An Annotation Scheme for Social Interaction in Digital Playgrounds

Alejandro Moreno, Robby Van Delden, Dennis Reidsma, Ronald Poppe, Dirk Heylen

HAL Id: hal-01556166

https://inria.hal.science/hal-01556166

Submitted on 4 Jul 2017

An Annotation Scheme for Social Interaction in Digital Playgrounds⋆

Alejandro Moreno, Robby van Delden, Dennis Reidsma,
Ronald Poppe, and Dirk Heylen

Human Media Interaction, University of Twente,
P.O. Box 217, 7500 AE Enschede, The Netherlands
[a.m.moreno, r.delden, d.reidsma, r.w.poppe, d.k.j.heylen]@utwente.nl

Abstract. This paper introduces a new annotation scheme, designed specifically to study children's social interactions during play in digital playgrounds. The scheme is motivated by analyzing relevant literature, combined with observations from recordings of play sessions. The scheme allows us to analyze how key social interactions are related to different stages of play and the physical activity levels associated to them. We can use this information for two goals. First, we can identify relations between social interactions and the impact that changing game dynamics have on their occurrence. Second, it facilitates the analysis of automatic recognition of these social behaviors. Results obtained are useful for both goals. They show that it is possible to identify social interactions and their relation to game dynamics. Finally they also allow for further analysis into the possibility of their automatic recognition.

Keywords: Observation scheme, children's play, social interaction, digital playground

1 Introduction

"... civilization is, in its earliest phases, played. It does not come from play... it arises in and as play, and never leaves it." Huizinga [9] stated that play has a major role in shaping human behavior. Even though people of every age play, children's play in particular has attracted a lot of interest in research fields such as health and well-being, psychology, medicine, among others. Some studies have identified a relation between obesity and children's current play habits [21]. Another study showed that children who fail to engage in social interaction during play, or show irregularities while doing so, can be linked to several specific mental or social impairments such as autism or mental retardation [7].

Children's play often takes place at playgrounds. Playgrounds can take many forms ranging from the traditional school or park playground to novel interactive digital playgrounds. Digital playgrounds differ from traditional ones in that they are augmented with different kinds of technologies that can enhance engagement,

---

⋆ This publication was supported by the Dutch national program COMMIT.

entertainment, social interaction, physical exertion and immersion [3,15]. For instance, Wyeth et al. [23] designed an interactive floor to promote social and physical interaction among mentally challenged people. Soler and Parés designed an interactive slide to counter the lack of physical activity and socialization of children [18]. Seitinger et al. developed the Interactive Pathway to research how children's play patterns could be affected by modifying the playground itself [16].

We aim to develop a new type of interactive playground that is able to automatically sense social behavior during children's play and respond to it. To accomplish this we need data to train algorithms for the recognition of social behavior. This means that reliable ground truth needs to be obtained through annotation schemes. These schemes serve three main purposes. First, they would allow digital playgrounds to be capable of online, automated sensing of interesting behavior, trained from prior offline annotation. Second, they facilitate the evaluation of reaching higher end goals such as learning social skills, facilitating play for specific target groups [23] or increasing physical interaction [12]. Third, they would help in understanding the impact of the responses of the playground on social behavior. However, due to the lack of annotation schemes, progress in the field is hampered [1]. Moreover, existing observation schemes are either very general (applicable to different settings but lacking depth) or are tailored and designed with specific settings in mind (only applicable in the intended setting).

We propose an observation scheme designed specifically to evaluate children's social interactions during play in digital playgrounds. It was designed by analyzing theories and schemes from literature and recorded play sessions. For the latter, we used the Interactive Playground, a digital playground developed to encourage physical activity, competition and cooperation [20]. Our scheme provides an in-depth observation methodology focusing on the social domain that allows the annotation of key social interactions along with useful information such as physical activity. From a bottom-up processing standpoint, it would allow for online, automatic social signal processing by identifying low level features related to specific social behaviors (modeled from offline annotations). From a top-down processing standpoint, it would facilitate evaluating the impact of modifying game dynamics such as success in promoting/reducing specific behavior, enhancing/reducing levels of engagement, immersion and fun. It must be noted that we present a scheme to describe social behavior, not a user experience evaluation scheme, which will help design socially aware digital playgrounds.

2   Existing Observation Schemes

Below we present a list of some of the widely used or recently created observation schemes focused on children's play in playgrounds. For more schemes to measure (anti-)social behavior from a clinical standpoint, please refer to [4,10].

2.1   Play Observation Scale

The Play Observation Scale (POS) was proposed by Rubin in 1989 and revised in 2001 [14]. Rubin tried to combine the two mainstream theories used in studying
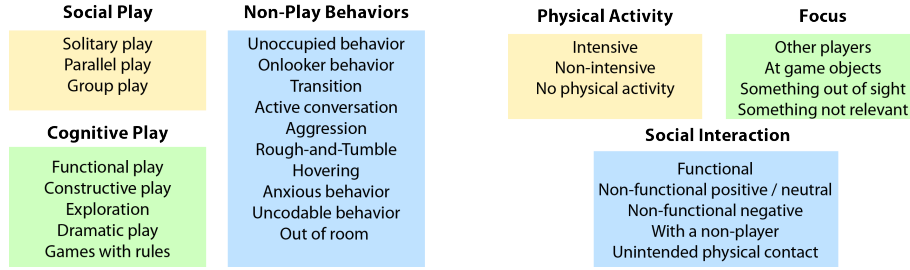
Social Play

Solitary play
Parallel play
Group play

Non-Play Behaviors

Unoccupied behavior
Onlooker behavior
Transition
Active conversation
Aggression
Rough-and-Tumble
Hovering
Anxious behavior
Uncodable behavior
Out of room

Cognitive Play

Functional play
Constructive play
Exploration
Dramatic play
Games with rules

Physical Activity

Intensive
Non-intensive
No physical activity

Focus

Other players
At game objects
Something out of sight
Something not relevant

Social Interaction

Functional
Non-functional positive / neutral
Non-functional negative
With a non-player
Unintended physical contact

Fig. 1. Play Observation Scale composition.

Fig. 2. Outdoor Play Observation Scheme composition

children's play: the social one proposed by Parten and the cognitive one promoted by Piaget. The purpose of this scale is to assess children's free play preferences. The scale is separated into three main play categories, each of which contains different aspects of play (Figure 1). The social play category deals with grouping and children's interactions, the cognitive play category is related to the type and role of play, and finally the non-play behaviors category addresses a broad range of items including social interactions as well as items outside of the actual play. The cognitive play category is nested into the social play category e.g. the child can show solitary-functional play, group-constructive play, etc.

The annotation is done by observing a child for 10 seconds, and then selecting the predominant behavior within this time lapse from a list of predefined actions. No child should be observed for more than five successive minutes to avoid subjective annotations. In total at least 15 minutes should be analyzed, thus several sessions are needed to be able to evaluate a single child.

According to Rubin, the scale is useful in identifying age and sex differences in play, effect of ecological setting in play and identifying children 'at risk' of developing psychological problems. This scheme was conceived to be very general, which has the advantage of making it applicable under different contexts to study children's play. However, when in need of more detailed behavior analysis, the categories are too coarse. For instance the distinction between competitive and collaborative play is not clearly present in the scheme. Also, there is no measurement for physical activity. Non-play behaviors do include some relevant social behaviors, such as aggression, rough-and-tumble and onlooker/unoccupied behavior.

2.2   Outdoor Play Observation Scheme

The Outdoor Play Observation Scheme (OPOS) was proposed by Bakker et al. in 2008 [1] and can be seen in Figure 2. They stated it was imperative for the HCI community to develop a structured methodology for observation of play. They were specifically interested in an observation scheme applicable in the evaluation of Head-Up games (HUGs) which are instrumented games for children that can be played without the need to look down [19].

In OPOS, two different types of approach are considered: event coding (frequency-based approach) and state coding (duration-based approach). Event coding is related to event-sampling; i.e., registering specific point-events such as yelling. On the other hand, state coding refers to actions that are continuous in nature, or at least have a measurable duration such as running. They also decided it was beneficial to analyze every child individually to avoid missing important information not measurable when taking the whole group as the basic unit.

OPOS is composed of four different behavioral categories. One of the categories was added just for practical purposes such as annotating children being out of sight. This category, named "general", is therefore not present in Figure 2. The physical activity category refers to the level of physical exertion exhibited by the players, the focus category indicates the player's gaze object of interest, and lastly the social interaction category deals with communication (verbal and non-verbal) in the broadest sense of the word.

OPOS was tested by evaluating a game called Lighthouse. The authors obtained a multi-rater κ measure of .7 for physical activity, .45 for focus and .24 for social interaction when comparing individual classified items among three coders. When comparing the total amount of occurrences of any of the items, all κ's increased (.9, .74, .73 respectively). Even though the scheme proves useful in finding differences in play, we find that the social interaction category is too coarse (functional vs non-functional only) for our purpose. For instance, there is no way to indicate social interactions such as competition/collaboration, rough-and-tumble or leadership. The physical activity category, however, is useful.

2.3   Manchester Inventory Playground Observation

Manchester Inventory Playground Observation (MIPO) was proposed by Gibson et al. in 2011 with the goal of evaluating children's play for clinical analysis [6]. They tried to determine social skills problems in children during play since previous studies showed that developmental disorders were associated to social functioning difficulties [7]. Playgrounds, which are fairly challenging social environments that require complex social interactions, lend themselves for useful studies of real time social information processing.

MIPO was designed to be a real-time social interaction observation scheme usable both for clinical analysis and research in general. It was structured so that it could identify several different patterns of social impairments across several disorders. It is composed of four main categories of peer socialization; each one containing different observations resulting in 28 different items as seen in Figure 3. The pro-social skills category mainly deals with grouping behavior and inclusion. The conflict management deals with types of aggression. Confiding and care concerns helping behavior. Atypicality deals with indicators of unusual, unwanted social behavior such as shouting out for no reason.

Each of the items was rated on a 4-category, rank-ordered rating scale that went from socially competent in this domain to marked difficulties. The annotators first observed the behavior of a randomly chosen child from a control group with no psychopathologies.
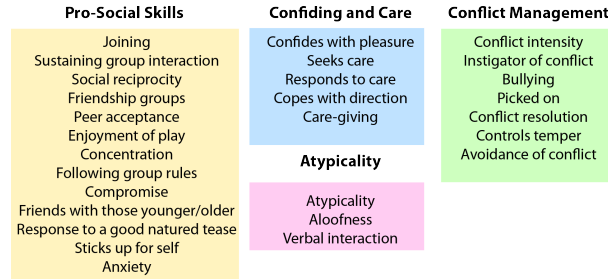
Pro-Social Skills	Confiding and Care	Conflict Management
Joining	Confides with pleasure	Conflict intensity
Sustaining group interaction	Seeks care	Instigator of conflict
Social reciprocity	Responds to care	Bullying
Friendship groups	Copes with direction	Picked on
Peer acceptance	Care-giving	Conflict resolution
Enjoyment of play		Controls temper
Concentration	**Atypicality**	Avoidance of conflict
Following group rules		
Compromise	Atypicality	
Friends with those younger/older	Aloofness	
Response to a good natured tease	Verbal interaction	
Sticks up for self		
Anxiety		

Fig. 3. Manchester Inventory Playground Observation composition

We found that MIPO was one of the most complete observation schemes in terms of social interactions. However, it neglects other areas relevant to children's play such as the phases of play and the level of physical activity. Evidently, since the scheme was designed specifically for clinical identification of social skill impairment, there was no need to observe other type of actions or behaviors.

2.4   Overall Conclusions

We found that current schemes provide several useful categories and items such as physical activity and some non-play behaviors (rough-and-tumble, aggression, etc). However, we found that the social categories did not suffice for our purpose. We are interested in studying a wide array of social behaviors (similar to MIPO), but also how they are related to play itself. We noticed that current schemes are built around play implicitly; they observe play through related categories but never address it as is. Morrison et al. proposed a set of terms to categorize phases that may emerge during play in the context of interactive art installations [13]. They proposed terms such as speculative play, interactive play or comprehension stages that we found applicable to describe play in interactive playgrounds as well. We decided that studying real play sessions in a digital playground was necessary to define other important items that could be added to the scheme.

3   The Interactive Playground

The Interactive Playground was implemented by Tetteroo in 2010. It was designed to stimulate physical activity, collaborative and competitive behavior during children's play. It is composed of a top-down projector and a near-infrared camera located on the ceiling, several hats with infra-red reflectors on it, balls and wristbands worn by the children, both with wireless motion sensors (Sun SPOTS). The projection on the floor of the room has several shapes like triangles, circles and squares that can be interacted with. Music is played and several sound effects are also emitted when certain events happen. The most important system interactions are listed below:

– If a player with a ball enters the playground he is surrounded with a colored shape; if entering without a ball, a short trail of shapes will follow him.

- If a player chases a free shape long enough the shape is added to his trail.
- If a player shakes the ball it will produce 'venom', killing nearby player's shapes. For every shape killed, his colored shape will grow.
- If a player with a ball is being chased by another player with a ball, he will lose the last shape of his trail to the chasing player.

3.1   Experimental Setup

Children were instructed to play in groups of two to four befriended children. A play session was composed of two sub-sessions of about ten minutes and a short break between the two sub-sessions. During the break, the functioning of some of the system interactions was explained to the children. In total, 39 children (aged between six and twelve) and around four hours of play were recorded. The system has been evaluated in a darkened controlled environment. At the end of every play session, a group discussion was carried out to get a better insight in the children's level of understanding of the provided interactions and their experiences. A complete description can be found in [20].

3.2   General Behavioral Observations

Tetteroo noticed that although some interactions were targeting collaboration, most interactions witnessed were based on individual play with the system or were competitive by nature. Nonetheless, even when the children were engaged in individual play, some social interaction between players still occurred. For instance, explaining out loud how certain interactions worked, which sometimes led to responses from the other players.

Tetteroo concluded that most interactions with the system could be triggered easily and were attainable in several seconds mainly because of the limited size of the playground. This might have led to the relatively low occurrence rates of cooperation and intensive physical activity. Although children were observed sitting or standing still, the predominant behavior was non-intensive physical activity. Because of the diverse interactions, children took some time to realize / understand how the playground responded to their actions and explored with different possibilities. This, in turn, gradually led to further discussion amongst them and the discovery of new interaction forms.

An interesting observation is the remarkable number of times that rough-and-tumble play occurred. Children were seen pushing one another, chasing / kicking each other and frantically throwing balls at each other. Very often the children were trying to step on the shapes on the floor and, sometimes, on other children's feet. They were also observed dragging each other around when someone fell down or was tripped by a sliding tackle. Sometimes, but not often, violent behavior was exhibited; i.e., having the intention to physically harm another. Aside from the physical contact, some children executed performances like swimming on the floor, spins, tricks, and dances. They were sometimes followed by others that showed the same kind of behavior on their own initiative or after being told or hinted to do so by another player.

Phase of Play	Social Behavior
Not playing [not] Exploration [exp] Playing (with the system) [pws] Playing (not with the system) [pns]	Leadership [lea] Rough-and-Tumble [rat] Violence [vio] Helping [hel] Performance [per] Mimicking [mim] Cooperation [coo] Competition [com] Other social behavior [osb]
Physical Activity	
None [non] Normal [-] Intense [int]	

Fig. 4. Final annotation scheme composition with abbreviations between brackets.

4   Proposed Observation Scheme

The information gathered from observing play in the targeted context helped define and corroborate items that should be available in the scheme. We designed an observation scheme for interactive playgrounds with a focus on social interactions between children. Our goal is to analyze children's play from a social interaction standpoint; identifying the different interactions present and studying how playground characteristics influence their occurrence. The scheme borrows concepts from the previously discussed schemes, social and psychological theoretical background, but also proposes new categories and observable behaviors especially in the social domain. The scheme was refined iteratively in two rounds (a pilot and final evaluation). The proposed scheme can be seen in Figure 4.

The scheme is composed of three main categories covering the social interactions we want to focus on, the higher end goal of increasing physical activity and the relation to the phase of play a child is in. The social interactions included in the scheme were chosen because they occur often, their relation to play is important, can possibly be influenced by changing game dynamics and hopefully can be automatically detected.

Our observations were carried out on a per child basis to prevent generalizing and avoid the risk of missing important individual behavior such as normally inactive participants suddenly engaging in play, increasing physical activity or social interaction [1,23]. Also, one of the main annotators did not use the sound of the footage, neither in the pilot nor the final annotation, to analyze the suitability of using computer vision algorithms for automatic recognition in the future.

4.1   Units of Annotation

When annotating, the first decision to make is whether predefined time slots or manually annotating the exact duration of an action is more beneficial. Using time slots makes the annotation easier and faster, increasing the efficiency of the annotation. It also allows the annotation of several children concurrently. However, when only one label can be assigned to a time slot and two or more actions take place simultaneously, the dominant one should be chosen, which can lead to discrepancies. One could solve this by allowing multiple labels in one time slot when the nature of the category allows this. On the other hand, annotating the exact duration of an action gives more accurate labeling (useful

for machine learning) but is much more time consuming since each child needs to be annotated individually and the entire video needs to be watched for each of the children. If time slots are chosen, another issue would be to define the duration of the time slots.

We noticed that for phase of play, a longer duration is useful since phases will not change that rapidly (we chose eight second sequences). In the case of social behavior, smaller sequences are more appropriate since most are usually completed rather quickly (we chose two second sequences). Lastly, physical activity will mostly show normal intensity, with bursts of intense/no activity, and thus the duration of the bursts was manually annotated. This led to the problem of thresholding which will be explained in the next section.

4.2   Detailed Description

Phase of play refers to the different stages that a child can go through when playing and are based on [13]. This dimension gives an indication of how simple the game is and whether children will play it. The observable behaviors are annotated exclusively, which means only one of them can be labeled in each time slot. One might find relationships between the phase of play and certain social behaviors. These findings could help in preventing or triggering certain social behaviors and in improving social signal processing.

Social behavior is the category that identifies different social interactions that occur when children play. Arguably, this might be one of the most important aspects of play. According to Bandura's *Social Learning Theory*: "Learning would be exceedingly laborious, not to mention hazardous, if people had to rely solely on the effects of their own actions... most human behavior is learned observationally through modeling: from observing others..." [2, p22] . In the same line, Vygotsky's *Social Development Theory* encompasses that social interaction plays a fundamental role in the process of cognitive development [22]. Lastly, Erikson's *Psychosocial Development Theory* defines a specific stage (*macrosphere play*) where children master social interactions and realize they can be successful in the larger social world [5].

The social behavior dimension contains two items based on POS: *Violence* (based on their item of aggression) and *Rough-and-Tumble*. Making the distinction is important because, according to Hughes [8], aggressive play should not be discouraged as it allows children to release their anxiety. It should, however, be controlled. He also mentions that play should be nonliteral, which is related to the *performance* item (such as dancing or doing flips) which was observed quite often in the recordings and is also mentioned in [12]. *Helping* can be linked to the more detailed caring category of MIPO. *Mimicry*, besides being a specific interest of our research, is mentioned in [13]. Erikson mentioned that during the third stage of psychosocial development (Initiative vs. Guilt) children need to lead or direct play and social interactions in their attempt to assert their power over the world [5]. *Leadership*, *cooperation* and *competition* were witnessed in the recordings and therefore found relevant to code in relation to play. The last item, *Other social behavior* (OSB) refers to any interaction not covered by the other

items in the social behavior dimension, that would imply that 2 or more children are interacting and belong to a group for the duration of this interaction.

The social behavior dimension may inform whether the playground, or parts of the playground, increase these social interactions or if unwanted social behavior like violence could be diminished. At the same time, it paves the way for a more socially aware digital interactive playground. These behaviors are non-exclusive and thus multiple labels can be assigned simultaneously.

Physical activity categorizes the amount of physical effort that the children are doing at a given time. In [8], Hughes mentions that play should keep children actively engaged, involved psychologically and physically. This conveys why this category is one of the common goals in interactive playgrounds. One might benefit from automatic measurement of this dimension by using, for instance, the number of steps [11], motion energy images or skeleton tracking. This is an exclusive category.

The complete table with an explanation of the specific observable behaviors can be found in the supplemental material provided[1]. Potential uses of this scheme include possible identification of ambiguous social signals, physical cues related to certain social interactions, identifying how playground characteristics affect social interactions and its effect on the phases of play, among others.

5   Analysis and Evaluation

5.1   Training and Annotation Setup

We annotated a subset of the videos using our proposed scheme. The annotation was performed using ELAN [17] from the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands. Four annotators were involved in the coding. Two of the annotators, closely involved in the creation of the scheme, participated in both sessions. Two external annotators, one participating in the pilot and the other one in the final evaluation, had no participation in the creation of the scheme. The two main annotators had a training session of coding three minutes of footage. The two external annotators were given an explanation of the scheme and its use followed by a training session and a discussion of the labeled fragments. For the pilot scheme training, the explanation lasted 30 minutes and the training video lasted 30 seconds. For the final scheme training, the times were twice as much to provide a better explanation. The training sessions covered most items; examples of the unseen items from the scheme were shown by fast forwarding through the training video.

For the pilot annotation, all annotators went through 10 minutes of video, separated into three videos of three minutes and 20 seconds each. The subsets of the three videos covered the beginning, middle and end of different sessions, respectively. For the final evaluation, only three minutes and 20 seconds were annotated. As explained before, one of the main annotators did not use the sound of any of the footage.

---

[1] `hmi.ewi.utwente.nl/playground-annotation-scheme-manual.pdf`

5.2   Shortcomings of the Pilot Scheme

Comparing annotators in the pilot study resulted in rather low Cohen's κ-values for all the dimensions. For a dimension such as social behavior, which can be quite ambiguous or subjective, this could be expected. However, the physical activity [11] and grouping dimension are considered objective dimensions and should have obtained a high agreement. Phase of play, on the other hand, should have obtained at least a modest agreement.

We identified some problems both in the scheme design and evaluation procedure. In regards to the design, we noticed that some items were not defined explicitly which led to ambiguities when annotating. For example, it was not stated if *rough-and-tumble* was to be used for every physical contact, intentional or not, or if it should be labeled for both children or only the one doing the contact. This problem was solved with a more comprehensive description of each item in the scheme manual and a more thorough training session. Besides this, we had also included a grouping dimension which created a lot of confusion due to distance restrictions between children and amount of interaction required to be labeled a group. Since physical grouping can be obtained accurately using computer vision, manual annotation seemed redundant. We decided to move towards a higher level meaning of grouping where intention, and not location, was the focus. We added the *other social behavior* item to the social behavior dimension and removed the grouping dimension.

In regards to the evaluation procedure, the physical activity dimension was first coded using defined time slots of eight seconds. However, agreeing on which was the dominant behavior for some time slots was problematic. To try to overcome this problem, we manually labeled bursts of *intensive* or *none* physical activity in the final version.

5.3   Results and Analysis of Final Scheme

Phase of Play Dimension had a fair agreement between the two main annotators (κ of .41). However, it had a low reliability when comparing the external annotator (κ of .08 and .09). When grouping together *playing with the system* and *exploration*, the resulting κ with the external annotator went up to .53 and .51 respectively (see Table 1). Tetteroo's Digital Playground is designed to allow children to explore and interact freely with the system, find new interactions, explore and interact with the system again. Our results show that differentiating *playing with the system* from *exploration* will be challenging in playgrounds that promote free play and emerging interactions due to the numerous phase transitions in the sessions. In these settings, combining both items would yield a higher inter-observer reliability while still providing useful information about the phase of play. For games with more strict or structured phase transitions it will be easier to annotate this dimension.

It is noteworthy that audio played some role in the ability to correctly annotate this dimension. For instance, sometimes children were asking "how can you get the Pacman?", clearly indicating it was an exploratory endeavor that

Table 1. Confusion matrix for the category of phase of play of the two main annotators, κ is .41

	not	exp	pws	pns	-
not	**0**	0	0	0	0
exp	0	**35**	15	1	0
pws	0	8	**28**	0	0
pns	0	0	0	**0**	0
-	0	7	3	0	**3**

Table 2. Confusion matrix for the physical activity category of the two main annotators, κ is .32

	non	int	-
non	**23**	0	2
int	0	**30**	4
-	26	21	

Table 3. Confusion matrix for the physical activity category of the external annotator and main annotator A, κ is .0.5

	non	int	-
non	**24**	0	2
int	0	**30**	7
-	11	6	

Table 4. Confusion matrix for the physical activity category of the external annotator and main annotator B, κ is 0.29

	non	int	-
non	**31**	0	5
int	0	**29**	6
-	24	23	

could not be coded correctly without the use of audio. Nonetheless, the degree to which audio affected the results was not high.

Physical Activity Dimension had a low agreement between the main annotators (κ of .32), especially considering physical activity ought to be an objective measurement, shown to have high agreement when observers are intensively trained using strict protocols [11]. When comparing the external annotator to the main ones, we obtain κ values of .5 and .32. Two important observations were seen i)*intensive* and *none* items were never confused (see Tables 2 and 4) and ii) whenever either annotator A or the external annotator labeled a burst of activity, annotator B almost always agreed (92% for none and 88% for intensive with annotator A, 86% and 83% for the external annotator). Thus, we can conclude that the problem is the threshold at which the annotators separate intensive activity from normal, and normal from none. For very intensive or obvious lack of physical activity, the agreement is extremely high, but in cases where the activity is somewhere in between, agreement is lacking.

Agreeing on a threshold for physical activities proves quite challenging: first, because expressing this with words is not trivial, second, because listing every action the children could execute is near to impossible. Using for instance up to two steps as threshold for no physical activity would not suffice as children might be intensively shaking using their entire upper body while standing still. The reliability of this dimension will depend on the type of playground, or on being able to define strictly the threshold between the level of activities before the annotation. For playgrounds with a lot of freedom of movement it will be hard to annotate this dimension.

Social Behavior Dimension had a high agreement between the main annotators with a κ of .67 (see Table 5). On the other hand, agreement between

the external annotator and main annotators A and B was κ 0.28 and 0.37 (see Table 6) respectively. Agreeing on when a behavior happened was problematic, thus we also analyzed, when both annotators agreed some behavior was underway, if they agreed on which one it was. The results showed that the agreement was quite high considering this is a subjective dimension (External annotator - main annotator A 0.66, External annotator - main annotator B 0.67, two main annotators 0.92).

Table 5. Confusion matrix for the social category of the two main annotators, κ is .67 (when taking out the empty slots, κ is 0.92)

	-	lea	rat	vio	hel	per	mim	coo	com	osb
-	**204**	0	3	0	0	9	0	1	12	6
lea	8	**2**	1	0	0	0	0	1	0	0
rat	3	0	**67**	0	0	0	0	0	0	0
vio	0	0	1	**1**	0	0	0	0	0	0
hel	5	0	0	0	**2**	1	0	0	0	2
per	5	0	0	0	0	**18**	0	0	2	0
mim	1	0	0	0	0	0	**5**	0	0	0
coo	5	0	0	0	0	0	0	**3**	0	0
com	2	1	0	0	0	0	2	0	**35**	0
osb	27	0	0	0	0	0	0	1	1	**22**

Table 6. Confusion matrix for the social category of main annotator B and the external annotator, κ is 0.37 (when taking out the empty slots, κ is 0.67)

	-	lea	rat	vio	hel	per	mim	coo	com	osb
-	**186**	6	1	0	3	3	3	1	12	14
lea	5	**5**	0	0	1	0	0	1	0	0
rat	34	1	**27**	0	1	2	0	3	0	2
vio	0	0	2	**0**	0	0	0	0	0	0
hel	4	0	0	0	**6**	0	0	0	0	0
per	9	1	1	0	1	**9**	0	0	2	2
mim	6	0	0	0	0	0	**0**	0	0	0
coo	1	0	0	0	0	0	0	**7**	0	0
com	19	1	0	0	0	0	0	0	**19**	1
osb	27	2	0	0	3	0	2	3	6	**8**

Rough-and-tumble was the most frequently annotated behavior and it was often accompanied by signs of *competition* (although the external annotator failed to see several occurrences of quick physical contact) which shows that identifying relations between social interactions is possible. *Performance* was also seen often during various parts of the game where the kids wanted to show off or just have more fun. There were some discrepancies in *leading* but when the videos were analyzed, all but one could be accounted to not using audio by one of the annotators. It is interesting that actually all but one coded instance of leadership were accompanied with speech. The influence of speech was also seen in *help*, where helping verbally could not be annotated without audio.

It can be seen in Tables 5 and 6 that interestingly, *competition* and *cooperation* were not confused even without the use of sound. In many cases speech contained information about the correct social signal, however it was often ac-

companied with clear physical cues. We noticed that various types of social interactions can be identified accurately without using audio but could benefit from it nonetheless. This seems promising for the automatic recognition of social signals in digital playgrounds using computer vision. For instance, sudden invasion of personal space by other children could signal aggressive behavior or competition, whereas synchrony in movement and actions could signal cooperation.

6   Conclusions

We have proposed a new annotation scheme for observational studies in digital playgrounds. The scheme is composed of three dimensions and focuses on children's social interactions during play, and their relation to play itself. It was designed using relevant elements from current schemes, psychology and sociology theories and observations from recorded play sessions. On average, the agreement was fair which means the scheme still needs to be refined. Nonetheless, results show that the scheme is useful in identifying relations between different social interactions. This provides information directly related to the impact that changing certain game dynamics to promote a specific kind of interaction can have on the game. They also show that, for certain social interactions, automatic recognition could be possible with a reasonable accuracy even without the use of audio. The scheme was designed having our own goals in mind, however it remains generic enough to be used in other settings. We regard the current state of the scheme as a good starting point, either as a base scheme for more specific ones, or for further development in creating a general socially focused annotation scheme for the evaluation of digital playgrounds.

References

1. Bakker, S., Markopoulos, P., de Kort, Y.: OPOS: an observation scheme for evaluating head-up play. In: Proceedings of the Nordic conference on Human-computer interaction. pp. 33–42. NordiCHI '08, New York, NY, USA (2008)
2. Bandura, A.: Social Learning Theory. Prentice Hall (1977)
3. Bichard, J.P., Waern, A.: Pervasive play, immersion and story: designing interference. In: Proceedings of the international conference on Digital Interactive Media in Entertainment and Arts. pp. 10–17. DIMEA '08, New York, NY, USA (2008)
4. Cook, F., Oliver, C.: A review of defining and measuring sociability in children with intellectual disabilities. Research in Developmental Disabilities 32(1), 11–24 (2011)
5. Erikson, E.H.: Childhood and Society. W. W. Norton & Company (1993)
6. Gibson, J., Hussain, J., Holsgrove, S., Adams, C., Green, J.: Quantifying peer interactions for research and clinical use: the Manchester Inventory for Playground Observation. Research in Developmental Disabilities 32(6), 2458–66 (2011)
7. Guralnick, M.J., Hammond, M.A., Connor, R.T., Neville, B.: Stability, Change, and Correlates of the Peer Relationships of Young Children With Mild Developmental Delays. Child Development 77(2), 312–324 (2006)
8. Hughes, F.: Children, Play, and Development. SAGE (2009)

9. Huizinga, J.: Homo Ludens: A Study of the Play Element in Culture. Roy Publishers (1950)
10. Leff, S.S., Costigan, T., Power, T.J.: Using participatory research to develop a playground-based prevention program. Journal of School Psychology 42(1), 3–21 (2004)
11. Loprinzi, P.D., Cardinal, B.J.: Measuring Children's Physical Activity and Sedentary Behaviors. Journal of Exercise Science & Fitness 9(1), 15–23 (2011)
12. Lund, H.H., Jessen, C., Klitbo, T.: Playware technology for physically activating play. Artificial Life and Robotics 9(4), 165–174 (2005)
13. Morrison, A., Viller, S., Mitchell, P.: Building sensitising terms to understand free-play in open-ended interactive art environments. In: Proceedings of the conference on Human factors in computing systems. pp. 2335–2344. CHI '11, New York, NY, USA (2011)
14. Rubin, K.H.: The Play Observation Scale (POS). Published by the Center for Children, Relationships, and Culture of the University of Maryland. (2001), `http://www.rubin-lab.umd.edu/CodingSchemes/`
15. Seitinger, S.: An ecological approach to children's playground props. In: Proceedings of the conference on Interaction Design and Children. pp. 117–120. IDC '06, New York, NY, USA (2006)
16. Seitinger, S., Sylvan, E., Zuckerman, O., Popovic, M., Zuckerman, O.: A new playground experience: going digital? In: CHI '06 on Human factors in computing systems. pp. 303–308. Montréal, Québec, Canada (2006)
17. Sloetjes, H., Wittenburg, P.: Annotation by Category: ELAN and ISO DCR. In: LREC (2008), `http://www.lat-mpi.eu/tools/elan/`
18. Soler-Adillon, J., Parés, N.: Interactive slide: an interactive playground to promote physical activity and socialization of children. In: Proceedings of the international conference on Human factors in computing systems. pp. 2407–2416. CHI EA '09, New York, NY, USA (2009)
19. Soute, I., Markopoulos, P., Magielse, R.: Head Up Games: combining the best of both worlds by merging traditional and digital play. Personal Ubiquitous Comput. 14(5), 435–444 (2010)
20. Tetteroo, D., Reidsma, D., van Dijk, Nijholt, A.: Design of an interactive playground based on traditional children's play. In: Proceedings International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2011). Genoa, Italy (2011)
21. Vandewater, E.A., Shim, M., Caplovitz, A.G.: Linking obesity and activity level with children's television and video game use. Journal of Adolescence 27(1), 71–85 (2004)
22. Vygotsky, L.S.: Mind in Society: The Development of Higher Psychological Processes. Harvard University Press (1978)
23. Wyeth, P., Summerville, J., Adkins, B.: Stomp: an interactive platform for people with intellectual disabilities. In: Proceedings of the International Conference on Advances in Computer Entertainment Technology. ACE '11, New York, NY, USA (2011)