. *Med Sci Sports Exerc* 42: 513–519. <https://doi.org/10.1249/MSS.0b013e3181cea27a>
20. Cox A, Loebach J, Little S (2018) Understanding the nature play milieu: Using behavior mapping to investigate children’s activities in outdoor play spaces. *Child Youth Environ* 28: 232–261. <https://doi.org/10.7721/chilyoutenvi.28.2.0232>
21. Loebach J, Cox A (2020) Tool for Observing Play Outdoors (TOPO): A new typology for capturing children’s play behaviors in outdoor environments. *Int J Environ Res Public Health* 17: 5611. <https://doi.org/10.3390/ijerph17155611>
22. Play Outside NS (2020) Outdoor Loose Parts Play—Event Recap [Video]. [cited 2026 June 15]. Available from: <https://www.youtube.com/watch?v=4RjvBc-G3Hg>.
23. Creswell JW, Creswell JD (2018) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5 Eds, Los Angeles, CA: SAGE.
24. IBM Corp (2021) IBM SPSS Statistics for Windows, Version 28.0, Armonk, NY: IBM Corp.
25. Pallant J (2016) *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*, 6 Eds, Maidenhead, UK: Open University Press.
26. McIver KL, Brown WH, Pfeiffer KA, et al. (2016) Development and testing of the Observational System for Recording Physical Activity in Children: Elementary School. *Res Q Exerc Sport* 87: 101–109. <https://doi.org/10.1080/02701367.2015.1125994>

27. Brown WH, Pfeiffer KA, McIver KL, et al. (2006) Assessing preschool children's physical activity: The Observational System for Recording Physical Activity in Children-Preschool Version. *Res Q Exerc Sport* 77: 167–176. <https://doi.org/10.1080/02701367.2006.10599351>
28. Puhl J, Greaves K, Hoyt M, et al. (1990) Children's Activity Rating Scale (CARS): Description and calibration. *Res Q Exerc Sport* 61: 26–36. <https://doi.org/10.1080/02701367.1990.10607475>
29. McHugh ML (2012) Interrater reliability: The kappa statistic. *Biochem Med* 22: 276–282. <https://doi.org/10.11613/BM.2012.031>
30. Lim C, Donovan AM, Harper NJ, et al. (2017) Nature elements and fundamental motor skill development opportunities at five elementary school districts in British Columbia. *Int J Environ Res Public Health* 14: 1279. <https://doi.org/10.3390/ijerph14101279>
31. Montoya-Fernández C, Gil-Madróna P, Losada-Puente L, et al. (2024) Play-based assessment: Psychometric properties of an early childhood learning and development assessment battery. *Educ Sci* 14: 1240. <https://doi.org/10.3390/educsci14111240>
32. Eddy LH, Bingham DD, Crossley KL, et al. (2020) The validity and reliability of observational assessment tools available to measure fundamental movement skills in school-age children: A systematic review. *PLoS One* 15: e0237919. <https://doi.org/10.1371/journal.pone.0237919>
33. Halpin SN (2024) Inter-coder agreement in qualitative coding: Considerations for its use. *Am J Qual Res* 8: 23–43. <https://doi.org/10.29333/ajqr/14887>
34. Monahan T, Fisher JA (2010) Benefits of “observer effects”: Lessons from the field. *Qual Res* 10: 357–376. <https://doi.org/10.1177/1468794110362874>



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)