Fiddling, Pointing, Hovering, and Sliding: Embodied Actions with Three Evaluation Tools for Children

Cristina Maria Sylla
University of Minho
Braga, Portugal
cristina.sylla@ie.uminho.pt

Elena Márquez Segura
Uppsala, Sweden
Uppsala, Sweden
elena.marquez@im.uu.se

Akeiylah DeWitt
University of California, Merced
Merced, California, USA
adewitt4@ucmerced.edu

Ahmed Sabbir Arif
University of California, Merced
Merced, California, USA
asarif@ucmerced.edu

Eva Irene Brooks
Aalborg University
Aalborg, Denmark
eb@learning.aau.dk

ABSTRACT
In user studies with children, it is important to use age appropriate evaluation tools to better understand their preferences, opinions, and thoughts. Here, we studied two accepted evaluation tools: The Five Degrees of Happiness, and the Sticky Ladder rating scale; together with the Paper Ladder, a paper version of the latter. Thirty-six preschoolers rated two creative and play activities ("Painting" and "Construction Blocks") and a game ("Musical Chairs") in terms of difficulty, enjoyment, and preference. Drawing from theories of embodied and distributed cognition, we performed a video analysis of the children's interactions with these tools, focusing on how each tool supported the children's cognitive processes and communication with the researcher. Here, we first describe children's embodied behavior and discuss how these were supported by design features and affordances of the tools. Then, we discuss strengths and shortcomings of each evaluation method. Last, we provide recommendations for their design, appropriation, and usage by researchers developing and evaluating playful solutions and games for children.

Author Keywords
Children; Evaluation; Survey Methods; Rating Scales; Likert Scale; User Studies; User Feedback; User Experience.

CSS Concepts
• Human-centered computing~HCI design and evaluation methods • Human-centered computing~User studies

INTRODUCTION
Survey methods are important to collect users' opinions, and preferences on usage and experience of products and activities [6]. In the case of children, and given that they perceive and act in the world differently from adults, it is important to ask them directly about their preferences and experiences rather than merely observing them [3,7,20,31,50]. Yet collecting young children's opinions in user studies is challenging [1,7,28]. Depending on age and cognitive development, children may find it difficult to understand interview or survey questions and to clearly communicate their thoughts verbally [1]. Also, children's responses may be influenced by the desire to please adults [1,7,27,28]. Therefore, special care must be taken when selecting an empirical evaluation method involving children.

In our research, we focus on the capacity of evaluation tools for children to capture their preferences and opinions, and elicit elaborated answers about their experience with products and activities. Here, we studied three evaluation tools: The Five Degrees of Happiness rating scale [27], the Sticky Ladder [1] and the Paper Ladder, a paper version of the later [62]. Our previous work on children's preference for different types of feedback in a math app [62] showed the potential of the Paper Ladder to be as effective as its predecessor, the Sticky Ladder [1]. Here, we extend this work investigating the capacity of these tools: i) to support children's behavior, their thinking and reflection processes, as well as the communication between the children and the researchers; and ii) to evaluate different kinds of play-related activities for children that vary in openness and flexibility: from the play activities of "Paining" and playing with "Construction Blocks", to the game of "Musical Chairs".

We report on a study conducted with 36 preschoolers aged three to six, in a Portuguese preschool, using the three evaluation tools to evaluate the above-mentioned play activities in terms of difficulty, enjoyment, and preference.

Inspired by theories of embodied and distributed cognition (EC and DC respectively) [32,39]; and methodologically informed by cognitive ethnographies and interaction analysis (IA) [32,35,29], we performed a video analysis of

CHI PLAY '19, October 22–25, 2019, Barcelona, Spain

ACM ISBN 978-1-4503-6688-5/19/10$15.00
https://doi.org/10.1145/3311350.3347170

the children's interactions with these tools, focused on verbal and non-verbal behavior.

Our work is relevant for the children, and play & games design and research communities for several reasons: first, our study focuses on the use of three evaluation tools to evaluate three play activities for children.

Then, these tools can be used to evaluate products and activities specifically designed for children, many of which will likely involve play. Here, we describe how each tool differently supported thinking and reflection processes and behavior of the children, and their communication with the researcher. We also help designers and researchers interested in these tools highlight design aspects of each of them, like affordances, that made a positive impact on the children's responses. We also discuss strengths and shortcomings of each tool, and provide recommendations for their appropriation and usage.

Last, we surface and illustrate interesting non-conscious play-related behavior that the tools supported, which helped children reflect on, and articulate their preferences. For example, the Sticky Ladder and the Paper Ladder supported fidgeting with the tools' tokens representing the play activities under evaluation, and playfully moving them on the ladders' rungs.

RELATED WORK

For long, methods in design and research have been adapted to suit and include children (e.g., [19,28,52,77]), design ideations [17,42,43] and even whole design processes of novel child-centered technologies [19]. Some of these methods are adaptations of commonly applied user-centered design methods for adults [28], while others are specifically designed for children [1,5,50,52,77]. Regardless of the method of choice, it is important to not reinforce adult representations of children, and to not restrict understandings of the children's experiences and knowledge [33].

Survey methods have been used with children to collect information on user interface design, particularly to evaluate children's preferences [1,77], explore the "appeal" and "fun" of a product [50,52], and to understand children's mental models and processes [7,51]. They have been modified to better capture the responses of children. For example, Likert scales have been modified to reduce the number of response choices, or the wording of survey responses modified to match the vocabulary and cognitive skills of the children [47].

Another widely-used approach are graphical response options. For example, the Visual Analogue Scale (VAS) seeks an agreement response located within a continuous line between two opposite end points [40]. Some researchers have also used smileys as response choices. For instance, the Smileyometer is a smiley version of a 5-point Likert scale that measures different dimensions of fun, and has been used with five to ten years old children [50,52]. The Funometer consists of a vertical ruler with a happy face on top and a sad face at the bottom [52]. These methods can be as effective as Likert scales. Further, they bypass the need to deal with complex language. Instead, they use familiar concepts and symbols for children, like emoticons.

Yet, evidence suggests that young children often have difficulties to differentiate between smiley faces expressing degrees of preference of the same valence, like "especially happy" and "slightly happy" [1,27,50,52]. Also, young children tend not to use the middle of a scale [15]. Several studies suggested that young children tend to bias their answers towards the extreme positive side [1,7,15,27,50, 52,54]. Further, six to ten-year-old children tend to score most products as "brilliant" [50,77], presumably because children of this age usually engage in dichotomous assessments. Children also have difficulties to make fine-grained discriminations of emotions [15]. These issues could affect both the classical Likert scale as well as its expressive smiley variants.

Relevantly, in a comprehensive study involving over 300 children aged nine to eleven-year-olds, Hall et al. [27] found out that children usually use two points of the smiley scale: "positive" and "very positive". Hence, to encourage the children to make full use of the scale, they designed the Five Degrees of Happiness scale, one of the evaluation tools investigated in this paper. This tool only presents positive emotions in the form of smileys. A user study revealed that this method supported nine to eleven-year-olds to use the full rating scale in their assessments [27].

We included the Five Degrees of Happiness scale, as smiley-based Likert scales are arguably the most popular method for collecting children's opinion in user studies. We picked it over other smiley-based Likert scales based on Hall et al.'s study [27], which indicated higher data variability of this method compared to others.

To summarize, evaluation methods for children exist and are valuable. However, they face a series of shortcomings: children gravitate towards using responses on the extreme sides of a scale [1,7,15], not often using scores in the middle of the scale [27]; also, children find it difficult to differentiate between similar ratings (e.g. "especially happy" and "slightly happy") [1]. Both the Sticky and the Paper Ladder [1,62] have shown potential to address these issues.

Physical Engagement and Tangible Evaluation Tools

Some evaluation tools designed for children emphasize acting instead of verbal articulations. They typically circumvent the need for complex instructions from the researcher, and support children to express their views without having to verbalize them. Examples of such methods are the This-or-That [77] evaluation tool, and Airey's et al.'s Sticky Ladder rating scale [1]. This-or-That supports pairwise comparisons of items representing products or activities through selective choice, i.e. children

are asked to choose their preferred item between two through pointing at it. This method has been used with four- to six-year-olds [77]. The Sticky Ladder is a tangible version of a Likert scale in the form of a felt ladder. Children rate activities or products represented by physical tokens that have Velcro to stick them onto the ladder: items placed on the top rung are scored higher (preferred; most positive score) than those on the lower rung (less positive score). The Sticky Ladder scale has been used with children aged four to ten years old with successful results [1].

The visual and tangible aspects of both evaluation tools align well with the principles of multiple resources and redundancy gain [70]. They can support the researcher during the instruction and explanation phase, and the children to better understand the evaluation method [5]. In addition, they can help the children to concentrate on the task and remember what is being evaluated [5].

Here, we focus on the Sticky Ladder rating scale, which presents advantages over other methods. For example, differently from This-or-That, the Sticky Ladder allows the comparison between multiple products at the same time. Further, it enables rating multiple different factors, including "enjoyment," "liking," "fun," and "difficulty" [1]. In addition, the Sticky Ladder supports rating on a scale, rather than a dichotomous Yes/No response. We argue that the cultural metaphor and visual representation of a ladder may also help the children to understand scale rating through projection and visual representation of the activities/ products and their ratings (i.e. position on the ladder).

Despite these advantages, the Sticky Ladder tool has not been widely used in research, exceptions notwithstanding (e.g. [74]). This is partially due to the logistics involved in creating the scale [62], which could be circumvented with a simplified version of this tool, like a paper version of it: The Paper Ladder [62]. The Paper Ladder is printed on a paper sheet, and its items or tokens (i.e., the products and/or activities under evaluation) are printed cutouts (Figure 1). This paper variant of the Sticky Ladder presents several advantages compared to its predecessor: it is low-cost and more easily deployable, and creating multiple variants or representations of products and(/or) activities to evaluate is easier. This can facilitate the appropriation of the tool to suit the needs of a target group. E.g. different representations of the tokens might emphasize different particularities of the evaluated products and(/or) activities, which can potentially evoke and focus the evaluation on different aspects of the product and(/or) activity.

A previous study [62] showed that this scale was as effective as its Sticky predecessor. It addressed common challenges that children face with evaluation tools (summarized at the end of the subsection above), e.g. children's assessments covered an unusual degree of variability for children [62]. In addition, the children felt comfortable with the Paper Ladder and enjoyed using it.

Embodied and Situated Cognition

Here, we briefly introduce key concepts from the domain of embodied and distributed cognition, which influenced our choice of evaluation tool, and the analytical lenses that we employed to investigate the different tools in this paper. Before, thinking was believed to happen in the brain and gestures and bodily actions were considered accessories to thought; a way of "externalizing thought but not part of creating it" [39]. They were seen as "expression of thought, proof that thinking was already taking place on the inside" [39]. Yet, theories of embodied, situated, and distributed cognition have shown that "thought is not confined to the brain;" it involves our whole body, which is tightly and inseparably coupled with our mind [39]. Hence, our bodies and our gestures can be considered "cognitive components, partially shaping how we think" [39]. Likewise, cognitive processes and reasoning extend "beyond the individual to encompass interactions between people and with resources and materials in the environment" [32]. From a social constructionism perspective, people co-construct meaning through their interactions [26,12,9,57,16]. From this standpoint, evaluation tools like questionnaires cannot be seen as mechanisms to "fish out" pre- existent thoughts and opinions, they also trigger and influence thoughts and opinions. The questions that they pose, and the way they are phrased can impact thinking and its outcome, this is why best and worst practices in questionnaire/interview design exist.

In this work, we focus on how evaluation tools trigger and support situated and embodied reasoning [38], as well as social construction of meaning [12]. We study and discuss how these phenomena relate to design properties of the evaluation tools. In this section, we present key conceptual underpinnings that impact our theoretical and methodological stance.

"(T)he environment people are embedded in is, among other things, a reservoir of resources for learning, problem solving, and reasoning." [32]. Thinking, recalling, reasoning, and processing involve not only internal cognitive processes, but also external ones that are deeply intertwined, such as manipulations of objects and data representation in the environment [32]. There is a complex relation between the internal and external processes and resources. Actions may be individual and social as well as internal and external [66,68,69]. Internal resources, such as "memory, attention, executive function" impact and are impacted by external resources, such as "objects, artifacts, and at-hand materials constantly surrounding us" [32]. From an embodied cognition (EC) perspective, both "the human body and the material world take on central rather than peripheral roles" [32] in cognitive processes.

Physical Tokens, Projection, and Affordances

Traditional psychology explains how we process information from symbols, which are understood as "tokens that refer to something other than themselves" [32]. Tokens embedded in multiple forms and representing diverse

content can benefit cognitive processes even in early ages. For example, Fröbel, the 'inventor' of kindergarten (in 1837) developed a set of physical objects, the "gifts," which were shaped after objects present in nature and in every-day life (e.g. balls, strings, sticks and blocks). They supported children's playful exploration, understanding abstract concepts, and to express ideas [11]. Tools and representations have a strong impact on people's actions and thoughts, and support creative thinking [53]. Visual representations are especially interesting for the purpose of our work. Visual material appears to have stronger saliency, a stronger emotional impact, and it is more memorable than words referring to a same concept [34]. Vygotsky [65,66], and Werner and Kaplan [67] explained how we learn to distinguish an object from what that object refers to (referent and referee) from early ages [10,76]. In the context of this paper, this highlights the potential of the smileys (the Five Degrees of Happiness Scale, or the Smileyometer [27]) to speak of the emotions that they represent; and the tokens (Ladders) to refer to the creative and play activities that the children evaluate.

Traditional psychology has focused less on how people "exploit the physical properties of the representing tokens themselves" [32]. This is important in the context of the elements in our evaluation tools, like the tokens in the Sticky and the Paper Ladder. The distinction between represented and representation can often blur, and valuable cognitive processes emerge from shifting "back and forth between attending to the properties of the representation and the properties of the thing represented" [32]. This shift provides "a range of cognitive outcomes that could not be achieved if representations were always only taken as representations of something else, and not as things in themselves" [32]. In this paper, this becomes relevant when attending to, and surfacing interaction patterns of the children with the tokens of the Ladders. In addition, spatial and material elements help perceive, process and compute things, and simplify choice [38,32]. People make "mental tools of things in the environment" [38], i.e. rather than using abstract computations, they use contextual elements "to help draw conclusions and solve problems" [38].

We relate this to Kirsh's concepts of perception, imagination, and projection [39]. *Perception* is understood as the factual existence of an external object in relation to existing others (e.g. a mark written down in a particular smiley box in the Five Degrees of Happiness Scale, or a token placed on a rung in the Ladders); *projection* refers to the mental process where an image, object, or token (e.g. the Ladders' tokens) is imagined or visualized attached to a physical structure or another object (e.g. the ladders): "When we project onto an object, whether kinaesthetically or visually, we experience ourselves intentionally augmenting the object. The object anchors our mental image*"* [39]. Kirsh explains *imagination* as a cognitive process that lacks a physical structure onto which an object is anchored.

In the context of this paper, we argue that the structural elements (such as the ladders and the scale in the Five Degree of Happiness scale) and the objects in our evaluation tools (tokens and pen, respectively) can help the children to visualize and to consider their responses. Through manipulating these objects (tokens and pen) in relation to an underlying structure (ladder, and scale), the children may engage in a reflective conversation with these materials [56] and their meaning. In particular imagining, projecting, and perceiving these objects over those structures can positively impact emergent cognitive processes, and the quality of rating outcomes.

Gestures and Bodily Action

The role of movement, gestures, and touch in cognitive processes is often overlooked. Ecological psychologist Gibson, one of the biggest contributors to the field of visual perception, explained the tight connection between manipulation, vision, and haptic feedback [25] and how we tend to be more aware of, and pay more attention to visual feedback. However, the role of sensorimotor action in cognition is well established in multiple disciplines and strands of thought, from evolutionary biology and neurology to phenomenology [71,72,58,59]. Neurologist Frank Wilson explains the strong interdependence of hand and brain function and the evolutionary origins of that relationship [71]: "Our fingers and hands are highly active and important means of perception and exploration, representing an access to our lifeworld which in some cases could not have been established by any other sense modality." Phenomenologist Sheets-Johnstone goes a step beyond explaining the primacy of movement for cognitive development [58]. Movement precedes perceptual-cognitive relations of ourselves and the world, and our linguistic capability [58]; through moving, "we discover ourselves" and "we embark on a lifelong journey of sense-making" [58]. Through people's life "a semantically congruent relationship" is maintained "between movement and meaning" [59].

The way people "manipulate icons, objects, and emergent structure is not incidental to their cognition; it is part of their thinking process [...]" [32]. This has been leveraged in the domain of learning. For example, Maria Montessori, well known for her revolutionary pedagogical methods and education curricula, contended that touch is the most precious sense for exploration and learning [48]. Montessori developed a set of learning tools, the Didactic Materials, that addressed every sense. She considered that children learn through exploration, and mistakes: "It is precisely the error that makes the tool important, for the child has to observe and try out different possibilities [48]. This resonates with works in Human-Computer Interaction, and learning. E.g. in the domain of educational games, Melcer and Isbister explored the impact of tangible "programing blocks" in a programming game, which positively impacted performance and learning outcomes, perception of own abilities, interest, and enjoyment [46].

This potential of tangible objects and their manipulation to support cognition is also exploited by evaluation methods using tangible items like the Sticky ladder. This method was created under the premise that children are better at expressing their opinions by manipulating physical objects rather than through words [1].

A particular bodily behavior that emerged as important in the context of our work is fiddling or fidgeting. These are typically small and repetitive body movements with and without objects that are (apparently) non-instrumental [73] to an ongoing task or situation. Examples of fidgeting are tapping on the floor (or on a table), or playing with worn objects (jewels, glasses), and handy tools (e.g. a pen). Although not purely hedonic, fidgeting has been classified as play behavior [37,63]. This aligns with Caillois's concept of *paidia* to refer to "spontaneous manifestations of the play instinct," which are exemplified through actions such as scribbling, and by young children's inclinations to "touch, grasp, taste, smell, and then drop any accessible object" [13].

In addition, fidgeting or fiddling can have an instrumental value. It has been explained as a displacement activity to deal with affective states [64], discomforting sensations or demanding situations [23,21], such as increased stress [4], and fatigue in classes [21] or exams. Although some works associate fidgeting to a decrease of attention, many others suggest the opposite (e.g. for a summary see [21]). In fact, fiddling has been proposed "as an 'embodiment' of the act of sustaining attention" [21]. It increases energy expenditure [41] and arousal [64,24], which can help the individual focus and sustain attention [2,24,18,44,60]. It is used as an embodied self-regulation practice towards achieving calm, focus, and creativity [37].

Fiddling is recognized to assist learning. It is listed within the kinesthetic category of Galbraith and James's [22] taxonomy of perceptual modalities relevant to learning and it explains how some students process information and learn best while moving, e.g. while "pacing around the room" or "moving some part of the body (e.g., tapping a pencil, fidgeting, kicking a leg, etc.)" [30]. It has also been related to increase retention [2]. Fidget toys have been specifically studied as tools to have for the classroom [61].

For children, fiddling can help retrieve and articulate thoughts; support attention and learning [18]; engage in cognitive tasks, regulate emotion, and deal with energy excess [14].

Cultural Influence and Social Construction of Meaning

The way the children perceive the three evaluation tools in particular and the world in general, does not occur in a vacuum. Culture shapes both behavior and cognitive processes transcending "the boundaries of individuals" [32]. Artefacts that we use are always part of historical and cultural contexts [32]. Culture provides us with tools that help us to understand things that we otherwise would have difficulties to understand without them and, thereby, extend a person's boundaries [32]. This puts the concept of *mediation* [69] at the core to conceptualize how the children in our study interacted with the evaluation tools. Particularly relevant in the context of this study is the cultural relationship between higher and better (which the ladder exploits).

The importance of social interaction for the construction of meaning is well established [12,9,57,16], this is relevant in the context of the study protocols that we focus on, including a researcher interacting with the children, and using evaluation tools as prompts to further inquire about the children's perspectives. Although this may contradict the traditional research value of non-interference, we contend it is mostly applicable to other kinds of quantitative analysis.

USER STUDY

In a previous study, the Paper Ladder showed to be as effective as its original counterpart [62]. We also found intriguing preliminary differences in how the children interacted with the evaluation tool. For example, the children slid tokens in the Paper Ladder, which we related to its non-adhesive nature and its tokens (each representing a different feature). Here we further investigate if and how this can be advantageous from an embodied and situated cognition perspective. This is, how and if the "slideability" of the paper tokens over the paper ladder's rungs can support children's *perception* [38,39] and help them consider different answers. In addition, we investigate if this feature can help the children to perceive their answers as more provisional compared to the original Sticky Ladder, were the answers are more "locked" once the tokens are affixed to the ladder. The observations from our preliminary study [62] seem to support this; the children often reordered the pieces on the Paper Ladder, which may indicate that they re-evaluated their preferences [62]. These observations motivated the conceptual and methodological stance in this work.

Play as a Tool and Object of Study

In this study, we focus on play-related activities instead of particular technologies or products for several reasons: the former are easily replicable and transferable to other domains. Also, this avoids potential biases regarding technology use. Last, play-related activities are likely the focus of user studies involving children.

For increased relevance for researchers and practitioners working with diverse kinds of play-related designs for children, we evaluated three play activities that are positioned at different points of the play-game continuum [55]. On one end of that continuum is structured and rules-bound play, which Callois's called *ludus*; on the other end is more improvisational and open play, also called *paidia* [13,55].

Here, we evaluate the creative and open-ended play activities of "Painting" and playing with "Construction Blocks", and the more structured play activity of the "Musical Chairs" game. These activities are well known by many children. The children in our study used the rating scales to evaluate three key aspects in play and games: difficulty, enjoyment, and preference.

Last, the play-related behavior of fidgeting was at focus in our analysis. Some previous work classified fidgeting as play behavior [37,63]. Fidgeting is also often equated with play by parents and children [14]. Yet, in this study, we see fidgeting as a less mindful and *autotelic* [55] play-related activity than other play activities children engage with, like those under evaluation.

Participants

The study took place at a Portuguese local public preschool, with the participation of 36 preschoolers, aged between 3 and 6 years (avg.= 5.2; SD = 0.86). With the help from the kindergarten teachers, we divided the children into three groups (12 children each) with comparable average age, cognitive development, verbal and motor skills, and gender ratio. The evaluation sessions were held in three consecutive days and each group used one of the three rating scales (Sticky Ladder, Paper Ladder, and Five Degrees of Happiness) to rate all the three activities ("Painting"/Drawing, "Construction Blocks", and the "Musical Chairs" game) in terms of difficulty, enjoyment, and preference. We received informed participation consents from the children's parents or legal guardians.

Instruments

To create the Paper Ladder, we printed a drawing of a ladder on an A4 paper. We prepared paper cards with representative drawings of the activities under evaluation, and glued each of them onto 50×50 mm cardboard cutouts for durability and ease of manipulation (see Figure 1, 2). Then, we replicated Airey's et al. Sticky Ladder by crafting a felt ladder that we glued on an A4 cardboard to facilitate its manipulation (see Figure 1, 2). We glued a piece of Velcro onto similar cutouts to those used for the Paper Ladder, and onto the center of each rung. Last, we used Hall's et al. [27] Five Degrees of Happiness rating scale, which consisted of three printed pages per child (one page per activity) (see Figure 2). We explained to each child how to rate the activities using each rating scale respectively.

The data was collected through observation, written notes, and audio-video recordings using a video camera on a tripod placed behind the children, set with fixed focus on and zoomed into the manipulation of the evaluation tools.

Procedure

The evaluation was conducted during regular class hours, in a separate room with one child at a time. Each child was presented with the respective rating scale. For the Paper Ladder and the Sticky Ladder groups, the rating scale and the tokens were placed on the table. For the Five Degrees of Happiness group, a printed questionnaire was placed on the table. The researcher conducting the intervention, called the child's attention to the tokens and asked if s/he knew the

Figure 1. The original Sticky Ladder [1], the Paper Ladder, and the Sticky Ladder used in our study, each ladder depicting the products under evaluation respectively.

activities represented. This assured that the child was familiar with all the activities. Then, the researcher focused on the ladder (Figure 2), explaining the meaning of placing the tokens on each rung: items on the bottom rung were the least preferred; those on the rung above were liked a little more, and so on; last, items on the upper rung were the most preferred. For the Sticky Ladder group, the researcher also held the instrument vertically to evoke a real ladder (this was only possible with the Sticky Ladder as it was glued on cardboard). For the Five Degrees of Happiness group, the researcher named the three activities and asked the child if s/he knew them. She then explained thoroughly the meaning of each of the different smiley faces, and how to fill in the scale according to their preference.

Figure 2. Children using their respective evaluation tool: The Five Degrees of Happiness; the Sticky and the Paper Ladder.

After assuring that the child understood the activities and how to use the respective scale, the researcher asked him/her to rate the activities one by one: e.g. Is it fun to play with Construction Blocks? /to Paint/Draw? /to play the Musical Chairs game? (and similarly, with the measures "difficulty", and "preference"). For all groups, the order of the measures and activities were randomized to avoid bias.

Analysis

Relevant aspects towards studying cognitive processes of children are object manipulation and the concepts of perception, projection, and imagination above presented. They influenced the analytical lenses of this work. While we acknowledge that we do not have access to all cognitive processes of the children, we contend that we do have access to important other ones, i.e. those that are leveraged and manifested verbally and non-verbally through gestures and bodily actions. In particular, we analyzed video data looking at the interactions with the three evaluation tools and coding gestures that reflected projection and

perception, such as positioning tokens or marks on the ladders or the Five Degrees of Happiness scale respectively; sliding tokens on the ladders; or hovering these tokens or the pen over the ladders and the Five Degrees of Happiness scale respectively. We also note how some of these actions are supported by particular properties of the evaluation tools, which in HCI are described as affordances [45,49]. For example, while the smooth surfaces of the ladder and the tokens in the Paper Ladder version afford sliding of the former on the latter; the Velcro on the Sticky Ladder does not; the pieces get stuck on the Velcro ladder, supporting perception rather than projection. Yet, we will discuss affordances in terms of action possibilities as they are mediated not only by the physical properties of the evaluation tool, but also by other contextual elements and cultural aspects [36].

The qualitative Interaction Analysis (IA) of the video material involved three authors: a first scan was conducted by two authors, independently noting relevant interactions, bodily expressions and nonverbal behavior that accompanied a rating and indicated thinking. These were then discussed, refined, and classified in overarching groups and in relation with concepts and theories in the background. This resulted in the creation of codes that then served to analyze the video data by one of those authors and another author. They first discussed the content of each code and category and jointly analyzed and discussed video samples to attain the maximum consensus. Afterwards, each author codified the videos independently.

The overarching groups and particular examples that were coded were: General Thinking Gestures (TG), e.g. hands on lips, and head resting on hands; Hovering (H), e.g. hovering the items or the pencil over the evaluation tool; Sliding (S), e.g. sliding the items, or the pencil, on the evaluation tool; Fiddling (F), e.g. flipping the token or passing them between the fingers, tapping the table with them, etc. We registered and counted their frequency during the children's use of each of the evaluation tools.

STUDY RESULTS

The mean time of interaction (including explanation and evaluation) per child and evaluation instrument was 4,38min for the Sticky Ladder, 4,33min for the Paper Ladder and 3,86min for the Five Degrees of Happiness.

In the following, we present relevant and illustrative behavior coded through nine representative scenes or vignettes (V1-V9). To each we add a brief annotation referring to pertinent concepts in the background. We conclude with the results from our IA in the form of frequency counts of relevant behavior.

Interaction Vignettes

Fiddling, Projecting, and Thinking Gestures

V1 (Sticky Ladder): The researcher asks the child if "Painting" is difficult. The child takes the "Painting" card, holds it against her face and says: "No," shaking her head. The researcher then asks "where are you going to place the card?" The girl says: "Humm…" holds the card on the right hand at eye level and twists and slightly moves the Sticky Ladder on the table, looking at it. Then she looks at the "Painting" card, which she continues holding on the right-hand, smiles to the researcher, and says "in the bottom" places the item card on the 1st rung of the Sticky Ladder and looks at it (this takes 0.24 seconds).

V2 (Paper Ladder): The researcher asks the child to rate "Painting" in terms of "enjoyment". The child takes the item card, brings it to his chin, then holds it in front of his face with both hands, looks at it while slowly rotating the card on his hands. Then, he leans his body to the front, coming near the Paper Ladder, and slowly moves to place the "Painting" card on the Paper Ladder (upper rung). He then hesitates moves back always holding the "Painting" card in both hands, looks at it, holds it with just one hand turning the item back and forth in the hand while simultaneously looking at the Ladder. He then moves the hand down to place the "Painting" card on the Ladder, lifts it a little bit and places it between the 4th and 5th rung on the left side of the rung. The total time for thinking and answering took 25 seconds. He then leans back with his chin on the fist and looks at the Paper Ladder.

Relevance: The projection behavior in V2, where the child looked at the token and subsequently at the ladder, indicates considering different possibilities, and a high degree of concentration. The vignette V1 (like V6 later) exemplifies fiddling and thinking behavior. Additionally, it shows consistency of verbal and non-verbal behavior, indicating that the child had understood how to operate the Ladder.

Sliding

V3 (Paper Ladder): The researcher asks the child to rate each of the activities in terms of "difficulty". The boy slides the "Musical Chair Game" token on the table dragging it into the ladder, and onto the second rung from top. Then, the researcher asks him to rate the "Painting" activity. The boy slides the "Painting" card on the right of the ladder, onto it, in a curve trajectory across the middle of the ladder and towards the top rung, above the "Musical Chair" token. Then, he rates the "Construction Blocks" activity. He slides that token drawing a curve from the middle of the ladder to the bottom onto the rung below the "Musical Chair Game."

Relevance: Here the focus is on the sliding movements into and out of the ladder. These gestures were frequent and well supported by the Paper Ladder; the children could slide tokens over different rungs and visualize their choices. From an embodied and situated cognition perspective, the "slideability" of the paper tokens over the paper ladder's rungs can support the children's *perception* and help them consider different answers.

Hoovering

V4 (Paper Ladder): The researcher asks the child to rate the "Musical Chair Game" in terms of "enjoyment". The child

takes the item card and fidgets with it drawing circles in the air over the ladder, looking at the token and the ladder. He then places the token on the second rung from top. He repeats the same pattern of movements with the next two tokens.

V5: (Sticky Ladder): The researcher asks the child if "Painting" is difficult. He takes the token, holds it with both hands, arms stretched, over the two upper rungs. He then moves the hands down to almost touch the third rung, stops for a fraction of a second and moves the hands with the token down to place it on the second rung of the ladder.

Relevance: Here, the focus is on the manipulation of the token and how the children hovered them over the underlying ladder structure. This accompanied their thinking and might have helped them to engage in a reflective conversation with the evaluation tools. This can positively impact emergent cognitive processes, and the quality of the rating outcomes.

Comparing and Considering Items Simultaneously

V6 (Paper Ladder): The researcher asks the child to rate the "Construction Blocks" in terms of "difficulty." The child says that the "Construction Blocks" is very easy and places the token on the 1st rung. The researcher asks her to rate the "Musical Chair Game." She says that it is also very easy, placing its token on the 2nd rung. The researcher clarifies that she can place more than one token on the same rung. The child then moves the "Musical Chair Game" token to the 1st rung, where the "Construction Blocks" token is. Next, the researcher asks the child to rate "Painting." She says it is easy and places its token on the 2nd rung. The researcher reminds her that she can place the token with the others if she thinks that "Painting" is as difficult as the other activities. The child looks at the Paper Ladder, and shakes her head: "No, 'Painting' is a little more difficult than 'Construction Blocks' or the 'Musical Chair Game.'"

Relevance: Here the focus is on the simultaneous visual comparison of tokens (some of them on the same rung). This behavior was only observed with the Paper Ladder. Although unplanned, we realized that our Sticky Ladder did not support such behavior well. The single piece of Velcro in the middle of each rung did not support placing tokens on the same rung. We noted this affordance as important to consider for future Sticky Ladder tools.

Identifying Mistakes and Changing Responses

V7 (Sticky Ladder): The researcher asks the child to rate "Painting," and the child takes the "Construction Blocks" token. The researcher points out that the question was about "Painting." The child then places the "Construction Blocks" token back on the table and picks the "Painting" token. From there on, the child picks the right token.

V8 (Paper Ladder): The researcher asks the child if playing with "Construction Blocks" is difficult. He shakes the head, saying "No," and moves to place the token on the top of the Paper Ladder. The researcher then reminds him how the Ladder works. Then, the child moves the token to the bottom of the Paper Ladder.

Relevance: V7 illustrates an instance of rating the wrong activity; V8 shows a verbalization mismatching the rating. Situations like this were caught by the researcher.

Conversation

V9 (Paper Ladder): The researcher asks if the "Musical Chair Game" is "difficult." The child answers "Yes" and starts explaining how the game goes. After rating that activity, the researcher tells the child that she is going to rate the activities in terms of "enjoyment." The child says that he likes to play with "Construction Blocks" and paint with his father.

Figure 3. Children using the Sticky Ladder.

Relevance: the tangible tokens used with both ladders often triggered a conversation between the child and the researcher, creating a friendly atmosphere where both felt at ease. Conversations and the actual use of the tokens (see e.g. V7 and V8), helped identify and clarify possible misunderstandings or lack of concentration. This did not happen with the Five Degrees of Happiness.

Coded behavior

Overall, we coded a total of 173 gestures for the Sticky Ladder; 96 for the Paper Ladder and 51 for the Five Degrees of Happiness (see table 1). The Sticky Ladder supported more fiddling (F) and thinking gestures (TG) - which are interpreted here as a positive sign of focus and reflection- than the Paper Ladder and the Five Degrees of Happiness. Sliding behavior only happened for the Paper Ladder, which we relate to the lack of such affordance for the other tools. This made the responses in those other tools more permanent. We relate this permanency of responses to the higher fiddling (F) and hovering (H) behaviors, and thinking gestures (TG) of the Sticky Ladder in comparison to the Paper Ladder. Hovering was used equally for the Five Degrees of Happiness, and the Paper Ladder, doubling its number for the Sticky Ladder. We consider hovering behavior to be closely related to sliding, since both support projecting or visualizing the position of responses.

METHOD	TG *	F *	S*	H*	TOTAL
Sticky Ladder	73	53	0	47	173
Paper Ladder	27	26	21	22	96
5D Happiness	23	6	0	22	51

Table 1. Number of coded gestures for the three methods

Interestingly, the sum of hovering and sliding for the Paper Ladder is close to hovering behaviors for the Sticky Ladder.

DESIGN IMPLICATIONS

Here, we focus on the evaluation tools. We discuss interesting behavior with respect to their affordances, reflect on their strengths and shortcomings, and conclude with design recommendations for researchers interested in using these evaluation tools.

Relevant affordances

Hovering, and Thinking with Items from the Evaluation Tools

We observed this behavior with all three evaluation tools, with either the tokens (ladders), or the pen (Five Degrees of Happiness). Yet we see advantages of hovering and thinking with the tokens. They embody a symbolic representation of the activity, which can help the children to recall the activity being evaluated, and reduce cognitive load (not having to recall the activity). Also, the empty ladders can trigger the children's projections, which can help them to consider different positions (and hence rating) for the products under evaluation before making their final decision and locking or placing their tokens onto the ladder.

Fiddling

Fiddling seemed to help the children retrieve and articulate thoughts, process information, and keep attention and focus on the task. It happened with all three tools (mostly with the ladders' tokens, and the pen). Similar to the behavior above, we find more valuable fiddling with items that embody a symbolic representation of the activity evaluated.

Sliding Tokens

This behavior was only possible with the Paper Ladder, given the non-adhesive nature of the ladder and its tokens. As seen in the background, this behavior can be advantageous from an embodied and situated cognition perspective; the "slideability" of the paper tokens over the Paper Ladder's rungs can support children's *perception* [39] and help them to consider different answers. This feature can also help the children to perceive their answers as more provisional compared to the original Sticky Ladder, were the answers are more "locked" once the tokens are affixed to the ladder. This assumption is also empirically supported by observations in our previous work related to how the children changed their mind and their responses [62].

Strengths and Shortcomings of the Evaluation Tools

Overall, we found advantages of the ladders over the Five Degrees of Happiness. Here we summarize the main ones.

Identification of the Activity under Evaluation

All the children successfully identified the activities represented on the item cards and understood how to use the respective rating scales. However, the children seemed to understand the functioning of the Sticky Ladder and the Paper Ladder more easily than the Five Degrees of Happiness rating scale. This may align with previous research pointing out that fine-grained discriminations of emotions, and choosing between slightly different degrees of happiness, is often challenging for children [1,15,50]. In addition, the similarity of the depicted evaluation scales corresponding to different questions confused some children. After turning the page to continue with other questions, a child said quickly: "We have done that already." Also, several children tried to use the same scale to mark more than one answer.

This did not happen with the ladders. After rating all the activities for a question, the children would "reset" the ladder, removing all tokens from it. No children raised the question of whether they had already rated the activities.

Handling the Tools' Items

As for the Five Degrees of Happiness scale, several children had issues when marking it. For some, especially the youngest ones, it was clear that writing a mark on a particular space was beyond their skills (see Figure 4).

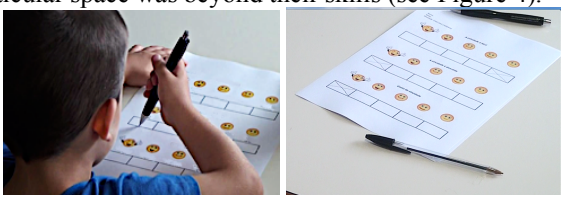

Figure 4. A child using the Five Degrees of Happiness scale.

One child asked the researcher to draw the marks for her, saying she could not yet do it. A child (male, 6.02 years) commented: "It's more difficult to answer a question [using a pencil]". In general, the children had to concentrate to draw their marks, and some had difficulties holding the pen upright. We observed several recurring issues, including children: trying to use all the space in the box assigned to a particular scale point; drawing ambiguous (+ or X) or inconspicuous symbols in a box; taking a long time to mark the boxes; and having general difficulties with writing marks in the response boxes (e.g. the children would hold the questionnaire paper with the left hand, draw a small line, then rotate the paper to the other side to draw another line composing an "X").

Revocable Responses

The ladders also afforded changing one's mind, as shown in V7 and V8. Here, the slideability of the Paper Ladder can offer advantages over the tokens with Velcro in the Sticky Ladder. In both cases, the children can (and did) (re)consider their preferences and (re)organized the items. These re-arrangements are valuable to the researcher to prompt the children to think aloud and discuss their actions and thinking [5]. This did not happen with the Five Degrees of Happiness; once the children marked one choice, they never tried to change it. The written mark first, followed by the tokens with Velcro second, are answers more "locked" to the scales than sliding tokens.

Altogether, the representation of the ladder and the tokens seemed to have enabled the children to consider and reconsider different answers before making up their mind. However, with the Sticky Ladder, the children tended to not move tokens once they were stuck onto the ladder. This can be in part because removing a token with Velcro was not straightforward, and at times it was even slightly difficult,

resulting in longer interaction times (on average the evaluation sessions with the Sticky Ladder took about one third longer than with the Paper Ladder). In turn, the absence of a gluing mechanism and the "slideability" of the paper tokens in the Paper Ladder seemed to support an understanding that choices were more provisional, encouraging both intended and unintended re-positioning.

Items Manipulation: Explanation, Engagement, and Visibility of Thought

Touch and manipulation were central in the ladder groups. The researcher used them to introduce the study and demonstrate usage of the evaluation tool; at the same time, the children gestured and touched the tokens and the ladder, e.g. climbing the rungs with the fingers. This indicates exploration, perception and processing [71]. Some children associated the tokens with the activities immediately, even before they were asked about them. Their engagement with the tangible rating scales was also notable. Most of them actively placed and removed the tokens, resembling a game play. The visual representation of the activities in the tokens helped the children to recall which was under evaluation.

Cultural Metaphor

Real world associations *mediated* by the Sticky Ladder and the Paper Ladder seemed to help the children to develop a better understanding of the tool. Ladders are part of the children's everyday lives. A child (5.09 years, male) commented, *"*I like the ladder! I don't have ladders at home, just one to climb to my bed.*"* Some children are familiar with relevant cultural associations, like attributing added value to "climbing up," or "being on top" which we found advantageous to explain and understand the tool. Furthermore, and based on prior research in embodied and situated cognition [38,39], we speculate that the physical structure of an empty ladder activates children's mental image of positioning and evaluating activities, making the evaluation easier. The ladder, specially the one we used with the Sticky Ladder allowed holding it vertically, thus facilitating understanding the ladder metaphor, and relevant cultural associations about ladders.

Sustainability

The Paper and the Sticky Ladders are more sustainable than the Five Degrees of Happiness scale, which required three pages per child. Hence, the more children, the more paper waste. Once the Paper or the Sticky Ladder and their tokens are created they can be reused with as many children as necessary, as long as they last in good state.

Design Recommendations

Here we list a series of design recommendations that result from our experience using these tools with our target group:

Paper Ladder

- Print a ladder that fits on an A4 paper with wider rungs. This will allow placing tokens together on the same rung;
- Glue the A4 paper with the printed ladder onto a cardboard. It will allow holding the ladder vertically to evoke associations to real live ladders, explaining and understanding the rating method;
- Draw the floor below the ladder, to also strengthen real-world ladder associations.
- Glue the paper pieces with the printed activities onto cardboard cutouts, which are easier to handle, more like objects, more robust, and durable than paper pieces.

Sticky Ladder

- Create rungs that are "velcroable" in their entirety (and not just the center) and long enough to allow several tokens on the same rung.
- Avoid gluing the Velcro onto the cloth as this will not be sturdy enough to endure several placing/removing of the items under evaluation.

Five Degrees of Happiness Rating Scale

- Visualize the activity being rated either through a drawing next to the questionnaire scale, or using tokens (e.g. placing them next to the questionnaire);
- For sustainability reasons, use a tablet to capture answers. Alternatively, a re-usable notebook sized whiteboard.

CONCLUSION

We studied embodied behavior of preschoolers with several evaluation tools: The Five Degrees of Happiness, the Sticky Ladder; and the Paper Ladder. They rated three creative and play activities in terms of difficulty, enjoyment, and preference. We discussed how the children's embodied behaviors were supported by affordances of the tools; strengths and shortcomings of each evaluation method, and presented recommendations for their design, appropriation, and usage by researchers working with children. Due to its affordances, we consider that the Ladders are especially suited for carrying out game user research with children. For example, the Sticky and the Paper Ladders can be used to evaluate design game elements, such as fun and enjoyment [8,75], preference for particular game characters, narrative turns of a digital game, or key game core mechanics [74]. While this remains future work for us, we will encourage other design researchers to explore this possibility. In future work we will use the Paper Ladder to access the children's preferences along the design of an interactive tool that we are currently developing.

Finally, we have made the Paper Ladder freely available for download under following link:
http://mobeybou.com/evaluation-tools/

ACKNOWLEDGMENTS

We thank all the children, their parents, the teachers, and the school board. The first author acknowledges the financial support from CIEC (Research Centre on Child Studies), with the Strategic Project UID/CED/00317/2013, co-financed through national funds from the FCT (Foundation for Science and Technology) and European Regional Development Funds (FEDER) through the Competitiveness and Internationalization Operational Program (POCI) with the reference POCI-01-0145-FEDER-007562.

REFERENCES

[1] Sharon Airey, Lydia Plowman, Daniel Connolly, and Rosemary Luckin. 2002. Rating children's enjoyment of toys, games and media. In Proceeding of the 3rd World Congress of the International Toy Research Association on Toys, Games and Media, 1–7.

[2] Jackie Andrade. What does doodling do? Applied Cognitive Psychology 24, 1: 100–106.

[3] Ahmed Sabbir Arif and Cristina Sylla. 2013. A comparative evaluation of touch and pen gestures for adult and child users. In Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13, 392–395. https://doi.org/10.1145/2485760.2485804

[4] David P. Barash. 1974. Human Ethology: Displacement Activities in a Dental Office. Psychological Reports 34, 3: 947–949.

[5] Wolmet Barendregt, Mathilde M. Bekker, and Ester Baauw. 2008. Development and evaluation of the problem identification picture cards method. Cognition, Technology & Work 10, 2: 95–105. https://doi.org/10.1007/s10111-007-0066-z

[6] Ida Bark, Asbjørn Følstad, and Jan Gulliksen. 2006. Use and usefulness of HCI methods: results from an exploratory study among nordic HCI practitioners. In People and Computers XIX — The Bigger Picture. Springer London, London, 201–217. https://doi.org/10.1007/1-84628-249-7_13

[7] Alice Bell. 2007. Designing and testing questionnaires for children. Journal of Research in Nursing 12, 5: 461–469. https://doi.org/10.1177/17449871079616

[8] Francesco Bellotti, Bill Kapralos, Kiju Lee, Pablo Moreno-Ger, Riccardo Berta. 2013. Assessment in and of Serious Games: An Overview. In Advances in Human-Computer Interaction. http://dx.doi.org/10.1155/2013/136864

[9] Peter L. Berger, Thomas Luckmann. 1991. The social construction of reality. London: Penguin Books.

[10] Chris J. Boyatzis and Malcolm W. Watson. 1993. Preschool Children's Symbolic Representation of Objects Through Gestures. Child Development 64, 3: 729–735.

[11] Norman Brosterman. 1997. Inventing Kindergarten. New York: Harry N. Adams Ine.

[12] Vivien Burr. 1995. An Introduction to Social Constructionism. London: Routledge. https://doi.org/10.4324/9780203133026

[13] Roger Caillois. 1961. Man, play, and games. University of Illinois Press.

[14] Suzanne B. da Câmara, Rakshit Agrawal, and Katherine Isbister. 2018. Identifying Children's Fidget Object Preferences: Toward Exploring the Impacts of Fidgeting and Fidget-Friendly Tangibles. In Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18), 301–311. https://doi.org/10.1145/3196709.3196790

[15] Christine T. Chambers. 2002. Developmental differences in children's use of rating scales. Journal of Pediatric Psychology 27, 1: 27–36. https://doi.org/10.1093/jpepsy/27.1.27

[16] Kathy Charmaz. 2000. Grounded theory objectivist and constructivist method. In Denzin, N. and Lincoln, Y. (Eds.), Handbook of Qualitative Research (pp. 509-535). Thousand Oaks, CA: Sage.

[17] Yoram I. Chisik, Alissa N. Antle, Brian Birtles, Elena Márquez Segura, and Cristina Sylla. 2014. The Kathmandu kids entertainment workshops. In Entertaining the Whole World, Adrian David Cheok, Anton Nijholt and Teresa Romão (eds.). Springer, London, 5–21. https://doi.org/10.1007/978-1-4471-6446-3_2

[18] John Cloud. 2009. Better Learning Through Fidgeting. Time 173, 14: 61–61.

[19] Allison Druin. 1999. Cooperative inquiry: developing new technologies for children with children. In Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99, 592–599. https://doi.org/10.1145/302979.303166

[20] Allison Druin. 2002. The role of children in the design of new technology. Behaviour & Information Technology 21, 1: 1–25. https://doi.org/10.1080/01449290110108659.

[21] James Farley, Evan Risko, and Alan Kingstone. 2013. Everyday attention and lecture retention: the effects of time, fidgeting, and mind wandering. Frontiers in Psychology 4. https://doi.org/10.3389/fpsyg.2013.00619

[22] Michael W. Galbraith, and Waynne B. James. 1987. The relationship of education level and perceptual learning styles. Journal of Adult Education, 15, 2: 27-35.

[23] Traci L. Galinsky, Roger R. Rosa, Joel S. Warm, and William N. Dember. 1993. Psychophysical Determinants of Stress in Sustained Attention. Human Factors 35, 4: 603–614. https://doi.org/10.1177/001872089303500402

[24] Stephen Garger. 1990. Is there a link between learning style and neurophysiology. Educational Leadership 48, 2: 63–65.

[25] James J. Gibson. 1979. The Ecological Approach to Visual Perception. Houghton Mifflin.

[26] Annette M. La Greca and Wendy L Stone. 1992. Assessing children through interviews and behavioral observations. In Handbook of Clinical Child

Psychology (2nd ed.), Clarence Eugene Walker and Michael C. Roberts (eds.). John Wiley & Sons, Oxford, England, 63–83.

[27] Lynne Hall, Colette Hume, and Sarah Tazzyman. 2016. Five degrees of happiness: effective smiley face Likert Scales for evaluating with children. In Proceedings of the The 15th International Conference on Interaction Design and Children - IDC '16, 311–321. https://doi.org/10.1145/2930674.2930719

[28] Libby Hanna, Kirsten Risden, and Kirsten Alexander. 1997. Guidelines for usability testing with children. interactions 4, 5: 9–14. https://doi.org/10.1145/264044.264045

[29] Christian Heath, Jonathan Hindmarsh, and Paul Luff. 2010. Video Analysis and Qualitative Research. Sage. Retrieved June 7, 2015 from https://kclpure.kcl.ac.uk/portal/en/publications/video-analysis-and-qualitative-research(3159d1c9-6d67-4513-9615-be695f6fbe33)/export.html

[30] Jeanne L. Higbee, Earl J. Ginter, and William Douglas Taylor. 1991. Enhancing Academic Performance: Seven Perceptual Styles of Learning. Research and Teaching in Developmental Education 7, 2: 5–10.

[31] Malcolm Hill. 2006. Children's voices on ways of having a voice: children's and young people's perspectives on methods used in research and consultation. Childhood 13, 1: 69–89. https://doi.org/10.1177/0907568206059972

[32] James Hollan, Edwin Hutchins, and David Kirsh. 2000. Distributed Cognition: Toward a New Foundation for Human-computer Interaction Research. ACM Trans. Comput.-Hum. Interact. 7, 2: 174-196. https://doi.org/10.1145/353485.353487

[33] Jean Hunleth. 2011. Beyond on or with: questioning power dynamics and knowledge production in "child-oriented" research methodology. Childhood 18, 1: 81–93. https://doi.org/10.1177/0907568210371234

[34] Hélène Joffe. 2008. The Power of Visual Material: Persuasion, Emotion and Identification. Diogenes 55, 1: 84–93. https://doi.org/10.1177/0392192107087919

[35] Brigitte Jordan and Austin Henderson. 1995. Interaction Analysis: Foundations and Practice. The Journal of the Learning Sciences 4, 1: 39–103.

[36] Victor Kaptelinin and Bonnie Nardi. 2012. Affordances in HCI: Toward a Mediated Action Perspective. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12), 967–976. https://doi.org/10.1145/2207676.2208541

[37] Michael Karlesky and Katherine Isbister. 2016. Understanding Fidget Widgets: Exploring the Design Space of Embodied Self-Regulation. In Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16), 38:1–38:10. https://doi.org/10.1145/2971485.2971557

[38] David Kirsh. 1995. Complementary strategies: why we use our hands when we think. In Proceedings of the 17th Annual Conference of the Cognitive Science Society - CogSci '95, 212–217.

[39] David Kirsh. 2013. Embodied cognition and the magical future of interaction design. ACM Transactions on Computer-Human Interaction 20, 1: 1–30. https://doi.org/10.1145/2442106.2442109

[40] H. van Laerhoven, Hester J. van der Zaag-Loonen, and B H F Derkx. 2004. A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. Acta paediatrica 93, 6: 830–835. https://doi.org/10.1111/j.1651-2227.2004.tb03026.x

[41] James A. Levine, Sara J. Schleusner, and Michael D. Jensen. 2000. Energy expenditure of nonexercise activity. The American Journal of Clinical Nutrition 72, 6: 1451–1454. https://doi.org/10.1093/ajcn/72.6.1451

[42] Elena Márquez Segura. 2015. Co-creating embodied sketches playing as a method to design with children. In Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology - ACE '15, 1–7.

[43] Elena Márquez Segura, Laia Turmo Vidal, Asreen Rostami, and Annika Waern. 2016. Embodied sketching. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16, 6014–6027. https://doi.org/10.1145/2858036.2858486

[44] Malia F Mason, Michael I Norton, John D Van Horn, Daniel M Wegner, Scott T Grafton, and C Neil Macrae. 2007. Wandering minds: the default network and stimulus-independent thought. Science 315, 5810: 393–395.

[45] Joanna McGrenere and Wayne Ho. 2000. Affordances: Clarifying and Evolving a Concept. Proceedings of the Graphics Interface 2000 Conference, p. 179-186.

[46] Edward F. Melcer and Katherine Isbister. 2018. Bots & (Main)Frames: Exploring the Impact of Tangible Blocks and Collaborative Play in an Educational Programming Game. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, Paper 266, 14 pages. DOI: https://doi-org.ezproxy.its.uu.se/10.1145/3173574.3173840

[47] David Mellor and Kathleen A. Moore. 2014. The use of Likert scales with children. Journal of Pediatric Psychology 39, 3: 369–379. https://doi.org/10.1093/jpepsy/jst079

[48] Maria Montessori. 1917/2008. Spontaneous Activity in Education. http://www.gutenberg.org/files/24727/24727-h/24727-h.htm. Retrieved 29 April 29th, 2017.

[49] Donald A. Norman. 1999. Affordance, conventions, and design. interactions 6, 3 (May 1999), 38-43. DOI=http://dx.doi.org.ezproxy.its.uu.se/10.1145/301153.301168

[50] Janet C. Read and Stuart MacFarlane. 2002. Endurability, engagement and expectations: measuring children's fun. Interaction Design and Children 2: 1–23. https://doi.org/10.1.1.100.9319

[51] Janet C. Read, Stuart MacFarlane, and Chris Casey. 2003. What's going on?: discovering what children understand about handwriting recognition interfaces. In Proceeding of the 2003 conference on Interaction design and children - IDC '03, 135. https://doi.org/10.1145/953536.953555.

[52] Janet C. Read, Stuart MacFarlane. 2006. Using the Fun Toolkit and Other Survey Methods to Gather Opinions in Child Computer Interaction. In Proceedings of the 2006 Conference on Interaction Design and Children (pp. 81-88). New York, NY: ACM Press.

[53] Mitchel Resnick, Brad Myers, Kumiyo Nakakoji, Ben Shneiderman, Randy Pausch, Ted Selker, and Mike Eisenberg. 2005. Design Principles for Tools to Support Creative Thinking. NSF Workshop Report on Creativity Support Tools, (Creativity Support Tools), 25–35.

[54] Christine M. Rubie-Davies and John A. C. Hattie. 2012. The dangers of extreme positive responses in Likert scales administered to young children. The International Journal of Educational and Psychological Assessment 11: 75–89.

[55] Katie Salen and Eric Zimmerman. 2003. Rules of Play: Game Design Fundamentals. The MIT Press, Cambridge, MA, USA.

[56] Donald A. Schön. 1984. The Reflective Practitioner: How Professionals Think in Action. Basic Books, New York, NY, USA.

[57] Thomas A. Schwandt. 2003. Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics and social constructionism. In Denzin, N. and Lincoln, Y (Eds.), The Landscape of Qualitative Research: Theories and issues. (pp. 292-331). Thousand Oaks, CA: Sage.

[58] Maxine Sheets-Johnstone. 1999. The Primacy of Movement. John Benjamins Publishing, Amsterdam, the Netherlands.

[59] Maxine Sheets-Johnstone. 2011. Embodied minds or mindful bodies? A question of fundamental, inherently inter-related aspects of animation. Subjectivity 4, 4: 451–466. https://doi.org/10.1057/sub.2011.21

[60] Donald Slater. 2011. An Innovative Use of Fidget Toys in a University Classroom. In SoTL Commons Conference, 2. Retrieved from https://digitalcommons.georgiasouthern.edu/sotlcommons/SoTL/2011/45

[61] Donald Slater and Jean French. 2010. Fidget toys in the classroom: refocusing attention.

[62] Cristina Sylla, Ahmed Sabbir Arif, Elena Márquez Segura and Eva Irene Brooks. 2017. Paper Ladder: A Rating Scale to Collect Children's Opinion in User Studies. In Proceedings of the 19th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '17). ACM, New York, NY, USA. https://doi.org/10.1145/3098279.3122139

[63] Cathy Treadaway, Gail, Kenning. Designing Sensory e-Textiles for Dementia. 2015. In Proceedings of the Third International Conference on Design Creativity (3rd ICDC). http://hdl.handle.net/10369/7470

[64] Alfonso Troisi, Sergio Belsanti, Anna R. Bucci, Cristina Mosco, Fabiola Sinti, and Monica Verucci. 2000. Affect regulation in alexithymia: an ethological study of displacement behavior during psychiatric interviews. The Journal of Nervous and Mental Disease 188, 1: 13–18.

[65] Lev S. Vygotsky. 1967. Play and its role in the mental development of the child. Soviet Psychology, 5: 6-18.

[66] Lev S. Vygotsky. 1978. Mind in society: The development of higher psychological processes. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, Massachusetts: Harvard University Press.

[67] Heinz Werner and Bernard Kaplan. 1963. Symbol formation. New York: Wiley.

[68] James V. Wertsch. 1985. Vygotsky and the social formation of mind. Cambridge, MA: Harvard University Press.

[69] James V. Wertsch. 1998. Mind as Action. Oxford University Press.

[70] Christopher D. Wickens, Sallie E. Gordon, and Yili Liu. 2004. An Introduction to Human Factors Engineering. Pearson Prentice Hall, London, UK.

[71] Frank R Wilson. 1998. The hand: how its use shapes the brain, language, and human culture (1st ed.). New York: Pantheon Books.

[72] Margaret Wilson. 2002. Six views of embodied cognition. Psychonomic Bulletin & Review 9, 4: 625–636. https://doi.org/10.3758/BF03196322

[73] Harry J. Witchel, Carlos P. Santos, James K. Ackah, Carina E. I. Westling, and Nachiappan Chockalingam.

2016. Non-Instrumental Movement Inhibition (NIMI) Differentially Suppresses Head and Thigh Movements during Screenic Engagement: Dependence on Interaction. Frontiers in Psychology 7. https://doi.org/10.3389/fpsyg.2016.00157

[74] Annika Wolff, Paul Mulholland, Zdenek Zdrahal, and Richard Joiner. 2007. Re-using digital narrative content in interactive games. International Journal of Human-Computer Studies 65, 3: 244–272. https://doi.org/10.1016/j.ijhcs.2006.10.003

[75] Geogios N. Yannakakis and John Hallam. 2004. Evolving opponents for interesting interactive computer games. In Proceedings of the International Conference on Computer Games: Artificial Intelligence, Design and Education.

[76] Li-Yuan Wu. 2013. Children's Play and Symbolic Representation. Review of Global Management and Service Science 3:8.

[77] Bieke Zaman, Vero Vanden Abeele, and Dirk De Grooff. 2013. Measuring product liking in preschool children: an evaluation of the Smileyometer and This or That methods. International Journal of Child-Computer Interaction 1, 2: 61–70. https://doi.org/10.1016/j.ijcci.2012.12.001.