Predicting Child Behavior and Mental Health with Toy Block Play Data

著者	Wang Xiyue
学位授与機関	Tohoku University
学位授与番号	11301甲第20111号
URL	http://hdl.handle.net/10097/00135832

TOHOKU UNIVERSITY

SENDAI, JAPAN

Predicting Child Behavior and Mental Health with Toy Block Play Data

A dissertation submitted in conformity with the requirements for the degree of Doctor of Philosophy

in the

Research Institute of Electrical Communication

Graduate School of Information Sciences

by

Xiyue WANG

Supervisor

Yoshifumi KITAMURA, Professor

Aug 18, 2021

Abstract

The advances made in ubiquitous mental states inference and the prevalence of child mental disorders push the need for seamless integration of child mental health prediction, monitoring, and awareness in daily life. Motivated by the use of toy blocks as a kind of classic toy that contains rich information to decompose, and the development of affective computing, this thesis establishes a novel method to address the above need by using data captured from playing with toy blocks, as one of children's favorite activities, to predict their behavior and mental health.

This thesis contains in-depth and in-breadth investigations to establish a method of using data captured from toy block play to predict a range of child mental health measurements. The initial step established the connection between child mental health and block play by assessing the short-term stress, measured in-situ, after the 2011 Tohoku Earthquake and Tsunami, using computed actions and video-coded behaviors extracted from play sessions with sensor-embedded toy blocks. The data analysis results indicate that play passively expressed stress. It next investigated the approach of predicting a range of child behavioral problems from automatically extracted block play features and sequential patterns. Internalizing problems, total problems and aggressive behavior were predicted, and the positive predictors were interpreted as inactive, indecisive, or drastic styles of play.

In addition, this thesis discusses the lessons learned toward improving daily mental state prediction and support. It proposes the next steps needed to increase the method's robustness using motion and structure features of block play, design guidelines toward real world applications, and future work that could extend and generalize the proposed method.

***Keywords*:** Human-computer interaction, tangible user interface, stress, behavior problems, assessment, prediction, feature extraction, machine learning, field study, free play, well-being

Contents

Chapter 1

Introduction and Motivation

This thesis proposes a novel concept to predict child mental health in daily life using data captured from toy block play. It investigates an interdisciplinary field encompassing human-computer interaction (HCI), mental health, and data science. It is an exploration of research opportunities that lie at the interdisciplinary interface between developing a tangible user interface (TUI)[1], specifically, smart daily objects, and investigating new possibilities of bringing child mental health awareness closer to daily life, inspired by affective computing[2]. This thesis is inspired from, and aims to contribute to, both computer science and psychology.

Chapter 1 Overview

This thesis begins in Section 1.1 with a discussion of several key concepts that both motivate and contextualize the research objectives. More specifically, Section 1.1.1 introduces blocks in two aspects - (1) blocks as visual-spatial constructive play objects, as well as their affordance and (2) their applications in analyzing child behavior. It highlights why block play is worthy of investigation. Affective computing is introduced in Section 1.1.2

[1]At a high-level, TUI is a sub-discipline within human-computer interaction in which everyday physical objects play a central role as both physical representations and controls for digital information.

[2]At a high-level, affective computing is an interdisciplinary field spanning computer science, psychology, and cognitive science. It is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects.

to outline its status as a crucial trend in healthcare, its development, and how it calls attention to the unaddressed needs of children. With this background information, a thesis overview is provided in Section 1.2. My specific contributions to the work presented are stated in Section 1.3.

1.1 Thesis Motivation

1.1.1 Blocks: Visual-Spatial Constructive Play Objects

Toy blocks (also building blocks) are wooden, plastic, or foam three-dimensional objects with various geometric shapes (cube, cuboid, cylinder, triangular prism, bridge, arch, etc.) and colors (red, green, blue, yellow, natural wood color, etc.) Often holding a place in the game corner of kindergartens and households, they are classic and popular visual-spatial constructive play objects because they provide a set of unique affordance that is not comparable with other toys and daily objects in terms of creativity, self-expression and self-reflection.

VCPOs and Affordance

The term visual-spatial constructive objects (VCPOs), purposed by Ness et al. [1], specifically denotes constructive objects that require spatial cognition. According to them, VCPOs include blocks (e.g. standard wood blocks, plastic blocks and foam blocks), bricks (e.g. LEGO bricks and Mega Bloks), and planks ($1 \times 3 \times 15cm$ thin rectangular wooden cuboids) that either contact each other or snap together and remain positioned by force of gravity. Interacting with VCPOs "involves the use of smaller objects as a means of building larger and often more elaborate structures" [1]. VCPOs are often used by individuals to model something they imagine and may actually construct in the real world.

To investigate the unique characteristics of VCPOs, the role of affordance needs to be considered. As Ness et al. introduced it, "affordance alludes to the qualities of an object that define its possible use or make clear how it can or should be used" [1]. High affordance refers to the specific affordance of an object in which one can easily perceive the object's operational properties and presume how others might interact with it. Low and diverse affordance refers to the qualities of an object with ambiguous constraints and unspecific usage. According to Ness et al., "affordance is essential in characterizing and analyzing VCPOs", because it "refers to the meaning of an object in terms of what it provides users that allows them to maximize their potential in constructions and related spatial behaviors" [1].

Blocks and bricks, targeting chilren above 1 and 1.5 years old, respectively, have a wider tolerance for age, compared to planks, which target children who are older than 5 years. Thus blocks and bricks are often seen around young children. Even though a common understanding is that they are interchangeable constructive toys, in fact they offer different sets of affordances and consequently should serve different purposes. Blocks have a lower and more diverse affordance due to their ambiguous constraints. The smooth surfaces do not allow blocks to lock into place. This characteristic offers a variety of ways to place a block. While each placement provides the foundation for the next one, some placements and actions may cause structural imbalance and affects the stability of inertia. As a result, for blocks, the diverse ways of potential use are recognized after their physical manipulations. Through contemplation, the play behavior becomes "reflective abstraction", which affect the conceptual knowledge and next move. This process "may enhance their self-regulation during constructive play and may be more likely to engage in creative tasks that involve synthesis and higher-order thinking" [2, 3]. On the contrary, the pips of bricks provide clear constraints that the blocks have to be locked to construct a structure. It provides ease of use (less sensitive to inertia, mass and balance) but less space for empirical manipulations, contemplations and reflections. Another specific affordance is that bricks are most often sold with themes and instruct the user to follow scripts for recommended constructions. Due to the clear and specific purposes they serve, to some extent they fail to provide children "the opportunities for creative play" [1].

According to a body of literature [1, 2, 3, 4, 5, 6], high and specific affordance impedes creative processes, problem-solving, spatial thinking, and cognitive development in general, and it may lower the level of interest, enjoyment, and satisfaction in completing the activity. Thus, the low and diverse affordance of blocks allows for creativity, self-regulation, and higher-order thinking.

Analyzing Behavior

Two forms of play activities, structured play and free play, are often constructed in therapies and assessments to analyze play behavior.

Structured, thematic, formulaic or scripted play are those play activities with specified narratives or plots that are often accompanied by detailed directions. According to Ness et al., such play "possesses a prescribed outcome evident in a set of specific instructions that must be followed to obtain a desired result" and "necessitates repetitive actions with no latitude for variation in the activity and no freedom for a child to express his or her imagination" [1]. Due to this trait, structured play is often designed to assess a child's abilities. For example, structured block play is used in screening for developmental disorders in the routine health check-up for 3-year-olds in Japan [7]. A child is asked to mimic a prepared truck structure following the scripts. Being able to follow the instructions by the assessor and build the complex structure autonomously is one of the factors of normal development they assess. In the domains of human-computer interaction and ubiquitous computing, novel approaches are proposed to computerize assessments of abilities in clinical diagnosis with three-dimensional shaped user interfaces, and these approaches are proved to be effective. Sharlin et al. made use of Active Cube [8], a brick-type tangible user interface, to build a system for constructional assessment [9]. Through a series of tasks with and without step-by-step instructions, the system provides the target structure and computationally compares the similarity between the participant's structure and the target structure, as well as the time taken. Their results show that the data obtained from this structured play could assess the constructional ability and Alzheimer's disease (AD), a neurodegenerative impairment. A similar system was later adopted to assess the constructional ability in children and proved to be effective [10]. Recently, Jiang et al. proposed a "Shape-color conflicting" game with blocks to test the ability of a child to place the blocks following a rule of "same color, same shape or both." This instructed play is a tangible version of the cognitive state detection paradigm Stroop Effect and Wisconsin Card Sorting Test [11]. It demonstrated the power to assess inattention and impulsivity symptoms of ADHD. The above approaches illustrated that structured play that augments or resembles the existing clinical diagnosis was effective in automating the assessment. Structured play serves as an effective method to test the abilities in children.

Meanwhile, research suggests free play provides opportunities to observe an intrinsically motivated action [2, 3, 4]. It allows for variation in behavior and the freedom to express one's own imagination and thoughts. Free play helps the child act out unconscious material; thus it is used to aid diagnostic understanding. In play therapy, free-play sessions are frequently constructed for healing through explorations and creativity. The free play also helps to relieve the accompanying tension and provides a medium for working through defenses and handling anxieties [12]. Axline summarized that free-play is needed because in an effective therapy, "the therapist is nondirective and the therapy is client-centered; the clients are the source of living power that directed the growth from within themselves" [13].

In play therapies for children, due to mental disorders often lying along a spectrum [14], the observations of characteristics are more crucial than the ability to perform the tasks. As an example, free block play has applications to two childhood disorders: social withdrawal and Attention-Deficit/Hyperactivity Disorder (ADHD) [15]. Co-constructivist theory suggests that children who have a good understanding of how to act during a specific play situation are more successful in their social interactions than children without a clear understanding of the play situation [16]. The characteristic of social withdrawal observed in block play includes the inability to negotiate play interactions. ADHD can also be reflected in free play. The therapeutic goal for children with ADHD is to teach them to be reflective, that is, to stop, think, plan ahead and weigh alternatives and consequences before acting. Thus, a therapist usually asks questions such as "What would you like to build" or "what could we do to keep the tower from falling" and encourages a longer attention span in order to nudge an ADHD child to practice being reflective and less impulsive. These examples demonstrate that free play can help a child be expressive, self-regulated, and intrinsically motivated. The above examples also demonstrate the observation of free play can be used to find intrinsic characteristics. Meanwhile, the main criticism of play interventions has been that this field in general lacks rigorous research design and data analytic methods [17, 18]. The observations and interactions are manually processed, which suffers from low generalizability. Thus, although challenging, it is valuable to automatically, computationally, and quantitatively extract free block play behavior and transform it into structured data.

1.1.2 Affective Computing: Ubiquitous Mental State Recognition

Affective computing is the study and development of computational systems that possess the ability to recognize, understand, and even to have and express emotions [19]. It is a modern branch of computer science, proposed by Picard in 1995, as "computing that relates to, arises from or influences emotions" [20]. Nowadays, it has become an interdisciplinary field spanning computer science, psychology and cognitive science [21], and its approaches have been expanded "from finding new ways to forecast and prevent depression; to inventing new solutions to help exceptional people who face communication, motivation, and emotion regulation challenges; to enabling robots and computers to respond intelligently to natural human emotional feedback; to enabling people to have better awareness of their own health and well-being; to giving people better control and protection over their most sensitive, private, personal data" [22].

The computational recognition, prediction, or inference of mental states and mental health is the backbone of seamless and interactive applications in affective computing. Its development alleviates the severe problems of the current diagnosis-based mental health approach: "the shortage of mental health specialists, the limited resources available, arduous close monitoring of symptoms, delaying optimal treatment, and potentially prolonging suffering" [23]. It accelerates the efficacious provision of healthcare, which is expanding from treatment to prevention and from clinics to daily life.

In affective computing, data are collected through various resources with different forms. They are processed, embedded, and encoded into higher-level knowledge, such as behavioral models and affect predictions, for monitoring mental health, preventing deterioration, and supporting well-being. As an example, passive recording of behavioral data (gathering information without an individual's direct input) has been identified as a potentially feasible method for long-term monitoring of depression [23]. The combination of sensor technology and machine learning enables detailed measurement in real time to capture a range of behaviors for predicting variations of depression.

Affective computing also provides possibilities of gathering the new data necessary for advancing research and applications [20], and novel interfaces have been developed to serve this purpose. Wrist-worn biosensors have been proposed to collect daily physiological signals [24, 25], and some have developed into market-available products [26, 27, 28]. SPRING, the Smart Platform for Research, Intervention, and Neurodevelopmental Growth, is a hardware and software system that integrates the tangible Shape Sorter and other games and multiple sensing methods [29]. It aims to "(1) automate quantitative data acquisition, (2) optimize learning progressions through customized, motivating stimuli, and (3) encourage social, cognitive, and motor development in a personalized, child-led play environment". This play system can also be paired with wearable sensors to probe the physiological underpinnings of motivation, engagement, and cognition. Inspired by these novel interfaces, this thesis investigates methods to develop free block play into a novel interface for extracting play behavior data that can probe a child's mental health and well-being.

Accuracy and Interpretability

Since affective computing closely involves different fields, such as novel sensing, analytics and enabling techniques, there is no standard or unified evaluation metric for an affective computing system. For systems that perform prediction or inference, the effectiveness of the system should not be evaluated soley by accuracy in light of two key considerations: (1) high accuracy is difficult to achieve due to the challenges of sample acquisition, data processing, and the lack of absolutely unbiased ground truth; and (2) since such work usually involves sensitive, vulnerable, and at-risk groups, building trust is a crucial part of using these systems in real-world settings. Explanations and justifications based on empirical knowledge from psychological and cognitive domains are needed to develop trust in such novel systems. Thus, for most affective computing systems, not only accuracy but also interpretability are crucial evaluation metrics to consider.

A body of literature has provided rich interpretations of their results. These interpretations not only provide foundations for further improving accuracy but also generalize knowledge and build trust. For example, Intarasirisawat et al. explored using touch and motion features extracted in mobile games (Fruit Ninja, Tetris and Candy Crush) to assess cognitive functions (Attention, Memory, Visuospatial ability, etc.) [30]. Although this work did not predict cognitive functions from game-play features, their correlation analysis and in-depth interpretations provide strong evidence that game-related metrics have potential use as proxies for conventional cognitive measures. Their investigations show promise in predictions and provide generalizable knowledge to other game-based assessment systems. Moving closer to real-world applications, Nosakhare et al. proposed methods to predict health and well-being from a range of behaviors collected from college students' daily activities and, moreover, to recommend behavior change for better health conditions and well-being [31]. Their models using sLDA (supervised Latent Dirichlet Allocation) and LASSO (Least Absolute Shrinkage and Selection Operator) predicted Stressed-Calm (57.0-58.4% accuracy, and 0.41-0.66 F1 score), Sad-Happy (56.3-56.7% accuracy, and 0.42-0.68 F1 score) and Sick-Healthy (49.4-54.6% accuracy, and 0.24-0.68 F1 score). They also provided interpretations of LASSO model coefficients and patterns found in sDLA models, as well as evidence-based insights in two case studies to illustrate that the model results can be used to recognize unhealthy behaviors and recommend behavior change. Wampfler et al. proposed a Semi-Supervised Learning method to predict affective states from smartphone touch data [32]. It predicted three principal dimensions of affective states, Valence (56-67% accuracy, 0.75-0.84 AUC micro), Arousal (53-63% accuracy, 0.73-0.82 AUC micro) and Dominance (61-65% accuracy, 0.78-0.82 AUC micro). Their methods used three feature sets separately extracted from the texting heat map: Pressure, Down-down, and Up-down, as well as a combination of the three feature sets. In addition to prediction models, they provided interpretation of features extracted from heat maps and demonstrated that performance was improved when predicting a sequence of constant affective states. Their interpretations demonstrate that more accurate predictions can be made when the affective states do not alternate within a short time.

One approach that should not be confused with predictions based on affective computing using data captured from seamless and undirected events is diagnostic assessments. Diagnostic assessments augment the well-established assessment paradigms with data and tangible user interfaces. For example, Jiang et al. proposed, WeDA, a wearable diagnostic assessment system for children with ADHD [11]. Ten fun tasks such as Catching Grasshoppers, Shape-color Conflicting and Keeping Balance were proposed to capture behavioral data for assessing ADHD in children. The tasks are based on a range of cognitive state assessment paradigms used as auxiliary diagnosis methods for ADHD. As a result, the tasks predicted ADHD with high accuracies (0.88-0.98 F-score for individual tasks and 1.0 combining all tasks).

Compared to Nosakhare and Wampfler's affective computing approaches, WeDa provided remarkably high accuracy. However, the system should be considered a diagnostic tool rather than predictions using behavioral data gathered and extracted passively and seamlessly. Because it mainly tests ability, the potential for discovering and interpreting behavior is low. Thus, it could not provide rich generalizable knowledge in the manner of Nosakhare and Wampfler's systems. Since the goal of this thesis is to explore the possibilities of developing a novel mental health prediction interface rather than augmenting existing diagnostic methods, it examines both accuracy and interpretability toward developing a robust prediction system.

1.2 Thesis Overview

This thesis was born out of an initial demand to connect smart daily objects and mental health. As explicated in Section 1.1, the block is a suitable entry point for such investigation. Considering the limited knowledge of this methodology in affective computing, my in-depth and in-breadth investigation serves as a solid way to establish such a method. For the high-level target, mental health, my specific targets from specific to general; it spanned from short-term, post-disaster PTSD-related stress to long-term behavioral problems used in screening at clinics. In terms of the methods used to decompose and quantify block play, I investigated quantitative actions, video-coded play behavior, and play patterns extracted from sequential actions. While the methods moved from manual to automated, the modalities and complexity increased in the investigations of ways to achieve higher robustness. To predict child mental health with toy blocks, two approaches were proposed in sequence: establishing the association, and exploring the predictions.

In Chapter 2, I present the initial step of building the connection between child mental health and block play. It assessed the short-term stress in-situ measured after the 2011 Tohoku Earthquake and Tsunami, with actions computed and behaviors video-coded from data of children playing with sensor-embedded toy blocks. Moving a step forward toward the real-world applications, the investigation of predicting a range of child behavioral problems from automatically extracted block play features is presented in Chapter 3. Finally, Chapter 4 discusses in depth the lessons learned toward designing TUIs for daily mental state inference and support, from the next steps in improving robustness to future works that further develop the approach or apply the knowledge to other approaches.

1.3 Collaborations & Contributions Statement

A large portion of the work reported in this thesis derives from papers that are products of my collaborations. During my PhD studies, I have had the honor to collaborate with human-computer interaction experts and psychologists. These individuals have broadened my horizon, and helped me to grow as a researcher. Their support and our discussions, especially around the target mental health measurements, have been invaluable. Thanks to their help and support, this thesis approaches its challenges within the interdisciplinary field of computer science and psychology. The thesis also largely benefited from the explorations of previous students. Their sensor-embedded block design and data collection had laid the groundwork of my own exploration. At the mean time, I have done the majority of the work associated with the aforementioned papers. Furthermore, it is my research narrative and journey that connects the papers together into the step-by-step investigations that establish the method of predicting child mental health with block play.

My specific contributions are as follows. My main contributions are in target definition, data processing, association or predicting, and encapsulating insights. These contributions constituted the major part of Chapter 2, Chapter 3, and Chapter 4. I partially contributed to system development and raw data collection. The system and raw data presented in Chapter 2 and Chapter 3 contain the implementation and field work of students who were previously working on this project. I have conducted system design and data collection, and these works and reflections upon them are presented in Chapter 4. I will specify my specific contributions when I come to a Chapter containing content from the aforementioned papers.

Chapter 2

Assessing Stress with Play Actions from Sensor-embedded Toy Blocks

Preface: This Chapter contains modified content from one of my previously published papers [Xiyue Wang, Kazuki Takashima, Tomoaki Adachi, Patrick Finn, Ehud Sharlin, Yoshifumi Kitamura. AssessBlocks: Exploring Toy Block Play Features for Assessing Stress in Young Children after Natural Disasters. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4, 1, Article 30 (March 2020), 29 pages.] ©2020 ACM. I adapted the cited paper and reorganized its content to integrate it into the thesis. I performed the majority of the work associated with the aforementioned paper. I contributed to: 1) establishing the presented story, 2) selecting the method for the analysis and conducting the analysis, 3) encapsulating the result and discussion, and 4) writing the paper in general.

2.1 Introduction

Natural disasters are occurring frequently [33, 34], and take a terrible toll: disasters such as the 2004 Indian Ocean Earthquake and Tsunami (228,000 casualties), 2008 Sichuan Earthquake (88,287 casualties), 2010

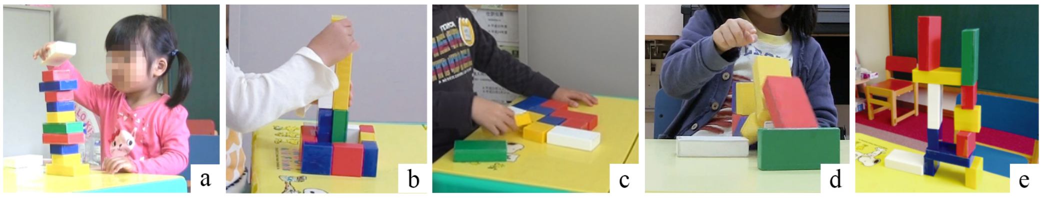


Figure 2.1: Post-disaster children's toy blocks play patterns, from left to right: ***(a)*** construction ***(b)*** fence building ***(c)*** playing flat ***(e)*** destruction **(e)** unstable structures

Haiti Earthquake (222,570 casualties), and the 2011 Tohoku Earthquake and Tsunami (22,626 casualties) left behind a complex array of problems, many of which are extremely difficult to solve [35]. Some of the most serious long-term challenges for survivors are mental health problems such as Post-Traumatic Stress Disorder, (or PTSD), conditions which can be particularly serious among young children [36]. Children's mental health problems are reported to be accentuated after significant natural disasters [36], and are often not healed by time [37, 38]. Children are also harder to treat than adults when using traditional, talk-based methods [39]. Young children's linguistic expression and cognitive development are not fully-fledged, compared to adults, so the understanding of children's internal psychological state and related mental health issues requires more delicate observation, and often calls for a different approach.

Researchers have developed computer-based interactive tools for addressing complex issues in children's mental health. In areas such as cognitive impairment, autism, and dyslexia interventions with Tangible User Interfaces (TUIs) have proven particularly effective [40, 41, 42]. While this work makes an important contribution to the study and treatment of mental health in children, less has been done on user interface research that could help those suffering from trauma caused by natural disasters. Given the number of people impacted by these events, and that such events are increasingly common, we propose a novel method to assess children's stress with a TUI approach. TUIs, especially in their simplest form, physical toy blocks, can potentially support play activities that captures some of the physicality of natural disasters - e.g. the physical destruction of real structures. Physical construction and destruction actions are inherent to playing with toy blocks, providing a non-verbal actions that can be mapped directly to the child's inner responses to the traumatic experience.

This paper presents our initial design and evaluation work towards a long-term goal of developing quantitative play feature-characterizing toy blocks for automatically assessing children's mental health. The long-term research question we pursue is:

> Can sensor-enabled toy blocks assess post-disaster stress in children?

Our research questions began to form after observing new patterns in PTSD-affected children playing with toy blocks in kindergarten after 2011 Tohoku Earthquake and Tsunami (see for example [43, 44]). The new patterns included cycles of building block construction followed by intense destructive actions. Six months later, both PTSD symptoms and destructive behavior seemed to diminish. Based on frequent but anecdotal observations, we wondered if the building, destruction, and rebuilding process reflected children's mental and psychological states and might be helping them come to terms with the destruction they witnessed. We consider whether the physicality of the blocks, and the freedom to create structures became a simple medium capturing the children's stress and allowing them to express their anxiety and fears. These observations led us to investigate the connection between children's toy block play patterns and mental state, especially in post-disaster stress. If block-play contained relevant information, perhaps we could automate the block play assessment approach for children's mental healthcare. Automated, computer-based assessment promises to reduce the need for professional assessors' time, improving access by lowering training requirements for assessors, eliminating some forms of bias, and improving the reliability of testing (as demonstrated in [9, 30]). Children's post-disaster automated stress assessment helps extend the quality, affordability, and access to critical health care necessary for many children in a world facing increased exposure to natural disasters.

To realize this vision, we designed and prototyped sturdy, simple, automated blocks with IMU capable of capturing basic play actions (see an example of play and sensor raw data in Figure 2.2). From 2013 to 2015, two years after the 2011 megathrust in the Tohoku region of Japan, the blocks were deployed in a set of studies with 52 pre-school children, aged 2.7 to 6.9 (see snapshots of studies in Figure 2.1). Among the participants, 15 (aged 5.9 – 6.5) experienced the most damaging impacts of the tsunami firsthand, when they were between the ages of 3.0 and 4.0. The unprecedented and devastating tsunami placed this group at a higher risk than the other participants for stress-related illness. We sampled approximately 20 minutes of play activity with our blocks, and manually evaluated each child's behavior during the session, noting areas such as concentration or for children who felt lost and required support. We measured each child's stress before and after play, using bio-marker sAA (Salivary Alpha-amylase Activitity), and evaluation form OSBD (Observation Scale of Behavioral Distress) and VAS (Visual Analogue Scale of Anxiety).

Our analysis showed that some of our block features, play behavior evaluations, as well as traumatic experience, related to children's stress measurements. Our findings indicate that the block-play features approach is promising for automatically predicting stress in children.

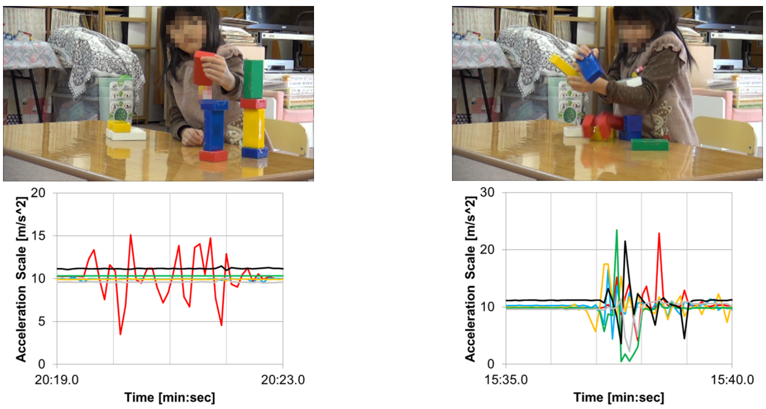


Figure 2.2: Children's play and snapshots of data from IMU embedded in our blocks

Our contributions are summarized as follows.

- The design of AssessBlocks, a computer-augmented toy using sensors-embedded toy blocks that document children's block-play features;

- A protocol and procedure for using AssessBlocks to study block play features, behavior evaluations, and children's stress measurements;
- Bivariate analysis of the data we collected with AssessBlocks, revealing correlations between block features, play behavior, trauma experience and stress measurements;
- Discussion of the potential and limitations of AssessBlocks and our toy block assessment approach; a roadmap for further iterations and future research.

2.2 Related Work

Our research is directly motivated by the record-setting Great East Japan Earthquake and its' psychological and social impact on children. The work is built on pediatric, psychological, and social studies of children's mental health after experiencing large-scale traumatic events, HCI research on playful and interactive user interfaces for assisting, treating, and assessing the health of children, and on knowledge of children's activity sensing and characterizing techniques.

2.2.1 Natural Disasters and Their Impact on Children

On March 11, 2011, a record-setting earthquake and tsunami hit the Tohoku (northeast) region of Japan. The Great-East Japan Earthquake is the fourth largest in modern record. At 14:46 JST, a magnitude 9 to 9.1 earthquake struck the coast. Tsunami waves of up to 40.5 meters followed the quake and gave people less than an hour to evacuate.

Japanese buildings are designed to be resilient during earthquakes, and citizens are well-trained for emergency evacuation; however, the scale of this event was a shock to the system. At least 22,626 people lost their lives. Witnesses confirm that children, even those in pre-planned shelters saw people die [45]. The impact of the disaster continues: the children affected by the natural disaster are now teenagers many of whom still suffer from latent PTSD and its long-term symptoms, often exacerbated by repeated exposure to elements of the trauma since major earthquakes are not uncommon in the region.

Surviving a major natural disaster can cause stress and illness in most people but is particularly hard on children [37]. After the 2008 Sichuan earthquake, PTSD and depression rates among children aged 8 to 16 years were 12.4% and 13.9% respectively when measured 15 months after the event [38]. Over 50% of school-aged children in Haiti were reported to have severe post-traumatic stress one year after the 2010 earthquake [37]. Becker's research on trauma in children who experienced the Asian Tsunami on December 26, 2004 showed a tendency towards regressive behaviors [46]. In the study, children between 4 and 7 exhibited typical regressive behaviors such as clinging, bedwetting, fearfulness, sleep disorders, and elevated reactions to stimuli.

Recently, more research attention is focused on needs that arise after survival requirements for food, clothing and shelter are met. In particular, mental health care is receiving increased focus [46, 36, 47, 48, 49]. Coping with disaster-induced mental health conditions is not easy, but timely access to appropriate mental health care is crucial to reduce the risk of developing PTSD [38]. Unfortunately, mental health interventions are usually short-term, and even these are hard to provide and hard to access. They require considerable time, professional expertise, and resources [47]. Many factors further complicate practitioners' ability to target and monitor the appropriate mental health care for children. Natural disasters cause simultaneous systemic shocks, and research shows that cumulative and conjoined traumatic events can amplify behavioral problems [50, 51]. After the 2010 earthquake in Haiti, Blanc relocated children to a center where psychosocial supports were in place. Yet, even with a systematic intervention Blanc's team were unable to show significant reductions in either PTSD or depression when compared to a control group [37]. Pynoos's analysis of PTSD in children following the 1988 Armenian earthquake suggests that girls sustained more serious and long-lasting suffering than boys [52]. While our work focused on helping children who survived the Tohoku earthquake and tsunami, its overarching goal is to promote research that can help all children who experience trauma following natural disasters.

2.2.2 Playful, Interactive Health Assessment and Treatment

Playful or play-based therapies are a well-established means for treating mental health. Creative play approaches such as Sand-Play and Painting Therapy are commonly used to treat chronic stress and PTSD [53, 12]. Block play has shown therapeutic results for social withdrawal and ADHD in children [15, 54]. Pullman has stated that with maturation, young children transition from transporting blocks to stacking them, and then move to three-dimensional composition [55]. As a result, blocks have been used in three-year old children's cognitive development checkups in Japan [7]. Traditionally, play therapy and assessment is conducted by an on-site therapist using observation, followed by question-and-answer interviews with the child, and sometimes involving the analysis of video recordings of their gameplay. These approaches are effective, but are profoundly time-consuming, require advanced therapeutic or psychological assessor expertise, and are often incapable of capturing nuanced differences in play actions.

Computer-assisted health assessment and therapy is increasingly common. Automated or computer-aided assessment can potentially reduce on-site professional time required, and reduce training requirements for caregivers or assessors. For example, "Cognitive Cubes" [9] proposes a method to measure spatial cognitive ability using a tangible interface called "Active- Cube" [8]. The Active-Cube researchers assessed construction ability and dementia by automatically presenting target shapes, then supporting the shape-building process using a 3D construction TUI, and automatically analyzing participants' shape similarity outcomes over time. Intarasirisawat et al. show that touch and motion features collected from three popular mobile games, Tetris, Fruit Ninja and Candy Crush, have potential to be used as proxies for conventional cognitive assessment of such elements as attention and memory[30]. An overview of PTSD diagnoses and treatment suggest that computational technologies can support information gathering and provide more objective PTSD assessments when compared to traditional paper and talk-based methods [56] Exposure therapy using computer simulations shows promising results for prevention and therapy in trauma-related disorders. Botella and Rizzo show the potential of adaptive displays and Virtual Reality Games (VRG) when treating combat-related PTSD [57, 58]. Several systems where caregivers manually track patients have shown that data-driven approaches improve quality of life for those suffering from PTSD and depression [59, 60].

The potential use of TUIs in children's healthcare is being explored in a growing number of projects. In one example, Fan et al. showed that working with tangible letters helped children with dyslexia learn to read and spell [40]. Westeyn et al. created augmented toys called, "Child'sPlay," using Inertial Measurement Units (IMU) and other sensors to support automated recording, recognition, and quantification children's play behaviors for subsequent analysis [61]. While adults use language and various abstractions and representations as their primary means of communicating with the world, TUIs create a unique space for children to express themselves since they are "easier to learn and use", as well as "draw upon physical affordances" and "support cognition through physical representation and manipulation" [62].

Blocks are the most widely accessible play object in early childhood classrooms [63, 55],and a popular form for creating playful interactions among children. Various TUIs were designed to assess and treat children using gameplay with automated blocks. For example, Vonach et al designed "MediCubes" that measure children's physiological parameters during play [64]. Jacoby proposed PlayCubes that assessed children's construction ability using a TUI [10]. StackBlock is a block-shaped interface that detects flexible stacking by embedding a matrix of infrared LEDs and phototransistors [65]. Our approach builds on these past projects and works towards providing young children at-risk of mental health problems after natural disasters a non-verbal TUI-based medium that would allow them to relate to and directly communicate physical elements of their traumatic experience.

2.2.3 Activity Sensing and Detection Techniques

Nowadays, one of the mainstream techniques used in action detection in children is image processing with cameras. Wang, Liu and Yang have presented various techniques for children's activity analysis using distributed cameras and Machine Learning algorithms [66, 67, 68]. While generally effective, there are some common difficulties when retrieving activity information using camera-based methods. First, in the preparation stage, the location of the cameras needs to be well-designed to establish camera views that capture quality information. Second, during data collection, children's actions are highly flexible and unpredictable, which makes occlusions by objects in the space and on children's bodies difficult to avoid.

Another trend in activity-detection is embedded sensors inside tangibles using Machine Learning methods to model data acquired by the sensors. "Child'sPlay" by Westeyn et al. uses a SVM (support vector machine) to enable the automatic recording, recognition, and quantification of play behaviors in children [61]. Hosoi et al. created IMU-embedded toy blocks, and modeled the raw acceleration data into actions with SVM, to recognize and assess building processes during play with toy blocks [69].

A common problem among all the above Machine Learning-based action-characterization methods happens after data acquisition. It takes considerable time to annotate and label the raw data [61, 68], and the acquired data is often imbalanced since it is extremely hard to ask children to perform the certain tasks [66, 67, 68, 61]. The result is an approach that does not generalize well among all children.

To avoid the above problems, we use sensor-embedded blocks with a state-machine algorithm to characterize different actions. The are several benefits to this approach. Technically, the time and expertise needed for site-specific setup are low, individual blocks are easily identified, and the blocks are durable. The state machine structure frees us from acquiring and labeling the balanced actions. Using simple, durable structures allowed our field studies to progress without excessive preparation and interruptions.

2.3 AssessBlock Design and Implementation

In order to enable the investigation of a possible connection between children's toy block play pattern and mental states we used an approach to capture data directly from the play activity as unfolded. We needed to design toy blocks that would retain the familiarity of commonly used wooden toy blocks, but at the same time provide intrinsic tracking of movements and information on the overall playing activity state. The goal was to design a set of simple, sturdy toy blocks that could be quickly deployed in real-world settings, such as kindergartens and used by caregivers, without the need to set up trackers or cameras. We propose a set of specific design guidelines focused on three aspects: appearance, tactile properties, and data acquisition.

2.3.1 Design Criteria

Appearance: the blocks must retain the size and appearance of traditional wooden building blocks to preserve the familiarity of the block play experience. Design should follow the traditional toy blocks' color scheme: using primary colors (red, blue and yellow) first, then secondary colors (green, orange, and violet). Blocks also need to be hollow and large enough to embed sensors. Embedding sensors must be completely hidden, including lights and sounds, to avoid causing children to become curious about the inside of the blocks.

Tactile properties: the blocks need to be affordable and sturdy enough to endure shaking, dropping and throwing. To preserve the tactility of wooden blocks, the build material should not only look like, but also provide sensory properties, such as hardness, sensory warmth, and thermal behavior of wooden blocks as characterized by Sadoh el al. [70]. The interactive blocks should also match the weight of well-designed commercial products.

Data acquisition: the system needs to allow quick and simple deployment in daycares, clinics, or evacuation centers. The sensor embedded in the blocks needs to unobtrusively and robustly capture and transmit human activity data. A good battery charge needs to be maintained during each

Figure 2.3: Toy blocks with the embedded Bluetooth-IMU-sensors

play session. Even it is not absolutely necessary, we suggest real-time data transmission and monitoring during play sessions. Since the available number of children for study is limited, ensuing optimal sensor operation, and accurate data gathering is crucial.

Our design goals and operational constraints led us to a relatively simple approach: we implemented a set of sturdy Bluetooth IMU-embedded toy blocks (Figure 2.3), retaining the appearance of familiar wooden toy blocks, allowing for real-time capture of basic play behaviors.

2.3.2 Physical Specification

Our block prototypes are made with PVC Foam Board providing a warm, hard tactile feel. Each wall of a block is sturdily glued. We prototyped two basic shapes: a big block, measuring $100mm \times 50mm \times 25mm$ and $90g$; and, a small block measuring $50mm \times 50mm \times 25mm$, and weighing $45g$. We used paper clay filling to achieve traditional wooden block's weight. The dimensions and the mass (including internal sensor) followed Nichigan Original's Wooden Tsumiki [71], one of the widely available toy block sets on the Japanese market.

2.3.3 Sensor

A wireless IMU sensor is fixed inside each block using Velcro. That, combined with the pressure from the lid, ensures the sensor will not move when

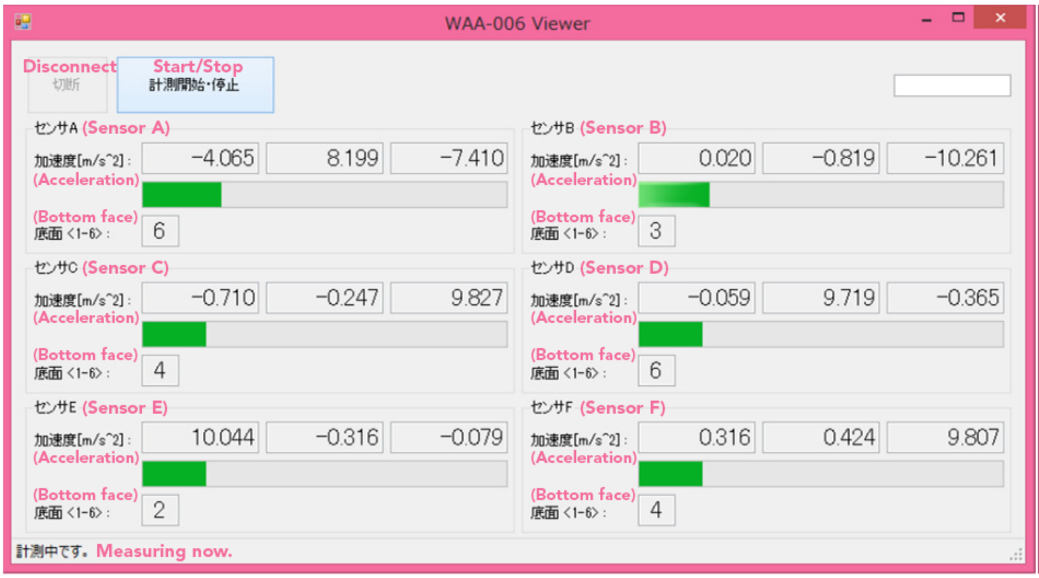


Figure 2.4: Software interface on the host computer

shaken or thrown (Figure 2.3). The wireless IMU sensors (TSND121, ATR-Promotions [72]) hidden in each block have the following specifications:

- Triple axis accelerometer and gyroscope with 3 working axes (X, Y, Z) providing a maximum sampling rate of 1000 Hz;
- Triple axis magnetometers with 3 working axes (X, Y, Z) and maximum sampling rate of 100Hz;
- Maximum acceleration detection range per axis of $\pm 16G$
- Binary format data output;
- Bluetooth communication;
- Built-in battery.

The raw sensor data included x, y, z-axis accelerometer and gyroscope values, which were sent in real-time to a host computer via Bluetooth using a 50Hz frequency, providing data transmission as well as monitoring and inspection capabilities during play sessions. The interface used for data collection is shown in Fig 2.4.

(a) Move (b) Hold (c) Stand (d) Lay

Figure 2.5: Block play actions characterized from pilot study

2.3.4 Block Play Features

We then calculated the numerical activity data from the raw data. To better understand the fundamental play activities that characterize behavior, we conducted a pilot study and observed free block play of 30 kindergarten children. From this study, we derived the following 7 fundamental activity features representing how active and constructive a play session is:

Time: total time between the start and stop of the program;

HoldTime: total time at least one block is held in hand but not moving. An example of holding a block can be found in Figure 2.5b;

MoveTime: total time at least one of the blocks is moving. An example of moving a block can be found in Figure 2.5a;

Movement: A sum of the magnitude of all three-axis acceleration values within a play session. This value is not equal to velocity since the sensor introduces accumulated error over time, however, it provides data indicating speed variations which are sufficient for comparison between subjects.

We further define two states during which a block is placed. When the largest face of a block contacts the ground, we call it "laying" (see Figure 2.5d), while "standing" refers to the state when any other face contacts the ground (see Figure 2.5c).

StandTime: the total time when a block is in a "standing" state;

StandCount: the number of events when the placing is classified as "standing";

LayCount: the number of events when the placing is classified as "laying".

2.3.5 Preprocessing and Play Features Detection

As illustrated in Figure 2.6, we implemented a threshold-based state machine algorithm to extract total counts and total time for the actions described above.

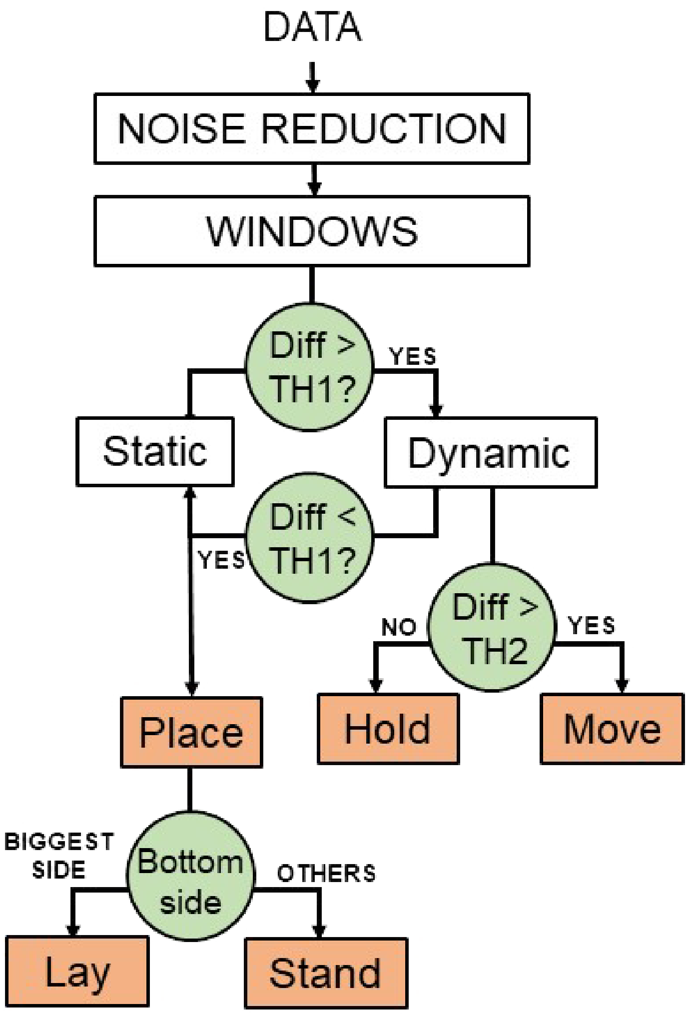


Figure 2.6: A threshold-based state-machine structure that recognizes different actions

We first processed raw data to extract difference (Diff) and "Bottom Side". We applied a moving-average filter to the raw accelerometer data using the unweighted mean of 5 data points to filter out high-frequency background noises. We then extracted Diff by (1) applying a sliding window of 20 data points with 50% overlap with the previous window; and, (2) computing Diff as the differences of the average of two adjacent sliding windows. This approach has proven effective as a motion data processing methods [73, 74]. We pre-assigned an ID to each face of a block and used the term Bottom Side to refer to the side facing the ground when a block is placed. Since the raw acceleration data includes gravity, determining the axis gravity was pointing toward allowed us to identify the ID of the Bottom Side for each placement. We then extracted the numerical counts and time of each action using the structure shown in Fig 2.6.

We next evaluated the accuracy of each action. Time and Movement were not evaluated since data were captured by the sensors and system directly. For count-related actions (StandCount, LanCount), we performed the action 100 times and recorded the count of detected actions. The accuracy is calculated as follows:

$$accuracy = \frac{count_{detect}}{count_{total}} \tag{2.1}$$

The accuracy is 98.0% for StandCount and 96.0% for LayCount. For time-related action features, we performed the action 10 times, for 10 seconds each time. We compared seconds detected with the ground truth in each trial to establish the error rate for each trial. Accuracy is calculated as follows ($n = 10$):

$$error\,rate = \left|\frac{time_{detect}}{time_{total}}\right| \tag{2.2}$$

$$accuracy = 1 - \frac{1}{n}\sum_{i=1}^{n} error\,rate_i \tag{2.3}$$

The accuracy is 98.4% for StandTime, 80.0% for MoveTime, and 66.2% for HoldTime. Play feature detection achieves a high degree of accuracy in Stand and Lay related features, and satisfactory accuracy in MoveTime detection. Accuracy is notably low for Hold Time, for which we suggest the following reasons: HoldTime is hard to detect precisely using acceleration data, with thresholds between movement and stasis. From our observations, a Hold tends to be classified as static, especially during the latter half of a 10-second hold when the holding arm leans against the table making movement minimal. However, since all participants suffer from this offset and children generally do not hold blocks statically for long periods, we consider these anomalies manageable during initial investigations and will investigate methods for improvement.

2.4 Area-based Field-Study Design

Based on our research question - can sensor-enabled toy blocks assess post-disaster stress in children? we designed a play study that collected children's block play features (from AssessBlocks), children's play-related behavior (through video-coding), and stress measurements measured (on-site and through video-coding) from each play session.

2.4.1 Stress Measurements

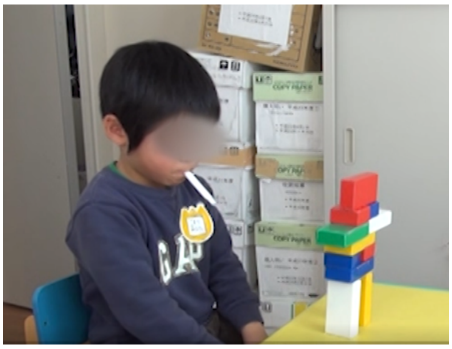
Currently, the best method for collecting stress measurements is self-reporting [75, 76, 77]. However, this method is ineffective with young children who are in the early stages of cognitive and communicative development. As a result, evaluations by professionals and caregivers, combined with biomarkers that can be captured and measured by instruments are often used together to indirectly capture stress in children [78, 79, 80, 81].

In our studies, we captured data using 3 established measurements and biomarkers related to stress:

- Salivary Alpha-amylase activities (sAA);
- Observation Scale of Behavioral Distress (OSBD);
- Visual Analogue Scale of Anxiety (VAS).

Salivary Alpha-amylase Activities (sAA) is recognized as a sensitive, but non-invasive biomarker for stress-induced changes in the body connected to activity in the sympathetic nervous system [82]. Alpha-amylase production in the salivary glands increases in response to psychological and physical stress, and has been shown to be an accurate marker of activity

(a) Instrument for measuring sAA

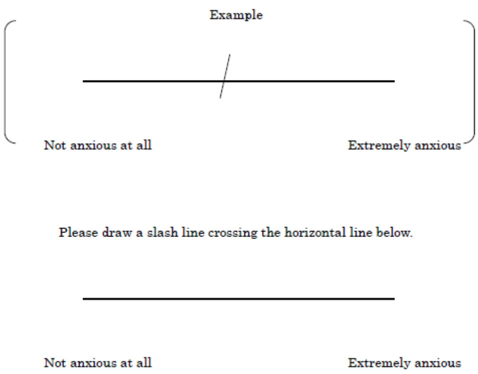


(b) A child is getting sAA measured

Figure 2.7: sAA measurments

		Point
1.	Searching for something	1.5
2.	Crying	1.5
3.	Crying and Screaming	4.0
4.	Physical resistance	4.0
5.	Linguistic resistance	2.5
6.	Seeking for emotional support	2.0
7.	Saying painful	2.5
8.	Aggressively shaking the body	4.0
9.	Saying scary	2.5
10.	Body is stiff	2.5
11.	Nervous behavior	1.0
12.	Neutral facial expression	0.5
13.	Smile	0.0

(a) OSBD form



(b) VAS for Anxiety form

Figure 2.8: Forms for evaluating children's stress level

in the autonomic nervous system. This approach is commonly used for PTSD-related assessments [83, 84]. In our study, we measured sAA before and after the experiment by asking participants to hold the measurement paper in their mouths for 10 seconds (see Figure 2.7).

Observation Scale of Behavioral Distress (OBSD) is a scale developed to measure children's behavioral responses to events that impact health and wellbeing. OSBD scores can be correlated with ratings of pain, anxiety, and physiological measures before, during, and after such events, and have been effective for evaluating children [85]. We use OSBD for stress evaluation because of its ability to capture subtle changes in a short time [85] since we cannot subject children to long study periods yet need to capture data efficiently. The OSBD measurement is presented as a form as shown in Figure 2.8a. It takes 13 measurements to capture severe stress responses such as physical resistance, mild stress responses such as asking for help or support, as well as calm play states. The sum of these measurements indicate behavioral stress in a moment of play. During the experiment, the OSBD was measured by an on-site psychologist three times; once at the beginning, once in the middle (approximately 10 minutes later), and then at the end of the play session.

Visual Analogue Scale of Anxiety (VAS) measures a characteristic or attitude ranging across a continuum of values. As shown in Figure 2.8b, VAS captures measurements by asking the observer to draw a vertical line across a horizontal scale indicating the value. The left side of the horizontal scale indicates the minimum value and the right the maximum. VAS is often used in epidemiological and clinical research to measure the intensity or frequency of various symptoms [86]. Its measurement has been used to collect adult's self-evaluation of mood, stress and health across different times of a day [77] as well as parent and staff reports on children's fear, pain and stress [87, 78]. In our study, we asked the child's caregiver to measure the child's anxiety before and after the experiment session.

Our selection of stress measurements was inspired by previous work using these three measurements in combination when evaluating stress in children undergoing medical procedures [78]. The combination of sAA and VAS has also been studied for measuring pain perception in children [79].

Together with sAA, Salivary Cortisol is also commonly used to capture stress bio-markers in children [88, 81, 89, 90], however, we preferred sAA because: (1) it is less invasive, more comfortable, and uses simpler equipment helping us maintain an orderly environment and avoiding a more medical setting which might increase stress in the children [90, 81]; and, (2) Cortisol and sAA are often not correlated [88, 81, 89] while sAA has been shown to be related to other stress measurements such as HR [81], and negative behavior measurements[90]; and, (3) sAA is well-suited for measuring mild to moderate stress responses [81, 80].

Following on previous work [78, 81, 89, 88], we were able to document time-based measurements at before (sAA, OSBD, VAS), during (OSBD only without interrupting the study) and after play (sAA, OSBD, VAS) to collect comprehensive measurements for each session without introducing interruptions.

2.4.2 Play Behavior Measurement

Based on the anecdotal evidence of a change in play behavior in children after the earthquake and tsunami, and on our observations from AssessBlock's pilot study, we designed 4 behavior measurements we speculated would correlate with a child's stress:

- Concentrated Time;
- Lost Time;
- Stacking Time;
- Flat Time.

Concentrated Time is calculated by accumulating the time a child is concentrated while playing with blocks, instead of running around, talking, or performing actions that are not block-play related.

Lost Time is calculated by accumulating the time a child does not know what to do, seeks help, or looks for encouragement.

Stacking Time is the accumulated time that a child is concentrated in stacking and building 3-dimensional structures. Examples are found in Figure 2.1a and Figure 2.1b.

Flat Time denotes time that a child is merely placing blocks flat down, with the largest side contacting the table. An example is showing in Figure 2.1c. This behavior was reported as happening with greater frequency immediately following the 2011 event.

The first two categories, Concentrated Time and Lost Time, are often used when examining behavioral conditions such as ADHD in children [15, 54]. The ability to build is commonly used in cognitive development checkups for three-year-olds in Japan [7]. We could not be certain whether we could use these features to assess stress but the approach seemed promising. We worried our approach might over-complicate the implementation of AssessBlock to detect these features, so to account for this we asked psychologists to video-code when measuring OSBD while watching videos of the play session. If proved to be significant identifiers of stress, it would be possible to use AssessBlock to capture data in the future, but we needed to ensure accuracy in the short term.

2.4.3 Participants and Timeline

Six months after the 2011 Earthquake and Tsunami, we started to contact the kindergartens to gather evidence of altered block play behavior, while the development phase of AssessBlock began. Working with childcare and community workers, we designed an area-based field study in one of the most affected areas - Sendai city, Japan. After getting Ethics agreements approved by affiliated organizations and formal agreements with the parents of each participant, our experiments took place between September 2013 and November 2015.

The participants we recruited were preschoolers aged 2.1 to 6.9 years old, among whom toy blocks are known to be particularly popular [63, 91, 92]. The participants were recruited from the following three locations.

Coastal kindergarten. From Oct 2013 to Feb 2014, the play study was conducted with 15 children (aged 5.9 to 6.5) at a kindergarten located in a coastal town. Participants were 3.0 to 4.0 years old on the day of the disaster,. This kindergarten was the one most damaged by the tsunami in the area. On that day, shortly after evacuating to a hill, the first floor of the school building flooded. While waiting for rescue, children watched the tsunami approaching, carrying debris and washing away almost everything in its path. After the event, the kindergarten was closed for two months. Among children lived in the coastal area, many families lost their houses and jobs, and children and their families had to live in temporary shelters for periods of several months to 5 years. The children from this kindergarten were noticeably nervous, passive, and commonly had difficulty concentrating.

Inland kindergarten. From Jan 2014 to April 2014, 17 children (aged 5.1 to 6.9) from an inland kindergarten participated in the study. These children were aged 2.3 to 4.0 on the day of disaster. Children in this group experienced the earthquake, but not the tsunami. Significantly, the building housing their kindergarten was not damaged or interrupted by the 2011 natural disaster.

Inland children's center. From Sep 2013 to Nov 2015, 20 children (aged 2.1 to 3.8) were recruited from a children's play center, where children came with their parents for group play and socialization. These children were 0 - 1.3 old at the time of disaster and were out the reach of tsunami, and thus less impacted by the combined effects of the earthquake and tsunami experienced by those from coastal areas.

Whatever the specific experience on that day, all experienced continued aftershocks. As of 16 March 2016 there were 869 aftershocks of over magnitude 5.0; 118 over magnitude 6.0; and, 9 over magnitude 7.0 [93]. The number of aftershocks experienced was shown to be associated with decreased health across Japan [49]. Studies have suggested that the earthquake itself might not as traumatic for the Japanese residents who are resilient to earthquakes, while the unexpected and record-setting tsunami that caused so much death and destruction triggered greater stress and sorrow [45]. Given that direct exposure to traumatic events is a predictor for higher levels of post-traumatic stress, we felt an area-based approach was a good starting point for our research.

Figure 2.9: Kindergarten rooms for children's study

2.4.4 Environment, On-site Procedure and Data Collection

All experiments were conducted in the children's familiar environment. Inside the room where the children usually play, a child's desk and chair were prepared and a set of 12 blocks was placed on the desk (see Figure 9). Studies with younger children included a parent, and those in the kindergarten included the students' regular teacher. We kept rooms quiet and well-lit to reduce potential for stress. A pediatric psychologist, and a psychology student were in the room for on-site support and observation. Two HD cameras aimed in different directions captured an audiovisual record of the children playing.

Each child was invited to the desk to play with the AssessBlock toys for a free play session of approximately 20 minutes. Children were free to stop early or continue longer if they wished. Before playing, a parent or a teacher completed consent forms, and filled in a VAS of Anxiety form. The pediatric psychologist completed the OSBD form, and conducted the sAA measurement which includes instructing a child to hold a small paper test strip above the tongue for 10 seconds (Figure 2.7). The play session then started, and a student research assistant started the AssessBlock program remotely to document the child's block play features. OSBD was evaluated again by the psychologist 10 minutes into the session. After a child stopped, the AssessBlock program was wirelessly stopped, and sAA, OSBD, and VAS were again measured and evaluated.

When the on-site experiments were complete, four more professionals rated each child's OSBD before, during, and after a session by observing the recorded video while video-coding play behaviors for each child. Data from the on-site psychologist, and the four external evaluators were averaged to arrive at a final score.

2.5 Results

In this section, we first present a profile of the data we collected, including demographic factors of the sample, and an examination within each group of data. We then explore the relationships between demographic factors and stress measurements, as well as the association between block-play related features and stress measurements. Spearman's Rank Correlation is used to assess the relationship between a pair of variables since most are either on a ratio or ordinal scale.

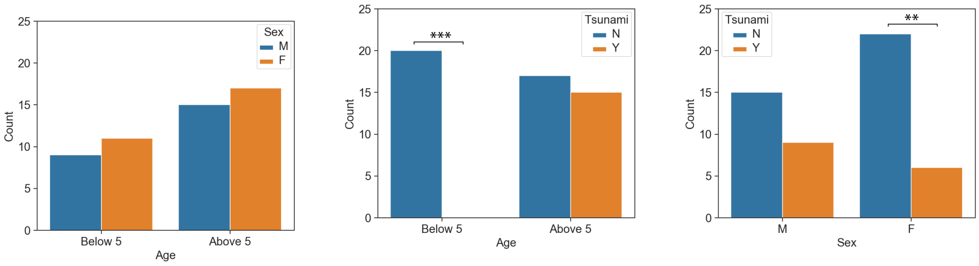
2.5.1 Data Profile

Participants Profile and Trauma Events

We conducted our experiment with 52 preschoolers (24 male, 28 female), aged 2.1 to 6.9 years old (mean = 5.0, SD = 1.74), from three locations in the Tohoku region of Japan. 15 children (9 male, 6 female) aged 5.9 to 6.6 (mean = 6.4, SD = 0.30) came from the coastal kindergarten. They were aged 3.0 to 4.0 (mean = 3.54, SD = 0.32) when the 2011 events took place. 17 children (6 male, 11 female) aged 5.1 to 6.9 (mean = 6.3 SD = 0.52) were from the inland kindergarten inside the city. They were aged 2.3 to 4.0 (mean = 3.36, SD = 0.52) at the time. Another 20 children (9 male, 11 female), aged 2.1 to 4.4 (mean = 3.0, SD = 0.73) were from an inland children's center and were aged 0 to 1.3 (mean = 0.51, SD = 0.59) at the time of the event. There is a differences in age scale since children from the first two locations are all above 5 years old, while those from the third group are all under 5 years of age (Figure 2.10a).

At the time of the study, 5 to 6 year olds were the youngest pre-school tsunami victims we could recruit, because those that were older at the time of the tragedy were now in elementary school and were no longer used to playing with blocks.

Studies have shown that traumatic events emerge as a significant con-



(a) Age (below 5/above 5) of male and female children

(b) Age (below 5/above 5) of children with and without tsunami experience

(c) Gender of children with and without tsunami experience

Figure 2.10: Demographic characteristics and trauma experience of the participants. Bars show counts of participants under different demographic characteristics. Statistically significant differences between the counts of different subgroups is denoted by asterisks (** indicates p < 0.01, *** indicates p < 0.001)

tributor to PTSD [50, 51, 83, 94]. In this study, we emphasize the tsunami experience as a traumatic event for participants. Those from our first field study location directly experienced the tsunami. On March 11, 2011, these children were rushed out of their classrooms and evacuated to a hill behind their school to find safety. Children from the other two locations might had experienced the 2011 earthquake but were out the reach of the tsunami.

In the sample, the traumatic event - tsunami experience, is associated with age when age is classified into Above 5 and Below 5 years old ($\chi^2(1) = 10.99, p < 0.001$). The trauma experience varied significantly among children below 5 years old ($\chi^2(1) = 20.0, p < 0.001$) (see Figure 2.10b) due to none of those under 5 having witnessed the tsunami.

Among 24 boys (9 with tsunami experience) and 28 girls (6 with tsunami experience), we did not observe a significant association between tsunami experience and gender; however, the traumatic experience varied significantly within the girls' cohort ($\chi^2(1) = 9.14, p = 0.002$) (see Figure 2.10c)

Stress Measurements

The stress biomarker sAA is measured at the beginning and end of each session, resulting in two features - sAA Before and sAA After. VAS from a caregiver or a kindergarten teacher is collected at the beginning and end,

Table 2.1: Descriptive profile of stress measurements

Feature	Average	Standard deviation	Range
sAA Before	67.09	55.93	4.00 - 240.00
sAA After	69.92	57.44	3.00 - 267.00
sAA Ave	68.51	48.33	5.50 - 237.00
sAA Diff	33.50	109.13	-96.00 - 563.64
OSBD Before	0.85	1.05	0.00 - 4.00
OSBD Middle	0.65	0.93	0.00 - 4.00
OSBD After	0.45	0.67	0.00 - 4.00
OSBD Ave	0.65	0.69	0.00 - 2.67
OSBD Diff	-0.48	1.19	-3.50 - 4.00
VAS Before	4.32	2.83	0.00 - 9.60
VAS After	2.10	2.24	0.00 - 8.20
VAS Ave	3.21	2.13	0.30 - 8.60
VAS Diff	-2.22	2.81	-7.90 - 7.30

Table 2.2: Correlations among stress measurements

	sAA Before	sAA After	sAA Ave	sAA Diff	OSBD Before	OSBD Middle	OSBD After	OSBD Ave	OSBD Diff	VAS Before	VAS After	VAS Ave	VAS Diff
sAA After	**0.520*****												
sAA Ave	**0.824*****	**0.889*****											
sAA Diff	**-0.348***	**0.536*****	.157										
OSBD Before	0.026	0.009	0.051	0.085									
OSBD Middle	-0.263	-0.131	-0.216	0.095	**0.527*****								
OSBD After	**-0.348***	0.010	-0.148	**0.309***	**0.457*****	**0.819*****							
OSBD Ave	-0.174	-0.057	-0.114	0.160	**0.784*****	**0.889*****	**0.813*****						
OSBD Diff	**-0.274***	-0.097	-0.208	0.027	**-0.700*****	-0.207	0.107	**-0.396****					
VAS Before	-0.205	-0.131	-0.139	-.009	**0.314***	0.167	0.216	0.229	-0.136				
VAS After	-0.107	0.166	0.078	0.248	0.150	**0.380****	**0.396****	**0.278***	-0.001	**0.479*****			
VAS Ave	-0.199	-0.019	-0.076	0.109	**0.275***	**0.346***	**0.368****	**0.333***	-0.100	**0.878*****	**0.802*****		
VAS Diff	0.183	0.243	0.212	0.140	-0.143	0.098	0.036	-0.001	0.059	**-0.756*****	0.057	**-0.475*****	

Note: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, bold values indicate significant correlations.

resulting in two features - VAS Before and VAS After. OSBD is evaluated at the beginning, middle, and end, both on-site and using the video coding, resulting in three features OSBD Before, OSBD Middle and OSBD After.

We then average the features of each stress measurements, to obtain a stress indicator for the entire session - sAA Ave, OSBD Ave, VAS Ave (Ave: an abbreviation for Average). We also calculated the changes of each measurement over the play session for each individual. For OSBD and VAS, we define the differences as after-values minus before-values, resulting in OSBD Diff and VAS Diff (Diff: an abbreviation for Differences). For sAA, we calculated the percentage change by taking the difference between post- and pre-play and dividing by the post value, resulting in sAA Diff. In general, a positive Diff indicates an increased stress measurement, while a negative Diff shows a decrease. The average, standard deviation and range values of the 13 stress measurements across 52 participants are presented in Table 2.1.

To briefly investigate the association between three stress measurements, which includes 4 sAA features, 5 OSBD features, and 4 VAS features, Spearman's rank correlation coefficients are computed between pairs of variables, as shown in Table 2.2.

The measurements within each stress category are highly correlated. Before and After measurements are positively correlated in sAA ($r = 0.52, p < 0.001$), OSBD ($r = 0.46, p < 0.001$) and VAS ($r = 0.48 p < 0.001$). The results indicate that a child with a relatively high stress measurement before a session tended to have a relatively high stress measurement after a session. Averages of all three stress measurements are positively correlated to their Before and After measurements, which matches expectation since Ave is a combination of Before and After and these two are shown to be correlated. Diff is negatively correlated to Before value in sAA ($r = -0.35, p < 0.05$), OSBD ($r = -0.70, p < 0.001$) and VAS ($r = -0.756, p < 0.001$), indicating that those with a high stress value before the session reduce more stress during the session. Diff is positively correlated to After value in sAA ($r = 0.54, p < 0.001$), indicating that those who have a low sAA value after the session show greater reductions during the session. Meanwhile, Diff is negatively correlated to Ave value in OSBD ($r = -0.40, p < 0.01$) and VAS ($r = -0.48, p <$

0.001), indicating that those with greater OSBD and VAS reductions in the session tend to have a low average value.

We then notice that OSBD and VAS are positively correlated in Before ($r = 0.31, p < 0.05$), After ($r = 0.40, p < 0.001$), and Ave ($r = 0.33, p < 0.05$), while similar correlations are not found in the other two pairs - OSBD and sAA, VAS and sAA. None of VAS measurement is correlated to sAA, while some OSBD measurements are correlated to sAA indirectly: OSBD After is correlated to sAA Before ($r = -0.35, p < 0.05$) and sAA Diff ($r = 0.30, p < 0.05$), and OSBD Diff is correlated to sAA Before ($r = -0.28, p < 0.05$).

Fundamentally, sAA measures physiological and psychological stress while OSBD and VAS measures observable behavioral stress. Based on the above observations, two behavioral stress measurements OSBD and VAS agree in Before, After, and Ave, while the same pattern is not observed between stress biomarker sAA and behavioral stress measurements.

OSBD is evaluated in a more objective and unbiased manner in comparison with to VAS. To simplify the dimension of our targets, in the following analysis we use OSBD measurements to indicate behavioral stress, and we use sAA to indicate physiological and psychological, or inner stress. To further reduce dimensions, we also omit OSBD Mid value since it is highly correlated to OSBD Before ($r = 0.53, p < 0.001$), After ($r = 0.82, p < 0.001$), Ave ($r = 0.89, p < 0.001$), and is included in the calculation of Ave. A total of 8 measurements, including stress biomarker, sAA Before, After, Ave, Diff, and behavioral stress, OSBD Before, After, Ave, Diff, are used in the following analysis.

Block Features and Video-Coded Play Behavior

Block features, in time and count, and play behavior, in time, are computed across a play session. Since the total length of a session differs between individuals, we further divide each feature, excepting the total time, by the total time in minutes, to obtain features per minute. The average, standard deviation and range values of the 7 block play features and 4 play behaviors across 52 participants are presented in Table 2.3.

The association among play behaviors, among block play features, and

Table 2.3: Descriptive profile of block and play behavior features. Time feature is documented in min, and other features are documented in $/min$

Feature	Average	Standard deviation	Range
P.Conc[a]	0.93	0.16	0.26 - 1
P.Lost	0.30	0.28	0 - 1
P.Stack	0.82	0.21	0 - 1
P.Flat	0.35	0.27	0 - 0.95
B.Time[b]	19.97	3.26	8.18 - 26.54
B.MoveTime	0.79	0.20	0.30 - 1
B.HoldTime	0.17	0.13	0.01 - 0.55
B.StandTime	0.58	0.27	0.05 - 1
B.StandCount	12.87	5.51	3.88 - 27.92
B.LayCount	11.59	4.23	2.41 - 22.96
B.Movement	23.22	8.63	6.61 - 48.98

[a] P. = Play behavior. Conc = Concentration Time
[b] B. = Block feature

Table 2.4: Correlations among play behavior, block features, and between them

	P. Conc	P. Lost	P. Stack	P. Flat	B. Time	B. MoveTime	B. HoldTime	B. StandTime	B. StandCount	B. LayCount	B. Movement
P. Lost	0.022										
P. Stack	0.248	0.156									
P. Flat	0.169	0.147	**-0.642*****								
B. Time	0.065	0.024	0.247	-0.197							
B. MoveTime	-0.185	-0.116	-0.197	0.109	-0.173						
B. HoldTime	-0.046	-0.168	0.221	-0.248	**0.348***	0.076					
B. StandTime	0.086	0.204	0.142	-0.107	-0.140	**-0.541*****	-0.241				
B. StandCount	0.068	**0.298***	**0.423****	**-0.277***	0.113	**-0.389****	**0.368****	**0.481*****			
B. LayCount	0.077	0.107	**0.450*****	**-0.402****	0.255	-0.157	**0.701*****	0.048	**0.794*****		
B. Movement	0.071	0.190	**0.557*****	-0.270	**0.314***	-0.129	**0.649*****	0.061	**0.654*****	**0.745*****	

Note: *: p <0.05, **: p<0.01, ***: p<0.001, bold values indicate significant correlations.

between them are investigated using Spearman's rank correlation. Their bivariate coefficients are shown in Table 2.4.

The 4 play behaviors are relatively independent, except that Stack is negatively correlated with Flat ($r = -0.64, p < 0.001$). This indicates that a child who dedicates more time to stacking spend relatively less time playing flat and vice versa.

In 7 block features, several correlations were found. First, StandCount is positively correlated to LayCount ($r = 0.79, p < 0.001$). Contrary to correlations of play behaviors Stack and Flat, counts of standing and laying blocks are highly correlated in the same direction. StandCount is positively correlated to Movement ($r = 0.65, p < 0.001$) while negatively correlated to MoveTime ($r = -0.39, p < 0.01$). This indicates that children who stand blocks more also moving them faster though not more frequently. StandCount is also positively correlated to StandTime ($r = 0.48, p < 0.001$) and HoldTime ($r = 0.37, p < 0.01$). HoldTime is positively correlated to LayCount ($r = 0.70, p < 0.001$), Movement ($r = 0.65, p < 0.001$), and Time ($r = 0.35, p < 0.05$). Additionally, LayCount and Movement ($r = 0.75, p < 0.001$), StandTime and MoveTime ($r = -0.54, p < 0.001$), and Time and Movement ($r = 0.31, p < 0.05$) are shown to be correlated.

Between block features and play behavior features, we observe that among 4 play features, Stack and Flat are correlated to some block features. The play behavior Stack is positively correlated to LayCount ($r = 0.45, p < 0.001$) and StandCount ($r = 0.42, p < 0.01$), while Flat is negatively correlated to both LayCount ($r = -0.40, p < 0.01$) and StandCount ($r = -0.28, p < 0.05$). Movement is positively correlated to Stack ($r = 0.56, p < 0.001$), indicating that children who moves faster tend to do more stacking. Besides Stack and Flat, StandCount is found positively correlated with Lost ($r = 0.30, p < 0.05$).

In general, many block features are correlated. Those showing the most correlation with others include StandCount and HoldTime. We also found that block features are commonly related to the play behaviors stacking and playing flat, but not with concentrated time.

Table 2.5: Correlations between demographic factor, trauma event, and stress measurements

	sAA Before	sAA After	sAA Ave	sAA Diff	OSBD Before	OSBD After	OSBD Ave	OSBD Diff
Tsunami Exp	**-0.356****	-0.235	**-0.306***	0.211	**0.313***	0.190	0.209	-0.104
Age Above5	-0.259	-0.257	-0.252	-0.016	-0.057	0.119	-0.031	0.119
Age	-0.216	-0.227	-0.244	0.004	-0.080	-0.028	-0.127	0.087

Note: *: p <0.05, **: p<0.01, ***: p<0.001, bold values indicate significant correlations. Tsunami Exp = tsunamp experience (0 = No, 1 = Yes)

2.5.2 Correlations with Stress Measurements

Demographic Factors, Trauma Events and Stress

Previous studies show that socio-demographic factors, such as age and gender operate as significant mediators of PTSD [52, 95, 96], and trauma events emerge as a significant contributor to PTSD [50, 51, 83, 94]. Therefore, we investigated whether or not correlations exist between socio-demographic factors and stress measurements.

We did not observe any significant differences in stress measurements between male and female children using a one-way ANOVA test. From our participant profile, we note that in the data set, those with tsunami experience are concentrated among children over 5 years of age. Thus, we speculate that if a stress measurement is found to correlate with tsunami experience, there is a chance that the correlation is influenced by to the stress measurement's correlation with age, notably, with differences in stress below and above the age of 5. As a result, we added an additional ordinal demographic variable, Age Above 5, to account for this condition. To investigate the relationship between age, tsunami experience, and stress measurements, Spearman's rank correlation coefficients are computed between Tsunami Exp, Age, Age Above 5, and the 8 stress measurements.

As shown in Table 2.5, Tsunami Exp is negatively correlated to sAA Before ($r = -0.36, p < 0.01$) and sAA Ave ($r = -0.306, p < 0.05$), and positively correlated to OSBD Before ($r = 0.313, p < 0.05$). No correlations are found between age and stress measurements, and between Age Above 5 and stress measurements, indicating that the correlation between traumatic event, tsunami experience, and stress measurement are not due to the correlation between age and stress measurements. Notably, Tsunami Exp contribute differently to sAA and OSBD - where correlations indicate that those with the tsunami experience have a lower sAA Before and sAA Ave, and a higher OSBD Before value.

Block Features, Play Behavior, Demographic Factors and Traumatic Events

Previous research suggests that block playing behavior differs across age and gender [55, 7], and boys and girls have different block playing behavior [97, 98]. Since demographic factors and traumatic events may work as mediators for stress, we briefly investigate the correlation between block features, play behavior and those factors.

We first test whether there are significant differences in block features and play behaviors between genders using one-way ANOVA. We do not observe significant differences between genders in any play behaviors. Among block features, HoldTime (F(1,50) = 7.822, p < .001), Movement (F(1,50) = 7.088, p < .01), and LayCount (F(1,50) = 4.575, p < .05) are significant different between genders.

As shown in Table 2.6, among play behaviors, Age is positively correlated to Conc ($r = 0.318, p < 0.05$), Lost ($r = 0.443, p < 0.01$) and Stack ($r = 0.410, p < 0.01$). Among block features, Age is positively correlated to StandCount($r = 0.302, p < 0.05$) and Movement ($r = 0.371, p < 0.01$). Predictably, StandCount, which indicates the 3-dimensional construction, and Movement, which indicates moving speed, increases with the maturation of a child.

Among play behaviors, Tsunami Exp is positively correlated to Conc ($r = 0.283, p < 0.05$), Lost ($r = 0.350, p < 0.01$). Among block features, Tsunami Exp is only positively correlated to Movement ($r = 0.321, p < 0.05$).

Table 2.6: Correlations between demographic factor, trauma event, and block features, play behavior

	P. Conc	P. Lost	P. Stack	P. Flat	B. Time	B. MoveTime	B. HoldTime	B. StandTime	B. StandCount	B. LayCount	B. Movement
Tsunami Exp	**0.283***	**0.350***	0.117	0.208	-0.075	0.061	0.021	0.027	0.177	0.115	**0.321***
Age	**0.318***	**0.443****	**0.410****	0.026	-0.107	-0.152	-0.183	0.232	**0.302***	0.164	**0.371****

Note: **: *: p <0.05, **: p<0.01, ***: p<0.001, bold values indicate significant correlations..

Block Features, Play Behaviors and Stress

Spearman's rank-order correlation coefficients between 7 block features, 4 play behavior, and stress measurements are presented in Table 2.7. Between behavior features and physiological stress measurements, Flat is positively correlated to sAA Diff ($r = 0.28, p < 0.05$). Notable correlations exist between Conc and sAA Before ($r = -0.27 p = 0.053$), and Flat and sAA Before ($r = -0.26 p = 0.06$) though they are not statistically significant since the p-values are slightly greater than 0.05. Interestingly, the negative correlations between Flat and sAA Before, and positive correlation between Flat and sAA Diff indicate that a child who has more time playing without construction tends to have a lower sAA starting value, and greater increase of sAA or less of a reduction after the session.

Among block features, sAA is negatively correlated to StandTime ($r = -0.303, p < 0.05$), and StandCount ($r = -0.345, p < 0.05$). While StandCount and StandTime both point to active construction, above finding indicates that sAA After are lower among those who construct; however, the same observation is not found in other sAA measurements.

For the behavioral stress measurement OSBD, there were on correlations with play behaviors. OSBD After is negatively correlated to the block features Time ($r = -0.44, p < 0.001$) and MoveTime ($r = -0.28, p < 0.001$). OSBD Ave is negatively correlated to Time ($r = -0.32, p < 0.05$). Surprisingly, none of the construction-related block features are correlated to the behavioral stress measurement OSBD.

In general, 2 out of 7 block features, and 2 out of 4 play behaviors show correlation with the physiological and psychological stress sAA. Only 2 from

Table 2.7: Correlations between block features, play behavior, and stress measurements

	sAA Before	sAA After	sAA Ave	sAA Diff	OSBD Before	OSBD After	OSBD Ave	OSBD Diff
P. Conc	-0.269!	-0.221	-0.243	-0.011	0.102	-0.066	-0.114	-0.039
P. Lost	-0.089	-0.092	-0.055	-0.017	0.101	0.153	0.080	-0.022
P. Stack	0.078	-0.133	-0.003	-0.248	0.009	-0.222	-0.229	0.003
P. Flat	-0.259!	0.010	-0.125	**0.280***	0.002	0.168	0.110	0.071
B. Time	0.219	0.160	0.226	0.014	-0.060	**-0.444*****	**-0.317***	-0.242
B. MoveTime	-0.095	0.107	0.014	0.118	0.068	**-0.281*****	0.245	0.076
B. HoldTime	0.094	-0.056	-0.017	-0.073	-0.095	-0.208	-0.251	0.089
B. StandTime	-0.082	**-0.303***	-0.255	-0.151	0.113	-0.010	0.099	-0.137
B. StandCount	-0.013	**-0.345***	-0.220	-0.163	0.010	-0.190	-0.138	-0.035
B. LayCount	-0.050	-0.254	-0.190	-0.103	-0.075	-0.185	-0.220	0.042
B. Movement	0.076	-0.132	-0.039	-0.087	-0.066	-0.235	-0.247	-0.021

Note: *: p <0.05, **: p<0.01, ***: p<0.001, !: p<0.1,bold values indicate significant correlations.

block features, show significant correlation with behavioral stress OSBD. Though not optimal, the results indicate the potential for assessing stress, and particularly internal stress as captured by sAA using toy block play features.

2.6 Discussion

Our study explores a novel approach using sensor-embedded blocks to assess children's stress two years after a large-scale earthquake and tsunami. Our findings provide an encouraging refection on the potential of the approach and help us form and propose guidelines for further research into the assessment of children's mental health using sensor-embedded block approaches. Here we discuss the potential and challenges of this work, the lessons learned, and provide guidelines for future exploration of this area of research.

2.6.1 Potentials

Addressing the long-term RQ: can sensor-enabled toy blocks assess post-disaster stress in children?

This exploratory study and correlation analysis show the potentials of block play features for assessing physiological and behavioral stress in children.

Focusing on block features, we find the 3D construction-related features, StandTime and StandCount, are negatively correlated to sAA After, indicating that fewer standing-related actions are associated with a higher sAA After, and vice versa. In play behavior features, Flat is negatively correlated to sAA Before and positively correlated to sAA Diff, indicating that more "playing flat" is associated with a relatively lower sAA Before value and an increased sAA during the play session. Simultaneously, from our block and play behavior feature correlations, we find Flat is negatively associated with Stack and StandCount. These combined findings indicate that relatively passive play - with more playing on a flat surface, and less 3D construction actions such as standing blocks up, tends to indicate a lower sAA Before, and a higher sAA After. In such cases, children's physiological stress sAA tends to increase.

Our findings also indicate that children who were directly exposed to the tsunami also exhibit lower sAA Before and Ave. In previous studies, Feldman et al. found that children exposed to war who were diagnosed with PTSD exhibited a low-level of stress biomarker sAA, at baseline, following a challenge, and during recovery [83]. Their results though focused on children exposed to a different kind of traumatic event, seem to align with our findings of sAA value among those with direct tsunami exposure, which indicate those who directly experienced the tsunami as a trauma event might be at greater risk for PTSD. Feldman et al. also discovered that children without PTSD employ comfort-seeking strategies while children with PTSD withdraw [83]. Their findings with regard to withdrawal in children with PTSD also seems to align with the "passive play" behavior we observed, and the low sAA Before value associated with "passive play". This result could indicate that "passive play" might be an indicator for children suffering from latent PTSD.

We also observe that the play behavior Concentrated Time is negatively correlated to sAA Before, similar to Flat. Meanwhile, Concentrated Time does not exhibit a positive correlation to sAA Diff as "playing flat" does. Without correlation to Flat, Concentrated Time may not relate to the "passive play" behavior mentioned above. Thus, the length of time a child is concentrated during play may work as a general indicator of physiological and psychological stress, but not as an indicator of "passive play" and the corresponding bio-marker tendencies.

For behavioral stress - OSBD, correlations between its measurements and play behavior features were not found. The only correlations found were between OSBD and computed block features. While most construction-related features do not exhibit a correlation with OSBD, Time and MoveTime are both negatively correlated to OSDB After. Since block features Time and MoveTime are not significantly correlated, their negative correlations with OSBD measurements should be examined independently. We set up our play sessions to run for 20 minutes, but left flexibility for children to choose how long they engaged in order to be as gentle with them as possible. Predictably, some left early and others played longer with total experiment time varing from 8.18 minutes to 26.54 minutes. The negative correlation between OSBD After and Time indicates that those who ended early, probably due to losing interest, tended to have a higher OSBD after play and a higher average OSBD, while the opposite was true for those still immersed in play after 20 minutes. Additionally we noted that those who had longer periods during which at least one block was moving indicated a lower OSBD measure at the end, and vice versa, revealing a negative relationship between active play and a high level of behavioral stress OSBD.

While in the early stages, our exploratory findings suggest a connection between block play actions and stress in children. There seems to be significant evidences to support the possibility of an automated block system that support assessment and maybe even predictions of children's stress in children after natural disasters.

Stress Measuring Method

In this early exploration related to children's mental health, and in particular to PTSD related health issues in children, we used three stress measurements. Distinct from questionnaire and self-evaluation method usually used in studies with adults [77], we used biomarker measurement and behavior evaluation by a third party to approximate stress levels in child. We felt that these were appropriate and relatively unobtrusive methods for working with children and believe they induces less stress fluctuations as a result of running the experiment itself.

We found no convergence between three measurements; however, OSBD and VAS seemed to agree in Before and After values, which validates the credibility of our behavioral measurements. We also found a negative correlation between the Before value of physiological stress sAA and Before value of behavioral stress OSBD. With that, it is hard to simply say the best ground truth for stress is OSBD.

In this study, we consider that OSBD, and particularly OSBD After, captures a child's external and observable stress in daily life, since OSBD Before values may be impacted by unavoidable environmental factors such as being asked to step into an experimental setting. We believe sAA is a promising internal measurement, which might be highly relevant for to traumatic experiences leading to an increased risk of PTSD. sAA reflects on the internal nervous system activity and was found to be related to the reactions to trauma experience and latent stress disorder [83]. In our study, sAA correlated to certain behaviors such as "passive play"; however, the dissonance between sAA and OSBD, and the hidden relationships between external behavioral stress and internal physiological stress requires future investigation.

Mediator and Moderator Effects

A mediator is a bridge between a pair of variables while a moderator regulates the size and direction of the association between two variables [99]. In PTSD research, demographic factors and exposure to traumatic events are often shown to be mediators, moderators, or both [52, 95, 96, 50, 51, 83, 94].

Our findings show that tsunami exposure correlates to both sAA and OSBD, and to block features and play behavior, indicating that tsunami experience may operate as a mediator for stress. Some block and play behavior features also correlate to age and gender, but age and gender do not exhibit the mediation effect since no correlation with stress biomarker or behavioral stress was found; however, these features may work as moderators influencing the correlation between play and stress differently (*i.e.,* among different ages, or genders). Moreover, other mediator and moderator factors such as trauma history, family social status, parental socio-demographic factors, and other relevant demographic details may exist.

Block Features and Play Behavior Features

From our analysis, we find both block features and play behavior features are correlated to stress measurements; however, human-annotated play behaviors did not outperform block features as we expected. Nevertheless, play behavior features did uncover a "passive play" behavior, and indicated that stress may be possibly related to concentration. Thus, play behaviors seem to perform an important role in stress assessment. Extracting these features automatically from AssessBlocks, either by investigating the computational algorithm, or by incorporating new sensors will be necessary to continue this work.

The Predictive Capability of Features

Using the same set of data, previous work shows that block features are reliable for predicting age [100], confirming previous studies of differences in block play between age groups [55] and demonstrating the predictive power of automated blocks. With correlations shown between block features and stress measurement, we see potential for predicting stress and other mental health factors using block features. Even though correlations between block features and stress measurements are not strong, we believe the aggregation of several sets of features may provide the required predictive power. Stress might be predictable using block features and play behaviors, mediated and moderated by demographic factors and trauma experience.

A practical obstacle to examining predictions, particularly with machine learning techniques, is the relatively small data size. Sampling psychological measurements to attain the large datasets necessary for constructing advanced machine learning models is challenging by nature. One way to continue to explore the prediction of stress measurement is to maintain focus on simple and interpretable methods such as linear models and simple nonlinear approaches such as Decision Trees. Another possibility is to explore Multitask Learning, where multiple correlated learning tasks can be solved simultaneously [77]. Since our stress measurement targets are highly correlated, particularly on Before, After and Ave, we are able to multitask targets to obtain a model for prediction.

Block Assessment Approach

All of the children showed interest in playing with our block prototypes, and none tried to open or break them indicating that our design approach achieved its goals. No child demonstrably rejected our study method indicating a level of comfort with our approach to field studies. These small signs seem to validate our toy block-based data collection and assessment approach as non-invasive and child-friendly which are crucial for both ethical and methodological reasons.

2.6.2 Limitations and Challenges

While our results are promising, more work is needed in order to use AssessBlocks as a tool to support mental health assessment. A number of limitations need to be addressed in further iterations.

Data Sampling

In this work, the data size is relatively small, and is unbalanced with regard to exposure to the traumatic event. The traumatic event in our study is particularly important, since Tsunami Experience is significantly correlated to both physiological stress and behavioral stress. In our data set, 15 among 52 had a tsunami experience. As the result of several constraints, all those with tsunami experience came from one location, and all were above the age of 5. This group is the youngest tsunami victim pre-schoolers we could access under the conditions we faced. The coastal kindergarten group was unique, as the only one in Sendai area damaged by the tsunami, and comparable collaborators with similar circumstances would be difficult to find and engage. While we cannot reasonably access comparative data sets, we believe collecting the data from a representative place affected by the natural disaster was crucial. With the data we had, we observed that tsunami experience is significantly associated with stress and age is not. Balanced data, with more children who experienced the tsunami from a wide range of locations would further validate our findings of correlations between tsunami experience and stress, and would help evaluate the predictive capability of our approach. At the same time, given the sensitivity of the topic and those affected, broadening the sampling size presents unique challenges.

Block Features and Play Behavior Features

Some block features, such as HoldTime, demonstrated relatively low accuracy (66.2%) which needs to be addressed. We believe this rate exposes a weakness in threshold-based motion extraction. One of the potential solutions could be introducing sensing modalities to the surfaces of the blocks. As previously illustrated, many block parameters are correlated, with significant Spearman's Rank Order coefficients from 0.31 to 0.79. While it is natural that some actions are related, such highly correlated parameters are usually considered detrimental to statistical analysis and could falsely increase the fit if used in linear models. Thus, a feature selection process is needed for stress prediction with our current set of features. In order to capture play events from different angles, we can explore combining IMU and other sensing modalities, to obtain multi-modal features that are not highly correlated.

2.6.3 Other Observations

Blocks and Therapeutic Effects

From the correlation analysis, we found Diff of both biomarker and behavioral stress are negatively correlated to Before value for children over 20 minutes of play. Does this indicate block play's effect on stress reduction? More long-term follow-up studies and rigorous analysis are needed to answer this question as well as to verify block play's potential therapeutic benefits. For now, we feel our results are sufficient to warrant further research.

Behavior Pattern

From our on-site and video-based observations, some unusual patterns were observed among children who had tsunami exposure and an increased sAA. One such pattern was a behavior of stacking followed by destroying followed by play restricted to a flat surface ("playing flat") (Coastal group P1). This pattern can be related to the "passive play" observed from block features which seemed to indicate a high risk for PTSD. We also discovered that some children who witnessed the tsunami and increased their sAA showed fence building behaviors (Coastal group P14) (Figure 2.1b). Some aggressive and destructive behaviors such as flicking the blocks was also observed (Figure 2.1d) among children who did not witness the tsunami (Inland group P5). Among children who increased sAA, many were observed to lack concentration, confirming the negative correlation between the human-annotated play behavior Concentrated Time and sAA Before measurement. Non-concentrated behaviors, including leaving the blocks aside to explore the room, getting attracted by others nearby, as well as tense or passive body movement were observed. In contrast, concentrated children had many trial-and-error sessions and became more engaged with construction if their structures collapsed.

While the above play patterns, such as fence-building, stack-collapse–flat play, stack–collapse-stack play cannot be formally analyzed using our current play features, they could be captured using sequences of block actions, and the relationship between stress and these patterns should be further examined.

2.6.4 Next Steps

Blocks Design

Based on our preliminary study results and our observation of specific and unique patterns, we propose a new block design framework to capture several categories of data (Table 2.8). The first type is individual block states. This type captures the block side that a child puts down, such as stand up (small side on bottom) or lay down (big side on bottom). The second is hand-to-block interactions, including holding, moving, and shaking. The third is the interaction between blocks in a group, in categories such as stacking, play flat, and start a new location. The final category is the manner of disassembly, including normal, free fall, and aggressive destruction. It is necessary to output these data in counts, time, and sequences. In order to accurately detect actions reflecting block surface connections and involving subtle movement, we propose enhanced sensing methods deployed on the surface of the blocks. Utilizing IMU and the screen of a Smartwatch as

Table 2.8: Activities need to be captured by blocks

Categories	Play Actions
Individual State	stand up
	lay down
Hand-to-block Interaction	hold
	move
	shake
Block-to-block Interaction	stack
	flat
	new location
Disassembly	disassemble
	free fall
	destroy

described in [101], motion data and surface connection can be integrated to detect surface contact-related actions such as Hold, Stack, Flat and Disassemble.

Experiment Design and New Assessments

We also propose investigating other mental health targets such as aggressive behavior, depression, and attention, with AssessBlocks. With enhanced block features and more data, we are eager to explore different kinds of predictive methods to improve reliability and the power of assessment.

Chapter 3

Predicting Behavior Problems with Play Actions and Pattern Engineering

Preface: This Chapter contains modified content from one of my previously published papers [Xiyue Wang, Kazuki Takashima, Tomoaki Adachi, Yoshifumi Kitamura. Can Playing with Toy Blocks Reflect Behavior Problems in Children? In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 540, 1–14.] ©2021 ACM. I adapted the cited paper and reorganized its content to integrate it into the thesis. I performed the majority of the work associated with the aforementioned paper. I contributed to: 1) establishing the presented story, 2) prepossessing data, 3) developing the feature engineer techniques and the prediction models, 4) propose and performed the qualitatively analysis on the model coefficients, 5) encapsulating the result and discussion, and 6) writing the paper.

3.1 Introduction

The last decade has seen a growing trend of behavioral and mental problems in young children, including ASD, ADHD, and PTSD. Such increased challenges in the mental health of children are largely affected by multi-dimensional external factors and such traumatic events as wars and natural disasters [102, 103], family relationships [104], and the influence of media [105]. Behavior problems in children became prevalent with many under-defined latent cases [106], however, assessing and diagnosing children's mental health and behavior problems remain challenging. Since young children's linguistic expressions and cognitive development have not completely matured, traditional self-check and questionnaire-based assessments do not apply to them. Investigating young minds is time-consuming and requires empirical knowledge, subtle observations, and persistent support from caregivers. Daily monitoring and the assessment of children's mental health remain to be less explored.

Among standardized methods assessing children, Child Behavior Checklist (CBCL) [107] appears to be widely-used, affordable, and reliable. It is a multi-axial empirically-based set of measurements that contain three broad groups of behavior problems: Internalizing, Externalizing, and Total Problems. It also carries eight specific syndromes, including Anxiety/Depression and Aggressive Behavior. Within each measurement, three ranges are defined based on age and gender: normal, borderline, and clinical. CBCL creates a profile that gives clinicians an overall picture of the variety and the degree of the behavior problem of children. However, as one of the first step screening tools in the clinical settings to create a behavioral and mental health profile, CBCL is not normally accessible to children and their caregiver in daily settings such as preschools and households. It also requires a caregiver's elaborate knowledge of a child's behavior over the past six months. Therefore, implementing CBCL as a daily assessment tool for every child is impractical.

As a preventive method to support wellness, attaining awareness of personal health and affects in a non-clinical setting is gaining attention in HCI research. Previous work predicted affective states with smartphone touch data [32], and mental well-being with a set of daily activity data including phone calls, sleep-wake patterns, and social activities [31]. Nevertheless, such adult activity data are neither applicable for children nor highly relevant to their high-level health status. For young children, the most basic element of their daily activities is playing. Free play with toy blocks, which is fundamental among preschoolers, is available in most preschools and households [108, 109]. With simple forms and minimal instructions, blocks provide children a space for exploration and expression. Thus, blocks has been used in children's cognitive development checkups and therapies [7, 15, 54]. Certain block-play actions captured by sensors, such as more of laying blocks flat and less of stacking blocks, are found to be correlated to high levels of physiological stress in a child before and after free-play sessions [110]. These prior literature demonstrated a connection between children's mental health and behaviors in a free-block-play session, and indicated that data automatically captured from block-play might be able to replace the observations to infer health status.

Motivated by prior work, we propose a TUI (Tangible User Interface) approach that deeply explores the relationship between free-play with toy blocks and prolonged behavior problems beyond stress. We explore whether and to what extent the quantitative data captured from a block-play session reflects and predicts a range of clinical behavior issues, including Internalizing, Externalizing, and Total behavior problems, as well as such specific syndromes as Aggressive Behavior, all of which can be measured by CBCL. If block-play predicts clinical behavior problems, it can be a powerful supportive tool for monitoring the daily health of children. Our proposed system can be useful in non-clinical and clinical scenarios where (1) CBCL or the knowledge required for assessing child behavior is inaccessible and/or (2) child behavior problems need to be further validated or frequently monitored.

We embedded IMU (Inertial Measurement Unit) sensors into toy blocks to collect children's play data and classified the following basic play actions: *static* (including *stand* and *lay*), *hold*, *move*, *shake*, and *fall*. From 2016 to 2017, our study took place in the area that was devastated by the 2011 Great East Japan Earthquake and Tsunami. This area has a higher prevalence of behavior problems among children due to its devastation that persisted for several years [102, 111]. As a preliminary investigation, we used a population-based approach and examined children from three preschools. We collected the quantitative data of a roughly 20-minute toy-block free-play session from 78 children as well as their CBCL measurements. The results found children with and without clinical behavior problems differed in play actions *hold*, *fall*, *shake*, *lay*, and in total play *time*, and suggested our block approach's potentials in predicting Total Problems, Internalizing Problems and Aggressive Behavior.

The following are our paper's specific contributions:

- We proposed a sensor-augmented free-block-play approach to predict a child's behavior problems in a controlled setting which can be easily constructed in daily lives.
- We quantified and classified play actions with real-world data (50%-88% accuracy) and leveraged sequential play patterns to predict behavior problems (82%-90% accuracy).
- We interpreted the prediction model features and presented insight into three styles of play discovered from the features among children with behavior problems.

Our results suggest initial promise for reflecting clinical behavior in children from a short play session with toy blocks. Currently, insights can be used to support observations and assessments, especially who and what play styles need further attention. Our approach and analysis methods may benefit future researches toward an ultimate goal of predicting, monitoring, and assessing the behavior problems of children in their daily lives.

3.2 Related Work

3.2.1 Children's Mental Health, Behavior Problems and Assessment

Over the past decades, mental disorders are significantly affecting children and adolescents. In 2001, the worldwide prevalence of child and adolescent mental disorders was approximately 10-20% [112], and in 2015 it was 13.4% among 6-18 years old [113]. Among a wide range of mental disorders, the prevalence of anxiety disorders, depressive disorders, attention-deficit hyperactivity disorders (ADHD), and disruptive disorders were the highest, ranged from 2.6% to 6.5% [113]. These mental and behavior disorders are often a comorbidity of such more severe psychiatric disorders as Autism Spectrum Disorders (ASD) [114], Bipolar Spectrum Disorder (BSD) [115], and Post-traumatic Stress Distress (PTSD) [116, 117, 118]. Childhood mental and behavior problems are affected by an aggregation of environmental factors such as negative, inconsistent parental behavior and parental disorder [119], high levels of family adversity [104], stressful social circumstances [120, 121], media usage [122, 105] and trauma events [102]. Findings also suggested that environmental factors indirectly affect children's mental health. The traumatic events, such as earthquake and war, may cause anxiety disorders and PTSD in parents and induce children's behavior problems [111, 103].

Scientific evidence argues that childhood mental and behavior disorders tend to persist into adolescence and adulthood [123, 124, 106], and some deteriorate into much more disabling disorders [112, 125] due to such complex reasons as lack of knowledge about childhood mental disorders, relatively weak advocacy, and insufficient training and resources [106]. When the health problems of children evolve into a global crisis, significant attention must focus on preventive methods, especially since the prevalence is often higher than estimates [126] and include a large number of underdiagnosed cases [116, 127]. Many needs remain unmet in many parts of the world [128, 127].

Although early detection and intervention prevent children's mental and behavior problems, the diagnosis of children is complicated. The standard clinical diagnosis, the Diagnostic and Statistical Manual of mental disorders (DSM), requires physician administration, structured clinical interviews, and consultations with external psychiatrists [129, 130, 131]. As an empirical and questionnaire-based screening method, CBCL and its different translations' reliability have been verified in a large body of literature [112, 132, 133, 134, 135]. CBCL is a pencil and paper test completed by caregivers. It asks about a child's behavior over the past six months and aggregates these data into behavior problem T-scores [107]. The long-term stability of CBCL clinical abnormal behavior was also found in a 4-year follow-up study [136]. Other research has shown that CBCL is predictive and supportive for the diagnosis of DSM symptoms, such as ADHD, bipolar disorder, and anxiety disorder [130, 137, 131, 129, 134]. CBCL has also been extensively used in epidemiological and longitudinal studies as an efficient screening method that creates a behavior and mental disorder profile of the children of a population, such as post-natural disasters [102, 111], post-war crises [103], and life in foster care [138]. Despite CBCL's efficiency, it is not generally used outside of clinical and research situations.

3.2.2 Playful, Interactive Healthcare for Children

Playful or play-based methods are well-established means for supporting children's mental health and well-being. Creative play approaches such as Sand-Play and Painting Therapy are commonly used to treat chronic stress and PTSD [53, 12]. Playing with toy blocks has shown therapeutic results for social withdrawal and ADHD in children [15, 54].

The potential use of TUIs to automate and advance children's healthcare has been explored. Spiel et al. reviewed a body of tangible and playful systems for autistic children that targeted behavior analysis, including diagnosis, monitoring, and therapeutic reviews [139]. Examples include motion-based interactive systems [140], emotional robots [141] and participatory design of smart tangible objects [142]. Playful systems have also effectively supported ADHD children. Quantitative evaluations have used gestures to detect behavior patterns to distinguish ADHD children [143, 144]. WeDA combined touchscreens, tangible objects, and a wearable-based system to diagnostically assess children with ADHD [11]. Besides ASD and ADHD, Fan et al. showed that working with tangible letters helped dyslexic children learn to read and write [40]. Westeyn et al. developed a Child'sPlay system with Inertial Measurement Units (IMU) and other sensor-embedded augmented toys, including puppies, blocks, and rings to support the automated recording, recognition, and quantification of children's play behaviors for development analysis [61]. Although adults use language as their primary means of communicating with the world, TUIs create a unique space for children to express themselves since they are "easier to learn and use", "draw upon physical affordances" and "support cognition through physical representation and manipulation" [62].

Blocks, which are the most widely accessible play object in toddler classrooms [63, 55, 109], are popular forms for creating playful interactions among children. Pullman argued that with maturation, young children transition from transporting blocks to stacking them and then three-dimensional composition [55]. As a result, blocks are used in the cognitive development checkups of three-year olds in Japan [7], and block-shaped interfaces were proposed for health assessments. Vonach et al. embedded sensors in MediCubes to non-invasively measure such children's physiological parameters as pulse, temperature, and blood oxygen saturation during interactions [64]. Jacoby et al. proposed PlayCubes, a children's instruction-based construction ability assessment [10], using a cube-shaped tangible interface called Active-Cube [8]. Hosoi et al. implemented IMU-embedded smart building blocks and demonstrated their ability to classify play actions using lab-collected data [69]. Our approach builds on the designs and implementations of these block-shaped interfaces. Specifically, we aim to provide young children who are at-risk of mental health problems a non-verbal TUI-based medium that allows them to directly communicate the physical elements of their behavior.

3.2.3 Daily Activity Data and Non-intrusive Health Monitoring

Leveraging quantitative daily activity data to imply meaningful health, behavior, and affect information has been getting increased attention. Previous literature forged a link between health and daily activity data from mostly mobile and wearable devices. Daily activity data include smartphone usage, for example calls and text messages [32, 31], as well a other meaningful activity information processed from sensors, for example how many steps a person has walked [145]. They can suggest a broad range of psychiatry phenotypes, such as depression, moods, social connectedness, self-reported health [146, 147, 31, 148]. However, the same scenario is generally not applicable to preschool children.

A large body of work that monitors and predicts children's health is comprised of specifically designed tasks and specific assessment goals, e.g., cognitive ability [10] and ADHD [11]. They are effective with high sensitivity and precision; but the test-like tasks are too specific to merge into daily lives. To integrate the data collection and assessment seamlessly into children's daily settings, the system should balance the specificity and ambiance for efficiency and acceptance. One thread used video and audio recording to ambiently capture activity data in daily settings [149, 150, 144], although they might face such obstacles as occlusion and a vast amount of unspecific information. Others put wearable devices on children as an activity-data collector [151, 152, 153]. Although such devices were effective for data collection, the tolerance of children (especially those at-risk) has been questioned [139].

Another promising method is to examine the data collected with the interfaces they normally interact. Intarasirisawat et al. described how the touch and motion features collected from three popular mobile games (Tetris, Fruit Ninja, and Candy Crush) have the potential to be used as proxies for the conventional cognitive assessment [30]. Mironcika et al. demonstrate that motion data captured from sensors-embedded tokens in the board game play is promising to assess fine motor skills [154]. By discovering the correlations between temporary stress during play and quantitative data captured from toy-block-play, Wang et al. showed the potential for evaluating children's stress with block-play activity [110]. These sensors-embedded interactive devices show promises for health-related uses. Our work further investigates the data collection and analysis methods that can be embedded in the daily lives of children, to infer their mental health and behavior.

3.3 Approach

3.3.1 Toy Blocks Design

We implemented a set of sturdy Bluetooth IMU-embedded toy blocks, Assess-Blocks (Fig. 2.3), resemble the dimensions, the mass (including the sensors' paper clay filling), and firm, warm tactile feelings of Nichigan Original's

Wooden Tsumiki [71], a widely available toy-block set on the Japanese market. Our block prototypes were assembled with PVC form board in primary and secondary colors: red, blue, yellow, green, and white. We developed two types: big blocks, which measured $100 \times 50 \times 25mm^3$ and weighed $90g$; and small blocks, which measured $50 \times 50 \times 25mm^3$ and weighed $45g$. Inside each block, we fixed in the center a Bluetooth IMU sensor (Fig. 2.3) that is resilient to shaking and throwing. Wireless IMU sensors (TSND121, ATR-Promotions [72]) hidden in each block contain a three-axis accelerometer and gyroscope, a Bluetooth, and a built-in battery. The raw sensor data included x-, y-, and z-axis accelerometer and gyroscope values were sent in real-time to a host computer by Bluetooth using a 50-Hz frequency, which was sufficient to distinguish fundamental play actions, validated in our previous studies [69, 110]. During the study, 12 blocks were prepared for each child, and the data were received by two laptop computers on-site (each of which was connected to six blocks with Bluetooth) as the play unfolded.

3.3.2 Experiments Design

Participants

As a preliminary investigation into the relationship between block-play and child's behavior problems, instead of looking for test and control groups of a specific disorder, we sampled children on a large scale, in an area with a high prevalence of behavior problems.

From January 2016 to February 2017, we invited 88 children to join our play study after getting ethics agreement approved from the affiliated organizations and formal agreements from the parents of each participant. The recruited participants were 4.11 to 6.11 years old preschoolers, an age cohort among whom toy blocks are particularly popular [63, 91, 92]. They were recruited from three preschools in Miyagi prefecture, which was devastated by the 2011 earthquake and under reconstruction for years [155]. A population-based report shows that after the disaster, the area's children had a high prevalence of behavior problems [102], and the prevalence persisted even three years after the disaster [111].

After collecting all the data, ten participants were removed from the

Figure 3.1: Preschool rooms for children's block play

analysis due to incomplete CBCLs and accidental sensor failure in either the battery or Bluetooth connection. A total of 78 children (30 girls), 4.11 to 6.11 years old (mean = 5.78, SD = 0.51), were included in the final dataset.

Behavior Measurements

The parents of each participant completed the Japanese version of CBCL for ages 4 to 18 years (CBCL/4-18), [1] which contains 122 items concerning behavior or emotional problems over the past six months. The responses are formatted into 0, not applicable; 1, somewhat or sometimes true; 2, very true or often true. Different items are combined into eight individual syndrome scales: Withdrawn, Somatic Complaints, Anxiety/Depression, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior. All individual syndrome scales are summed into a Total Problems scale. Withdrawn, Somatic Complaints, Anxiety/Depression form an Internalizing Problems scale, while Delinquent Behavior and Aggressive Behavior provide an Externalizing Problem scale. Raw scores are converted to gender and age-standardized T-scores to permit comparisons across gender, age, and scales. It takes about 25 to 30 minutes to complete the checklist.

Procedure

The experiments were conducted during regular school hours inside the preschools. The room where the children usually play included a child's chair and a desk on which a set of 12 blocks was placed (Fig. 3.1). We kept the room quiet and well-lit to reduce any potential stress.

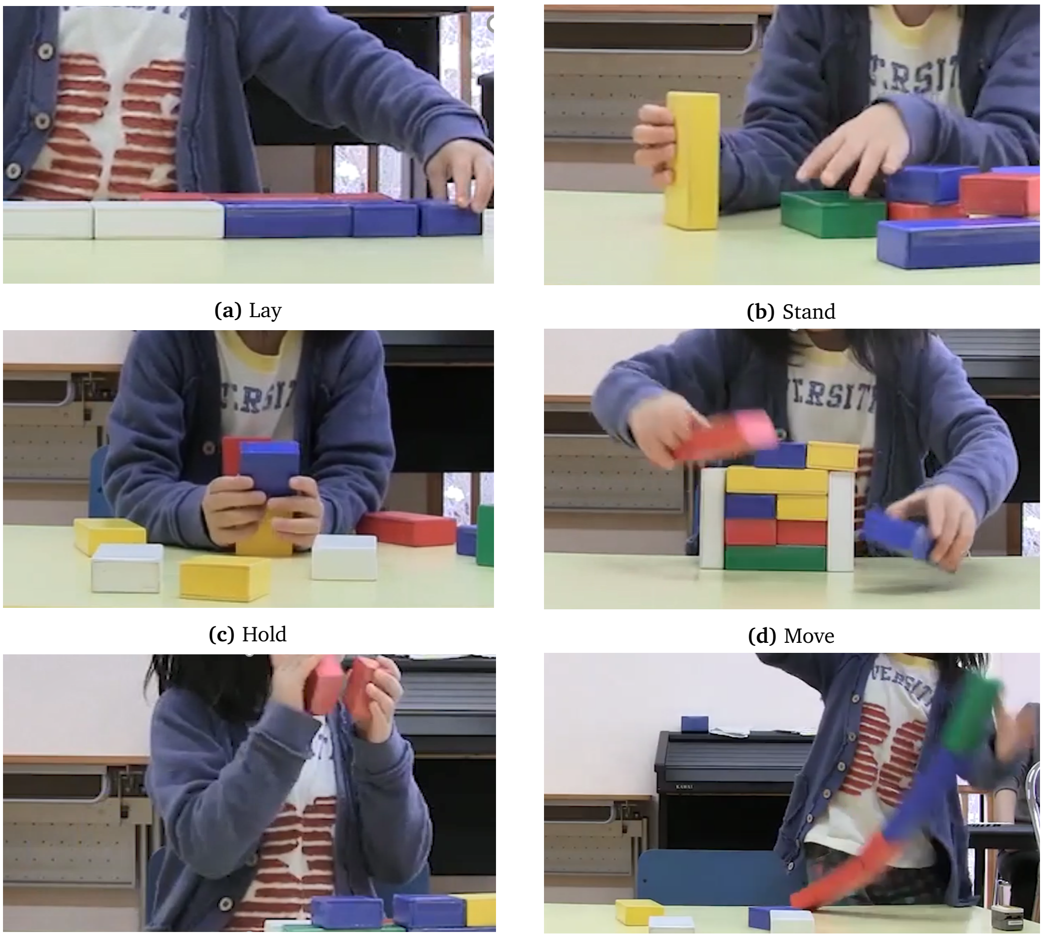
Each child was invited to play with AssessBlocks for approximately 20 minutes, a time frame based on the children's regular playtime. This length of time also reflects a period during which most children can concentrate. In the study, the child could stop early or continue slightly longer if they wished. The child's regular teacher was sitting nearby. The free-play session started when she encouraged the child to play with the blocks. A student research assistant remotely started the AssessBlock program to receive the IMU data. The teacher provided no instructions, tasks, or help. Minimum interactions happened when the child was actively searching for social-emotional support such as attention or when the child was idle for a long time. A child development psychologist and a psychology student were in the room for on-site support and observation. Two HD cameras in different directions captured audiovisual records of the children's play. After the child ended the play session, the AssessBlock program was wirelessly stopped.

3.3.3 Quantitative Data Classification and Extraction

Quantitative Action Definition

We defined actions to quantify the play sessions based on previous literature and a pilot study. A rich body of literature assessing children's emotional and cognitive development has focused on observing, interpreting structure, and identifying atypical play behavior. Knocking down and shaking toys are the most common atypical emotional responses [156, 7, 110]. Movements, holds, pauses, and different ways to place a block also provide information such as motor skills, concentration levels and challenge levels [109, 108]. Although quantifying the structure remained difficult, we broke down the play sessions into a sequence of actions to categorize the children's behavior.

[1] CBCL's use, scoring, and pricing information are accessible at: http://www.aseba.org/.



(a) Lay

(b) Stand

(c) Hold

(d) Move

(e) Shake

(f) Fall

Figure 3.2: Block play actions characterized in pilot study. Note that *lay* and *stand* are both derived from *static*.

With the knowledge and experience of two psychologists who specialize in child development and play therapy, we conducted a pilot study that observed the free-block-play of 30 healthy preschool children. From the pilot study we derived the following nine play features, including two characteristics and seven fundamental actions:

- **Time:** total amount of time between the start and stop of the play session.
- **Movement:** a sum of the magnitude of all three-axis acceleration values within a play session for capturing personal differences of moving speed.

  The following refer to actions in the static state:
- **Static:** the state after a block is placed on the table. It can be further classified into *lay* and *stand*.
- **Lay:** performed if the largest face of a block contacts the ground when being placed (Fig. 3.2a).
- **Stand:** the state when any other face contacts the ground (Fig. 3.2b).

  The following are the actions in the dynamic state:
- **Hold:** when the block is being held without substantial displacement (Fig. 3.2c).
- **Shake:** moving or swaying a block with quick and irregular vibratory movements (Fig. 3.2e).
- **Move:** performed when the amount of movement is in between *hold* and *shake* (Fig. 3.2d).
- **Fall:** movement caused by gravity when a structure collapses or is knocked down (Fig. 3.2f).

Labeling and Preprocessing

We classified five actions, *static*, *hold*, *move*, *shake*, and *fall*, from raw IMU data. *Lay* and *stand* were distinguished accurately from *static* by checking the axis to which the acceleration's gravity portion is pointing. A previous approach built a rule-based, three-class classifier with adult data collected in the lab to classify the fundamental actions of *static*, *hold*, and *move* [110]. However, models built with adult data that classify complex actions may not generalize well to children. Westeyn et al. built binary classifiers to categorize each of 34 actions for playing with toys and found the sensitivity (true positive rate) drops from 78.6 to 55.7% when switching the test dataset from adult's to child's. *Shake* and *fall*, which achieved a high sensitivity with adult data, performed poorly among children (50-75% sensitivity) [61].

To improve the generalization among children, we built an action classifier from the children's data collected during the experiment. Three graduate students acted as coders to exhaustively label portions of the data using ELAN software [157]. The data for the labeling were selected from nine participants (female = 5, five had at least one clinical behavior problem). These nine participants (10% of the original 88) were chosen based on observations to ensure they represented almost all play styles, and both normal and clinical children. We found data collected in the field were highly unbalanced in a large portion of *static* and *move*. Within each participant, we selected on average 4-minute play segments in which more *hold*, *fall*, and *shake* actions were performed to balance the corpus. In total, 38.5 minutes were selected for coding. The coders were trained by a professional (the third coder) for 1.5 hours to recognize each action and to familiarize themselves with the software. They reported that it took roughly 1 hour to code 2.5-minute of data, and found almost no distinct new play actions. The labels provided by the first two coders had a Cohen's kappa of $k = 0.774$, which indicated a substantial agreement among them [158, 159]. The professional (third coder) checked their coded data thoroughly and found that the disagreements mostly were at the start and end of some actions. She compared the labeled data first two coders agreed-upon with the videos, and fine-tuned the start and end of each action, to obtain a set of labeled actions.

We preprocessed the raw IMU data following the data processing pipeline proposed by previous work [73, 74, 61]. The feature space included a 3-axis accelerometer and 3-axis gyroscope values. We combined the magnitudes of each and produced eight features. A moving-average filter of three data points was then applied to each feature to remove any high-frequency noise. Next a half-second sliding window without overlapping was applied to each of the features. The mean, variation, and power spectral density were computed over each window.

Classification

Among the labeled data of the nine participants, six were used for training and three for testing. This participant-based testing was structured to validate the performance of unseen participants. By comparing a range of feature selections and classification models, we found that applying Logistic Regression with balanced class weights on the windows of the means of eight features best predicted the labels. The accuracy was maximized at 85.5% in the testing data, and 50.0 to 88.2% for each classes (baseline 25%). The classification result on the test data can be found in Fig. 3.3.

The classifier linearly captured general rules from the data. Classifying the unstructured children data in the field was harder than the adult structured data [61]. Most misclassifications were in *shake*, which may due to the limited sample size in the corpus. However, further fine-tuning towards *shake* might degrade the performance of the other more common actions (especially *fall*) in this five-class model. Since this is the first step and the importance of each action is unknown, we believe our model overall is acceptable.

Extract Timeline and Quantitative Features

We next applied the classifier to all 78 participants. Each block's entire play session can be presented by a time-series comprised of 0.5-second long actions. Next we computed an *all* timeline by aggregating every block's timeline to represent an overview of the session regardless of the scale of

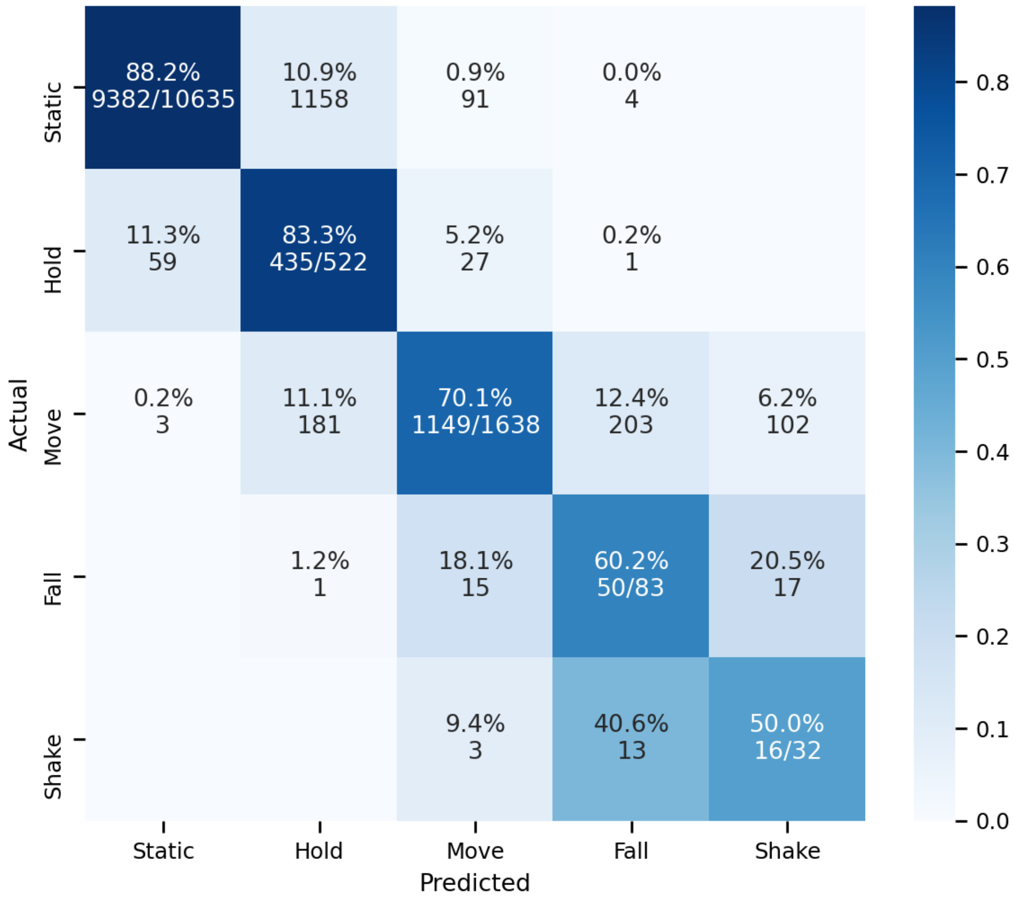


Figure 3.3: Confusion matrix for classification of five actions on test data.

the blocks. At each moment, the most representative action was chosen from all 12 blocks using an order of importance from drastic to static: *fall*, *shake*, *move*, *hold*, *stand* and *lay*. The visualization of the 12 timelines of *each* block and one *all* timeline of a 6-minute play session are found in Fig. 3.4. We observed that in *each* timelines, small blocks were relatively inactive with long *stand* and *lay* periods. However, the combined *all* timeline was active throughout the session with a few short pauses. Although both types of timeline capture the play behavior during a session, the *each* and *all* timelines can exhibit quite different characteristics.

Next, 23 quantitative features were computed from each play session. *Play time* and *movement* were accumulated from the raw data. We calculated

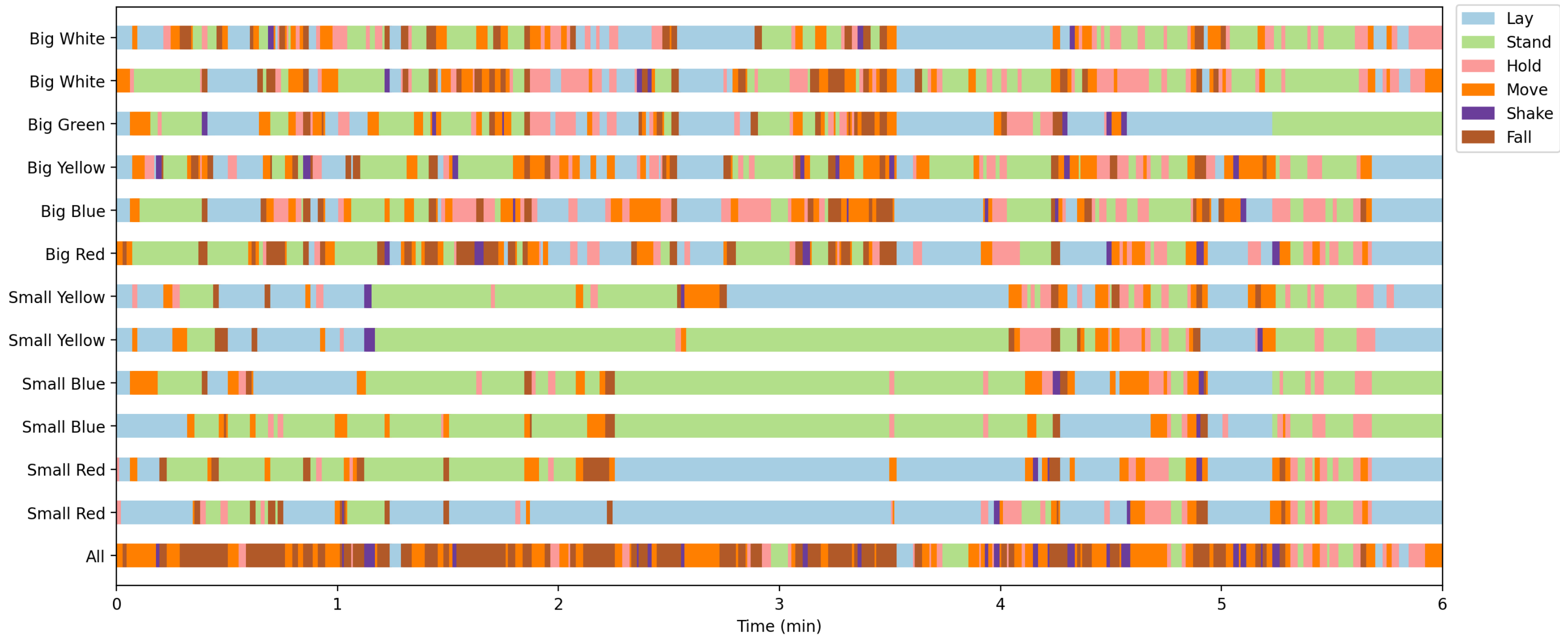


Figure 3.4: Example of timelines of a six-minute play: *each* blocks, and *all* which summarizes the main action at each moment.

from the timelines the quantitative representations of seven actions in two forms, *time* and *count*. Unlike the *time* form, which sums up the occurrence of one action, the *count* increments only when an action performed is different from the previous one. To investigate which timeline manifests more critical information, we computed the *time each* and *count* metrics by processing and totaling *each* timeline of 12 blocks, and the *time all* metric by processing the *all* timeline.

3.4 Result

3.4.1 Data Profile

CBCL

Table 3.1: Descriptive characteristics of three broad range behavior problems and four selected individual syndromes among our participants (N = 78).

Behavior Problem	Normal			Borderline			Clinical		
	n	%	95% CI	n	%	95% CI	n	%	95% CI
Total Problems	63	80.8	(72.0-89.5)	4	5.1	(0.2-10.0)	11	14.1	(6.4-21.8)
Internalizing Problems	67	85.9	(78.2-93.6)	0	0		11	14.1	(6.4-21.8)
Externalizing Problems	65	83.3	(75.1-91.6)	3	3.8	(0.0-8.1)	10	12.8	(5.4-20.2)
Anxiety/Depression	74	94.9	(90.0-99.8)	1	1.3	(0.0-3.8)	3	3.8	(0.0-8.1)
Social Problems	72	92.3	(86.4-98.2)	3	3.8	(0.0-8.1)	3	3.8	(0.0-8.1)
Attention Problems	58	74.4	(64.7-84.0)	5	6.4	(1.0-11.8)	15	19.2	(10.5-28.0)
Aggressive Behavior	72	92.3	(86.4-98.2)	2	2.6	(0.0-6.1)	4	5.1	(0.2-10.0)

The prevalence of clinical and borderline children among the participants is found in Table 3.1. The percentage of children with clinical problems in our sample was lower than in a previous study of children's behavior problems after the 2011 Earthquake in Japan (25.9%, 27.7%, and 21.2% for Total, Internalizing, and Externalizing Problems) [102]. Nevertheless, it exceeded the 2008 survey of mental problems among Japanese nursery school children (4.6%) and the prevalence of preschoolers in other parts of the world [160], indicating that children who are growing up in a post-disaster area are experiencing a higher risk of behavior problems.

Among eight individual syndrome scales, we included Anxiety/Depression, Attention Problems, Social Problems, and Aggressive Behavior in our study because they (1) contributed more to the broad scales and (2) contained more clinical and borderline children. We found that among children with borderline and clinical Total Problem cases, an average of 30.6% (SD = 14.2%) of the scores was comprised of the Attention Problems. The other leading contributing syndrome scales were Aggressive Behavior (27.8%, SD = 13.8%), Social Problems (12.7%, SD = 5.4%), and Anxiety/Depression (9.2%, SD = 8.4%). Aggressive Behavior (90.8%, SD = 6.6%) contributed the most to the Externalizing Problems, and Anxiety/Depression (68.3%, SD = 17.4%) contributed the most to the Internalizing Problems.

In this preliminary investigation, we omitted the borderline children in the following analysis since (1) the group size was small, with 1 or 0 cases in some measurements; and (2) it enabled us to draw a clearer line between normal and those with a high risk of behavior problems.

Play Action Features

A descriptive profile of the play features is shown in Table 3.2. Two play session features and seven action features in three metrics comprised a total of 23 quantitative features. Since the complete length of a session differed among the children, we normalized the features by dividing each feature (except the time) by time (in minutes) to obtain feature values per minute. The average, standard deviation, and range values of the features across the participants are presented in Table 3. To investigate which action metric better reflects behavior problems, we included all three (*time each*, *time all* and *count*) in the following analysis.

3.4.2 Relationships Between Behavior Problems and Each of the Play Features

To investigate whether each play action reflects on children's behavior problems, we first looked into the differences of play features between normal and clinical children. A Mann-Whitney U test was conducted on each play feature factored by each behavior problem.

Table 3.2: Descriptive profile of quantitative play behavior features. Time feature is documented in min, and other features are documented in $/min$ (N = 78).

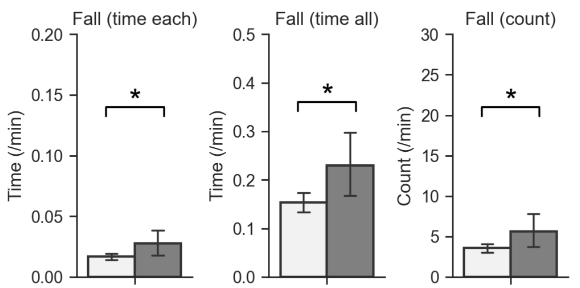
Feature	Average	SD	Range	Feature	Average	SD	Range	Feature	Average	SD	Range
Static (time each)	0.732	0.088	0.496-0.920	Static (time all)	0.094	0.093	0.007-0.515	Static (count)	23.072	6.440	6.796-38.872
Stand (time each)	0.236	0.147	0.0-0.540	Stand (time all)	0.034	0.041	0.0-0.244	Stand (count)	8.844	5.331	0.0-21.730
Lay (time each)	0.496	0.166	0.127-0.910	Lay (time all)	0.060	0.081	0.001-0.515	Lay (count)	14.228	6.494	4.415-33.064
Hold (time each)	0.143	0.049	0.034-0.265	Hold (time all)	0.170	0.089	0.042-0.628	Hold (count)	22.332	6.279	6.462-40.107
Move (time each)	0.100	0.035	0.013-0.219	Move (time all)	0.524	0.103	0.155-0.705	Move (count)	14.155	5.500	1.652-31.646
Shake (time each)	0.0060	0.0060	0.0-0.025	Shake (time all)	0.0500	0.039	0.0-0.186	Shake (count)	1.590	1.350	0.0-5.680
Fall (time each)	0.018	0.013	0.0-0.070	Fall (time all)	0.162	0.090	0.001-0.418	Fall (count)	3.816	2.448	0.053-13.583
Time	18.091	4.598	4.400-25.158	Movement	22.280	8.904	4.689-48.053				

For children with and without clinical Total Problems, we found significant differences in terms of *fall (time each)* ($U = 483$, $z = 2.074$, $p < .05$), *fall (time all)* ($U = 495$, $z = 2.256$, $p < .05$), and *fall (count)* ($U = 481$, $z = 2.044$, $p < .05$) (Fig. 3.5a). For Internalizing Problems, significant differences were found in *hold (time each)* ($U = 212.5$, $z = -2.240$, $p < .05$), *hold (count)* ($U = 216.5$, $z = -2.182$, $p < .05$), and *lay (count)* ($U = 229.5$, $z = -1.996$, $p < .05$) (Fig. 3.5b). For Anxiety/Depression, significant differences were found in *time* ($U = 33.0, z = -2.053, p < .05$) (figure 3.5c). For Aggressive Behavior, our results found significant differences in *fall (time each)* ($U = 237.0$, $z = 2.163$, $p < .05$), *fall (time all)* ($U = 236.0$, $z = 2.140$, $p < .05$), *fall (count)* ($U = 237.0$, $z = 2.163$, $p < .05$), as well as *shake (time each)* ($U = 239.0$, $z = 2.210$, $p < .05$), *shake (time all)* ($U = 230.0$, $z = 2.001$, $p < .05$), *shake (count)* ($U = 239.0$, $z = 2.210$, $p < .05$), and *time)* ($U = 51.0$, $z = -2.163$, $p < .05$) (Fig. 3.5d). No significant difference was found in any play features between normal and clinical children in Externalizing, Social, and Attention Problems.

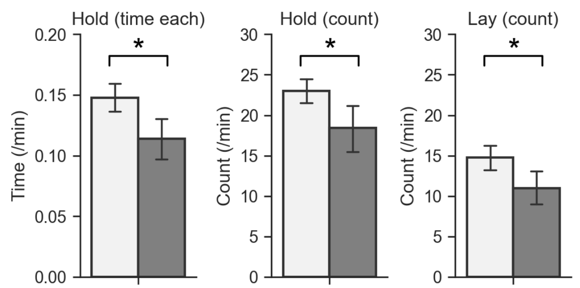
The result showed that among all the play features, *fall*, *shake*, *hold*, *lay*, and *time* are more representative phenotypes of different types of behavior problems. Children with Total Problems tend to perform more falls, and children with Aggressive Behavior tend to have more falls and shakes, indicating a more drastic style of playing. Children with Internalizing Problems tend to have shorter time holding the blocks and fewer *hold* and *lay* counts. Children with Anxiety/Depression and Aggressive Behavior tend to play for a shorter time, suggesting difficulties in concentrating or enjoying block-play. These results demonstrated that children with and without Total Problems, Internalizing Problems, Anxiety/Depression and Aggressive Behavior play with blocks differently.

3.4.3 Exploratory Prediction

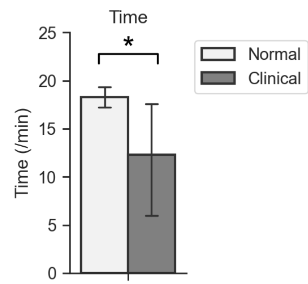
With a fairly small dataset, we explored simple models to investigate the predictive power of block-play. First, we used quantitative features and play patterns extracted from the timeline to predict the behavior problems. We then examined the features that were selected as the best predictors.



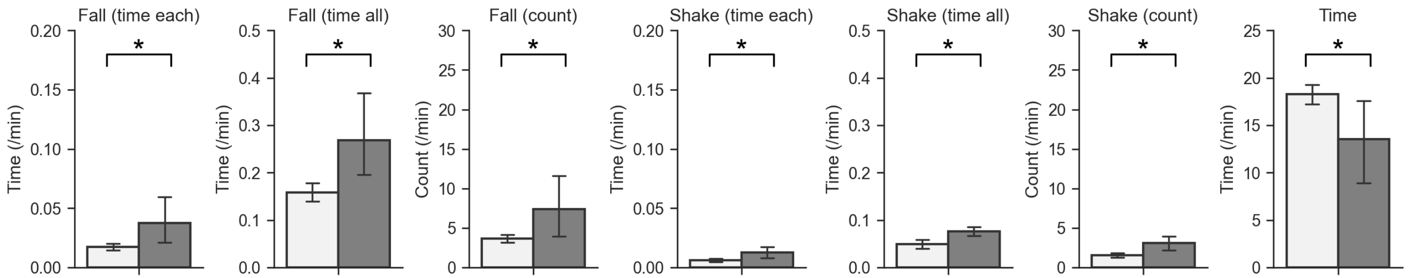
(a) Total Problems



(b) Internalizing Problems



(c) Anxiety/
Depression



(d) Aggressive Behavior

Figure 3.5: Quantitative feature values that showed differences between normal and children with clinical behavior problems. Within each plot, the feature name is marked at the top. Bars represent two groups: normal and with the denoted problem.

Next, the characteristics of the best predictors of behavior problems were summarized and confirmed with observations.

Feature Engineering and Model Selection

In the previous session, several quantitative features exhibited differences between children with and without clinical behavior problems. Previous literature also observed that some sequential play patterns, which are difficult to capture by time and count, might be relevant to the inner states of children, such as playing on a flat surface after the structure has collapsed [110]. Motivated by these findings, we explore the possibilities of extracting useful sequential action patterns from the entire play sequence.

Following the N-gram representation commonly used in sequence analysis in linguistics and biology [161], we produced pattern features by generating N-gram actions after downsampling the play sequences. The timeline of all 12 blocks were used since we found they outperformed the aggregated *all* timeline in the prediction. *Lay* and *stand* were uniformed to *static* to simplify the sequence into the composition of five actions: *static*, *hold*, *move*, *shake*, and *fall*. Downsampling creates non-overlapping windows of the sequence, and then selects the most frequent action within the window. Originally, each action in the timeline was 0.5-second long. As an example, the 1-gram resembles actions in the *time each* metric. The 2-grams creates many 1-second sequences of adjacent actions, which appeared to be too fine-grained. Thus, downsampling was conducted to find the length of action that best generated predictive pattern features. As the downsample rate increased, each action spanned a longer time and became coarser.

The N-gram representation also permutes the actions and drastically increases the feature dimension, as 5-gram can reach 3125 ($= 5^5$) features. To select the most important features, we employed L1-regularization (or LASSO), which is widely used in the tasks with high-dimensional features that require feature selection and the interpretability [162]. When the feature space contains a group of correlated ones, LASSO retains only one feature and sets the others in the group to zero. Although this retains the model's simplicity, the coefficients can be interpreted as associations.

We trained a number of 3-fold cross-validation L1-regularized models (scikit-learn implementation with Logistic Regression, penalty = l1, solver = liblinear) by sweeping 120 downsampling rates from 1 action per sec. to 1 action per 2 min., incremented 1 sec. each time. Each round, we went through a pipeline: (1) generating a downsampled sequence; (2) extracting 1-gram to 5-gram features from it; and (3) building a LASSO model and comparing the performance.

Prediction Performance

We investigated the predictions using the fundamental quantitative features as a baseline, and added the N-gram patterns to explore whether play patterns improved the performance of the prediction. In Total Problems, Internalizing Problems and Aggressive Behavior, we were able to build models with a sensitivity (true positive rate) higher than 0.5. The models using features alone and features plus patterns are presented in Table 3.3. The predictions with the best accuracy are highlighted.

We found the initial set of quantitative features exhibited difficulties predicting the behavior problems. With this highly unbalanced dataset, sensitivity was relatively low since the highest was 0.36 in the Internalizing Problems. Adding pattern representations of the play sequences increased the sensitivity and precision (positive predictive value) and maintained or slightly increased the specificity (true negative rate). In this imbalanced dataset with a small amount of true positives, the current sensitivity indicates that the models can identify 50 to 64% of the clinical children with three behavior problems. The precision shows that among all the predicted positives, 22 to 55% are true. The specificity shows that our predictions hold a relatively satisfactory true negative rate of 82 to 93%. Most normal children can be correctly identified.

Feature Coefficients and Interpretations

The non-zero coefficients from the models that best predict Total Problems, Internalizing Problems, and Aggressive Behavior are presented in Fig. 3.6. In each model, we interpreted the tendencies of the dominant features, and

Table 3.3: Performance for prediction of Total Problems, Internalizing Problems, and Aggressive Behavior. AUC represents micro-averaged and macro-averaged AUC. Se, Sp, and Pr denote sensitivity, specificity, and precision.

Prediction	Performance Metrics						
	Features	Accuracy	AUC	Se	Sp	Pr	F1 Score
Total	23 Features	0.70	0.45	0.10	0.81	0.08	0.08
Problems	**23 Features + Patterns**	**0.82**	**0.75**	**0.64**	**0.86**	**0.44**	**0.52**
Internalizing	23 Features	0.81	0.62	0.36	0.88	0.33	0.35
Problems	**23 Features + Patterns**	**0.87**	**0.74**	**0.55**	**0.93**	**0.55**	**0.55**
Aggressive	23 Features	0.89	0.59	0.25	0.93	0.17	0.20
Behavior	**23 Features + Patterns**	**0.90**	**0.71**	**0.50**	**0.92**	**0.25**	**0.33**

grouped them into distinct play styles. A mapped-out relationship of the target behavior problems, the main features, and the styles can be found in Table 3.4.

Total Problems: The play pattern features that best predict Total Problems have a rate of seven seconds per action. We first found the positive predictors involved with *fall* and *move*. On the contrary, negative predictors are mostly *static* and *hold*. This result indicates a more active, or even "drastic" style among clinical children, and a gentle style otherwise. Two features involving the *hold move* pattern were positive predictors of the Total Problems. The same pattern was not found in the negative predictors. This *hold move* style can be characterized into an "indecisive" play style, which holds the block for a while before deciding where to move it. Meanwhile, some features found to be hard to interpret. For example *static* related features appeared to be both positive and negative predictors.

We next ran a quick observational analysis to look for the occurrence of "drastic" (Fig. 3.7b) and "indecisive" (Fig. 3.7c) style among children with high and low Total Problems T-scores. Among 11 children with clinical Total Problems, seven were "indecisive," six played "drastically", and two exhibited both. Many (P19, P44, P50, P76) seemed to grasp the block tightly during the "hold" phase. We examined the children with the lowest Total Problem t-scores and found that 3 of 10 were "indecisive" and none were "drastic." No child seemed to grab the blocks hard.

Internalizing Problems: In this model, the play pattern features have a rate of 20 seconds per action, which is considerable long. We found among the pattern features, positive predictors all contained a long, 80-second *static*. It can be characterized into an "inactive" play style with long pauses. Other feature coefficients were inconclusive because similar features appeared as both positive and negative predictors. Although *time* was significantly shorter among children with an Internalizing Problem evaluated by a Mann-Whitney U test, longer time was a positive predictor in the model.

We examined the "inactive" (Fig. 3.7d) play styles from the recorded video. Among children with clinical Internalizing Problems (N = 11), seven were inactive with long pauses. Among children with the lowest Internal Problem t-scores (N = 10), six also showed long pauses. However, three of the six were excitedly explaining their structure during the pause (P31, P32, P56) after they finished building it.

Aggressive Behavior: The dominating positive predictors for Aggressive Behavior were *fall (time each)* and *shake (time each)*, which indicated a "drastic" style. The rest of the positive features, *static (time each)*, *stand (time each)*, and *hold (time each)*, slightly indicated an "inactive" style. The negative predictors were quite diverse, included *move*, *hold*, *stand* a block and *static*. A longer playtime was also associated with normal children, which we confirmed with a significant difference.

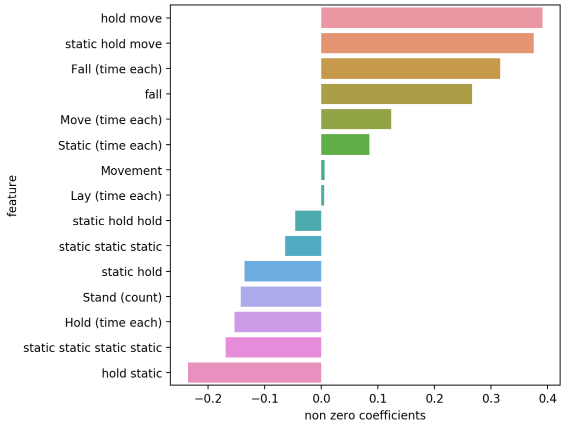
By analyzing the video, we observed that all the children with clinical Aggressive Behavior (N = 4) exhibited the "drastic" play style and were impatient, violent, and noisy with many intentional falls. One had many observable *shake* actions (P62), two boys flicked the blocks, built high towers and then repeatedly knocked them down (P62, P65). Among them, two were also "inactive." Among the children with the lowest T-scores (N = 10), two were "inactive," but none exhibited the "drastic" style. Another observed difference was that when the children with Aggressive Behavior disassembled their structures, they knocked them down (P46, P49, P62, P65). On the contrary, children with low Aggressive Behavior scores would gently take their block structures apart block by block to avoid a collapse (P34, P56).

We also observed that two "inactive" children out of four with Aggressive Behavior (P62, P65) were highly distracted by their environment when they saw or heard others pass by. One had clinical Attention Problems, and the other had borderline Attention Problems. Such "distracted" behavior wasn't found among children with low Aggressive Behavior t-scores. However, capturing such "distracted" behavior was complicated by the blocks since the children might or might not be holding a block when they were "distracted." Since the experiment did not include a designed distraction, the relationship between being distracted and Aggressive Behavior or Attention Problems cannot be verified yet.

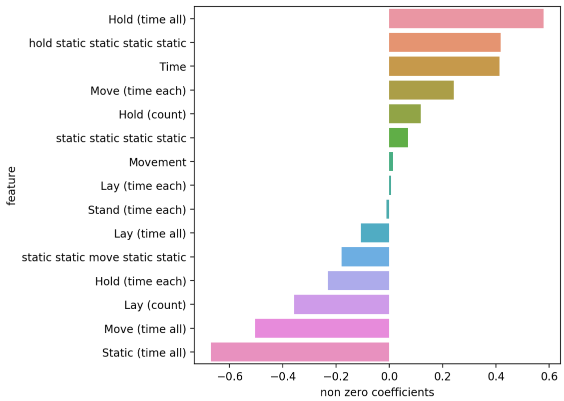
Overall, the feature interpretations and validations demonstrated that the predictions provided insights, which confirmed a majority of the observations and further induced observational hypotheses and discussions.

Table 3.4: Discovered mappings of target behavior problems, predictor features, and characterized play styles.

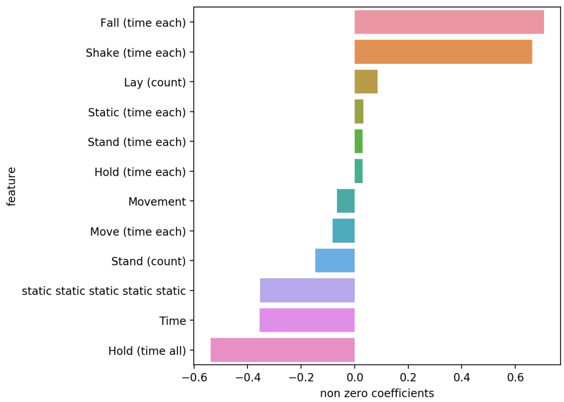
Target	Positive Predictors	Style
Total Problems	fall (pattern) fall (time each) move (time each)	drastic
	hold move (pattern)	indecisive
Internalizing Problems	static (pattern)	inactive
Aggressive Behavior	fall (time each), shake (time each), less time	drastic
	static (time each), stand (time each), hold (time each)	inactive



(a) Total Problems
(action length = 7 sec.)



(b) Internalizing Problems
(action length = 20 sec.)

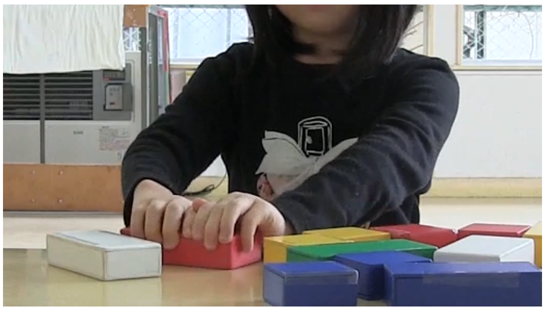


(c) Aggressive Behavior
(action length = 3 sec.)

Figure 3.6: Non-zero coefficient estimates for Total Problems, Internalizing Problems, and Aggressive Behavior. Positive coefficients are positively correlated with clinical problem, and negative coefficients are positively correlated with no problem.

(a) active construction, which is most common)

(b) Drastic play, which is related to Total Problems and Aggressive Behavior



(c) Indecisive play, which suggests Total Problems

(d) Inactive play, which suggests Internalizing

Figure 3.7: Toy-block-play styles

3.5 Discussion

3.5.1 Potentials

Addressing our crucial question: can playing with toy blocks reflect behavior problems?

Our multi-stage quantitative approach demonstrated that the individual free-block-play captured in the field reflects some behavior problems identified by CBCL. Significant differences were found in quantitative play features factored by Total Problems, Internalizing Problems, and specific syndromes Anxiety/Depression and Aggressive Behavior, indicating that children with and without these behavioral problems play differently. Although the performance isn't optimal, our exploratory prediction models with features and patterns showed the promises to estimate Total Problems, Internalizing Problems, and Aggressive Behavior.

By interpreting the features in the prediction models, we summarized three styles that indicate behavior problems: "drastic," "indecisive," and "inactive." We validated them as prevalent among more than half of children with three behavior problems. The same styles were not typically found in children without such a problem. Children with Total Problems and Aggressive Behavior tended to exhibit "drastic" styles, involving active knocking, flicking, and other destructive behaviors. Those with Total Problems also tended to be "indecisive", holding a block with a strong force before moving it. Children with Aggressive Behavior demonstrated an "inactive" tendency with long pauses. One exception is that "inactive" style prevailed in both children with and without Internalizing Problems. However, our observation suggests that normal children might "pause" to engage - communicate with others and share verbal opinions about their structures. Children with clinical behavior problems might "pause" due to disengagement and distractions. Furthermore, those with Aggressive Behavior and Attention Problems might be easily distracted.

The insights related block-play to behavior problems demonstrated the potential of our methods. Although our ultimate goal is to replace observations, the system's current role is to provide quantitative measurements and predictions to assist the observations of psychologists and caregivers and to direct what play actions and styles to observe and to focus more care on. Although our system cannot currently be used in a messy environment, it can be available in a setting with one child who is willing to play, one caregiver, no instructions, and minimal disruptions, all of which can be easily reconstructed in our daily life. Our system can also guide future works that deepen the connections between behavior problems and block-play with further quantitative and qualitative investigations.

Predictive Power

The behavior prediction with toy block play data was novel and challenging, especially with a highly imbalanced dataset whose positive rates were around 14.9, 14.1, and 5.2%, respectively, capturing the imbalanced nature of the behavior problems. Our exploratory predictions using quantitative features and N-gram patterns demonstrated the possibility of predicting Total Problems, Internalizing Problems, and Aggressive Behavior with 0.56-0.64 sensitivity, 0.86-0.93 specificity, and 0.25-0.55 precision. Even though the performance failed to reach the level of diagnosis, the current prediction is meaningful because (1) our F1 scores and AUCs are comparable to the state-of-the-art works that predicted adult mental health and affects [31, 32]; (2) the results were justified by professional observations, such as *shake* and *fall* are similar and indicate a "drastic" play style; and (3) the prediction does not largely cause unnecessary concerns since it predicts few false positives with relatively high specificity. Thus, we reported the models, and invested the predictor coefficients to provide insight. Although the current prediction utilized a simple linear model, the performance rose when sequential patterns were added to the quantitative features. It demonstrated the potential of building sequential models to predict behavior problems from block features. The different downsampling rates for three predictions also indicate that downsampling is necessary for performance. Our current prediction performance can be used as a benchmark for future explorations.

Actions, Timeline and Metrics

We classified five actions from raw IMU data: *static*, *hold*, *move*, *shake*, and *fall*. Since the classifier was built from the children's data gathered in the field, the accuracy of the children was more reliable than the play action classifiers built on adult data [61, 69].

Two timelines were transformed from raw data. *Time each* and *count* metrics were summarized from each block's timeline and the *time all* metric was summarized from the *all* timeline. Our results indicated that *time each* and *count* are slightly more related to problem behaviors identified by CBCL, since significant differences in *hold* action's *time each* and *count* values can be found with and without Internalizing Problems but not *time all*. The L1-regularized prediction models also selected more coefficients in *time each* and *count*, and our test showed that the predictions based on N-gram patterns generated from 12 timelines outperformed those from the *all* timeline. Thus, at the current stage *each* block's timeline revealed more information related to behavior problems than the aggregated *all* timeline. However, we cannot conclude that the separate timelines are superior, since the current *all* timeline might also aggregate the errors of each block's timelines. The current result demonstrated the requirement in developing a more informative *all* timeline and evaluating its predictive power.

3.5.2 Limitations

Sensitivity and Interpretability

We noticed that playing was not sensitive to some behavior problems, such as Externalizing Problems, Anxiety/Depression, Attention Problems and Social Problems. Sensitivity to Externalizing Problems and Anxiety/Depression might contain room for improvement since the correlated ones, Aggressive Behavior and Internalizing Problems, were reflected in the block-play. Their predictions might be improved by (1) examining more clinical children and (2) exploring longer study durations, such as conducting experiments over time to test whether more significant details can be captured. Meanwhile, the insensitivity to Attention Problems and Social Problems indicated that the block approach might not be effective for them. In our experiment, a small number of children were distracted while playing. Since distraction was not part of our protocol, we were unable to infer a relationship between being "distracted" and the Attention Problems. For the Social Problems, which involves such problems as “cannot get along with others” [107], our current experiment design, which wasn’t constructed around social play, might not be able to capture any signs of them.

Our current prediction showed that some features selected by the L1-regularization were hard to interpret. Similar actions in different metrics were associated to behavior problems in opposite directions. This demonstrated that not all of our feature coefficients align with our observations or knowledge. Perhaps the limitations on the accuracy of the actions and the data size restricted their interpretability. These counter-intuitive findings might be eliminated with an improved overall performance.

Action Accuracy

We built simple models to learn the linear rules from data collected in the field. Current data processing remains unable to achieve high accuracy on each action, especially the separation between *shake* and *fall*. While realizing it harms the conclusiveness of the prediction models, it might not be extremely detrimental since the professional observations also found that *fall* and *shake* were similar, and these two actions demonstrated a converged trend towards the behavior problems. Although manually coding the entire dataset could provide a set of reliable action labels, it is beyond the scope of our current work due to time and labor constraints. Since we discovered that *fall* and *shake* were crucial actions, further investigations around software and hardware designs can be implemented to improve their accuracy. In the software part, classifiers that are specialized in *fall* and *shake*. For the hardware part, other sensors and modalities, for example, a capacitive touch sensor, can be used to distinguish *fall* from other hand gripping actions. In the future, manually coding more such low accuracy actions can also be explored to improve the classifier’s accuracy.

Small and Imbalanced Data

Since we collected data in the field without controlling and testing groups, they are unbalanced toward a large number of negatives, or normal children. It shows the imbalanced nature of behavior problems, even though they were reported to be prevalent in the area [21, 27, 81]. Finding a significant number of clinical child participants for each of the 11 measures from CBCL was costly. Since excluding healthy children from our relatively small dataset was also risky, we leveraged it as is and provided various metrics (Table 3) to elaborate our system's pros and cons. In the future, more clinical children or repeated measures from them are needed to balance the data.

The current small dataset also made it difficult to apply complex ML algorithms. Moreover, the study was comprised of participants in one area, thus the demographic and cultural similarity and differences couldn't be examined. Although the cultural differences of block play were not mentioned in the previous literature, further bigger data from diverse participants are needed to validate, solidify, and generalize the approach.

3.5.3 Future Work

The current work described the potential of predicting children's behavior problems with a simple and interpretable quantitative method that uses motion data captured during free-block-playing sessions. Based on this foundation, our future work is three-fold. The blocks design needs to integrate multi-modal sensing to capture a range of important data, such as gripping force, surface touch, and even facial and verbal expressions. The next step of the data collection needs to expand the scope and depth. We need to include more diverse participants, special groups of children with specific clinical syndromes, and repeated experiments to deeply scrutinize their play behaviors. In terms of analysis, we can investigate more complex but less interpretable models, such as sequential ones, or use an end-to-end approach that does not involve several stages of data processing.

Chapter 4

Toward Daily Mental State Prediction and Support with TUIs

4.1 Connecting Two Studies: Post-analysis and Collective Findings

In Chapter 2, correlation analysis is performed to establish the association between the play features and on-site short-term stress measurements. The play features comprise quantitative actions automatically extracted from IMU-embedded blocks and the play behavior manually video-coded from video documentations. The target stress measurements include the physiological stress bio-marker SAA and the behavioral stress evaluation OSBD. The results show that play done in a flat style with less standing actions is correlated to a high SAA After value with an increase in SAA, and having less time playing with blocks and moving the blocks indicates a higher OSBD after the play session. In general, play "passively" indicates stress.

Moving beyond on-site stress, Chapter 3 explores the predictions of a range of prolonged behavior problems that are measured by CBCL. The limitations found in Chapter 2 include sequential patterns and some actions, which were observed in our participants but not captured in our dataset, along with unsatisfactory accuracies in some actions. These limitations motivated me to further develop play feature extraction techniques.

As a result, although both used blocks, Chapter 2 and Chapter 3 made use of separate sensing schemes, collected data from different groups, and aimed for distinct targets. Therefore, direct comparisons between the two studies would be difficult. However, indirect comparisons can be made to examine collective findings, consistencies and inconsistencies. Connections and limitations can be found through the following post-analysis.

4.1.1 Correlations Between On-site Stress and Prolonged Behavior Problems

The data presented in Chapter 3 not only collected participants' behavior problems, which is measured by CBCL questionnaires, but also on-site measured SAA and OSBD, using the same protocol as that presented in Chapter 2. Our approach to find the relationship between target mental health measurements is to analyze Spearman's rank correlation coefficients and partial correlations between pairs of variables (CBCL symptoms, SAA and OSBD measurements) from the dataset presented in Chapter 3. The correlation analysis is for providing a general understanding of the relationship, while the partial correlations are used to measure the linear correlation between two variables with the effect of other variables removed. Some indirect correlations may arise between two variables in correlation analysis, but they are eliminated in partial correlation analysis so that only direct correlations remain.

The variables included in the correlation analysis are Anxiety, Aggressive Behavior, Attention Problems and Social Problems (four behavioral symptoms used in Chapter 3), Before and After values of sAA, and Max and After values of OSBD (sAA and OSBD are on-site stress measurements). These variables are all measured directly, whereas Total Problems, Internalizing Problems and Externalizing Problems in CBCL, the average and percentage change of SAA and OSBD are excluded due to being compound variables calculated from those directly measured variables. These excluded variables exhibit multicollinearity with the direct variables, which makes partial correlation analysis unstable. The significant Spearman's rank correlations and partial correlations are shown with a Correlation Network (CN) (Fig. 4.1) and a Partial Correlation Network (PCN) (Fig. 4.2) built using the qgraph package available in R [163].The edges in this network reflect correlations with a significance level of $p < 0.05$. The width of the edge represents the strength of association between two correlated variables and the number on the edge denotes the coefficient of correlation. As evident from the Mental Problem CN, among the four behavior problems, Aggressive Behavior, Attention Problems and Social Problems hold strong correlations between each pair of them. In PCN, however, Aggressive Behavior is no longer correlated with the other two. This indicates correlations between Aggressive Behavior and Attention Problems, Aggressive Behavior and Social Problems in CN are most likely caused by indirect associations. Similarly, PCN revealed that the correlation between Anxiety and Social Problem is indirect. Neither direct nor indirect correlation was found between Aggressive Behavior and Anxiety.

The sAA and OSBD, two on-site measurements, also show significant correlations. In CN, SAA Before shows moderate correlation to OSBD Max and weak correlations to OSBD After. From the PCN, however, we can see the correlation between sAA Before and OSBD After is eliminated if OSBD Max and other variables hold constant, which likely indicates indirect associations between sAA Before and OSBD After.

Between long-term behavior problems and on-site stress measures, Anxiety shows weak direct correlations to OSBD Max and indirect correlations to OSBD After. No indirect correlation between Aggressive Behavior and sAA Before was found, but, they revealed weak direct correlations when other variables held constant.

The above findings indicate that Anxiety, one component of Internalizing Problems, and Aggressive Behavior, one component of Externalizing Problems, are not correlated. Meanwhile, there are direct correlations between Attention Problems and Social Problems. Anxiety has a weak positive correlation to OSBD, especially OSBD Max value, while Aggressive Behavior might weakly associate with a high sAA Before value. The correlations between sAA and OSBD are also shown implicitly, similar to what we discovered in Chapter 2.

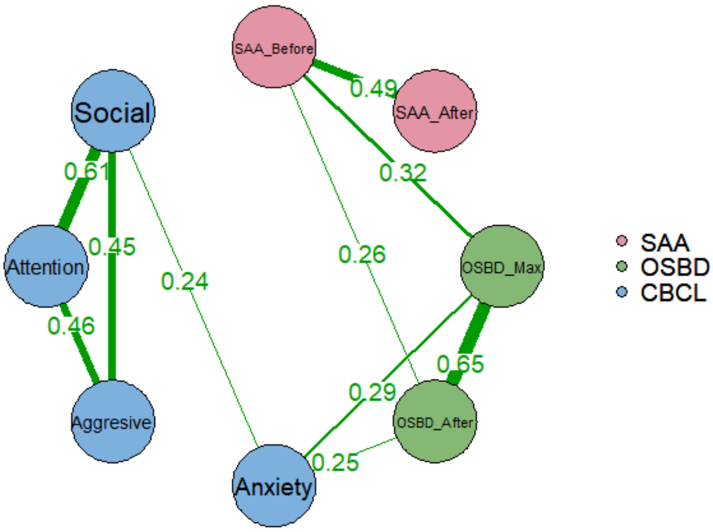


Figure 4.1: Mental Problem Correlation Network

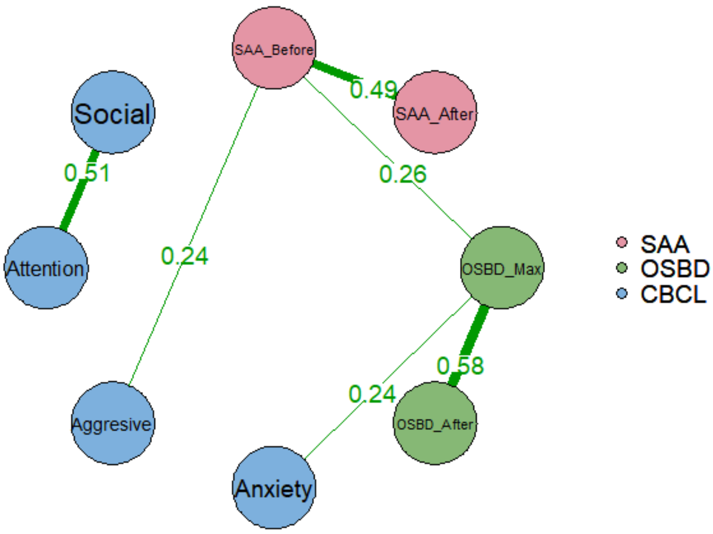


Figure 4.2: Mental Problem Partial Correlation Network

4.1.2 Correlation Between Mental Problems and Block Play Features

Spearman's rank correlation analysis and partial correlation analysis were conducted to analyze both direct and indirect correlations between the target health measurements and block features.

The CN and PCN between each of six target behaviors analyzed in Chapter 2 and 3 (sAA , OSBD, Total Problems, Internalizing Problems, Externalizing Problems, Anxiety and Aggressive Behavior) and the corresponding block features we found important can be seen in Appendix A.1.

In general, the comparisons between each pair of CN and PCN show that the direct correlations between targets and blocks are weak, and the partial correlations that hold other block parameters constant even weaken the association between them. This demonstrates that the block features are intrinsically correlated and hard to consider separately–a child that has more Fall actions generally has more Shake actions. On the other hand, it also reveals that the associations between features are stronger and more dominating than the associations between features and targets. This is not totally unexpected since all features are computed from IMU sensors and they all represent the same event-playing with the blocks. Meanwhile, the results of CN and PCN imply that to develop a robust system, we need more modalities that are not highly correlated, that is, ones reflecting different angles of play. It is also necessary to be cautious of unstable results and overfitting if there are highly correlated features. More samples are needed to validate and test how well the system fits all kinds of data.

4.2 Reflections

Chapters 2, 3 and Section 4.1.2 find a set of promises and limitations in capturing and processing block play into structured and quantitative data. These promises and limitations form the guideline for data processing and automated feature extraction toward capturing the entire scope of play behavior. Section 4.2.1 to Section 4.2.3 discusses the specific guidelines and how data preprocessing can address them.

4.2.1 Feature Extraction Based on Findings

The findings of previous sections proposed several important styles that are interpreted from, but not directly related to block features:

- Passive play: play flatly with few stacking actions.
- Indecisive play: cannot decide where to place the block. Holds a block before moving it.
- Inactive play: mainly hold actions or no actions.
- Drastic play: shakes the blocks. Builds tall structure intentionally hitting it or pushing it over. Knocks down the entire structure instead of taking the structure apart piece by piece.

The findings also present the need to address the limited sample size and to find more modalities. Thus, in order to capture the wide scope of play behavior from the diverse angles, both action features and structural features need to be further examined.

So far, the exploration mainly uses action features, including quantitative actions and sequential patterns to correlate and predict mental health. The structural features were merely examined through stacking time and flat time, the two play behaviors that are manually coded, which could not be objectively extracted and generalized when it comes to a new dataset.

To capture the multimodal data that reflect the above play styles, it is necessary to increase the sample size based on the limited nature of psychological observations, and automate the feature extraction process. Further design of motional feature and structural feature extraction are proposed in Section 4.2.2 and Section 4.2.3, respectively, to address the above needs.

4.2.2 Extract Motional Features

Currently, the feature extractions mainly focus on processing the raw data from accelerometers and gyroscopes into action features. The action features provide direct association with psychological and clinical knowledge, and they are easily interpreted. However, the process of classifying low-level raw data into high-level knowledge of actions lose some important information such as the magnitude. The classifier may compress a range of magnitude into a single feature and the nuance is therefore lost. For example, some children might hold a block firmly, while others might be involuntarily shaky due to anxiety and other inner fluctuations. Thus, the magnitude, shown in Figure 4.3, needs to be included in the analysis. The signal-processed magnitude, such as signal decomposition in frequency domains, may also be included as a features to capture actions' frequencies.

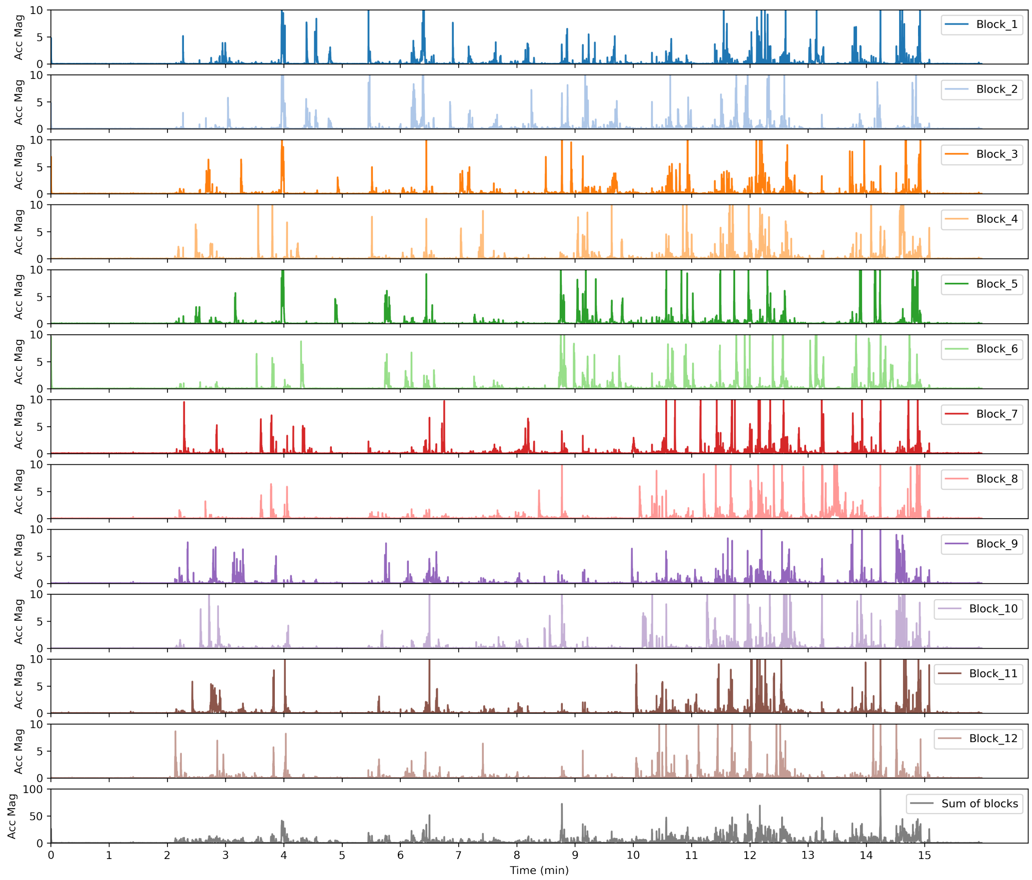


Figure 4.3: Raw data of acceleration magnitudes

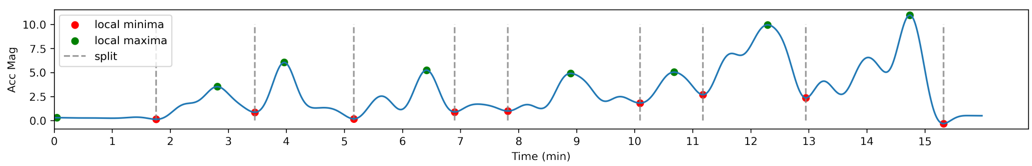


Figure 4.4: Filtered data of average acceleration magnitudes, and split points based on local minima

Other crucial information that can be extracted from the motion data includes the places to split the entire play session into smaller sub-sessions. In the studies presented in Chapter 2 and Chapter 3, one frequent observation is that a child plays for a short session of one to several minutes long, and then pauses before starting another short session. Our experiments are conducted with a rough time frame of 20 minutes for each child, but in reality, a child's play is composed of such small sessions. Thus, splitting the data at the places when the child finishes a small session might be the key to generating more data samples from the limited participant data.

The session can be split at the places where the magnitude of action is minimal. A splitting algorithm is proposed by computationally looking for the local minimum of the magnitude. First, the magnitudes of all blocks' accelerations are averaged. Next, forward and backward filtering with a Butterworth low-pass filter is performed on the average magnitude to filter out noise and smooth the motion. Then the local minima can be detected by calculating the relative extremes. Figure 4.4 shows an example of detected local maxima and minima with a distance longer than 45 seconds and the sessions can be split by the local minima (grey dashed lines.) In this figure, we can see that a 15-minuite play session (from P62, boy, 6 years old, clinical Externalizing Problems and Aggressive Behavior) is split into 10 sub-sessions, ranging from around 1 to 2 minutes. These split points are also confirmed in the video as places he stopped playing.

4.2.3 Extract Structural Features

Structural features have demonstrated importance in several styles we discovered, such as passive play and drastic play. Structures are also often observed as the explicit results of play [7, 164]. Meanwhile, it is difficult to extract from IMU solely without an external reference due to the problem of dead reckoning and drifting. Thus, external sensors are necessary to accurately detect the structure. In the current dataset, the video documentations captured by an RGB camera can be used to extract the structural information. In future design and data collection, capturing videos with RGB cameras can be replaced by using depth cameras, Lidar sensors, or

Lighthouse tracking systems (used by VR systems like HTC Vive,) to obtain accurate structure or position information.

Block Detection using Computer Vision

A computer vision-based multi-object tracking algorithm is developed to extract the blocks from each frame of the video. The details of the object tracking pipeline are as follows: (1) the frame applies a temporal filter to reduce noise. The temporal filter is a simplified version of the spatiotemporal filtering proposed by Richardt et al [165]. (2) The output is applied with five HSV range filters to extract the area within a color used in the blocks. (3) The contour of the areas feeds into a shape detector to filter out non-block-like shapes. (4) The centroid of each detected shape feeds into a Kalman filter for tracking. Figure 4.5 shows the results of the detection and tracking with metadata. The center and rectangular bounding box of each block can be distinguished in the frame. While the tracking algorithm loses the track occasionally due to occlusions and noises, it provides a stable result most of the time.

The results of block color detection and tracking can be found in Figure 4.6, and Figure 4.7 shows the same information on a black background. This simplified frame can be used to extract the structural features.

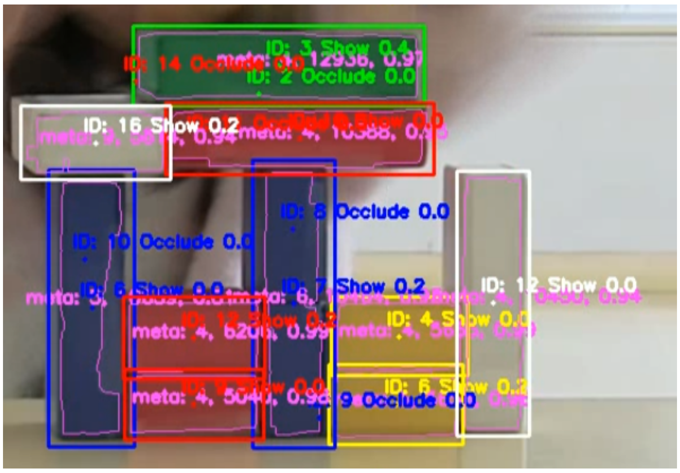


Figure 4.5: Results of block detection and tracking on the frame

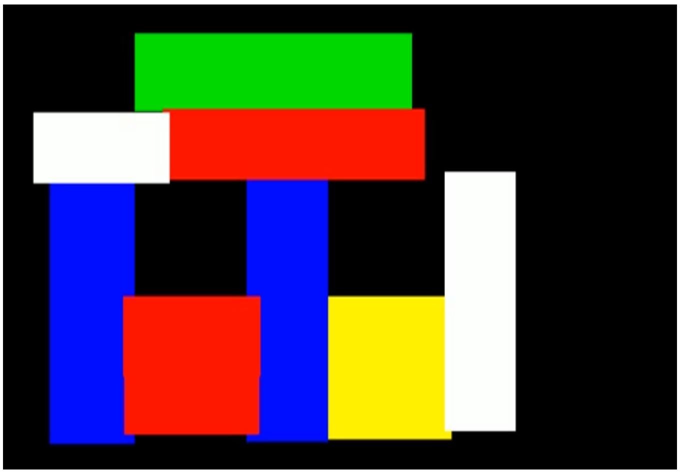


Figure 4.6: Visualization of detected blocks in color

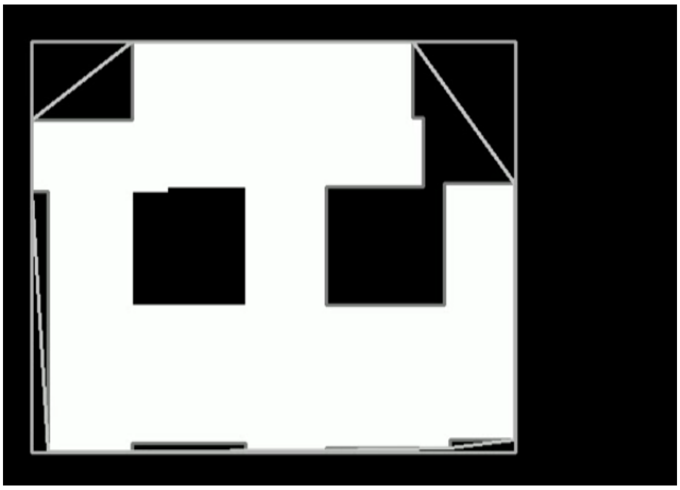
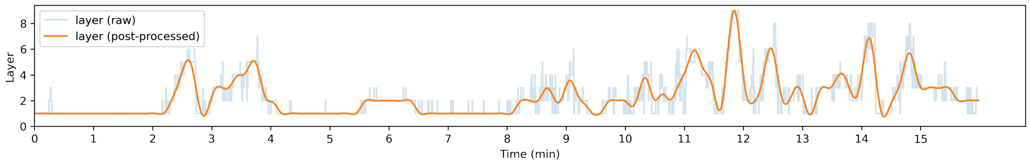


Figure 4.7: Visualization of detected blocks in grayscale, hull area, and bounding box

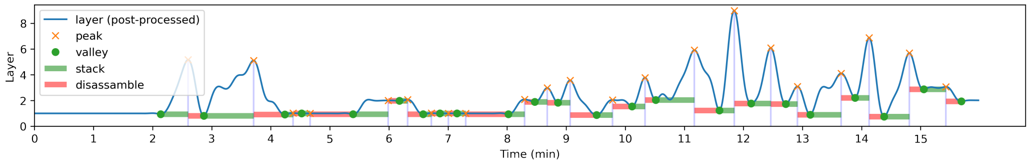
Feature Extraction

Taking into account the structure-related play styles and knowledge based on observation, three type of features can be extracted: layer, complexity and aspect ratio.

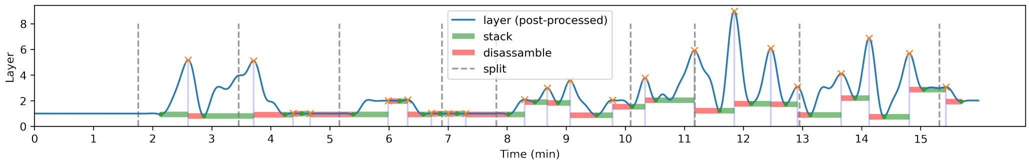
The layer feature reflects how tall the structure is by calculating how many blocks are stacked. The layer data calculated from one session (P62, boy, 6 years old, clinical Externalizing Problems and Aggressive Behavior) can be found in Figure 4.8a. The raw data's noise is reduced using forward



(a) Raw and noise-reduced layer sequence



(b) Detected peaks, valleys, stacking time and disassembly time



(c) Layer information with split points on the sequence

Figure 4.8: Layer of a play session

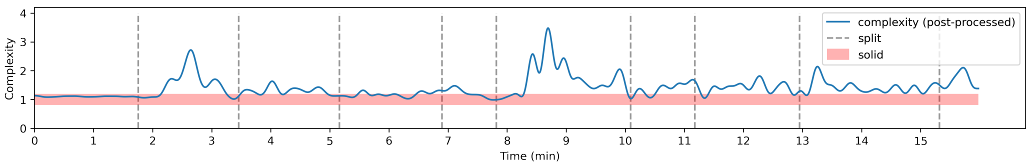


Figure 4.9: Complexity of a play session

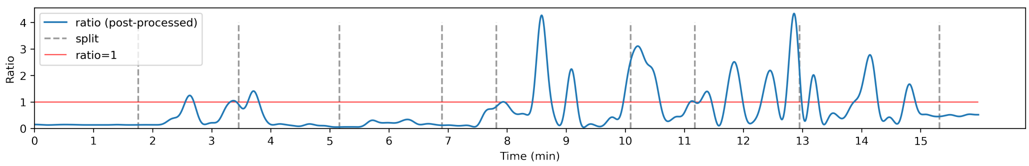


Figure 4.10: Aspect Ratio of a play session

and backward filtering with a Butterworth low-pass filter. From this timeline, the peak is detected using the Find Peaks function provided by scipy.signal library, detecting the peaks with a width larger than 8 seconds. Next, the valley is calculated by detecting the first minimum within two peaks. From the peaks and valleys, the times of the child's stacking and disassembly of the structures can be detected, as reflected in the timeline shown in Figure

4.8b. The split points extracted from the motional features can be applied to the sequence of layers, as shown in Figure 4.8c.

Complexity depicts how complex the structure is. Learned from knowledge and observations, a complex structure involves careful considerations and balancing. Thus, a complex structure usually has more space between the blocks. On the contrary, simple structures are those where one block is stacked exactly on top of another, or blocks line up to build ether a blob or a flat surface. Complexity can be calculated by dividing the Convex Hull area by the solid area. The Convex Hull of a shape is a tightly fitting convex boundary around the shape. An example of a Convex Hull area is shown with a polygon formed by bright gray lines in Figure 4.7, detected by finding the Convex Hull of the largest connected area in the frame using OpenCV. The solid area is the area filled with white color shown in Figure 4.7. The extracted complexity of one participant is shown in Figure 4.9. The complex value of 1 represents a solid structure. The more hollow space between the blocks, the more complex the structure becomes.

The aspect ratio reflects the general height-width ratio of the structure. This is calculated by dividing a detected structure's height by its width. The height and width are extracted from the bounding box of the biggest connected area. An example of a bounding box is shown with a rectangle formed by gray lines in Figure 4.7, which also encloses the polygon-shaped Convex Hull bounding box that is shown with a slightly lighter gray color. The aspect ratio of a play session can be found in Figure 4.10. An aspect ratio of 1 denotes a square structure, and an aspect ratio below 1 indicates a relatively flat structure while the ratio above 1 indicates a tall structure.

Play styles can be found through these three structural features. For example, drastic play can be seen from the latter half of the layer sequence where cycles of stacking high and disassembling in a short time are performed. In the earlier part, indecisive play might be reflected from the layer and complexity sequence, where the stacking takes a long time but the complexity is low. However, determination of the indecisive style needs to take into account the motional features as well. The aspect ratio sequence shows that around 4 to 7 minutes, the participant is performing the "flat play" style.

By simply observing the structural features, we find that they share commonalities but they are not largely similar. Such extracted data are likely capturing different aspects of the structural information.

4.3 Next Steps

Learning from collective findings and reflections, the next step is to build a robust child behavior mental health detection system. Chapter 2 provides interpretations but no predictions; Chapter 3 predicts the child behavior problems and interprets the result, but the accuracy awaits improvement and validation. Next, more advanced analysis and machine learning algorithms need to be investigated with the aim of developing a prediction system that achieves both high accuracy and interpretability.

4.3.1 Case Studies and Multimodal Learning

This step aims to utilize data presented in Chapter 3 to further improve the accuracy of the prediction using the new set of features presented in Section 4.2.1. While the above feature extraction techniques present stimulating possibilities, the optimized hyperparameters such as the distance to split the session and the objective way to define layers need to be further explored. The accuracies of feature extraction also need to be validated in detail with the participants' data.

Before moving on to predictions, case studies need to be conducted to validate the feature extraction. Exploratory data analysis (EDA), and especially statistical analysis, can be performed to validate the relationship of variables in the data. Feature extraction is a major part of developing a robust prediction system. The case studies must ensure the validity and usability of (1) extracted features and (2) augmented data with sub-sessions.

Next, machine learning methods will be investigated to predict the behavior problems measured by CBCL. Among diverse algorithms, special attention needs to be paid to Multimodal Learning [166], which is capable of inputting the multimodal signals such as video, audio, and text, and then creating the joint representations of different modalities for predictions.

A simplified example of a multimodal learning structure can be seen in Figure 4.11. In our case, we may be able to use motional features and structural features together or separately to build LSTM sub-networks, while also inputting structural video (see example in Figure 4.6) to build a CNN sub-network. The encoded features will then be concatenated to predict the result.

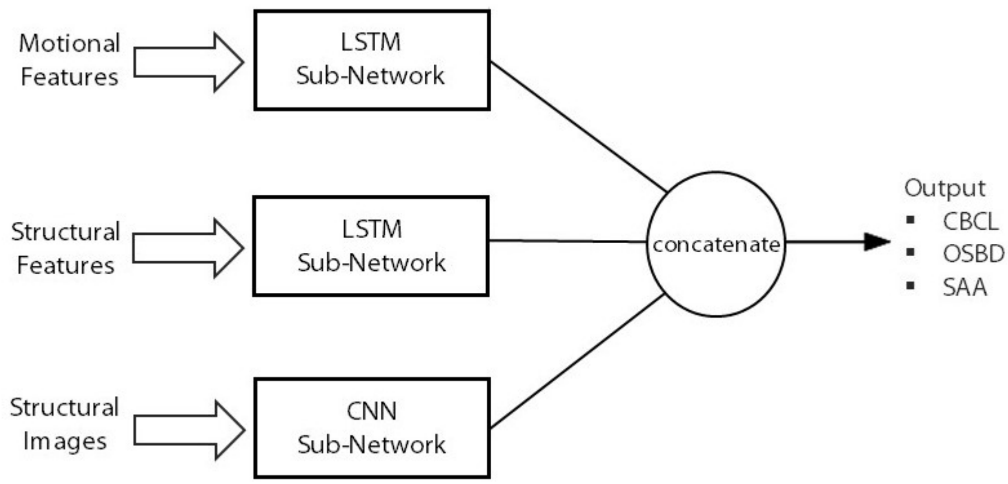


Figure 4.11: Example of multimodal deep learning that takes motional features, structural features, and structural images as input

4.3.2 Modalities: Extract Meaningful Behavior

This next step aims to revise and improve the block system to include more modalities in data collection, which may enhance the ability to extract meaningful behavior. The design of the system can be expanded to more shapes and objects, internal sensors, and environmental sensors.

Besides big and small cuboid-shaped blocks, shapes such as columns, cylinders, and triangle primes can be embedded with sensors (see example in Figure 4.12.) Note that these new shapes may provide higher and more specific affordance, which may encourage specific actions such as house building and rolling. Thus, a sufficient number of low-affordance blocks, such as cuboids, needs to be maintained. Pilot studies on normal children and at-risk children need to be run under the conditions of with and without new shapes, to analysis whether blocks with new shapes can extract more meaningful behaviors.

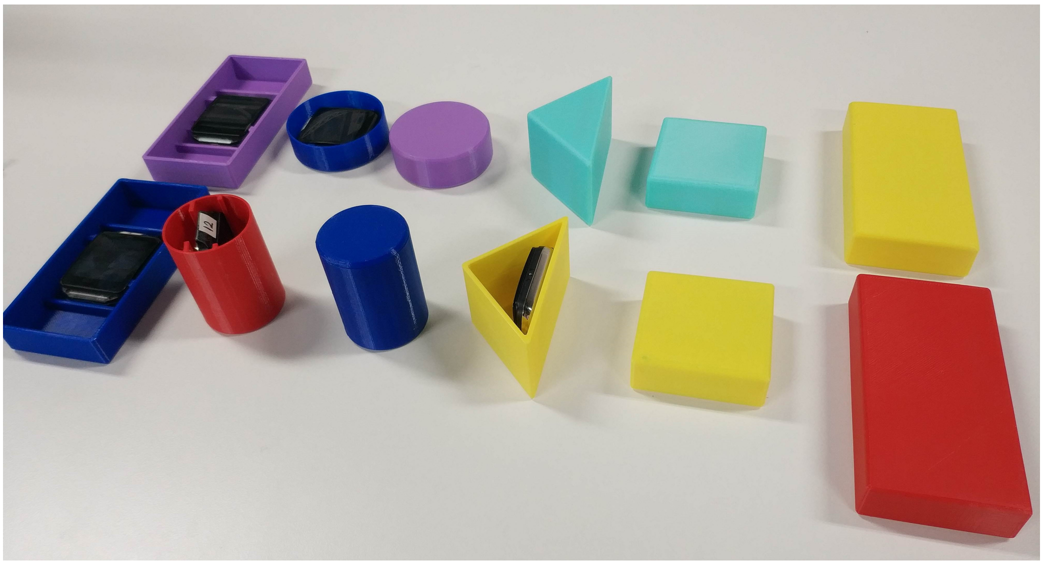


Figure 4.12: Blocks with five shapes embedded with sensors: big and small cuboid, column, cylinder, and triangle prime

Sensors embedded in each block can be further extended to collect structural and physiological information. For example, a Lighthouse positioning system [167] can be used on top of IMU to acquire the accurate location information of each block. A Lighthouse deck can be embedded inside each of the blocks, and externally based stations can be installed in the room. Sensors can also be embedded on the surface, such as IR sensors [65] or capacitive sensors [101], to detect the hand grasp as well as to indicate the structure. However, before these novel sensing techniques are adopted, high accuracy needs to be ensured for collecting valid data.

Environmental sensors, such as cameras, can be set up as a way to document the session, as well as a mean to extract structural information, interactions, and audio. From our observations, children in general are not as sensitive to cameras around them as adults. Most of them are not interested in the camera, and those whose attention is caught by the camera tend to forget its existence quickly once they start to play. Thus, the camera provides an objective means to capture a range of features, even though it suffers from occlusion and cannot capture various nuances. Instead of a conventional RGB camera, a depth camera integrated with a Lidar sensor might provide richer information. The location of the camera also needs to be designed so that generalizability is maintained by the videos that are captured from similar positions and angles.

4.4 Future Work

The studies and insights in this thesis provides a wide space for future work and demonstrates the need to bring the system closer to daily life. Future investigations can go either deeper or wider, and a wide range of applications can be developed to support health and well-being.

4.4.1 Future Directions in Depth and Breadth

The future directions can be summarized in two ways: depth and breadth.

In depth, the block play behavior detection can focus on specific disorders, such as Autism, ADHD, and PTSD. Our work provides promising signs for probing these disorders, while free play could be replaced by or combined with structured play that augments existing diagnosis methods to automate detection and observation with the aim of higher accuracy. Future work could also move from prediction to support, which could be used in play therapy, intervention and education. Our behavior-extraction structure provides the capability to replace existing therapies and intervention methods that require detailed observations and expertise.

In breadth, by utilizing the obtained knowledge, future work could move beyond blocks to develop new toys and other types of TUIs for behavior extraction. The concept could be generalized to other vulnerable groups, such as people with accessibility needs and the elderly. Extracted behavior could be used to detect the user's characteristics and thus to customize the designs and interfaces for each individual. Our behavior characterization could also be generalized to the majority of people. For example, an adult's behavioral characteristics could be extracted from their hand-activity data, which could be used to monitor and predict their affects and mental states.

4.4.2 Applications: From Detection to Interaction

Future work should not only focus on detection but also investigate the methods that use the detection results for interactions. Currently, our system mainly captures input from the user. A full loop of interaction could be achieved by designing output systems, or coordinating with existing output systems.

Specific examples such as augmenting play therapies could be developed. Nowadays, the play therapy for ADHD encourages the child to be less impulsive and more reflective. The therapist mediates the play by asking questions before building, when the structure falls, and when involvement level is decreased [15]. The system might be able to assist the role of the therapist by (1) detecting when the play is impulsive and has low engagement and (2) outputting feedback, such as digital feedback on a screen or tabletop. Social robots could be used to ask the questions and foster interactions.

The block system could also be used to augment existing play therapy for social withdrawal. A child with social withdrawal is known to be unable to negotiate play interactions with others. When playing socially, they cannot understand the language and actions of others or how to respond with appropriate language and actions in the context of the immediate play situation [15]. Such behavioral characteristics could be reflected in our extracted behaviors if we distinguished the users, using video or other tracking tools. If we detected a child exhibiting difficulties in interacting with others, digital feedback, a social robot, or move-able blocks could be designed and integrated to instruct, model, prompt, and reinforce the appropriate actions, which was previously done by a therapist.

Based on the scenario and context, other interactive applications could be developed to predict child behavior and thus support the healthier behavior.

Chapter 5

Conclusion

This thesis focused on establishing a novel method of predicting child behavior and mental health using data captured from toy block play. Step by step, the thesis delivered the following contributions:

- It proposed a range of methods to decompose a child's block play behavior into quantitative actions, sequential play patterns, structures and play styles.
- It discovered the link between children's post-disaster short-term stress, measured in-situ, and the fundamental toy block features, including actions and play behaviors.
- It developed child behavior prediction methods. Total problems, internalizing problems, and aggressive behavior can be predicted based on quantitative and sequential play features.
- Design and data-extraction guidelines were proposed for developing robust systems for mental health prediction and support through play behavior.

These contributions are connected through the goal of achieving a robust mental health prediction system using toy blocks. They also lead to a range of opportunities for future work. In depth, they suggest detecting specific disorders and building support systems. In breadth, they can be expanded to other TUIs for inferring mental states through interactions with daily objects.

Acknowledgments

I deeply thank my supervisor, Professor Yoshifumi Kitamura, for providing his guidance, concrete and constructive comments, and a myriad of opportunities to conduct research and engage in the research community. With great passion and dedication toward research, he has always been my model as a researcher.

I am also thankful to the Associate Professor of my laboratory, Kazuki Takashima, for his mentoring, advice, and support. I am grateful for the number of intellectually stimulating discussions we have had over the years, as well as for his support when I was constructing various research papers.

I thank Professor Tomoaki Adachi of Miyagi Gakuin University, Sendai, Japan, for his guidance as the main psychologist in the team. Without his support, I could have achieved my aims.

I am also grateful to Professors Ehud Sharlin and Patrick Finn of the University of Calgary, Alberta, Canada, for their guidance, support, and encouragement.

I am very thankful to the Assistant Professor of my laboratory, Kazuyuki Fujita, for his support and advice during the course of my studies.

I would also like to express my gratitude toward the professors and students who were working on the subject before I joined the team: Prof. Hiroyasu Kanetaka, Graduate School of Dentistry, Tohoku University; Prof. Yoshie Ohashi, Faculty of Human Welfare, Seigakuin University; and Yoshiki Hosoi, Hiromi Sato and Miteki Ishikawa, alumni of Interactive Content Design Lab, Tohoku University, for their contributions to the field study, data collection and analysis.

I would also like to thank all of the collaborating kindergartens for organizing their valuable participants and giving support in field studies and data collections.

I also thank all of my colleagues and fellow students in the Interactive Content Design Lab for our intellectual exchanges over the years and for enriching my master and doctoral studies.

I would like to express my appreciation to my lab's secretary Ms. Nobuyo Maejima, who has given me warm support and help, especially when I encounter difficulties as an international student.

I thank the Graduate Program of Data Science, particularly Professors Mitsuyuki Nakao, Kuzunori Yamada, and Takuro Nakayama, for not only helping me develop my skills as a data scientist but also financially supporting me and gave me opportunities for exchanges with international and interdisciplinary researchers.

I thank the other organizations that have provided resources for my doctoral studies. I thank the Japan Society for the Promotion of Science (JSPS)'s DC.2 program for their financial support, which gave me the research freedom and stimulated me to be a stronger researcher. I also appreciate the research support from the JSPS Kakenhi and Natural Sciences and Engineering Research Council of Canada (NSERC).

Last but not least, I would like to thank my family for supporting me throughout my life. I am especially thankful for their love and emotional support during my most difficult times of writing papers.

Without the above support, this work would not have been possible.

Publications

In Chapter 2:

[1] Xiyue Wang, Kazuki Takashima, Tomoaki Adachi, Patrick Finn, Ehud Sharlin, and Yoshifumi Kitamura. *AssessBlocks: Exploring Toy Block Play Features for Assessing Stress in Young Children after Natural Disasters.* Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4, 1, Article 30 (March 2020), 29 pages. https://doi.org/10.1145/3381016

[2] Xiyue Wang, Miteki Ishikawa, Kazuki Takashima, Tomoaki Adachi, Ehud Sharlin, Yoshifumi Kitamura. *Activity-Characterizing Toy Blocks for Behavioral Assessments.* 4th Ensemble Workshop for Young Researchers at Tohoku University. Sendai, Japan 2018 (Best Poster Award)

In Chapter 3:

[3] Xiyue Wang, Kazuki Takashima, Tomoaki Adachi, Yoshifumi Kitamura. *Can Playing with Toy Blocks Reflect Behavior Problems in Children?* In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 540, 1–14. https://doi.org/10.1145/3411764.3445119

[4] Xiyue Wang, Miteki Ishikawa, Kazuki Takashima, Tomoaki Adachi, Patrick Finn, Ehud Sharlin, Yoshifumi Kitamura. *Children's Blocks: Machine Learning and the Analysis of Motion During Play.* Asian CHI Symposium at CHI'18 2018 (Best Demo/Poster Award)

In Chapter 4:

[5] Xiyue Wang, Kaori Ikematsu, Kazuyuki Fujita, Kazuki Takashima, Yoshifumi Kitamura. *An Investigation of Electrode Design for Physical Touch Extensions on a Capacitive Touch Surface.* In The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19). Association for Computing Machinery, New York, NY, USA, 66–68. https://doi.org/10.1145/3332167.3357117

[6] Xiyue Wang, Miteki Ishikawa, Kazuki Takashima, Tomoaki Adachi, Ehud Sharlin, Patrick Finn, Yoshifumi Kitamura. *Designing Action-Characterizing Toy Blocks for Behavior Assessments.* In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18). Association for Computing Machinery, New York, NY, USA, Paper LBW513, 1–6. https://doi.org/10.1145/3170427.3188451

[7] Xiyue Wang, Kazuki Takashima, Tamoaki Adachi, Yoshifumi Kitamura. *Machine Learning Enhanced Novel Sensing with Smart Toy Blocks for Children's Action Recognition.* International Kick-off Symposium of Graduate Program in Data Science. Sendai, Japan 2018

Bibliography

[1] D. Ness and S. J. Farenga, “Blocks, Bricks, and Planks Relationships between Affordance and Visuo-Spatial Constructive Play Objects s,” *American Journal of Play*, vol. Vol. 8, no. Iss. 2, pp. 201–227, 2016.

[2] T. M. Amabile, B. A. Hennessey, and B. S. Grossman, “Social influences on creativity: The effects of contracted-for reward.,” *Journal of personality and social psychology*, vol. 50, no. 1, p. 14, 1986.

[3] T. M. Amabile and J. Pillemer, “Perspectives on the social psychology of creativity,” *The Journal of Creative Behavior*, vol. 46, no. 1, pp. 3–15, 2012.

[4] E. L. Deci and R. M. Ryan, “The empirical exploration of intrinsic motivational processes,” in *Advances in experimental social psychology*, vol. 13, pp. 39–80, Elsevier, 1980.

[5] S. Farenga, D. Ness, D. D. Johnson, and B. Johnson, *The importance of average: Playing the game of school to increase success and achievement*. Rowman & Littlefield Publishers, 2010.

[6] D. Elkind, *The power of play: Learning what comes naturally*. Da Capo Lifelong Books, 2007.

[7] N. Kimura, “How to do the screening for developmental disorder in the routine 18 month and 36 month health check up [in Japanese],” No. 24, pp. 13–19, 2009.

[8] Y. Kitamura, Y. Itoh, and F. Kishino, “Real-time 3D interaction with ActiveCube,” in *Proceedings of Extended Abstracts on the SIGCHI Conference on Human Factors in Computing Systems*, pp. 355–356, 2001.

[9] E. Sharlin, Y. Itoh, B. Watson, Y. Kitamura, S. Sutphen, and L. Liu, “Cognitive Cubes: A tangible user interface for cognitive assessment,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 347–354, 2002.

[10] S. Jacoby, G. Gutwillig, D. Jacoby, N. Josman, P. L. Weiss, M. Koike, Y. Itoh, N. Kawai, Y. Kitamura, and E. Sharlin, “PlayCubes: monitoring constructional ability in children using a tangible user interface and a playful virtual environment,” in *Proceedings of the IEEE Virtual Rehabilitation International Conference*, pp. 42–49, jun 2009.

[11] X. Jiang, Y. Chen, W. Huang, T. Zhang, C. Gao, Y. Xing, and Y. Zheng, “Weda: Designing and evaluating a scale-driven wearable diagnostic assessment system for children with adhd,” in *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, p. 1–12, 2020.

[12] B. A. Snyder, “Expressive art therapy techniques: healing the soul through creativity,” vol. 36, pp. 74–82, Blackwell Publishing Ltd, dec 1997.

[13] V. Axline, *Play Therapy*. Ballantine books, Ballantine Books, 1969.

[14] D. Adam, “Mental health: On the spectrum,” *Nature*, vol. 496, pp. 416–418, apr 2013.

[15] H. Kaduson and C. E. Schaefer, *101 favorite play therapy techniques. Volume III*. Jason Aronson, 2010.

[16] D. D. Ross and D. L. Rogers, “Social competence in kindergarten: Analysis of social negotiations during peer play,” *Early Child Development and Care*, vol. 64, no. 1, pp. 15–26, 1990.

[17] R. D. Phillips, “Whistling in the dark? a review of play therapy research.,” *Psychotherapy: Theory, Research, Practice, Training*, vol. 22, no. 4, p. 752, 1985.

[18] M. LeBlanc and M. Ritchie, "Predictors of play therapy outcomes.," *International Journal of Play Therapy*, vol. 8, no. 2, p. 19, 1999.

[19] R. W. Picard, *Affective computing*. MIT press, 2000.

[20] R. W. Picard, "Affective Computing,"

[21] J. Tao and T. Tan, "Affective computing: A review," in *International Conference on Affective computing and intelligent interaction*, pp. 981–995, Springer, 2005.

[22] R. W. Picard, "Overview ‹ affective computing — mit media lab."

[23] Abdul Latif Jameel Clinic for Machine Learning in Health (J-Clinic), "Leveraging artificial intelligence for the assessment of severity of depressive symptoms."

[24] D. Cogan, M. B. Pouyan, M. Nourani, and J. Harvey, "A wrist-worn biosensor system for assessment of neurological status," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5748–5751, IEEE, 2014.

[25] R. R. Fletcher, K. Dobson, M. S. Goodwin, H. Eydgahi, O. Wilder-Smith, D. Fernholz, Y. Kuboyama, E. B. Hedman, M.-Z. Poh, and R. W. Picard, "icalm: Wearable sensor and network architecture for wirelessly communicating and logging autonomic activity," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 215–223, 2010.

[26] R. W. Picard, "From icalm to q sensor to physiio to empatica."

[27] Empatica, "Embrace2,seizure & epilepsy watch."

[28] Empatica, "E4 wristband."

[29] K. T. Johnson and R. W. Picard, "Spring: Customizable, motivation-driven technology for children with autism or neurodevelopmental differences.," in *IDC*, pp. 149–158, 2017.

[30] J. Intarasirisawat, C. S. Ang, C. Efstratiou, L. W. F. Dickens, and R. Page, "Exploring the touch and motion features in game-based cognitive assessments," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, pp. 87:1–87:25, Sept. 2019.

[31] E. Nosakhare and R. Picard, "Toward assessing and recommending combinations of behaviors for improving health and well-being," *ACM Transactions on Computing for Healthcare*, vol. 1, Mar. 2020.

[32] R. Wampfler, S. Klingler, B. Solenthaler, V. R. Schinazi, and M. Gross, "Affective state prediction based on semi-supervised learning from smartphone touch data," in *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, p. 1–13, 2020.

[33] Center for research on the epidemiology of disasters, "The human cost of natural disasters: a global perspective," 2018.

[34] UNISDR, "Disaster Statistics," 2018.

[35] T. L. Holzer and J. C. Savage, "Global earthquake fatalities and population," vol. 29, pp. 155–175, Earthquake Engineering Research Institute, feb 2013.

[36] C. Kousky, "Impacts of natural disasters on children," vol. 26, pp. 73–92, Princeton University, 2016.

[37] J. Blanc, E. Bui, Y. Mouchenik, D. Derivois, and P. Birmes, "Prevalence of post-traumatic stress disorder and depression in two groups of children one year after the January 2010 earthquake in Haiti," vol. 172, pp. 121–126, Elsevier, feb 2015.

[38] Z. Jia, W. Tian, X. He, W. Liu, C. Jin, and H. Ding, "Mental health and quality of life survey among child survivors of the 2008 Sichuan earthquake," vol. 19, pp. 1381–1391, Springer Netherlands, nov 2010.

[39] T. M. Powell and T. Bui, "Supporting social and emotional skills after a disaster: findings from a mixed methods study," vol. 8, pp. 106–119, Springer US, mar 2016.

[40] M. Fan, A. N. Antle, M. Hoskyn, C. Neustaedter, and E. S. Cramer, "Why tangibility matters: a design case study of at-risk children learning to read and spell," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1805–1816, 2017.

[41] A. Hornof, H. Whitman, M. Sutherland, S. Gerendasy, and J. McGrenere, "Designing for the "Universe of One": personalized interactive media systems for people with the severe cognitive impairment associated with Rett syndrome," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2137–2148, 2017.

[42] K. Spiel, C. Frauenberger, E. Hornecker, and G. Fitzpatrick, "When empathy is not enough: assessing the experiences of autistic children with technologies," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2853–2864, 2017.

[43] T. Adachi, Y. Kitamura, K. Takashima, T. Hosoi, Y. Ohashi, Y. Ito, and H. Kanetaka, "Effect of playing with building blocks on young children with and without posttraumatic stress disorder [in Japanese]," No. 14, pp. 25–30, 2014.

[44] T. Adachi, Y. Kitamura, K. Takashima, and M. Ishikawa, "Effects of block play on salivary alpha-amylase activity in children who lived in affected or less affected area by the tsunami," 2017.

[45] A. Ishigaki, H. Higashi, T. Sakamoto, and S. Shibahara, "The great east-Japan earthquake and devastating tsunami: an update and lessons from the past great earthquakes in Japan since 1923," vol. 229, pp. 287–299, Tohoku University Medical Press, 2013.

[46] S. M. Becker, "Psychosocial care for adult and child survivors of the tsunami disaster in India," vol. 20, pp. 148–155, Blackwell Publishing Inc, aug 2007.

[47] B. Saraceno, "The WHO's mental health response to the Asian tsunami," vol. 4, pp. 66–7, World Psychiatric Association, jun 2005.

[48] WHO, "Mental Health Assistance to the Populations Affected by the Tsunami in Asia," 2012.

[49] T. Sugimoto, T. Shinozaki, and Y. Miyamoto, "Aftershocks associated with impaired health caused by the great east Japan disaster among youth across Japan: A national cross-sectional survey," vol. 15, dec 2013.

[50] C. M. Chemtob, Y. Nomura, and R. A. Abramovitz, "Impact of conjoined exposure to the World Trade Center Attacks and to other traumatic events on the behavioral problems of preschool children," vol. 162, p. 126, American Medical Association, feb 2008.

[51] E. Mullett-Hume, D. Anshel, V. Guevara, and M. Cloitre, "Cumulative trauma and posttraumatic stress disorder among children exposed to the 9/11 World Trade Center attack.," vol. 78, pp. 103–108, 2008.

[52] R. S. Pynoos, A. Goenjian, M. Tashjian, M. Karakashian, R. Manjikian, G. Manoukian, A. M. Steinberg, and L. A. Fairbanks, "Post-traumatic stress reactions in children after the 1988 Armenian earthquake.," vol. 163, pp. 239–47, The Royal College of Psychiatrists, aug 1993.

[53] J. Baggerly and H. A. Exum, "Counseling children after natural disasters: guidance for family therapists," vol. 36, pp. 79–93, Taylor & Francis Group, nov 2007.

[54] L. A. Reddy, T. M. Files-Hall, and C. E. Schaefer, *Empirically Based Play Interventions for Children*. Washington: American Psychological Association, 2005.

[55] M. J. Pollman, *Blocks and beyond : strengthening early math and science skills through spatial learning*. Paul H. Brookes Pub. Co, 2010.

[56] B. M. Landry, E. K. Choe, S. McCutcheon, and J. A. Kientz, "Post-traumatic stress disorder: opportunities challenges for computing

technology," in *Proceedings of the ACM International Health Informatics Symposium*, p. 780–789, 2010.

[57] C. Botella, R. Baños, B. Rey, M. Alcañiz, V. Guillen, S. Quero, and A. García-Palacios, "Using an adaptive display for the treatment of emotional disorders: a preliminary analysis of effectiveness," in *Extended Abstracts on SIGCHI Conference on Human Factors in Computing Systems*, pp. 586–591, 2006.

[58] A. Rizzo, A. Hartholt, M. Grimani, A. Leeds, and M. Liewer, "Virtual reality exposure therapy for combat-related posttraumatic stress disorder," vol. 47, pp. 31–37, jul 2014.

[59] G. Barish, E. Elbogen, P. Lester, and W. R. Saltzman, "Beyond sensors: reading patients through caregivers and context," in *Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 1273–1277, 2014.

[60] N. Yamashita, H. Kuzuoka, K. Hirata, T. Kudo, E. Aramaki, and K. Hattori, "Changing moods: how manual tracking by family caregivers improves caring and family communication," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 158–169, 2017.

[61] T. L. Westeyn, G. D. Abowd, T. E. Starner, J. M. Johnson, P. W. Presti, and K. A. Weaver, "Monitoring children's developmental progress using augmented toys and activity recognition," vol. 16, pp. 169–191, Springer-Verlag, feb 2012.

[62] A. Girouard, D. McGookin, P. Bennett, O. Shaer, K. A. Siek, and M. Lennon, "Tangibles for health workshop," in *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3461–3468, 2016.

[63] S. Cartwright, "Play can be the building blocks of learning," vol. 43, pp. 44–47, 1988.

[64] E. Vonach, M. Ternek, G. Gerstweiler, and H. Kaufmann, “Design of a health monitoring toy for children,” pp. 58–67, 2016.

[65] M. Ando, Y. Itoh, T. Hosoi, K. Takashima, K. Nakajima, and Y. Kitamura, “StackBlock: block-shaped interface for flexible stacking,” in *Adjunct Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 41–42, 2014.

[66] P. Wang, G. D. Abowd, and J. M. Rehg, “Quasi-periodic event analysis for social game retrieval,” in *Proceedings of IEEE International Conference on Computer Vision*, pp. 112–119, sep 2009.

[67] Y. Liu, X. Zhang, J. Cui, C. Wu, H. Aghajan, and H. Zha, “Visual analysis of child-adult interactive behaviors in video sequences,” in *Proceedings of the IEEE International Conference on Virtual Systems and Multimedia*, pp. 26–33, oct 2010.

[68] B. Yang, J. Cui, H. Zha, and H. Aghajan, “Visual context based infant activity analysis,” in *IEEE International Conference on Distributed Smart Cameras*, pp. 1–6, 2012.

[69] T. Hosoi, K. Takashima, T. Adachi, Y. Itoh, and Y. Kitamura, “A-blocks: recognizing and assessing child building processes during play with toy blocks,” in *SIGGRAPH Asia 2014 Emerging Technologies*, pp. 1–2, 2014.

[70] T. Sadoh and K. Nakato, “Surface properties of wood in physical and sensory aspects,” vol. 21, pp. 111–120, Jun 1987.

[71] Nichigan Original, “Unpainted Wooden Blocks,” 2017.

[72] ATR-Promotions, “TSND121/151,” 2017.

[73] D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. P. Cardoso, “Preprocessing techniques for context recognition from accelerometer data,” vol. 14, pp. 645–662, Springer-Verlag, oct 2010.

[74] A. M. Khan, Young-Koo Lee, S. Y. Lee, and Tae-Seong Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," vol. 14, pp. 1166–1172, sep 2010.

[75] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," pp. 385–396, JSTOR, 1983.

[76] A. Keller, K. Litzelman, L. E. Wisk, T. Maddox, E. R. Cheng, P. D. Creswell, and W. P. Witt, "Does the perception that stress affects health matter? the association with health and mortality," vol. 31, pp. 677–684, sep 2012.

[77] S. A. Taylor, N. Jaques, E. Nosakhare, A. Sano, and R. Picard, "Personalized multitask learning for predicting tomorrow's mood, stress, and health," pp. 1–1, 2017.

[78] Y. Shimomura, M. Fukasawa, and K. Takeda, "Stress evaluation in children with cancer undergoing invasive medical procedures [in Japanese]," vol. 22, pp. 112–118, 2010.

[79] P. Ferrara, G. Bottaro, S. Angeletti, A. Gatto, O. Vitelli, D. Battaglia, M. Del Re, A. Ruggiero, and G. Dicuonzo, "Salivary alpha-amylase: a new non-invasive biomarker for assessment of pain perception in epileptic children," vol. 113, pp. 279–283, Springer, 2013.

[80] H. Tsumura, H. Shimada, H. Morimoto, C. Hinuma, and Y. Kawano, "Effects of distraction on negative behaviors and salivary α-amylase under mildly stressful medical procedures for brief inpatient children," vol. 19, pp. 1079–1088, Sage Publications Sage UK: London, England, 2014.

[81] N. Furlan, M. Gaviäo, T. Barbosa, J. Nicolau, and P. M. Castelo, "Salivary cortisol, alpha-amylase and heart rate variation in response to dental treatment in children," vol. 37, pp. 83–87, 2012.

[82] G. K. Sahu, S. Upadhyay, and S. M. Panna, "Salivary alpha amylase activity in human beings of different age groups subjected to psychological stress.," vol. 29, pp. 485–490, Springer, oct 2014.

[83] R. Feldman, A. Vengrober, M. Eidelman-Rothman, and O. Zagoory-Sharon, "Stress reactivity in war-exposed young children with and without posttraumatic stress disorder: Relations to maternal stress hormones, parenting, and child emotionality and regulation," vol. 25, pp. 943–955, Cambridge University Press, nov 2013.

[84] U. M. Nater, N. Rohleder, J. Gaab, S. Berger, A. Jud, C. Kirschbaum, and U. Ehlert, "Human salivary alpha-amylase reactivity in a psychosocial stress paradigm," vol. 55, pp. 333–342, Elsevier, mar 2005.

[85] C. H. Elliott, S. M. Jay, and P. Woody, "An observation scale for measuring children's distress during medical procedures," vol. 12, pp. 543–551, Oxford University Press, dec 1987.

[86] VAS, "Visual Analogue Scale - Physiopedia," 2017.

[87] R. L. Blount, V. Bunke, L. L. Cohen, and C. J. Forbes, "The child–adult medical procedure interaction scale-short form (campis-sf): validation of a rating scale for children's and adults' behaviors during painful medical procedures," vol. 22, pp. 591 – 599, 2001.

[88] K. J. Koss, M. R. George, E. M. Cummings, P. T. Davies, M. El-Sheikh, and D. Cicchetti, "Asymmetry in children's salivary cortisol and alpha-amylase in the context of marital conflict: Links to children's emotional security and adjustment," vol. 56, pp. 836–849, Wiley Online Library, 2014.

[89] E. P. Davis and D. A. Granger, "Developmental differences in infant salivary alpha-amylase and cortisol responses to stress," vol. 34, pp. 795–804, Elsevier, 2009.

[90] M. J. P. dos Santos, D. G. Bernabé, A. C. d. M. S. Nakamune, S. H. V. Perri, S. H. P. de Oliveira, *et al.*, "Salivary alpha amylase and cortisol levels in children with global developmental delay and their

relation with the expectation of dental care and behavior during the intervention," vol. 33, pp. 499–505, Elsevier, 2012.

[91] J. Sarama and D. H. Clements, "Building blocks and cognitive building blocks: playing to know the world mathematically," in *American Journal of Play*, vol. 1, pp. 313–337, 2009.

[92] E. S. Hirsch, *The block book*. National Association for the Education of Young Children, 1996.

[93] Japan Meteorological Agency, "Number of aftershocks of M5.0 or more [in Japanese]," 2020.

[94] F. Shahid, W. Rahman, A. B. Islam, N. Paul, N. Khan, M. S. Rahman, M. M. Haque, and A. B. M. A. Al Islam, "Two tell-tale perspectives of ptsd: neurobiological abnormalities and bayesian regulatory network of the underlying disorder in a refugee context," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, (New York, NY, USA), pp. 101:1–101:45, ACM, Sept. 2019.

[95] B. L. Cardozo, A. Vergara, F. Agani, and C. A. Gotway, "Mental health, social functioning, and attitudes of Kosovar Albanians following the war in Kosovo," *JAMA*, vol. 284, pp. 569–577, 08 2000.

[96] Z. Steel, D. M. Silove, T. Phan, and A. Bauman, "Long-term effect of psychological trauma on the mental health of vietnamese refugees resettled in australia: a population-based study," vol. 360, pp. 1056–1062, 2002.

[97] R. A. Goodfader, "Sex differences in the play constructions of preschool children," vol. 52, pp. 129–144, Taylor & Francis Group, mar 1982.

[98] B. M. Casey, E. E. Pezaris, and J. Bassi, "Adolescent boys' and girls' block constructions differ in structural balance: A block-building characteristic related to math achievement," vol. 22, pp. 25–36, JAI, feb 2012.

[99] J. Mirowsky, *Analyzing associations between mental health and social circumstances*, pp. 105–123. 1999.

[100] H. Kamide, K. Takashima, M. Ishikawa, T. Adachi, and Y. Kitamura, "Quantitative evaluation of childrenapos;s developmental stage in building block play [in japanese]," vol. 20, pp. 107–114, 2018.

[101] X. Wang, M. Ishikawa, K. Takashima, T. Adachi, E. Sharlin, P. Finn, and Y. Kitamura, "Designing action-characterizing toy blocks for behavior assessments," in *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*, 2018.

[102] T. Fujiwara, J. Yagi, H. Homma, H. Mashiko, K. Nagao, M. Okuyama, *et al.*, "Clinically significant behavior problems among young children 2 years after the great east japan earthquake," *PLoS One*, vol. 9, no. 10, p. e109342, 2014.

[103] J. Parsons, T. J. Kehle, and S. V. Owen, "Incidence of behavior problems among children of vietnam war veterans," *School Psychology International*, vol. 11, no. 4, pp. 253–259, 1990.

[104] S. B. Campbell, "Behavior problems in preschool children: A review of recent research," *Journal of Child Psychology and Psychiatry*, vol. 36, no. 1, pp. 113–149, 1995.

[105] V. C. Strasburger, A. B. Jordan, and E. Donnerstein, "Health effects of media on children and adolescents," *Pediatrics*, vol. 125, no. 4, pp. 756–767, 2010.

[106] W. H. Organization, *Atlas: child and adolescent mental health resources: global concerns, implications for the future*. World Health Organization, 2005.

[107] T. M. Achenbach, "Manual for the child behavior checklist/4-18 and 1991 profile," *University of Vermont, Department of Psychiatry*, 1991.

[108] K. Wellhousen and J. E. Kieff, *A constructivist approach to block play in early childhood*. Cengage Learning, 2001.

[109] D. Ness and S. J. Farenga, “Blocks, bricks, and planks: Relationships between affordance and visuo-spatial constructive play objects.,” *American Journal of Play*, vol. 8, no. 2, pp. 201–227, 2016.

[110] X. Wang, K. Takashima, T. Adachi, P. Finn, E. Sharlin, and Y. Kitamura, “Assessblocks: Exploring toy block play features for assessing stress in young children after natural disasters,” *Proceedings of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, Mar. 2020.

[111] Y. Honda, T. Fujiwara, J. Yagi, H. Homma, H. Mashiko, K. Nagao, M. Okuyama, M. Ono-Kihara, and M. Kihara, “Long-term impact of parental ptsd symptoms on mental health of their offspring after the great east japan earthquake,” *Frontiers in Psychiatry*, vol. 10, p. 496, 2019.

[112] W. H. Organization, *The World Health Report 2001: Mental health: new understanding, new hope*. World Health Organization, 2001.

[113] G. V. Polanczyk, G. A. Salum, L. S. Sugaya, A. Caye, and L. A. Rohde, “Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents,” *Journal of Child Psychology and Psychiatry*, vol. 56, no. 3, pp. 345–365, 2015.

[114] E. Simonoff, A. Pickles, T. Charman, S. Chandler, T. Loucas, and G. Baird, “Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample,” *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 47, no. 8, pp. 921–929, 2008.

[115] K. R. Merikangas, R. Jin, J.-P. He, R. C. Kessler, S. Lee, N. A. Sampson, M. C. Viana, L. H. Andrade, C. Hu, E. G. Karam, *et al.*, “Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative,” *Archives of General Psychiatry*, vol. 68, no. 3, pp. 241–251, 2011.

[116] K. T. Mueser and J. Taub, “Trauma and ptsd among adolescents with severe emotional disorders involved in multiple service systems,” *Psychiatric Services*, vol. 59, no. 6, pp. 627–634, 2008.

[117] S. A. Husain, M. A. Allwood, and D. J. Bell, “The relationship between ptsd symptoms and attention problems in children exposed to the bosnian war,” *Journal of Emotional and Behavioral Disorders*, vol. 16, no. 1, pp. 52–62, 2008.

[118] R. R. Silva, M. Alpert, D. M. Munoz, S. Singh, F. Matzner, and S. Dummit, “Stress and vulnerability to posttraumatic stress disorder in children and adolescents,” *American Journal of Psychiatry*, vol. 157, no. 8, pp. 1229–1235, 2000.

[119] S. Arman, H. Salimi, and M. Maracy, “Parenting styles and psychiatric disorders in children of bipolar parents,” *Advanced Biomedical Research*, vol. 7, no. 1, p. 147, 2018.

[120] R. C. Whitaker, S. M. Phillips, and S. M. Orzol, “Food insecurity and the risks of depression and anxiety in mothers and behavior problems in their preschool-aged children,” *Pediatrics*, vol. 118, no. 3, pp. e859–e868, 2006.

[121] V. Patel and A. Kleinman, “Poverty and common mental disorders in developing countries,” *Bulletin of the World Health Organization*, vol. 81, pp. 609–615, 2003.

[122] B. J. Bushman and L. R. Huesmann, “Short-term and long-term effects of violent media on aggression in children and adults,” *Archives of Pediatrics & Adolescent Medicine*, vol. 160, no. 4, pp. 348–352, 2006.

[123] R. C. Kessler, P. Berglund, O. Demler, R. Jin, K. R. Merikangas, and E. E. Walters, “Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication,” *Archives of General Psychiatry*, vol. 62, no. 6, pp. 593–602, 2005.

[124] WHO, “Autism spectrum disorders,” 2019.

[125] R. C. Kessler, K. A. McLaughlin, J. G. Green, M. J. Gruber, N. A. Sampson, A. M. Zaslavsky, S. Aguilar-Gaxiola, A. O. Alhamzawi, J. Alonso, M. Angermeyer, *et al.*, "Childhood adversities and adult psychopathology in the who world mental health surveys," *The British Journal of Psychiatry*, vol. 197, no. 5, pp. 378–385, 2010.

[126] M. D. Kogan, S. J. Blumberg, L. A. Schieve, C. A. Boyle, J. M. Perrin, R. M. Ghandour, G. K. Singh, B. B. Strickland, E. Trevathan, and P. C. van Dyck, "Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the us, 2007," *Pediatrics*, vol. 124, no. 5, pp. 1395–1403, 2009.

[127] D. Vigo, G. Thornicroft, and R. Atun, "Estimating the true global burden of mental illness," *The Lancet Psychiatry*, vol. 3, no. 2, pp. 171–178, 2016.

[128] P. S. Wang, S. Aguilar-Gaxiola, J. Alonso, M. C. Angermeyer, G. Borges, E. J. Bromet, R. Bruffaerts, G. De Girolamo, R. De Graaf, O. Gureje, *et al.*, "Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the who world mental health surveys," *The Lancet*, vol. 370, no. 9590, pp. 841–850, 2007.

[129] S. E. Meyer, G. A. Carlson, E. Youngstrom, D. S. Ronsaville, P. E. Martinez, P. W. Gold, R. Hakak, and M. Radke-Yarrow, "Long-term outcomes of youth who manifested the cbcl-pediatric bipolar disorder phenotype during childhood and/or adolescence," *Journal of Affective Disorders*, vol. 113, no. 3, pp. 227–235, 2009.

[130] J. Biederman, M. Monuteaux, E. Kendrick, K. Klein, and S. Faraone, "The cbcl as a screen for psychiatric comorbidity in paediatric patients with adhd," *Archives of Disease in Childhood*, vol. 90, no. 10, pp. 1010–1015, 2005.

[131] S. V. Faraone, R. R. Althoff, J. J. Hudziak, M. Monuteaux, and J. Biederman, "The cbcl predicts dsm bipolar disorder in children: a receiver operating characteristic curve analysis," *Bipolar Disorders*, vol. 7, no. 6, pp. 518–524, 2005.

[132] P. W. Leung, S. Kwong, C. Tang, T. Ho, S. Hung, C. Lee, S. Hong, C. Chiu, and W. Liu, “Test–retest reliability and criterion validity of the chinese version of cbcl, trf, and ysr,” *Journal of Child Psychology and Psychiatry*, vol. 47, no. 9, pp. 970–973, 2006.

[133] N. Bilenberg, “The child behavior checklist (cbcl) and related material: standardization and validation in danish population based and clinically based samples,” *Acta Psychiatrica Scandinavica*, vol. 100, pp. 2–52, 1999.

[134] R. F. Ferdinand, “Validity of the cbcl/ysr dsm-iv scales anxiety problems and affective problems,” *Journal of Anxiety Disorders*, vol. 22, no. 1, pp. 126–134, 2008.

[135] T. Itani, “Standardization of the japanese version of the child behavior checklist/4-18,” *Psychiatr Neurol Pediatr Jpn*, vol. 41, pp. 243–252, 2001.

[136] F. C. VERHULST, H. M. KOOT, and G. F. BERDEN, “Four-year follow-up of an epidemiological sample,” *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 29, no. 3, pp. 440 – 448, 1990.

[137] J. Biederman, C. R. Petty, H. Day, R. L. Goldin, T. Spencer, S. V. Faraone, C. B. Surman, and J. Wozniak, “Severity of the aggression/anxiety-depression/attention (aaa) cbcl profile discriminates between different levels of deficits in emotional regulation in youth with adhd,” *Journal of Developmental and Behavioral Pediatrics*, vol. 33, no. 3, pp. 236–243, 2012.

[138] C. A. Heflinger, C. G. Simpkins, and T. Combs-Orme, “Using the cbcl to determine the clinical status of children in state custody,” *Children and Youth Services Review*, vol. 22, no. 1, pp. 55–73, 2000.

[139] K. Spiel, C. Frauenberger, O. Keyes, and G. Fitzpatrick, “Agency of autistic children in technology research—a critical literature review,” *ACM Transactions on Computer-Human Interaction*, vol. 26, Nov. 2019.

[140] L. Bartoli, F. Garzotto, M. Gelsomini, L. Oliveto, and M. Valoriani, "Designing and evaluating touchless playful interaction for asd children," in *Proceedings of Conference on Interaction Design and Children*, pp. 17–26, 2014.

[141] L. Boccanfuso, E. Barney, C. Foster, Y. A. Ahn, K. Chawarska, B. Scassellati, and F. Shic, "Emotional robot to examine different play patterns and affective responses of children with and without asd," in *Proceedings of ACM/IEEE International Conference on Human-Robot Interaction*, pp. 19–26, 2016.

[142] K. Spiel, C. Frauenberger, E. Hornecker, and G. Fitzpatrick, "When empathy is not enough: Assessing the experiences of autistic children with technologies," in *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, p. 2853–2864, 2017.

[143] H. Gilbert, L. Qin, D. Li, X. Zhang, and S. J. Johnstone, "Aiding the diagnosis of ad/hd in childhood: using actigraphy and a continuous performance test to objectively quantify symptoms," *Research in Developmental Disabilities*, vol. 59, pp. 35–42, 2016.

[144] M. A. Bautista, A. Hernández-Vela, S. Escalera, L. Igual, O. Pujol, J. Moya, V. Violant, and M. T. Anguera, "A gesture recognition system for detecting behavioral patterns of adhd," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 136–147, 2015.

[145] X. Li, J. Dunn, D. Salins, G. Zhou, W. Zhou, S. M. Schüssler-Fiorenza Rose, D. Perelman, E. Colbert, R. Runge, S. Rego, *et al.*, "Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information," *PLoS Biology*, vol. 15, no. 1, p. e2001402, 2017.

[146] J. Torous, M. V. Kiang, J. Lorme, and J.-P. Onnela, "New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research," *JMIR Mental Health*, vol. 3, no. 2, p. e16, 2016.

[147] A. Sano, S. Taylor, A. W. McHill, A. J. Phillips, L. K. Barger, E. Klerman, and R. Picard, "Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study," *Journal of Medical Internet Research*, vol. 20, no. 6, p. e210, 2018.

[148] S. Place, D. Blanch-Hartigan, C. Rubin, C. Gorrostieta, C. Mead, J. Kane, B. P. Marx, J. Feast, T. Deckersbach, A. Nierenberg, *et al.*, "Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders," *Journal of Medical Internet Research*, vol. 19, no. 3, p. e75, 2017.

[149] S. Vosoughi, M. S. Goodwin, B. Washabaugh, and D. Roy, "A portable audio/video recorder for longitudinal study of child development," in *Proceedings of ACM International Conference on Multimodal Interaction*, pp. 193–200, 2012.

[150] I. Chin, M. S. Goodwin, S. Vosoughi, D. Roy, and L. R. Naigles, "Dense home-based recordings reveal typical and atypical development of tense/aspect in a child with delayed language development," *Journal of Child Language*, vol. 45, no. 1, pp. 1–34, 2018.

[151] G. Marcu, A. K. Dey, and S. Kiesler, "Parent-driven use of wearable cameras for autism support: a field study with families," in *Proceedings of ACM Conference on Ubiquitous Computing*, pp. 401–410, 2012.

[152] M. Muñoz-Organero, L. Powell, B. Heller, V. Harpin, and J. Parker, "Automatic extraction and detection of characteristic movement patterns in children with adhd based on a convolutional neural network (cnn) and acceleration images," *Sensors*, vol. 18, no. 11, p. 3924, 2018.

[153] F. Nazneen, F. A. Boujarwah, S. Sadler, A. Mogus, G. D. Abowd, and R. I. Arriaga, "Understanding the challenges and opportunities for richer descriptions of stereotypical behaviors of children with asd: a concept exploration and validation," in *Proceedings of International*

ACM SIGACCESS Conference on Computers and Accessibility, pp. 67–74, 2010.

[154] S. Mironcika, A. de Schipper, A. Brons, H. Toussaint, B. Kröse, and B. Schouten, “Smart toys design opportunities for measuring children’s fine motor skills development,” in *Proceedings of International Conference on Tangible, Embedded, and Embodied Interaction*, p. 349–356, 2018.

[155] Reconstruction Agency, “Progress to Date.”

[156] R. R. Hansel, “Kindergarten: Bringing blocks back to the kindergarten classroom,” *YC Young Children*, vol. 70, no. 1, pp. 44–51, 2015.

[157] ELAN (Version 5.9), “Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive.,” 2020.

[158] J. Cohen, “A coefficient of agreement for nominal scales,” *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[159] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *Biometrics*, pp. 159–174, 1977.

[160] N. Kato, T. Yanagawa, T. Fujiwara, and A. Morawska, “Prevalence of children’s mental health problems and the effectiveness of population-level family interventions,” *Journal of Epidemiology*, vol. 25, no. 8, pp. 507–516, 2015.

[161] A. Tomović, P. Janičić, and V. Kešelj, “n-gram-based classification and unsupervised hierarchical clustering of genome sequences,” *Computer Methods and Programs in Biomedicine*, vol. 81, no. 2, pp. 137 – 153, 2006.

[162] E. Nosakhare and R. Picard, “Probabilistic latent variable modeling for assessing behavioral influences on well-being,” in *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 2718–2726, 2019.

[163] S. Epskamp, A. O. Cramer, L. J. Waldorp, V. D. Schmittmann, and D. Borsboom, "qgraph: Network visualizations of relationships in psychometric data," *Journal of statistical software*, vol. 48, no. 1, pp. 1–18, 2012.

[164] C. Ito and T. Takahashi, "Developmental characteristics seen in the building block play of an infant [in japanese]," *The Journal for the Association of Art Education*, vol. 32, pp. 41–53, 2011.

[165] C. Richardt, C. Stoll, N. A. Dodgson, H.-P. Seidel, and C. Theobalt, "Coherent spatiotemporal filtering, upsampling and rendering of rgbz videos," in *Computer graphics forum*, vol. 31, pp. 247–256, Wiley Online Library, 2012.

[166] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 2, pp. 423–443, 2018.

[167] D. C. Niehorster, L. Li, and M. Lappe, "The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research," *i-Perception*, vol. 8, no. 3, p. 2041669517708205, 2017.

Appendix A

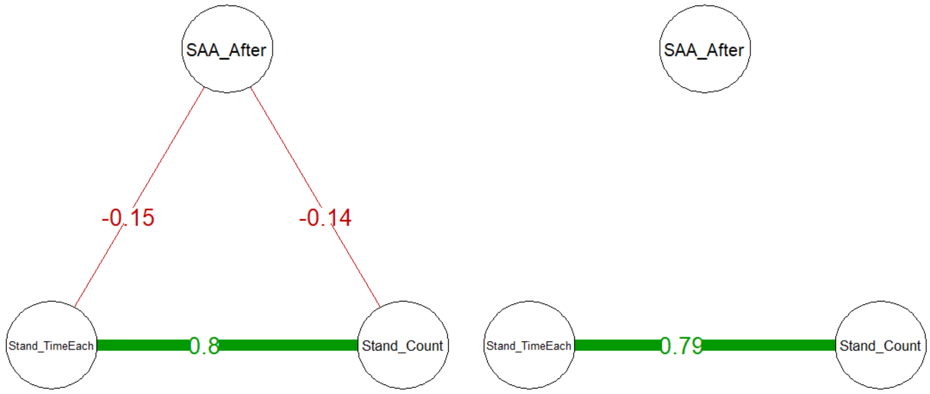
Appendix

A.1 Correlation Between Mental Problems and Block Play Features

This appendix contains Correlation Networks (CN) and Partial Correlation Networks (PCN) built between each pairs of target mental health measurements (SAA, OSBD, Total Problems, Internalizing Problems, Aggressive Behavior) and block features that we found to be important in Chapter 2 and Chapter 3, based on Spearman's rank correlation correlations and partial correlations. The data used to conduct this analysis is the same set data with what presented in Chapter 3 due to it captures variables that data presented in Chapter 2 do not contain.

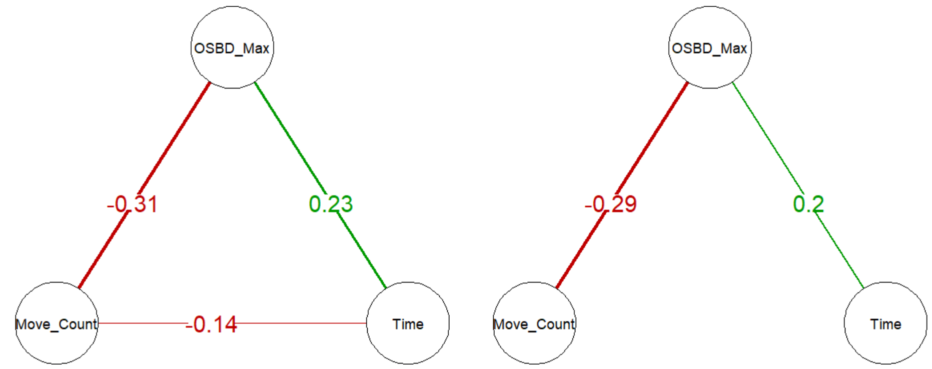
All networks are built using qgraph package available in R [163]. Due to the weak correlations found between target mental health measurements and block features, the edges in all networks reflect correlations with a value greater than 0.1. The width of the edge represents the strength of association between two correlated variables and the number on the edge denotes to the value of correlation.

A.1.1 Correlations and Partial Correlations



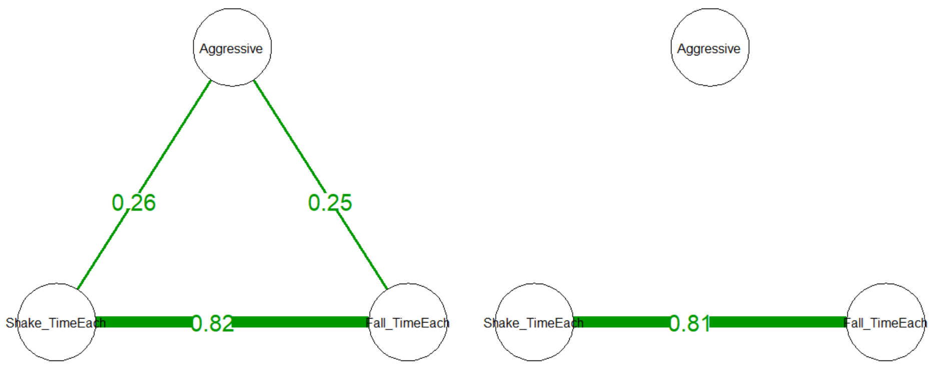
(a) CN **(b)** PCN

Figure A.1: Correlations of sAA and Block Features



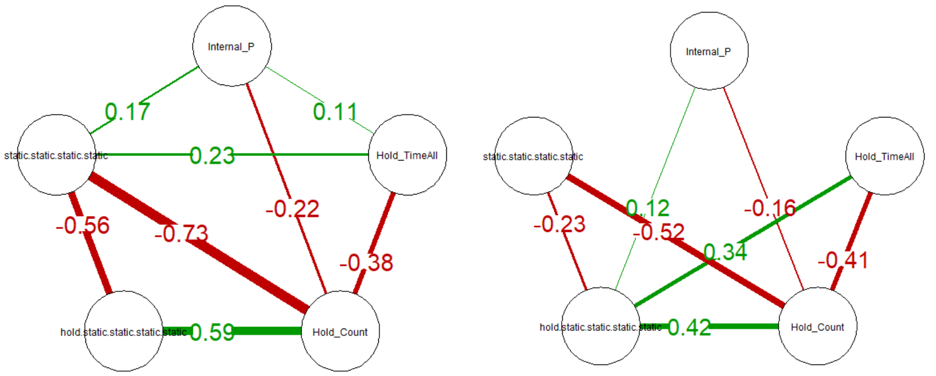
(a) CN **(b)** PCN

Figure A.2: Correlations of OSBD and Block Features



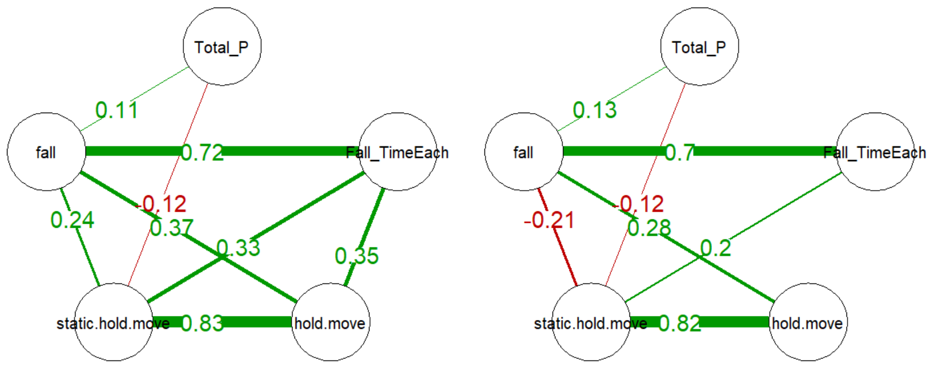
(a) CN **(b)** PCN

Figure A.3: Correlations of Aggressive Behavior and Block Features



(a) CN **(b)** PCN

Figure A.4: Correlations of Internalizing Problems and Block Features



(a) CN **(b)** PCN

Figure A.5: Correlations of Total Problems and Block Features

List of Figures

List of Tables