TOWARD A VALID AND RELIABLE MEASURE OF PLAY:

THE DIMENSIONS OF PLAY FRAMEWORK

By

Aubrey Henriksen

Amye R. Warren
Professor of Psychology
(Chair)

David E. Ferrier
Assistant Professor of Psychology
(Committee Member)

Jayne Griffin
Director of Education at Creative Discovery Museum
(Committee Member)

TOWARD A VALID AND RELIABLE MEASURE OF PLAY:

THE DIMENSIONS OF PLAY FRAMEWORK

By

Aubrey Henriksen

A Thesis Submitted to the Faculty of the University of
Tennessee at Chattanooga in Partial
Fulfillment of the Requirements of the Degree
of Master of Science: Psychology

The University of Tennessee at Chattanooga
Chattanooga, Tennessee

December 2020

ABSTRACT

While it is established that play is important for development, researchers do not have a measure that examines both play type and complexity. Existing measures are limited to examining play type without complexity, complexity of a specific play type, or ethnographic measures. Based on a combination of scales in the literature, I developed the Dimensions of Play Framework as a tool for observational quantitative measurement of play in children's museums. The purpose of this initial study was to investigate interrater reliability using a time-sampling method. Video tapes of 140 children interacting with either a simple or elaborate version of a clay exhibit were collected for analysis. Multiple rounds of coding by undergraduate assistants demonstrated moderate agreement yet failed to yield acceptable interrater reliability for hypothesis testing (Kappa > .70). Results indicate more should be done to structure the framework and determine the appropriate method of measurement to accompany it.

ACKNOWLEDGEMENTS

I want to thank my thesis chair, Dr. Amye Warren, not only for the support and guidance she has offered to me throughout this journey, but also for the seemingly endless energy and passion that she has for the entire psychology department. I would also like to express gratitude to my thesis committee members, Dr. David Ferrier and Jayne Griffin, for serving on my committee and offering insightful comments and questions that pushed me to new levels of understanding. I also want to express gratitude for my research assistants, Madeleine Condoure, Claudia Divincenzo, Hannah Kollmansperger, and Faith Humberd for their time spent and feedback throughout the coding process. Finally, I want to thank my partner, Chris Scoggin, and all of my friends for buoying me when I felt I was sinking and reigniting my fire when I wanted to give up. Their support has been an endless source of strength and resolve.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

While there is widespread agreement among psychologists and educators that play throughout childhood and beyond is important for learning and development (Bateson, 2015; Bergen, 2015; White, 2012), researchers do not currently have a clear framework for quantitatively measuring both play type and complexity (Fagen, 2015). Existing definitions and theories of play are limited in 1 of several ways:

1. They utilize a list of criteria to distinguish between playful behaviors and not playful behaviors ignoring play type and complexity.
2. They examine play type but not complexity.
3. They examine the complexity levels of a particular play type.
4. They conflate play type with complexity.
5. They rely on linguistic ethnographic methods.

Therefore, an empirical measure of the dimensions of play that can be applied across ages and contexts has yet to be developed and tested. The development and utilization of such a measure would be of particular interest to those who create play-based environments, such as children's museum professionals.

Children's museums are a conducive environment to study play for several reasons: children go to children's museums intending to play (Dooley & Welch, 2014), children's museums are designed to invoke high quality play behaviors, and children's museum

professionals could benefit from measuring play because it could help them to improve their exhibit designs (Krasnor & Pepler, 1980). According to Fagen (2015, p. 96), "Ecological resources affect play and in doing so they affect development." Children's museums design their exhibits to foster learning and development through play. The Association of Children's Museums advocates for play as the most effective approach to facilitate children's learning because it is engaging, experiential, and autolytic (Association of Children's Museums, 2012; Brown & Vaughan, 2009), and most children's museums feature play as a key part of their mission statement (Luke, Letourneau, Rivera, Brahms, & May, 2017). The intention to facilitate play-based experiences sets children's museums apart from other types of museums (Henderson & Atencio, 2007) and makes children's museums an ideal environment to study play in a natural context.

Children's museum exhibits are developed to facilitate play, but designers do not currently have an empirical method to measure play in exhibits. For example, an observational quantitative measure of play could aid decision making for exhibit designers during the prototyping phase, where exhibit designers are making choices between different designs and labels. Does an exhibit label with explicit step-by-step instructions or a label with creative prompts facilitate more high-quality play? How many props in an exhibit are enough to elicit high quality play? What are the differences in play quality for children who experience a perfectly set exhibit versus an exhibit that is a mess? An observational quantitative measure of play could also lead to the ability to demonstrate more explicit links between the play that occurs in an exhibit and the outcomes of that experience (Fagen, 2015). It could also help children's museum professionals to make the best use of museum resources. Therefore, children's museum

exhibits are an opportune environment for the development of an empirical measure of play. However, play has historically proved difficult to define and measure (Lillard, 2014).

Early Definitions and Theories

The earliest work relevant to this research is social play behavioral categories. In 1932, Parten focused on children's social development and defined the development of social play as progressing through 5 stages: solitary, onlooker, parallel, associative and cooperative (Parten, 1932). In solitary play, children play by themselves. In onlooker play, children observe the play of others. In parallel play, children play in close proximity to others and often with similar materials. In associative play, there is interaction between the child and others along with similar activity. And finally, in cooperative play children work together with a common purpose or within a shared reality. This scale has been successfully used in conjunction with other play behavioral category scales to form a more complete picture of play behaviors (Rubin, 1977).

After Parten, observational play research in humans fell out of fashion for a couple of decades as researchers turned to other methods such as questionnaires and interviews (P. K. Smith, 2015). A resurgence of observational play research in humans came after researchers were inspired by the ethological approaches of other disciplines. Observational research by psychologists began to pick up in natural settings such as labs, childcare centers, playgroups, schools and homes (P. K. Smith, 2015). The first major contribution came from Piaget (1962) who studied the development of play behaviors using systematic, detailed observations of children and their cognitive development. He defined play as progressing through 3 behavioral stages, sensorimotor, symbolic, and games with rules, corresponding with the first 3 stages of his theory of cognitive development (Piaget, 1962).

During the period of sensorimotor play, children use their senses and their body to explore and interact with their environment. Through playful repetition of actions, the child begins to separate the meaning of an action from the action itself, preparing the way for Piaget's second stage of play. During the symbolic play stage, children are able to engage in imitation or representation with a simple mental model. As the mental models become more complex, the child moves from being self-focused to incorporating different points of view with the play becoming more based in objective reality. Piaget's final stage of play is games with rules, where children play using established rules and regulations based in a shared reality with other players (Piaget, 1962).

Following Piaget's work, Smilansky (1968) conducted further systematic observations and elaborated on these 3 stages. She divided play behaviors into 4 categories: functional, constructive, dramatic, and games with rules. She used this scale to examine the differences between children from advantaged versus disadvantaged sociocultural backgrounds. Her functional play category corresponds with Piaget's sensorimotor play, in that it involves exploration of the environment through acting and sensing. However, she saw room for discrimination between 2 distinct types of Piaget's symbolic play: constructive play and dramatic play. Constructive play includes building and creating, while dramatic play includes pretending and symbolism. Games with rules corresponds with Piaget's stage by the same name. Smilansky's need to differentiate between constructive play and symbolic play suggests that it may be useful to further separate the dimensions of play type from play complexity.

Rubin (1977) took Smilansky's categories and used them in conjunction with Parten's social play categories to code preschool children's naturalistic classroom play. Rubin utilized the combination of the scales to investigate the relationship between social play level and cognitive

level and assessed which types of play children engage in while playing in different contexts. Of relevance is that he found preschool children operating mostly in the onlooker, solitary, and parallel categories, and that they engaged in less high-quality social play than kindergarten children. Preschool children also engaged in more functional play than constructive and dramatic play. Rubin's work is some of the most compelling evidence of the potential for a valid and reliable scale of children's play dimensions. Unfortunately, Rubin's work moved away from play after this era, and much of his research is inaccessible due to the rapid evolution of technology between 1977 and now (K. H. Rubin, personal communication, September 2018).

The Current State of Play Research in Humans

Currently, play is regularly in the news as a vital ingredient to healthy development. For example, the American Academy of Pediatrics recently published a clinical report instructing pediatricians to give parents information on how to encourage healthy development through play (Yogman, Garner, Hutchinson, Hirsh-Pasek, & Golinkoff, 2018). They cite several well-documented benefits to play including improved executive functioning, language, early math skills, social development, peer relations, physical development, health, and sense of agency.

Current methods of studying play vary across disciplines and contexts. They draw from many different fields including psychiatry, early childhood education, children's museum evaluation, and developmental psychology (Howes, 2015). A recent major contribution to play research emerged from the National Institute for Play (Brown & Vaughan, 2009), which focuses on studying play, utilizing play therapy to treat depression, and collecting and analyzing thousands of case studies. The founder of the National Institute of Play, Stuart Brown, has defined categories of play as body and movement play, object play, imaginative play, social play,

storytelling and narrative play, and transformative-integrative and creative play (Brown & Vaughan, 2009). A few of Brown's defined categories have overlapping qualities not ideal for observational quantitative coding research. For example, imaginative play, storytelling and narrative play involve many of the same behaviors, and social play can occur across all of the other play categories (Brown & Vaughan, 2009; Howes, 2015; Pellegrini, 2015; Rubin, 1977).

Some early childhood experts have examined the progression of play complexity within specific play types. For example, Leong and Bodrova (2012) defined stages for assessing imaginative play. They created a list of what they viewed as critical aspects of imagination-focused play: planning, roles, props, extended time frame, language, and scenario. They then outlined how each of these aspects evolve through 5 stages of dramatic, imagination-focused play. They proposed that by identifying the stage of play quality a child was operating within, early childhood educators could scaffold the children's experience, helping the child progress to the next level (Leong & Bodrova, 2012)

During the first stage of Leong and Bodrova's (2012) model the child's pretend actions are defined by an object in the play environment. For example, a child picks up a plastic duck and makes quacking sounds. During the second stage, the child takes on a role. For example, the child cradles a baby and says they are the mommy. In the final 3 stages, the child acts in coordination with play partners and the roles have rules that can be violated; there is a level of regulation required. In these more advanced stages, these rules are agreed upon by the players before the play scenario begins. The goal of Leong and Bodrova's article (2012) was to encourage early childhood educators to identify the level at which children under their purview were playing and aid them to reach the next stage by bringing in new materials that would extend their play scenarios and modeling higher stages of imaginative play.

The museum field has also wrestled with researching play, and observational quantitative research measuring play behaviors within museum exhibits is sparse (Van Schijndel, Franse, & Raijmakers, 2010). In 2012, a research summary on play and learning was created for the Minnesota Children's Museum. It defined play categories as social, object, pretend, physical, and media and has been widely cited in the children's museum field (White, 2012). Within the literature review however, several of the categories of play overlap, and no framework or scale emerged that could be a basis for measurement. However, some museum researchers and evaluators have examined museum exhibit behaviors in several diverse ways (Van Schijndel et al., 2010).

Some have investigated children's museum play behaviors using descriptive ethnographic measures. For example, Shine and Acosta (2000) examined parent-child social play using linguistic analysis of speech data to explore why parents and children did not fully engage in the imaginative play suggested by a grocery store exhibit. They found that conflicting goals and a lack of following each other's cues resulted in parents and children not fully engaging in the exhibit the way it was intended. Parents tended to guide children, frame the scene, and attempted to engage in logico-mathematical skills training rather than engage in the imaginative play of their children (Shine & Acosta, 2000). These linguistic ethnographic measures provide a detailed account of the exhibit play behaviors, but they do not allow for quantitative comparison between different exhibit designs. Moreover, such methods are not useful for children whose verbal language is not fully developed.

Another recent study examined play type but not quality. The Columbus Art Museum used Brown's typology of play behaviors to assess whether the Wonder Room exhibit met the established play objectives. They compared the various exhibit elements and found, for example,

that the mirror and costumes area evoked the highest amount of play behaviors and that pretend and social play were the most frequent types (Ancelet & Luke, 2015; Brown & Vaughan, 2009). This measure allowed for quantitative comparison between exhibit designs, but measurements were limited to type and not complexity of play.

Another recent study examined play quality of a single play type. Van Schijndel and colleagues (2010) used the Exploratory Behavior Scale to examine the effectiveness of different coaching strategies on children's play behavior at science museum exhibits. They had adults guide children through 2 exhibits using 1 of 3 coaching styles: minimal, explaining, and scaffolding. Minimal coaching involved very little interaction with the child at all. Explaining involved giving definitions or explanations about the processes of the exhibit activity. Scaffolding involved trying to push the child past their current abilities. They found, for example, that the minimal coaching style resulted in more advanced play behaviors at 1 exhibit, while the explaining coaching style resulted in more advanced play behavior at the other exhibit (Van Schijndel et al., 2010). This measure allowed for quantitative comparison between exhibit designs, but measurement was limited to the complexity of only 1 of several types of play, object play.

A recent review of the literature by psychologist Burghardt (2015) lists twelve different types of play: large motor play, small motor play, mastery play, rule-based play, construction play, make-believe play, symbolic play, language play, playing with the arts, sensory play, rough-and-tumble play, and risk-taking play. The categories are not mutually exclusive and conflate play type with complexity. For example, large-motor play can involve rule-based play and risk-taking play. He also synthesized the work of prior criteria-based approaches to differentiating between playful behavior and not playful behavior into the following 5 criteria.

1. “The performance of the behavior is not fully functional in the form or context in which it is expressed; that is it includes elements, or is directed toward stimuli, that do not contribute to current survival” (Burghardt, 2015, p. 13).
2. “The behavior is spontaneous, voluntary, intentional, pleasurable, rewarding, reinforcing or autotelic” (Burghardt, 2015, p. 14).
3. “It differs from strictly functional expressions of behavior structurally or temporally in at least 1 respect: incomplete, exaggerated, awkward, precocious, or involves behavior patterns with modified form, sequencing, or targeting” (Burghardt, 2015, p. 14).
4. “It is performed repeatedly in a similar, but not rigidly stereotyped, form during at least a portion of the animal’s ontology” (Burghardt, 2015, p. 15).
5. “It is initiated when an animal is adequately fed, clothed, healthy, and not under stress, or intense competing systems” (Burghardt, 2015, p. 16).

While this work is a thorough overview of the many forms that play can take and the criteria for determining whether behavior is play or not play, it is limited because the categories overlap and no framework for measurement is provided.

Creating the Dimensions of Play Framework

Existing measures cannot be used across children’s museum exhibit context because they are limited to examining play type without complexity, complexity of a specific play type, or using linguistic ethnographic measures. To address this problem, I developed the Dimensions of Play Framework as a tool for observational quantitative measurement of play in children's museums. It is impossible to examine all things at once. When a researcher chooses to examine play in a new context where there has been little prior observational quantitative research, a new

category system will likely need to be developed (P. K. Smith, 2015). The major obstacle that I had to overcome to create the dimensions of play framework was to disentangle the variety of overlapping definitions and theories and reorganize them into categories that were general enough to apply across exhibit contexts but specific enough to be useful for measurement (Lillard, 2014). I accomplished this by revising and adapting the prior models I have mentioned into a clear framework. Although I have made modest alterations to the labels and definitions used, the principles are the same.

Table 1

The Dimensions of Play Framework

Play Type	Play Complexity	Social Level
Object Focused	Exploration	Solitary
Body-Movement Focused	Imitation	Onlooker
Imagination Focused	Elaboration	Parallel
	Regulation	Associative
		Cooperative

The framework outlines 3 dimensions of play: type, complexity, and social level. Each dimension has multiple categories. The first dimension, play type, consists of 3 categories: body movement play, object play, and imagination play. In body movement play, the focus is on exploring how the body works and interacts with the world (Pellegrini, 2015; P. K. Smith, 2015). It begins in infancy with babbling and moving the arms and legs and continues to develop throughout childhood and into adulthood (Brown & Vaughan, 2009; Rubin, 1977; Smilansky, 1968; White, 2012) Toddlers walk, run, dance, and balance. Older children sing songs, play tag, do cartwheels, and climb (Pellegrini, 2015). As they grow into adulthood, they may take on

dance or marathons as a hobby. Through body-movement play, people develop their agility, spatial awareness, and knowledge of their body's abilities and limits (Brown & Vaughan, 2009).

In object play, the focus is on how objects work and interact with the world (Bjorkland & Gardiner, 2015). It begins in infancy with shaking rattles and squeezing teddy bears and continues to develop throughout childhood and into adulthood. Toddlers bang on toy instruments, fill and empty cups with sand, and make marks on paper with crayons. Older children build block towers, mold play dough into shapes, and sort play food into categories. As they grow into adulthood, they may take on engine building or sculpting as a hobby. Through object play, people develop conceptual knowledge such as cause and effect, measurement, classification, and quantification (Brown & Vaughan, 2009) as well as skills for tool use (Bjorkland & Gardiner, 2015).

In imagination play, the focus is on the use of symbolism to create roles and scenarios. It begins around the age of 2 or 3 with familiar roles such as mommy and baby. Toddlers pretend to feed and dress their baby dolls. Older children pretend to go grocery shopping, drive a fire truck, and build a campfire. As they grow into adulthood, they may take on writing stories or acting in plays as a hobby. Through imagination play, people develop their creativity, language, symbolic thinking, problem-solving skills, and social skills (Brown & Vaughan, 2009; Leong & Bodrova, 2012; Smilansky, 1968).

The second dimension of the framework, play complexity, consists of 4 categories: observation, replication, elaboration, and regulation. The first category is observation, where a person first uses their senses to explore and considers the opportunities and possibilities available to them (Vandenberg, 1980). They may observe the available materials and how other people are engaging with them. Their curiosity is sparked, or something resonates with them, and they make

a choice about what to engage with. They then explore the properties of the environment and materials and the possibilities for engagement using their senses (Rubin, 1977; Van Schijndel et al., 2010). For example, a child walks around a playground and sees a slide, monkey bars, and swings. They hear people laughing and the squeak of the swings. They see a person go down the slide, and it interests them. They walk over to the slide and touch it to see if it feels slippery. Then they watch another person go down. In this category of complexity, they observe the context and think about how they want to engage.

The second complexity category is replication, where a person follows prescribed instructions or imitates what another person is doing or has done. They may repeat something they did during their exploration to better understand what is happening. They further discover what is possible through repetition (Brown & Vaughan, 2009; Rubin, 1977; Smilansky, 1968; Van Schijndel et al., 2010). For example, the person follows the next person up the stairs and slides down the slide just as they did. They slide down over and over again in the same way. In this category of complexity, the replicate and repeat behaviors they have previously observed.

The third complexity category is elaboration. They alter their behavior to push the boundaries of possibility. This quality level consists of open-ended testing, creation, and innovation without restrictions (Brown & Vaughan, 2009; Rubin, 1977; Smilansky, 1968; Van Schijndel et al., 2010). For example, the person tries unusual ways of going down the slide. They go backwards, then laying on their belly, then laying on their back. They may test another way to the top such as climbing back up the slide. They may find a toy truck and roll it down the slide, first forwards, then backwards. In this category of complexity, they try out new ways and means.

The fourth and final complexity category is regulation. Individuals impose rules and restrictions on their play to create a more complex and challenging experience. They may say,

“The floor is lava. You can’t touch it when you get to the bottom of the slide or you die,” or “Let’s see who is faster, me or the truck.” During this category of complexity, they work towards a goal, or within a set of shared rules or a shared reality.

The primary objective of this study was to use inter-rater reliability to examine if it is possible to easily and effectively use this system of categories to measure play behavior within a children’s museum exhibit. My initial plan, if raters achieved good interrater reliability, was to also examine validity using a simple versus elaborate design manipulation of 3 exhibits, a body-movement play exhibit, an object play exhibit, and an imagination play exhibit. The simple design for each exhibit would include limited loose parts and little else while the elaborate design would include a larger quantity and variety of loose parts and parent prompts to scaffold their children from 1 play complexity level to the next. I hoped to use the comparison of the 3 types of exhibits as evidence of validity for the play type dimensions and the comparisons of the simple and elaborate designs as evidence of validity for the complexity dimension. However, due to unforeseen circumstances including a broken leg and a pandemic, I was only able to collect data at 1 exhibit, the object play exhibit.

CHAPTER II

METHODOLOGY

Participants

Participants were recruited during normal operations at Creative Discovery Museum in Chattanooga, TN. Parental consent was obtained for a total of 140 children between the ages of 1 and twelve. I collected video recordings from 10:00 AM to 12:00 PM on 7 weekend days between August of 2019 and February of 2020. When recording was taking place, I blocked the entrance to the exhibit with an informed consent checkpoint consisting of stanchions, ropes, a table, and a sign that read, “Please speak with a staff member before entering.” On the table I had a clock, a clipboard for writing down participant descriptions, and a stack of informed consent fliers. When a group approached the table I said, “Hello, today I am collecting data for my master’s thesis. It involves recording families while they normally interact with the exhibit. Would that be okay with you?” I then provided the informed consent letter with further details. See Appendices A, B and C for IRB approvals and informed consent letter. If they agreed to participate, I collected the children’s ages, whether or not they had visited the exhibit before, a description of their hair and clothing, and the time they entered the exhibit. If they did not wish to participate, they were invited to explore the exhibit when recording was over. Only 3 out of 140 declined participation, and none expressed frustration with the inconvenience. No signatures or any other identifying information was collected. They did not receive any incentive to participate.

A subset of 104 participants was selected for analysis using stratified random sampling. In order to appropriately balance the groups for each condition, participants were matched by age and gender as described in the procedure section. Overall, the sample was 60% female and their mean age was 4.67 years (*SD*=2.22), and 54% of participants hadn't visited the exhibit before while the other 46% had. Detailed demographic breakdowns by condition are shown in Table 2.

Table 2

Participant Demographics by Condition

Age in Years	Simple Condition		Elaborate Condition	
	Boy	Girl	Boy	Girl
1	3	0	3	0
2	3	2	3	2
3	4	6	4	6
4	4	5	4	5
5	1	6	1	6
6	3	4	3	4
7	2	4	2	4
8	0	2	0	2
9	0	2	0	2
10	0	0	0	0
11	1	0	1	0
12	0	0	0	0

Materials

The study took place at the clay table in the Visual Arts exhibit gallery at Creative Discovery Museum. In the simple design condition, materials at the clay table included 4 fist-sized balls of gray clay, 2 carving tools and a roller. In the elaborate design condition, materials included 6 fist-sized balls of clay in various colors, a 10-piece clay tool set, and 2 wooden busts for building clay onto.

I mounted 2 small wide-angle action cameras made by DBPower on the walls on either side of the clay table. Initially I transported recordings to UTC and transferred them to a hard drive which was kept in a locked cabinet in the lab. During this time research assistants coded the videos using Behavioral Observation Research Interactive Software, also known as BORIS (Friard & Gambo, 2012). When COVID-19 appeared and the labs were closed, the IRB granted permission to move the recordings to a password protected cloud folder. Only members of my thesis committee and my trained research assistants could view these recordings. Once I moved the videos to the cloud, they could no longer be accessed using BORIS because BORIS needed a local file in order to function. I then moved the coding to Excel.

Procedure

The first phase of analysis was the sampling procedure. I entered the potential participant's age and genders into a database and used SPSS to conduct random sampling in order to achieve an equal number of boys and girls of each age for each condition. I then edited the full video recordings down to participant specific videos where the video started when the participant touched the clay, table, or chairs and ended the moment that the participant left the

clay table. In the instance that the participant returned to the table multiple times, individual videos were made.

I trained the research assistants to code using 11 videos of children that were not a part of the participant pool. The training participant pool was 64% female and 36% male, 82% simple condition and 18% elaborate condition, and 36% had visited the exhibit gallery before. The average age of training participants was 6.27 (*SD*=3.35). I demonstrated how to code using 2 example videos, and the remaining 9 were assigned. The following week, coding differences were identified and discussed, and further clarification was given. The participant videos were then coded in 20% sections. The first 20% of participants were coded using BORIS. When COVID-19 appeared and the labs were closed, videos had to be moved to a secure cloud folder and could no longer be accessed using the BORIS software. Excel spreadsheets were created to facilitate coding from round 2 forward.

For each round I calculated Cohen's Kappa (1960) to determine consistency among raters. Cohen's Kappa examines percentage agreement corrected for chance agreements. Slight agreement is defined as Kappa = 0.00 to 0.20. Fair agreement is defined as Kappa = 0.21 to 0.40. Moderate agreement is defined as Kappa = 0.41 to 0.60. Substantial agreement is defined as Kappa = 0.61 to 0.80. Almost perfect agreement is defined as 0.81 to 1.00.

In the first round of coding, 2 undergraduate research assistants coded the first 20% of participant data. The round 1 participant pool was 50% female and 50% male, 100% simple condition, and 60% had visited the exhibit gallery before. The average age of round 1 participants was 4.45 (*SD*=2.14). Time-samples were recorded every 20 seconds throughout the duration of each video recording. For each time-sample, research assistants recorded a written

moment description and then used the ethogram to code the behavior occurring at the point of the time-sample. An interrater reliability analysis using the Kappa statistic was performed to determine consistency among raters for 1 behavior state, play, with 5 modifiers. Those modifier variables were play type, complexity, social level, parent role, and crowd size. Table 3 illustrates the ethogram for round 1 coding.

Table 3

Round 1 Coding Ethogram

Variable	Code	Code Description
Behavior state	Playing	The participant is engaged in playing with the exhibit such as exploring the clay using their senses, molding the clay with their hands, using tools to shape the clay, imitating how others are interacting with the clay and/or tools, creating something out of the clay.
	Not playing	The participant is not engaged in playing with the exhibit such as adjusting their chair or having an off-topic conversation.
	Not visible	View of the participant is obstructed.
Focus	Object	The play is focused on objects and their applications such as making, building, and creating.
	Body-movement	The play is focused on the body and its abilities such as singing, dancing, throwing, running, and tumbling.
	Imagination	The play is focused on imagination and symbolism such as object transformation, roleplay, pretend, being non-literal, and storytelling.
	Not playing	The participant is not exploring and/or interacting with the exhibit
	Not visible	View of the participant is obstructed.
Complexity	Exploration	The participant is primarily observing. For example, smelling, touching, tasting, listening.
	Repetition	The participant is copying the action of another, following directions, or repeating an action they've already done.
	Elaboration	The participant is doing a new action, changing something about an action, conducting open-ended experiments.

Table 3 Continued

	Regulation	The participant is playing within defined parameters, setting a goal and working towards it, challenging themselves/competing with others, or creating/following a new rule.
	Not playing	The participant is not exploring and/or interacting with the exhibit.
	Not visible	View of the participant is obstructed.
Social	Solitary	The participant is playing alone with their focus on their own activity.
	Onlooker	The participant is observing the play of another but not engaging.
	Parallel	The participant is playing with similar materials or in a similar manner in close proximity to the other's play.
	Associative	The participant is interacting with other's play in an unorganized way that lacks coordination or roles.
	Cooperative	The participant is interacting with other's play in an organized way with coordinated roles.
	Not playing	The participant is not exploring and/or interacting with the exhibit.
	Not visible	View of the participant is obstructed.
Parent	Player	The parent is playing near or with the participant.
	Facilitator	The parent is helping the child physically.
	Interpreter	The parent is helping the child verbally.
	Supervisor/Student	The parent is watching the child while seated at the table with the child.
	Not present	The parent is not present at the table.
	Not playing	The participant is not exploring and/or interacting with the exhibit.
	Not visible	View of the participant is obstructed.
Crowd	(Number)	Number of people at table including the participant
	Not playing	The participant is not exploring and/or interacting with the exhibit.
	Not visible	View of the participant is obstructed.

Because of the disagreement in parent role and after discussion with the coders, the ethogram was revised to include an “Other” category for parent role. In the second round of coding, the same 2 undergraduate research assistants coded the second 20% of participant data using this revised ethogram. The round 2 participant pool was 63% female and 37% male, 100%

simple condition, and 42% had visited the exhibit gallery before. The average age of round 2 participants was 5.05 (*SD*=2.57). I analyzed interrater reliability analysis using the Kappa statistic to determine consistency among raters.

In the third round of coding, 1 prior and 2 new undergraduate research assistants coded the third 20% of participant data using a new, simplified ethogram. The round 3 participant pool was 63% female and 37% male, 100% elaborate condition, and 26% had visited the exhibit gallery before. The average age of round 3 participants was 4.42 (*SD*=2.32). Time-samples were recorded every 20 seconds throughout the duration of each video recording. An interrater reliability analysis using the Kappa statistic was performed to determine consistency among the 2 of those raters who had the most experience with children. The 4 variables were play type, complexity, children present, and parent or guardian present. Table 4 illustrates the ethogram for round 3 coding.

Table 4

Round 3 Coding Ethogram

Variable	Code	Code Description
Complexity	Observing	The participant is watching, looking, listening, smelling, tasting, touching, or holding.
	Replicating	The participant is copying an action they observed another doing, doing an action they were instructed to do, repeating an action they did in a prior time sample
	Elaborating	The participant is experimenting with the materials without instruction, direction, or purpose, trying out new materials, or doing an action they haven't done in prior time samples.
	Regulating	The participant is setting a goal and working towards it, setting parameters and working within them, setting a challenge or competing, or showing off a result.
	Other	The participant is adjusting or moving their chair, having an off-topic conversation, leaving the table, or doing anything else that does not fit within the above categories.
Focus	Object	The play is focused on objects and their applications such as making, building, and creating.
	Body-movement	The play is focused on the body and its abilities such as singing, dancing, throwing, running, and tumbling.
	Imagination	The play is focused on imagination and symbolism such as object transformation, roleplay, pretend, being non-literal, and storytelling.
	Other	The participant is adjusting or moving their chair, having an off-topic conversation, leaving the table, or doing anything else that does not fit within the above categories.
Children	Present	There is at least 1 other child present at the table.
	Not present	There is NOT at least 1 other child present at the table.
Parent/ Guardian	Present	There is at least 1 parent or guardian of the child at the table.
	Not present	There is NOT at least 1 parent or guardian of the child at the table.

CHAPTER III

RESULTS

For the first variable, play or not play, the interrater reliability was substantial for round 1 of coding, Kappa = 0.64 (p = .000), 95% CI [0.40, 0.87] and moderate for round 2, Kappa = 0.48 (p = .000), 95% CI [0.34, 0.62]. Disagreement between playing versus not playing mostly involved moments when the participant paused to observe their surroundings. In round 3 this variable was eliminated and "not playing" was added as a category under the play type and play complexity categories.

For the second variable, play type, the interrater reliability was substantial for round 1, Kappa = 0.61 (p = .000) , 95% CI [0.37, 0.84], moderate for round 2, Kappa = 0.48 (p = .000), 95% CI [0.34, 0.62], and fair for round 3, Kappa = 0.34 (p = .000), 95% CI [0.11, 0.56]. Nearly all of the disagreement occurred when 1 rater coded the behavior as not playing while the other coded object play. Again, these were moments where the participant would pause and observe their surroundings. In 1 case, 1 rater coded the behavior as imagination play while the other coded the behavior as object play. During this particular instance, the participant was singing and dancing while molding the clay in her hands.

For the third variable, complexity, the interrater reliability was moderate for round 1 of coding, Kappa = 0.44 (p = .000) , 95% CI [0.33, 0.56], fair for round 2, Kappa = 0.39 (p = .000), 95% CI [0.29, 0.48], and moderate for round 3, Kappa = 0.50 (p = .000), 95% CI [0.41, 0.59].

Because 3 of the code labels for complexity were changed between round 2 and round 3, disagreement data below is listed in separate tables. The largest amount of disagreement in all 3 rounds occurred when 1 rater coded the behavior as repetition or replication while the other coded the behavior as elaboration or vice versa. Additional common disagreements involved the exploration or observation code. Further disagreement percentages are outlined in Table 5 and Table 6.

Table 5

Round 1 and 2 Coding Complexity Percentages

Category	Round 1 Percentage	Round 2 Percentage
Agreement total	62%	55%
Disagree - repetition vs elaboration	16%	15%
Disagree - exploration vs elaboration	6%	3%
Disagree - exploration vs repetition	5%	2%
Disagree - not playing vs exploration	4%	4%
Disagree - repetition vs regulation	4%	3%
Disagree - elaboration vs regulation	3%	3%
Disagree - not playing vs repetition	1%	13%
Disagree - not playing vs elaboration	0%	3%
Disagree - not playing vs regulation	0%	0%
Disagree - exploration vs regulation	0%	0%

Table 6

Round 3 Coding Complexity Percentages

Category	Percentage
Agreement	64%
Disagreement - replicating vs elaborating	13%
Disagreement - observing vs elaborating	7%
Disagreement - other vs observing	7%
Disagreement - observing vs replicatio0.4n	3%
Disagreement - other vs replicating	2%
Disagreement - observing vs regulating	2%
Disagreement - other vs elaborating	1%
Disagreement - elaborating vs regulating	1%
Disagreement - other vs regulating	1%
Disagreement - replicating vs regulating	1%

For the fourth variable, social level, the interrater reliability was substantial for round 1 of coding, Kappa = .62 (p = .000), 95% CI [0.50, 0.73] and moderate for round 2 of coding, Kappa = .42 (p = .000), 95% CI [0.32, 0.53]. The largest amount of disagreement in both round 1 and 2 involved the parallel play code. Further disagreement percentages are outlined in Table 7. In round 3 this variable was changed to a present or not present model, and interrater reliability was nearly perfect, Kappa = 0.92 (p = .000), 95% CI [0.85, 0.98].

Table 7

Round 1 and 2 Coding Social Level Percentages

Category	Round 1 Percentage	Round 2 Percentage
Agreement total	82%	67%
Disagree - solitary vs parallel	6%	3%
Disagree - onlooker vs parallel	5%	1%
Disagree - not playing vs parallel	4%	15%
Disagree - parallel vs associative	4%	6%
Disagree - not playing vs onlooker	1%	3%
Disagree - not playing vs associative	1%	1%
Disagree- solitary vs onlooker	1%	1%
Disagree - solitary vs cooperative	1%	0%
Disagree - not playing vs solitary	0%	1%
Disagree - not playing vs cooperative	0%	0%
Disagree - solitary vs associative	0%	1%
Disagree - onlooker vs associative	0%	1%
Disagree - onlooker vs cooperative	0%	0%
Disagree - parallel vs cooperative	0%	0%
Disagree - associative vs cooperative	0%	0%

For the fifth variable, parent role, the interrater reliability was moderate for round 1 of coding, Kappa = .57 (p = .000), 95% CI [0.46, 0.68] and substantial for round 2, Kappa = .60 (p = .000), 95% CI [0.51, 0.69]. The largest amount of disagreement involved the supervisor/ student and player categories. Further disagreement percentages are outlined in Table 8. In round 3 this variable was changed to parent or guardian present or not present and interrater reliability was nearly perfect, Kappa = 0.83 (p = .000), 95% CI [0.70, 0.95].

Table 8

Round 1 and 2 Parent Role Percentages

Category	Round 1 Percentage	Round 2 Percentage
Agreement total	71%	64%
Disagree - player vs supervisor/student	11%	3%
Disagree - facilitator vs supervisor/student	8%	2%
Disagree - not playing vs player	4%	8%
Disagree - not playing vs supervisor/student	2%	8%
Disagree - player vs facilitator	2%	1%
Disagree - supervisor/student vs not present	2%	0%
Disagree - not playing vs facilitator	0%	1%
Disagree - not playing vs interpreter	0%	1%
Disagree - Not playing vs not present	0%	0%
Disagree - player vs interpreter	0%	1%
Disagree - player vs not present	0%	0%
Disagree - facilitator vs interpreter	0%	0%
Disagree - facilitator vs not present	0%	0%
Disagree - interpreter vs supervisor/student	0%	0%
Disagree - interpreter vs not present	0%	0%

In rounds 1 and 2, the final additional variable was crowd size. This variable was eliminated in round 3 to reduce decision fatigue. In round 1 of coding interrater reliability was almost perfect, Kappa = .87 (p = .000), 95% CI [0.81, 0.94] and for round 2 it was substantial, Kappa = .70 (p = .000), 95% CI [0.60, 0.80].

CHAPTER IV

DISCUSSION

The purpose of this initial study was to develop a framework that could be used for observational quantitative measurement of play in children's museums. The first step in testing the framework was to investigate interrater reliability using the Dimensions of Play Framework in conjunction with a time-sampling method. Multiple rounds of coding by undergraduate assistants demonstrated moderate agreement yet failed to yield acceptable interrater reliability for hypothesis testing (Kappa > .70). However, in 2 out of the 3 rounds of coding, we were able to achieve at least moderate agreement in every dimension. Demonstrating moderate interrater reliability using inexperienced undergraduate assistants suggests that the Dimensions of Play may be a promising new framework for measurement of play in children's museum exhibits. To my knowledge, this is the first quantitative observational measure to include both the dimensions of play type and complexity. Results indicate more should be done to structure the framework and determine the appropriate method of measurement to accompany it. Several factors negatively affected reliability including the time sampling method, counter-intuitive code labels, and the complex relationship between observation and play.

The first hindrance to reliability was the time-sampling method. Some of the time-samples the research assistants were tasked with coding were puzzling because they were transitional moments, meaning that either multiple behaviors were occurring within that particular second or that the participant paused between 2 behaviors for the duration of the

second. For example, imagine that a participant is copying their guardian by rolling clay into a ball between their hands, and then the participant drops the ball on the table and smooshes it into a pancake with their hand. The time-sample to be coded is the second that stretches from the last rolling motion through the drop to the point that the child raises their hand to start the smooshing motion. This is a moment of transition between replication and elaboration. It was confusing and difficult for research assistants to code these kinds of moments.

Another issue that arose in reliability during the first 2 rounds of coding was an issue with the ethogram structure involving the relationship between play and exploration, leading exploration to be renamed observation in round 3 of coding. Throughout the recordings, there were moments when the participant paused their activity and observed their surroundings. For example, they would look at another person's activity, visually scan the available materials, inspect their own materials, or stare off camera. One could argue this behavior is not play because the participant is not actively manipulating the materials, but one could also argue this behavior is play because it is so entwined and wrapped up in the surrounding play behaviors. For example, a child was making a fish out of clay and she paused to visually scan the table for a tool to press into the clay to create scales. During this pause she was not actively manipulating the materials, but this moment of observation was an integral part of the play sequence towards the creation of the fish. She had not stopped playing; she was in the middle of playing.

The intention was that this type of behavior would be coded as play and that the complexity would be coded as exploration. However, there were a couple of factors that led to confusion. In rounds 1 and 2 of coding, the ethogram was structured so that the research assistant would first decide if the behavior was play or not play, and then they would code the level of

complexity. Play was interpreted by the research assistants to mean active manipulation, so it was not intuitive for the research assistants to code these observational behaviors as play. In addition, exploration was a poor code label for this type of behavior because it also implied active manipulation. A couple of changes were made to address the interaction between observation and play for the third round of coding. The play versus not play variable was eliminated, and behavior that was not play was thereafter recorded as other. Exploration was renamed observation to be more intuitive of the behaviors it was intended to represent.

Another issue with code labels also occurred within the complexity variable. The operational definition of repetition included copying the action of others, following direct instruction, or repeating an action they have already done. However, research assistants had trouble applying the term repetition to copying or following instructions. For the third round of coding, repetition was relabeled as replication to be more inclusive of the behaviors it was intended to represent.

Replication also had to be further operationally defined. Research assistants expressed a couple of frustrations. First, it was unclear at what point behaviors became a repetition of previous behaviors. For example, if a participant was banging on a piece of clay with the roller for an extended period of time, it was unclear whether it become repetition after 5 seconds, 10 seconds, or 20 seconds. Second this code required them to hold in mind all of the actions of the participant throughout the entire video. For example, 7 minutes into a video the participant starts doing a chopping motion. The research assistant is not sure if they have seen this behavior before, so they had to re-watch the entire video to determine if it was a new behavior or a repeated behavior. To address these issues, research assistants were instructed to only refer to

previous time-samples when determining if a behavior was repetitive. This gave them clear guidance for repetition and allowed them to look back through their moment descriptions rather than re-watch the entire video.

While changes to the ethogram solved many of the problems, substantial reliability was still not achieved for all variables in round 3 of coding. While my returning research assistant affirmed that the new pared down ethogram structure was easier to use and resulted in less decision fatigue, the new assistant admitted often feeling bogged down viewing the same moment over and over trying to discern what was happening. On average it took research assistants between 4 and 6 hours to code each round. My intention for this framework is for it to be useful and useable for children's museum exhibit designers, and in this study, it was too time consuming and cumbersome for this purpose. I believe the time-sampling method is likely the major culprit.

Limitations

There were several limitations to this study. First, adapting Parten's Social Play Scale (1932) and the Adult-Child Interaction Inventory (Beaumont, 2010) for this purpose was unsuccessful. Therefore, the framework as it stands is incomplete because of the missing social dimensions of play. The Social Play Scale was originally intended to be used in settings with other children. The Adult-Child Interaction Inventory was intended for use with parent-child dyads. Neither were intended for use in a setting with multiple parents and children, and they became conflated when both parents and other children were present.

For example, the original intention was that the social play categories would only be applied to other children. However, this did not make sense in practice. For example, a parent

was helping their child to roll out the clay. The parent was standing over the child and they were both pushing on the roller to flatten out the clay. Another child was sitting nearby not playing. According to the ethogram the parent role should be recorded as player and the social level should be coded as solitary, but that seems counter-intuitive and wrong since the child is engaged in cooperative social play with their parent. Another source of confusion was several instances where the parent of a participant was engaged with a different child at the table. For example, 1 child was watching their parent facilitate the play of their sibling next to them. It did not make sense to code this as facilitator as it applies to the participant; neither did it seem intuitive to code the parent role as not present.

The final strike for these scales was that research assistants found it to be highly taxing to juggle using so many variables and codes for every time-sample. It required them to consistently switch their focus and framing and made it difficult to get into a state of flow. This seemed to lead to some decision fatigue and extra time spent having to re-watch portions of the video multiple times. In order to address these issues, the social play and parent role variables were eliminated in the third round and replaced with a simplified present or not present model. Research assistants coded each time sample for other children present or not present and parents or guardians present or not present.

There were also limitations to the video recordings. First, even with 2 camera angles it was at times difficult or impossible to see what was needed to code the video. For example, if 1 camera was obstructed by someone standing in front of it while the participant's hands were obstructed by someone sitting next to them it was not possible to know for sure what the participant was doing. Also, if the participant was looking at something or someone off camera,

it was not possible to know if they were observing another person playing, looking at another activity, or staring into space.

A new set of limitations developed for round 3 of coding. At the end of the Spring 2020 semester COVID-19 caused the University to shut down, and 1 of my research assistants graduated. The new research assistants were not as experienced in working with children, and in addition all training had to take place entirely online. Also because of this transition to working from home, research assistants were being asked to watch and code these videos using their laptops. The research assistant would need to have the video open as well as the coding spreadsheet. This resulted in viewing the videos in a small window which made small actions using small materials difficult to discern. These issues resulted in inconsistencies, missing data, and mistakes. It also resulted in research assistants spending a lot of time playing, pausing, rewinding and re-watching, making this method more cumbersome than expected.

Future Directions

Because the Dimensions of Play framework is based on prior scales that were mostly developed through long durations of observations and more qualitative ethnographic measures, I think it would be prudent to use the framework in a phenomenological qualitative study to improve its structure. This study could involve more in-depth research into a smaller sample. For example, the researcher could record children with developed language skills and then review the recording with the participant using an in-depth interview to discover how the participant makes sense of their personal experience. This type of approach is flexible enough to reveal unanticipated dimensions, categories, and definitions of play (J. A. Smith, 2004). Such a process may help to develop more in-depth, thorough materials to train research assistants. It also may

help to identify and address idiosyncrasies or issues with the framework. For example, where are things are getting left out? What types of behaviors are difficult to classify and why? Are the categories overlapping or co-occurrent? Such research using the framework could also help to address how social complexity and parent role may be able to be integrated into the framework. At children's museums exhibits, children may sit down with or without other children, their guardian or guardians, other children's guardian or guardians, and museum educators. Whether it is possible to create a framework dimension or dimensions that could accommodate this level of variability is currently unknown.

Future studies should also explore other means of quantitative observational measurement of play besides the time-sampling procedure utilized in this study. A less time-consuming approach could be to mark the presence or absence of each category in each dimension over the course of the entire play sequence. However, it is still a worthwhile goal to explore methods that would indicate the amount of each play type and play complexity category. A way to potentially achieve this would be to use a one-zero sampling method where, "time is broken up into short intervals… Then, the occurrence of each behavior is noted, either as present or absent for each interval" (P. K. Smith, 2015, p. 145). This method may help to address the issues with coding transitional moments, and although it would not give a measure of duration or frequency, it would give a measure of the amount of occurrence for each category that could be utilized for hypothesis testing.

Conclusion

Play is important for learning and development, and children's museums serve their communities by not only offering those experiences to children but also demonstrating the importance of playful interactions (Dooley & Welch, 2014). Therefore, a method of measurement is needed that could empirically test hypotheses regarding play in children's museum exhibits. However, prior approaches to measuring play are limited to examining play type without complexity, complexity of a specific play type, or ethnographic measures. I developed the dimensions of play framework to examine both play type and complexity using a quantitative observational method, but more work needs to be done to develop it into a framework that will be useful for hypothesis testing.

A reliable framework for measuring play in children's museum exhibits should continue to be pursued because success would be mutually beneficial for both researchers and children's museum professionals. Researchers would benefit from access to a natural context that offers great validity, and children's museum professionals would benefit from being able to discover more about what facilitates high quality play (Dooley & Welch, 2014). This study has important implications for future attempts at quantitative observational measurement of play behaviors in children's museums and demonstrates that more theory development is necessary to be successful in such an endeavor.

REFERENCES

Ancelet, J. E., & Luke, J. J. (2015). *Wonder Room evaluation study preliminary findings*. Columbus Art Museum.

Association of Children's Museums. (2012). *Standards for professional practice in children's museums*.

Bateson, P. (2015). Theories of play. In A. D. Pellegrini (Ed.), *The Oxford Handbook of the Development of Play* (pp. 41-47). New York, NY: Oxford University Press.

Beaumont, L. (2010). *Developing the Adult Child Interaction Inventory: A methodological study*. Boston Children's Museum and Evergreene Research and Evaluation.

Bergen, D. (2015). Psychological approaches to the study of play. *American Journal of Play, 8*(1), 101-128.

Bjorkland, D. F., & Gardiner, A. K. (2015). Object play and tool use: Developmental and evolutionary perspectives. In A. D. Pellegrini (Ed.), *The Oxford Handbook of the Development of Play* (pp. 153-171). New York, NY: Oxford University Press.

Brown, S., & Vaughan, C. (2009). *Play: How it shapes the brain, opens the imagination, and invigorates the soul*. New York, NY: Penguin Group.

Burghardt, G. M. (2015). Defining and recognizing play. In A. D. Pellegrini (Ed.), *The Oxford Handbook of the Development of Play* (pp. 9-18). New York, NY: Oxford University Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 10. doi:10.1177/001316446002000104

Dooley, C. M., & Welch, M. M. (2014). Nature of interactions among young children and adult caregivers in a children's museum. *Early Childhood Education Journal, 42*(2), 125-132. doi:http://dx.doi.org/10.1007/s10643-013-0601-x

Fagen, R. M. (2015). Play and Development. In A. D. Pellegrini (Ed.), *The Oxford Handbook of The Development of Play* (pp. 83-100). New Yor, NY: Oxford University Press.

Friard, O., & Gambo, M. (2012). Behavioral Observation Research Interactive Software. Retrieved from http://www.boris.unito.it/pages/about.html

Henderson, T. Z., & Atencio, D. J. (2007). Integration of play, learning, and experience: What museums afford young visitors. *Early Childhood Education Journal, 35*(3), 245-251. doi:http://dx.doi.org/10.1007/s10643-007-0208-1

Howes, C. (2015). Social play of children with adults and peers. In A. D. Pellegrini (Ed.), *The Oxford Handbook or the Development of Play* (pp. 231-244). New York, NY: Oxford University Press.

Krasnor, L. R., & Pepler, D. J. (1980). The study of children's play: Some suggested future directions. In K. H. Rubin (Ed.), *Children's Play* (pp. 85-96). San Francisco, CA: Jossey-Bass.

Leong, D., & Bodrova, E. (2012). Assessing and scaffolding: Make-believe play. *Young Children, 67*, 7.

Lillard, A. S. (2014). The development of play. In *Handbook of Child Psychology and Developmental Science* (pp. 425-468). Hoboken, New Jersey: John Wiley & Sons, Inc.

Luke, J. J., Letourneau, S. M., Rivera, N. R., Brahms, L., & May, S. (2017). *Play and children's museums: A path forward or a point of tension?* Curator The Museum Journal.

Parten, M. (1932). Social participation among pre-school children *The Journal of Abnormal and Social Psychology, 27*(3), 243-269. doi:http://dx.doi.org/10.1037/h0074524

Pellegrini, A. D. (2015). The development and function of locomotor play. In A. D. Pellegrini (Ed.), *The Oxford Handbook of the Development of Play* (pp. 172-184). New York, NY: Oxford University Press.

Piaget, J. (1962). *Play, Dreams, and Imitation in Childhood* (C. Gattegno & F. M. Hodgson, Trans.). New York, NY: The Norton Library.

Rubin, K. H. (1977). Play behaviors of young children *Young Children, 32*(6), 16-24.

Shine, S., & Acosta, T. Y. (2000). Parent–child social play in a children's museum *Family Relations: An Interdisciplinary Journal of Applied Family Studies, 49*(1), 45-52. doi:http://dx.doi.org/10.1111/j.1741-3729.2000.00045.x

Smilansky, S. (1968). *The effects of sociodramatic play on disadvantaged preschool children*. New York, NY: John Wiley & Sons.

Smith, J. A. (2004). Reflecting on the development of interpretative phenomenological analysis and its contribution to qualitative research in psychology. *Qualitative Research in Psychology, 1*, 39-54.

Smith, P. K. (2015). Observational methods in studying play. In A. D. Pellegrini (Ed.), *The Oxford Handbook of the Development of Play* (pp. 138-152). New York, NY: Oxford University Press.

Van Schijndel, T. J., Franse, R. K., & Raijmakers, M. E. (2010). The Exploratory Behavior Scale: Assessing young visitors' hands-on behavior in science museums *Science Education, 94*(5), 794-809. doi:http://dx.doi.org/10.1002/sce.20394

Vandenberg, B. (1980). Play, problem solving, and creativity. In K. H. Rubin (Ed.), *Children's Play* (pp. 49-68). San Francisco, CA: Jossey-Bass Inc.

White, R. E. (2012). *The power of play*. Retrieved from https://mcm.org/wp-content/uploads/2015/09/MCMResearchSummary1.pdf

Yogman, M., Garner, A., Hutchinson, J., Hirsh-Pasek, K., & Golinkoff, R. M. (2018). The power of play: A pediatric role in enhancing development in young children. *Pediatrics, 142*(3). doi:10.1542/peds.2018-2058

APPENDIX A

IRB APPROVAL LETTER

THE UNIVERSITY OF TENNESSEE CHATTANOOGA

Institutional Review Board
Dept 4915
615 McCallie Avenue
Chattanooga, TN 37403
Phone: (423) 425-5867
Fax: (423) 425-4052
instrb@utc.edu
http://www.utc.edu/irb

TO: Aubrey Henriksen **IRB # 19-096**
Dr. Amye Warren

FROM: Lindsay Pardue, Director of Research Integrity
Dr. Amy Doolittle, IRB Committee Chair

DATE: 7/29/2019

SUBJECT: IRB #19-096: (Pilot) Toward a Valid and Reliable Measure of Play: The Dimensions of Play Framework

Thank you for submitting your application for exemption to The University of Tennessee at Chattanooga Institutional Review Board. Your proposal was evaluated in light of the federal regulations that govern the protection of human subjects.

Specifically, 45 CFR 46.104(d) identifies studies that are exempt from IRB oversight. The UTC IRB Chairperson or his/her designee has determined that your proposed project falls within the category described in the following subsection of this policy:

46.104(d)(1): Research in established or commonly accepted education settings that involves normal educational practices

Even though your project is exempt from further IRB review, the research must be conducted according to the proposal submitted to the UTC IRB. If changes to the approved protocol occur, a revised protocol must be reviewed and approved by the IRB before implementation. For any proposed changes in your research protocol, please submit an Application for Changes, Annual Review, or Project Termination/Completion form to the UTC IRB. Please be aware that changes to the research protocol may prevent the research from qualifying for exempt review and require submission of a new IRB application or other materials to the UTC IRB.

A goal of the IRB is to prevent negative occurrences during any research study. However, despite our best intent, unforeseen circumstances or events may arise during the research. If an unexpected situation or adverse event happens during your investigation, please notify the UTC IRB as soon as possible. Once notified, we will ask for a complete explanation of the event and your response. Other actions also may be required depending on the nature of the event.

Please refer to the protocol number denoted above in all communication or correspondence related to your application and this approval.

For additional information, please consult our web page http://www.utc.edu/irb or email instrb@utc.edu.

Best wishes for a successful research project.

APPENDIX B

IRB CHANGES APPROVAL LETTER

THE UNIVERSITY OF TENNESSEE CHATTANOOGA

Institutional Review Board
Dept 4915
615 McCallie Avenue
Chattanooga, TN 37403
Phone: (423) 425-5867
Fax: (423) 425-4052
instrb@utc.edu
http://www.utc.edu/irb

TO: Aubrey Henricksen
Dr. Amye Warren

IRB # 19-096

FROM: Lindsay Pardue, Director of Research Integrity
Dr. Susan Davidson, IRB Committee Chair

DATE: 1/31/2020

SUBJECT: IRB #19-096: Dimensions of Play: A Framework for Measurement

The University of Tennessee at Chattanooga Institutional Review Board has reviewed and approved the following changes for the IRB protocol listed above:

- Change of title to the one listed above.
- Change of end date to 12/3/2020
- Addition of two student research assistants from a list on the form B. (All those listed have completed the appropriate CITI training; if something changes and a different student will be working on the project, please submit their name to the IRB office before adding them.)
- Two additional exhibits at the Creative Discovery Museum are being added.
- Small change to the order of elements in the informed consent form.

Please keep in mind that all research must be conducted according to the proposal submitted to the UTC IRB. If changes to the approved protocol occur, a revised protocol must be reviewed and approved by the IRB before implementation. For any proposed changes in your research protocol, please submit an Application for Changes, Annual Review, or Project Termination/Completion form to the UTC IRB. Please bear in mind that significant changes could result in having to develop a new application for submission and approval. Your protocol will be automatically closed at the end of the proposed research period unless a change request application is submitted. No research may take place under a closed or expired protocol.

A goal of the IRB is to prevent negative occurrences during any research study. However, despite our best intent, unforeseen circumstances or events may arise during the research. If an unexpected

situation or adverse event happens during your investigation, please notify the UTC IRB as soon as possible. Once notified, we will ask for a complete explanation of the event and your response. Other actions also may be required depending on the nature of the event.

Please refer to the protocol number denoted above in all communication or correspondence related to your application and this approval.

For additional information, please consult our web page http://www.utc.edu/irb or email instrb@utc.edu.

Best wishes for a successful research project.

APPENDIX C

IRB INFORMED CONSENT FLYER

Exhibit Play Study Informed Consent

Dear Caring Adult,

Today I am conducting research about play and learning in exhibits as part of my master's thesis at the University of Tennessee at Chattanooga. The research involves using video recordings to watch how children and adults normally interact with exhibits in the Performing Arts gallery. I won't be collecting any personal information or asking any questions except your child's age.

Voluntary Participation & Withdrawal from the Study

Participation is completely voluntary. If you change your mind, you can withdraw from the research at any time by sending your participant ID number (#_____) to Aubrey Henriksen at aubrey-henriksen@utc.edu or calling 901-302-8795.

If you *don't* want to participate in the research or be recorded, you can avoid the Performing Arts gallery until recording is over.

Confidentiality
I am not collecting any personal information except your child's age. I will keep video recordings on a hard drive a locked cabinet in a research lab at the University. Only my research team (myself, two research assistants, and my three Psychology faculty advisors at the university) will be able to see the recordings. I will delete the video tapes when my thesis project is complete, approximately December 2020.

Procedures
You should explore the Performing Arts exhibit gallery as you normally would.

Risks and Benefits
There are no foreseeable risks to participating. There are no direct benefits to you or your child for participating, but your participation could help us understand more about learning through play in our exhibits.

Contact Person
If you have any questions about the project, you can contact Aubrey Henriksen at aubrey-henriksen@utc.edu or 901-302-8795 or Amye Warren at amye-warren@utc.edu or 423-425-4239 . This project has been approved by Creative Discovery Museum and the University of Tennessee at Chattanooga's Institutional Review Board (FWA00004149 Project #19-096). If you have any questions concerning UTC IRB policies, or you and your child's rights as a human subject, please contact the Office of Research Integrity at (423) 425-5867 or email instrb@utc.edu.

Thank you for considering participating in this research on learning and play in exhibits. This research will help the museum to understand more about how exhibit design can promote learning through play.

Sincerely,
Aubrey Henriksen
Graduate Student
University of Tennessee at Chattanooga
540 McCallie Avenue
Chattanooga, TN 37402

Dr. Amye Warren
Professor of Psychology
University of Tennessee at Chattanooga
540 McCallie Avenue
Chattanooga, TN 37402

VITA

Aubrey Henriksen earned a Bachelor of Arts degree in Anthropology and Sociology with a minor in Education from the University of Tennessee at Chattanooga in 2012. After graduation she spent time working as an early childhood educator at Shiny Penny Childcare Center, substitute teacher for Hamilton County Schools, and museum educator at Creative Discovery Museum. After 1 year of working for the Museum, she was promoted to exhibit development and evaluation assistant where she became interested in studying play behaviors in museum exhibits. A couple years later the Museum received a 3-year grant to develop an evaluation department, and she was again promoted to evaluation coordinator. During this transition she was accepted to the University of Tennessee at Chattanooga's Research Psychology Masters Program. During her graduate career, Aubrey taught undergraduate courses (including Introduction to Psychology and Research Methods Lab) and presented at museum conferences across the country (including the American Alliance of Museums, Association of Children's Museums, Association of Science and Technology Centers, and Visitors Studies Association conferences). Aubrey graduated with a Master of Science degree in Psychology in December 2020.