Quantifying the Coding Playground: A Pilot Study Creating and Attempting to Validate a Rubric for Positive Technological Development

A thesis submitted by

Tess Levinson

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Child Study and Human Development

Tufts University

August 2021

Advisor: Marina Bers

Abstract

When interacting with technology, children can engage in positive behaviors: creativity, collaboration, content creation, communication, community building, and choices of conduct. These behaviors form a framework known as Positive Technological Development (PTD). Earlier measures for quantitatively assessing PTD behaviors were difficult for teachers to administer to multiple students simultaneously and did not produce consistent results across assessors. This thesis used a design-based research framework to develop and pilot validation of the PTD Rubric, an observational rubric of PTD behaviors. Both teachers and researchers were able to administer the rubric in classrooms to multiple students simultaneously. Neither interrater nor stability reliability were strong, but construct validity was high for some items and moderate for others. This finding suggests PTD behaviors of multiple students can be assessed simultaneously using an observational rubric. PTD likely varies in relation to context. Future iterations of the PTD Rubric will have to refine the item anchors to improve validity and reliability of results.

Keywords: Coding, robotics, educational technology, early childhood, assessment, positive technological development

Acknowledgements

I am so grateful for the support of mentors, family, and friends as I have completed this master's degree and will move on to my PhD. To begin, I would like to express my gratitude for my committee. My mentor and advisor, Dr. Marina Bers, has encouraged me throughout this process and given me the support and confidence to explore new projects, take on new endeavors, and explore new research questions. I am so grateful for her support and look forward to continuing to learn from her in the years ahead.

I would also like to thank my second advisor, Dr. Rich Lerner, who has taken time to sit with me and discuss methods, theory, and more. I have learned so much from working with him, and I am grateful for all that he has taught me. Finally, I would like to thank my third committee member, Dr. Chris Rogers, for graciously agreeing to be part of this process.

Many members of the DevTech research group supported me through this process, as colleagues, mentors, and friends. Dr. Amanda Strawhacker has given me so much guidance, both in the specifics of Design-Based Research methodology and the general processes of applying for conferences, and the nuances of navigating a Doctoral degree. Dr. Emily Relkin has been an invaluable resource and mentor as I have explored measurement and validation, in addition to supporting me through the trials of data collection and thesis writing. Aim Unahalekhaka welcomed me to DevTech and Boston, and our own active lab meetings running on the Charles were the best sort of sounding board for this and other projects. This thesis would not have been possible without earlier work led by Dr. Madhu Govind in St. Louis, MO. Madhu, thank you for teaching me how to work with teachers and administrators, build KIBO curriculum, and run an in-person research program during a pandemic year. Finally, I need to give a special thank you to Jessica Blake-West, who has shaped my time at DevTech. Thank you for all that you have done to support this project – attending observations and assessments with me at the last minute, helping me submit my expense reports, encouraging me when I thought I would never finish, and being my coffee buddy when I needed a break. I'm so excited to continue to share an office with you next year!

I also need to thank everyone who assisted with the pilot study that was my thesis. Many students at DevTech assisted with data collection including Aim Unahalekhaka, Sam Schwamm, Aliza Bromberg Gaber, Jessica Blake-West, and Madeline Nievera. From Horizons for Homeless Children, this study would not have been possible without the work of Amanda Martinez, Jayd Rodrigues, Sheila O'Neil, and of the teachers and students who took part in the CAL-KIBO evaluation project.

Contents

List of Tables

List of Figures

Quantifying the Coding Playground: A Pilot Study Creating and Attempting to Validate a Rubric for Positive Technological Development

At this writing, the past five years have involved an increasing emphasis in the United States on computer science education for students in Grades K-12 (Code.org et al., 2021). This emphasis has included initiatives such as Code.org's Hour of Code, the K-12 Computer Science Framework, and the International Standards for Technology in Education (Code.org, 2021; *ISTE-S*, 2016; *K–12 Computer Science Framework*, 2016).

This education often takes the form of formal and structured pedagogies referred to as instructionist (Bers, 2020a; Papert, 1980). Seymour Papert describes instructionism as a theory of teaching, or instruction (Papert, 1980). Instructionist pedagogies are teacher centered and often involve presenting a piece of knowledge or skill according to a standard, followed by an assessment of whether the child has gained said knowledge or skill. Instructionist pedagogies are not unique to computer science. This theoretical model of teaching and learning is often prioritized in the United States because it provides the ability to assess learning and aligns with a national emphasis on standardized testing scores following the implementation of the No Child Left Behind Act in 2001 (Pederson, 2007).

The learning objective of an instructionist pedagogy is that a student learns a specific piece of information, with there being one possible correct answer to a proposed question or one correct expected learning outcome (Papert, 1980). Instructionist pedagogies have historically been aligned with cognitive and behaviorist understandings of knowledge; both instructionism and cognitive and behavioral sciences define learning as the acquisition of and ability to reproduce information (Ertmer & Newby, 1993). Researchers in these fields, to assess and measure student learning, have developed multiple measurements of students' abilities, including but not limited to observational assessments, standardized assessments, and in-classroom assessments of student learning (Dixson & Worrell, 2016).

In contrast, constructionist contexts and pedagogies are student-centered. The student's questions and interests guiding the direction of learning, and the teachers support students in this exploration (Papert, 1980; Resnick & Rusk, 2020). In technology contexts, learners complete self-directed projects using technology tools. The technology tools (such as a coding language or engineering materials) exist as tools facilitating the process, so the emphasis is on the creation process rather than the learning of a specific tool. The learning comes alongside the use of the tool, and children learn new skills and ways to use the tool as becomes necessary for their creative process. For example, Papert described children using LOGO, an original programming language for young children, as a tool through which children could explore mathematics, geometry, and art (Papert, 1980). Although the children were learning the language of LOGO and were able to create programs demonstrating a knowledge of geometry, the emphasis remained on the children's artistic creation rather than the specific math concepts being learned.

Papert referred to scaffolded learning contexts such as LOGO as microworlds, structured environments and realities in which children could learn through free exploration. Bers expanded on this concept of microworlds with "coding playgrounds" (Bers, 2020a). Coding playgrounds are microworlds centered on the powerful ideas of coding and robotics as compared to math and geometry. As with the original microworlds developed and described by Papert, coding playgrounds emphasize play, creativity, and exploration alongside computational learning.

Outcomes from constructionist pedagogies are difficult to assess in the classroom environment. Unlike with an instructionist pedagogy, where the objective of student learning is to learn standardized information that can be assessed with the correct answering of questions, a constructionist pedagogy proposes a student learns new skills as the student encounters them, with the pedagogy focusing on the self-directed process of creating artifacts. Whereas one student may use Repeat Loops to create a long, repeating animation, another student may investigate sensors to create an interactive doll. Rubrics have been created to evaluate complexity of students' creations, but these only measure a portion of the child's abilities, as a child may not choose to use every element they are capable of in each project (Govind & Bers, 2021; Unahalekhaka & Govind, 2021). No validated measures currently exist to evaluate the constructionist learning process.

Positive Technological Development

The Coding as a Playground pedagogy takes a constructionist approach to computer science learning in early childhood (Bers, 2020, 2018). Like other constructionist pedagogies, this approach is student centered, with skill-learning driven by learners' individual questions and interests. This approach to learning coding has also organically emerged in many online communities such as the communities around game-modification or fandom web design, where the existence of a creative community leads individuals to learn to code in order to create or co-create content for that community's enjoyment (Dym et al., 2021). This pedagogical model can also be explicitly implemented in learning communities with developmentally-appropriate, open-ended coding platforms or technological environments. Some examples of these learning environments include Scratch, a block-based coding language for children to create art, animations, and games; ScratchJr, a coding platform for young children; KIBO, a robotics tool with a tangible programming language consisting of wooden blocks; and maker spaces (Bers, 2018; Bers et al., 2019; Maloney et al., 2010; Strawhacker & Bers, 2018).

Bers describes the coding playground as mimicking the physical playground, particularly with regard to the open-ended play-based learning students engage in, the role of adults in scaffolding and supporting learning without direct instruction, and the multiple opportunities for engagement students from which students can choose (Bers, 2012; Bers et al., 2009). In the coding playground, learners take part in the Six Cs, six behaviors that also appear on the physical playground: *Communication, Collaboration, Curiosity, Creativity, Community Building, and Content Creation* (Bers et al., 2009). Within the context of technology environments, these behaviors compose a construct known as Positive Technological Development (PTD) (Bers, 2010; Strawhacker & Bers, 2018).

Bers coined the construct of PTD by examining how the Positive Youth Development (PYD) framework (Lerner et al., 2015, 2021) applies to experiences youth have with technologies and technological environments (Bers, 2012; Bers et al., 2009). PYD is a framework examining factors that lead to the development of thriving among youth. A PYD framework was developed in the 1980s and 1990s by social entrepreneur Rick Little; the framework focused on the presence of positive, as compared to the absence of negative, attributes (Lerner et al., 2021). This initial PYD framework centered around four strengths that adolescents might possess and that might be developed in adolescence: Competence, Confidence, Connection, and Character. Lerner and Lerner later added a fifth C to the model of PYD: Caring. Lerner and Lerner (e.g., see Lerner et al., 2015) argued that when the Five Cs of PYD were well developed, a youth would develop a sixth C, that is, and Contribution to self, family, community, and ultimately civil society (Lerner et al., 2015, 2021).

PYD has been measured in different ways (e.g., see Geldhof et al., 2014a, 2014b). One measure used to assess the PYD is the GPS to Success Rubric (Bowers et al., 2013; Napolitano et al., 2014). This tool is a self-report rubric for adolescents and consists of five items for each PYD construct. Each item is scored on a scale from one to five. The rubric is often paired with a mentor rubric, completed by an adult in a mentorship position with an adolescent, which is used to evaluate the mentor-mentee dyad relationship as well as to better understand the adolescent's PYD.

Another option for assessing PYD is using survey items to assess the Five Cs. In the 4-H Study of PYD conducted by Lerner and Lerner (e.g., Lerner et al., 2015), measurements of the Five Cs of PYD (Confidence, Competence, Connection, Character, and Caring) were measured (e.g., Geldhof et al., 2014a, 2014b; Bowers et al., 2010). These measures were self-report instruments.

The groundwork for PTD was initially proposed by Bers in 2006 as a model of how PYD could present in behaviors of children engaging with technologies (Bers, 2006). Bers's work at the time focused on identity construction environments, or technology platforms and environments that provided space for children to engage in identity development, community building, and narrative expression. Identity construction environments provided opportunities for children to engage in the 6 C's of PYD through specific behaviors, interactions, and activities in the technology environment. The framework for PTD was proposed in 2010 (Bers, 2010). According to the framework of PTD, the assets of PYD coact with technology-supported behaviors of PTD. This coaction, over the course of personal development, leads to self-improvement and contribution to society.

Initially, the PTD framework focused on the application of PYD constructs for adolescents within technology contexts (Bers et al., 2009). As examples, the PYD construct of competence appeared in a technology context when adolescents recreated familiar buildings in a digital world, and the PYD construct of connection appeared in a technology context when adolescents created projects alongside peers with similar interests. However, a revised model of PTD, introduced in 2012, focused on six behaviors arising from the presence of these constructs in technology contexts (Bers, 2012). It was at this time that Bers proposed the 6 Cs of PTD: Communication, Collaboration, Content Creation, Creativity, Community Building, and Choices of Conduct.

Communication

When writing code or building a robot, children need to be able to use their developing language skills to effectively ask for help, share a story with a friend, or work alongside a peer. As such, in the case of PTD and the coding playground, the construct of communication refers primarily to the domain of language pragmatics, or the use of language with respect to context and situation. This ability to use language to communicate, both expressively and receptively, develops throughout the early childhood years (Becker Bryant, 2009). By the pre-school years, children can take turns when conversations consist of two people but have difficulty with conversational turn taking in groups of more than two (Becker Bryant, 2009; Casillas et al., 2016).

In the coding playground, children communicate with both children and adults, including making requests, jointly working, and asking for help. Like older adolescents and adults, children communicate differently with peers than they do with adults, for example, using more formal and indirect language when making requests of adults than of peers (Becker Bryant, 2009; Ervin-Tripp, 1982).

Collaboration

In the coding playground, collaboration takes the form of the sharing of materials, ideas, and knowledge. Children might plan a project together, discussing their vision for the project and the story it will tell. Children can also share roles, taking turns physically programing their project, adding decorative details, or scanning KIBO blocks. Children also collaborate by helping each other, asking for and offering help in debugging a program or with the physical act of scanning. In early childhood, children are more willing to share ideas than they are to share physical resources, and children at that age expect to receive equal outcomes as others when working together (Slocombe & Seed, 2019). Although outcomes from coding projects are not tangible, children might still expect to receive equal recognition or equal time demonstrating the project.

The ability to collaborate with others, working together towards a shared goal, develops across early childhood. Whereas children as young as one year old can take part in joint attention, more advanced cooperative working behaviors develop as children age (Slocombe & Seed, 2019). By the age of three, most children can switch roles when working in a group, and by the ages of five or six, children are able to coordinate shared work by communicating. The development of collaborative and cooperative behavior is possibly related to moral development, with developing understandings of helping and fairness playing a role in children's behavioral decision making (Damon, 1990; Olson & Spelke, 2008).

The role of collaboration in creating a positive technological environment has been seen not only in early childhood, but across all ages. In a study examining the negative effects of playing violent video games on trust and cooperative behaviors, playing with a partner was found to improve trust and cooperative behaviors with new partners (Greitemeyer et al., 2012).

This finding suggests that when collaboration is added to a technological environment, the environment can lead to the development of positive behaviors.

Creativity

As will be noted again, in the fields of education and child development, the developmental construct of creativity is often considered to belong within the domain of arts education. However, with the reimagination of STEM (Science, Technology, Education, and Math) to STEAM (Science, Technology, Education, Arts, and Math), the role of creativity in science and technology education is becoming increasingly discussed and prioritized (Milgrom-Elcott, 2019). Both within the arts and sciences, creativity includes the production of or use of ideas, tools, or products in novel or unusual ways (Zimmerman, 2009). In the coding playground, creativity is seen in children's generation of ideas, as well as in the use of tools or materials. A child may try a new method of scanning the KIBO robotics programming blocks to improve scanning outcomes or may rearrange programming blocks in new ways to create a desired program. Children also show creativity in their created projects, as they plan and create projects that allow them to play games or tell stories in a new medium (Resnick et al., 1988).

Content Creation

Children's creation of content is a theme currently explored and emphasized across multiple domains of education, including art, writing, and more recently STEAM, as mentioned above (Garvis & Twigg, 2010; Milgrom-Elcott, 2019). In the coding playground, learning is centered around project and program creation. Children create personally meaningful projects and learn necessary skills as part of the process. The role of content creation on children's development has been most significantly studied within the context of arts education. Within this context, the process of content creation is emphasized over the resulting content itself. When creating content, children can experience success in their learning activities that can lead to increases in motivation or confidence (Alvino, 2000). When children are intending to create a project with one correct answer, that success can only be experienced in one manner. However, children's feelings about created content vary depending on the content, with children proud of and eager to display some items and less proud of and willing to display others (Garvis & Twigg, 2010).

Choices of Conduct

Positive behavioral choice making is related to several domains, including prosocial behavior, self-regulation, and moral decision making (Damon, 1990; Eisenberg et al., 2015; Garrigan et al., 2018; McClelland et al., 2015). Moral decision making is itself related to children's cognitive, affective, and social development (Garrigan et al., 2018). In the coding playground, children make behavioral choices regarding the handling of tools, response to frustration, and collaboration with peers. These skills develop throughout the early childhood and later years, but are influenced by classroom factors such as teacher behavior and social emotional learning curricula (Spivak & Farran, 2012; Viglas & Perlman, 2018). As on the physical playground, these factors can be more or less present in the coding playground dependent on the scaffolding behaviors of the present adults. In addition, children can also scaffold each other's prosocial behavior with feedback and responses (Eisenberg et al., 1981). In the coding playground, opportunities to collaborate and communicate provide children with opportunities to provide each other positive feedback on behavioral choices.

Community Building

As mentioned above, constructionist, playground-like coding environments have codeveloped alongside online communities in a variety of contexts (Dym et al., 2021). For young children, creating a coding community allows children to use coding as a bridge to connect with their peers and larger community (Bers, 2020a). This connectivity can be seen at a local scale, as children in a classroom work together on personally meaningful projects, or on a global scale, as children post and remix projects on the online Scratch Community (Resnick & Rusk, 2020).

In one robotics curricula, kindergarten classrooms from religious and secular schools in the United States and Argentina were prompted to create projects celebrating the “treasures” of their community (Hunt & Bers, 2021). With these projects, students celebrated the unique aspects of their community, including language, values, and culture. In addition to discussing and celebrating community, this program allowed for the inclusion of students with disabilities and provided a means by which children included their peers (Levinson et al., 2021). In addition, children who were actively engaged in the online Scratch Community referred to participation in the community and community involved behaviors as furthering their development both as coders and as community members (Roque & Rusk, 2019). These children understood the role of the online coding community as having played a part in the development of behaviors such as giving and receiving feedback, reaching out to others to invite collaboration, and taking on leadership roles within the community.

Assessment of Positive Technological Development

Previously, Positive Technological Development behaviors have been assessed using the Positive Technological Development Engagement Checklist (DevTech Research Group, n.d.; Strawhacker & Bers, 2018). This assessment consisted of two checklists: the Checklist for Children, to assess PTD behaviors demonstrated by children, and the Checklist for Environments/Facilitators, to assess classroom factors that would promote PTD behaviors. Each checklist consisted of three items for each C. Each item was scored on a Likert scale of 1 to 5 or by using the response option of not applicable. This checklist was entirely researcher administrated.

However, this assessment had difficulties in administration. Although researchers completed trainings to develop consistent scoring, there was still low interrater reliability on the assessment (Levinson, 2021b). In addition, some items were difficult to score on a 1 to 5 Likert scale. For example, items such as “children place work on display somewhere in the space” were binary, either present or absent with little area between, and did not lend themselves well to the Likert scale measure. For other items, the “not applicable” answer sometimes led to confusion. For example, when scoring the “children work on projects related to their local environment” item, it was unclear if children working on projects related to a fictional storybook would be given a score of 1 (because the behavior was not present) or scored as not applicable (because the behavior was not relevant to the assignment). For this reason, researcher responses were particularly inconsistent on the lower end of the Likert scale.

In addition, because the assessment was entirely researcher administered, it was difficult to assess the behavior of multiple children within each classroom or setting. For this reason, when the assessment was used in larger studies, such as the Coding as Another Language study in Norfolk, VA, only the classroom checklists were administered (Bers, 2020b).

Pilot Study: Longform PTD Checklist

In the initial pilot of the revised assessment of PTD behaviors (Levinson, 2021b), I developed a longform checklist from the original PTD checklist. These changes were made to improve usability and reliability. To determine what observational assessments were easy for teachers to use, I spoke with teachers and research partners from sites such as Head Start (Levinson, 2021b). I was told that three anchor assessments (Yes/Somewhat/No) checklists were preferable to multi-point Likert scales, and assessments with items using clearly defined behaviors were preferred to items describing behavioral constructs (Levinson, 2021b). In this revised longform checklist, items were rewritten to reflect single behaviors, with multiple behaviors per behavioral construct. In addition, item scoring was altered to be on a three-point scale (Present/In Progress/Absent). These assessments have a low number of anchors, thus reducing potential confusion, but also provide the ability to differentiate whether the student has and uses the skill, is working on it but only inconsistently uses the skill, or is not using the skill at all. However, by reducing the anchors to reflect if behaviors were present, in progress, or absent, there was a risk of losing nuance in measurement of the behaviors.

To create the revised longform PTD checklist, I divided the items within the original PTD checklist into specific behaviors. Items containing multiple behaviors were transformed into multiple items. For example, the item *“Children are observing and/or engaging each other’s work (Children watch as others work on a project, Children express themselves through their projects, Children touch or play with each other’s projects while they work)”* was transformed into the following three items: 1. “Child watches others work;” 2. “Child touches or plays with other child's project while working;” and 3. “Child expresses self through project.” I also removed the “not applicable” answer choice, including in the instructions that if a behavior is not present, it is absent, regardless of perceived applicability.

I went through and clarified which items belonged to the environment or teacher checklist as compared to the student checklist, ensuring that each checklist referred either to behaviors exhibited by a child, or teacher behaviors and factors that impact a child. Items in the student checklist that referred to teacher behavior were moved to the teacher checklist, and items referring to student behavior were moved to the student checklist. Duplicate items were removed.

I then compared the items on the checklist with the definitions of the constructs. Items that more closely resembled definitions of constructs other than the construct they were currently located in were moved. In the example mentioned above, the original PTD checklist item was in the Communication domain. In the longform checklist, I kept the first two items ("Child watches others work" and "Child touches or plays with other child's project while working") in the Communication domain, but I moved the third item (""Child expresses self through project") to the Content Creation domain. Finally, I revised the wording of items that described classroom behavior but did not include mention of technology to include language referencing technology learning environments.

I administered this checklist to one of our research collaborators, an early childhood administrator for a Head Start program, to receive feedback regarding the changes made above as well as to receive any additional suggestions from her. She suggested some rephrasing, and indicated which items were clear or unclear to her as an individual who regularly administers similar behavioral assessments.

I then completed a pre-pilot of the checklist in a classroom at the Eliot-Pearson Children's School in Medford, MA (Levinson, 2021a). This pre-pilot was administered virtually due to the COVID 19 pandemic. In this pre-pilot, I determined that the assessment was difficult to administer over video and would need to be administered in person. Moreover, the assessment took time to administer, and it was still only possible to administer to one child at a time. This procedure would lead to difficulty scaling the assessment, and difficulty administering the assessment if researchers would not be allowed to visit sites due to the pandemic or for other reasons.

Design Considerations for a Revised PTD Assessment:

The goal of this thesis was to create a revised assessment for Positive Technological Development behaviors that could be easily administered by both teachers and researchers. Based on the earlier work described above, as well as prior research in the field, the following design considerations for such an assessment tool were identified:

1: Speed in administration. The average preschool classroom has approximately 17 students in it, whereas the average early elementary classroom has approximately 21 students in it (Bowne et al., 2017; *Public School Teacher Data File*, 2017). These numbers mean that even an assessment that requires five minutes per student could take an hour for a teacher to complete for an entire classroom. Teachers have limited time, and lack of time for non-mandatory things has been found to be a limiting factor for teachers integrating new technology materials into their classrooms (Chen, 2008). Limited time affects many aspects of teachers' days, including a lack of uninterrupted or undesignated time in which teachers can plan, grade, or answer emails (Collinson & Fedoruk Cook, 2001). Teachers completing an assessment need to be able to complete the assessment while teaching or during a brief uninterrupted block immediately following teaching, so my goal in designing this assessment was to design a usable assessment for both research and classroom assessment of PTD behaviors that would be used to assess the behaviors of children in the entire classroom. For this assessment to be usable, a teacher would need to be able to complete it in the time available. As such, I aimed to design a rubric that could be completed in between one and five minutes per child.

2. Easy comprehension with minimal training. As mentioned above, teachers have limited time, and this issue includes limited time for professional development (Collinson & Fedoruk Cook, 2001). This situation would prevent teachers from taking time to complete a complex training for non-mandatory assessments (Collinson & Fedoruk Cook, 2001). As such, to create an assessment that would be viable for use in the classroom, I wanted to create an assessment that was easy to understand and administer with minimal training so that a professional development session would not be required for teachers to use the material. Teachers often seek out easy-to-use resources, for all domains including computational thinking instruction, on websites such as Teachers Pay Teachers, where reviews highlight ease of use, functionality, or ability of the resource to solve instructional problems (Xu et al., 2022). However, these computational thinking resources are not research based, and Xu et al. found that 30% of these materials were harmful to student learning, including with factual errors or misleading information. I aimed to create a free and research-based resource that met the functionality and ease-of-use desires of teachers while serving as an accurate and valid tool for assessment and research. However, reducing training increases risk of inconsistency in results if teachers do not understand materials. To improve comprehension with minimal training, the assessment would include all necessary information in the rubric itself, and any instruction materials would be short and easy to understand.

3. Specificity to the coding or technology environment. Many assessments already exist evaluating early childhood behaviors or development, such as the Teaching Strategies GOLD which is used in many Head Start and preschool classrooms nationally (Lambert et al., 2015). These assessments evaluate student behavior or development as it occurs in the classroom, but do not measure specifically behaviors of students engaging with technology.

Whereas the Teaching Strategies GOLD may examine how children communicate with their peers, the communication construct of PTD is specifically related to how children communicate when working in a technological environment such as with the KIBO Robot or in a Makerspace (Bers, 2012; Strawhacker & Bers, 2018). As such, the goal of the PTD Rubric was not to replicate existing rubrics assessing social emotional, language, or approaches to learning development, but to create a rubric specifically to examine PTD behaviors. To do so, the PTD Rubric needed to be specifically focused on how students behave in the technology settings.

4. Developmentally appropriate for early childhood settings. PTD behaviors exist for individuals at all ages, but these behaviors may look different dependent on the age and developmental level of the individual. For example, community building and content creation among young adults in online fandom communities emphasized micro communities centered around niche interests with high proportions of women, racial, and gender minorities creating WordPress themes or websites to host art and digitally created creative content (Dym et al., 2021). Community building and content creation for young children might be seen as children creating a project about their neighborhood (Sullivan & Bers, 2016). For researchers and teachers to identify and score the PTD behavior observed on a labeled anchor chart, the anchors must be developed to be developmentally appropriate to the PTD expectations of that age group.

5. Reliable results across assessors. This assessment was designed to provide reliable results for classroom and research use. For this purpose, the assessment should provide reliable results across assessors, both researchers and classroom teachers.

6. Universal design. The early childhood classroom is inclusive of children with multiple types of disabilities of developmental delays. Over half of children with disabilities spend over 80% of their time in the general education classroom, and over 80% of children spend over 40%

of their time in the general education classroom (National Center for Education Statistics, 2021). As such, general education within early childhood and early elementary classrooms often includes children with disabilities such as autism spectrum disorder, speech and language impairments, orthopedic impairments, or developmental delays (Congress, 1975). This fact means that an assessment of PTD behaviors for the early childhood classrooms needs to also be usable for children with disabilities in the same setting. Although children with disabilities will not be completing the assessment themselves, the principals of Universal Design for Learning state that the items and anchors should be written in a way that children can meet the standard in multiple ways such that their disability is not the limiting factor (Capp, 2017). For example, a child with a speech delay may not speak fluently, but may be able to demonstrate collaboration with their peers in the technology environment through active listening, responding, sign language, facial expressions, or augmentative and alternative communication methods such as text or images (Iacono et al., 2022; Lieber et al., 2008).

7. Linguistically and culturally unbiased. Only 47% of children in the United States are non-Hispanic, white but the majority of classroom assessments are designed for and normed to non-Hispanic white children (National Center for Education Statistics, 2021). When children in linguistically, racially, or ethnically minoritized communities are assessed using these measures, there may be item bias that leads to a misrepresentative underperformance, caused by differences between their cultural and linguistic experiences and those expected of children taking the assessment by the assessment designer (Marotta et al., 2022). This bias can also be present in early childhood computer science and computational thinking instruction, where the non-digital metaphors used to assess computational thinking knowledge can show bias just as in any other field (Harper et al., 2022). The CRRAFT project found that working in community partnerships allowed for the design of culturally responsive STEAM pedagogies and programs, including a model of culturally responsive computational thinking (Caudle et al., 2021; Harper et al., 2022).

Present Study

My primary goal in this thesis was to create an assessment of early childhood PTD behaviors that could be administered simultaneously to multiple children in the classroom setting. The research questions for establishing the reliability and validity of the measure were: 1. Do assessors agree on item scores when administering the PTD rubric (reflecting interrater reliability)?; 2. Do item and overall scores on the PTD rubric vary reliably across time within children (reflecting stability reliability)?; 3. Do item scores on the PTD rubric relate to each other of PTD (reflecting internal consistency)?; and 4. Do item and overall scores correlate with language and social emotional items on the Teaching Strategies GOLD, which measures developmental behaviors for preschool children (reflecting construct validity)?

Method

For this study, I used a design research method to develop and attempt to validate a rubric of PTD behaviors. Design-based research is a method of educational resource development and research that accommodates both rigorous research methods and the goal of providing tools and resources to teachers as they are being developed (Brown, 1992; Sandoval, 2014). This method allows teachers to access new tools and resources immediately rather than waiting until after the tool has been examined in a controlled laboratory setting or within a randomized control trial. Design-based research is often used as a method for studying children's learning alongside the development of curriculum or classroom tools, but this process has also been used to create and understand assessments of student knowledge and learning (de Ruiter & Bers, 2021; Dede et al., 2004; van Aalst & Chan, 2007).

I selected design-based research because it allows for the development of a tool that can be used in both research and practice settings (Brown, 1992). The purpose of this project was to create an assessment that produced valid responses (as operationalized by convergence with scores from other measures of positive behavior that are theoretically linked to PTD) and, as well, a reliable measure that provided information that may spur further research about the measure. I developed a conjecture map for this project (Figure 1), as described by Sandoval for design-based research, to conceptualize the conjectures of the project (Sandoval, 2014)

Measure Development

To meet the design goals described above, two observational rubrics were developed using a design-based methodology. Observational rubrics were selected as the means of assessment for a few reasons. To create an assessment that could be scaled for administration to multiple children, the measure was designed to be easily completed by classroom teachers. Rubrics are frequently used by teachers in the early childhood setting for other assessments, so early childhood teachers are familiar with completing these rubrics to measure behaviors (Lambert et al., 2014). Similarly, rubrics are designed with specific anchors. This design may allow for more precision in answers and may remove features that were unclear in the old version of the PTD checklist. For example, in the old rubric, items could be rated as either not applicable or as “0” if the behavior was not present, but there was not a clear determination as to in what context the behavior would be rated as “0” or as “not applicable.” In the new rubric, the behavior is specifically described, regardless of context. If the anchor were to describe the behavior as “Does not take part in group activities; Only works alone without talking to others,” the rating would be the same for a student who is disengaged from their class’s group activity, and a student in a classroom where all students are working silently alone.

In addition, observational rubrics can be completed by the teacher during a preparation or grading period, meaning that the completion of the assessment does not take away from instructional time. Furthermore, this feature also allows for the teachers to complete assessments for multiple children without needing to individually pull or distract children from their classroom work. Finally, teacher observational rubrics can provide a more holistic measure of behavior for younger children than other assessments, as they are completed by a teacher who knows the child in the child's natural environment (Maxwell & Queensland School Curriculum Council, 2001).

In revising the PTD Checklists, I developed two rubrics so that teachers would be able to assess student behavior in a measure separate from an assessment measuring the teacher behavior and classroom environment. The Positive Technological Development (PTD) Rubric was specifically designed to measure student behavior, whereas the Positive Technological Environment (PTE) Rubric was designed to measure PTD-promoting factors in the learning environment and remains researcher completed. This thesis focused only on the development and validation of the PTD Rubric. However, both rubrics are included in the appendices of this thesis.

The CAL KIBO PreK Study

The reliability and validity testing of the PTD Rubric were conducted as a subset of a larger study evaluating a tangible robotics curriculum in a preschool center. This larger study took place over the course of a school year and consisted of multiple measures of teacher and student performance. Only measures relevant to the reliability and validity testing are described within this thesis.

Study Site

The CAL KIBO PreK study took place at Horizons for Homeless Children in Boston, MA, a preschool site serving children experiencing homelessness. Six classrooms of children between the ages of three and five years old took part in the research program. Each classroom consisted of 18 children and three teachers, including one bilingual teacher in English and Spanish. Conducting the design and validation processes within a teaching setting allowed for the opportunity to design a measure that was not created and normed for white, middle-class children in the same way as are many educational assessments (D'ignazio & Klein, 2020; Gould & Gold, 1996). Although this setting did not allow for comparative analyses between white middle-class children and the minoritized or low-SES children with whom the measure was developed, Hall et al. (2016) posits that including children of the dominant culture for a comparative analysis is not necessary for all research, unless that research is operating under the assumption that white middle-class children reflect the standard of development (Hall et al., 2016). Although the pilot findings will not necessarily generalize to other children, future research can include children in other preschool settings in the use of this tool.

Three of the classrooms at Horizons for Homeless Children are part of the Boston Universal Pre-Kindergarten (UPK) program (Gray-Lobe et al., 2021). This program serves Boston children between the ages of four and five in the year before they enter kindergarten and is highly structured and evidence-based, having been evaluated positively in a randomized controlled trial (Gray-Lobe et al., 2021). These classrooms use the Creative Curriculum for daily instruction and, although they are part of the Horizons for Homeless Children and administered through Horizons for Homeless Children, they are also partially administered with oversight through Boston Public Schools (*Research Foundation: The Creative Curriculum®*, 2010).

Protocol

I embedded the PTD Rubric Pilot Validation within the evaluation study of the Coding as Another Language (CAL) KIBO PreK curriculum, a tangible robotics curriculum using the KIBO robot for children between the ages of three and five (Bers, 2019). The curriculum consists of 30, 30-minute lessons. Lessons are structured based on the Creative Curriculum and include whole group activities, small group or center activities, songs, games, and opportunities for creative expression (*Research Foundation: The Creative Curriculum*®, 2010).

Participants

Two classrooms (Room A and Room B) were selected for the PTD Rubric reliability and validity testing due to both the KIBO schedule and the teachers' willingness to take part in an additional element of the CAL-KIBO PreK research project. The two classrooms selected were both part of the structured UPK program (Gray-Lobe et al., 2021). As such, the two classrooms had similar age students, had teachers with similar certifications, and used the same daily curriculum (Lambert et al., 2014; *Research Foundation: The Creative Curriculum*®, 2010). Both classrooms used the CAL-KIBO PreK curriculum for robotics instruction.

Teachers in each of the two classrooms selected five children for participation in the PTD Rubric observations. In each classroom, teachers prioritized selecting students who were likely to be in attendance for KIBO observations, but attendance in the observed KIBO sessions varied (Table 1).

Location

CAL-KIBO PreK lessons were taught either in the classroom or during STEM time, when classes would visit Horizons for Homeless Children's STEM room. Lessons in the classroom were taught at a central carpet used for other group lessons including Morning Meetings and whole-class read-aloud activities (Figure 3). For the initial part of the lesson, children sat in a circle around the edge of the carpet in positions that were either pre-assigned or assigned the day of the lesson. During the lesson, children would move locations as needed to relocate for small-group work, for gross-motor activities (such as dancing or games), or for individual interactions with the KIBO robot. Potential distractors were available for children at a reachable height around the edge of the carpet, including calming materials, toys, and books. Children were not restricted to the carpet area and could leave to go to a calming area of the classroom or to request other centers, although they were encouraged to stay at the KIBO lesson.

The STEM room featured a large carpet and two tables (Figure 4). Whole group activities were taught at the central carpet with children encouraged to sit in a circle, although seats were not always assigned, and children were able to rotate seats. Small group activities were taught either on the central carpet or at child-sized tables. Children were encouraged to remain in the central lesson area but were allowed to leave the lesson to engage in other STEM activities in the room. The materials for these other STEM activities were on the opposite side of the room from the KIBO lesson and not in the line of sight for children but were at children's height for access. If children left the KIBO lesson but were not engaging in other STEM activities, children were prompted to return to the KIBO lesson. There was no calm-down area in this room, so if children needed a calming area, they were taken on a walk by an adult and left the room.

Data Collection

I identified five lessons in the CAL KIBO PreK curriculum (5, 12, 13, 14, and 19) as having small group activities with opportunities for PTD behaviors. In Lesson 5, the children explore the hardware of the KIBO robot and assemble the robot. In Lesson 12, the children write a KIBO program as a whole class and play an unplugged computational thinking game (Bell et al., 1998; Brackmann et al., 2017). Unplugged computational thinking is a method of non-digital instruction for the concepts of computational thinking and computer science. The name "unplugged computational thinking" is a contrast to "plugged-in," digital, traditional computational thinking and computer science instruction. In this lesson, the unplugged computational thinking activity is "Programmer Says." In Programmer Says, the teacher prompts children to do an action, and children respond to the prompt using their own bodies as if the teacher were programming the children like a computer or robot. For example, if the teacher were to say, "Programmer Says Shake" and holds up a picture of KIBO's shake block, the correct response would be for children to shake their bodies, but if the teacher were to say, "Programmer Says Spin" and holds up a picture of KIBO's spin block, the correct response would be for children to spin. In Lesson 13, the children explore the process of scanning KIBO blocks and experiment with strategies for scanning. In Lessons 14 and 19, the children engage in explorative free play using new coding blocks. These five lessons consist of similar, play-based learning experiences for the children while focusing on different powerful ideas of computer science and concepts of computational thinking instruction.

For each of these five lessons, two researchers independently observed and rated each child in the group on their PTD behaviors using the PTD rubric during that session. In addition, the teacher leading that lesson was provided with a copy of the rubric for each of the observed children and asked to rate the behaviors of the child. Because one of the design objectives for the PTD Rubric was to create a rubric that could be administered with minimal instruction, teachers were only given the instructions to complete the rubric, return it to the study coordinator at the next observation visit, and make a note with any feedback, comments, or questions.

This assessment process led to three ratings per child per activity, two from graduate student researchers and one from the teacher leading the small group activity. In Room A, one classroom teacher completed all five of the observations and either taught or co-taught the lessons. In Room B, one teacher taught and completed the observations for the first four lessons. A second teacher taught the final lesson and completed the observations for that lesson. Of these three teachers, two identified as Hispanic women and one identified as an African-American man. The teachers had an average of 13 years of teaching experience. One had prior experience doing coding with a family member, but none had formally taught coding or robotics or had any experience with the KIBO platform.

Attendance varied for students, with Student #9 attending all five lessons and Student #8 and Student #10 only attending two lessons each (Table 1). This variation in attendance was for a variety of reasons, including but not limited to COVID and quarantine related absences, non-COVID childhood illness, or absence from the KIBO lesson due to individual services such as therapy or individual assessments. No lesson had data collected for all 10 students. In addition, there were errors in data collection leading to missingness for Lessons 5 and 12. For Lesson 5, one teacher did not complete student rubrics for all students leading to missing teacher data for three students. For Lesson 12, a second researcher was not present in one classroom leading to missing second assessor data for four students.

Additional Measures

In addition to the PTD rubric, additional measures were collected to be correlated with PTD behaviors for the purpose of assessing validity.

Teaching Strategies GOLD. The Teaching Strategies GOLD is a standardized assessment evaluating children’s academic and non-academic development in nine domains (physical, cognitive, language, socio-emotional, mathematics, literacy, science and technology, social studies, and the arts) as well as English acquisition for English language learners (Lambert et al., 2014, 2015). Horizons for Homeless Children shared Teaching Strategies GOLD data for students enrolled in research from three timepoints –prior to curriculum implementation, mid-curriculum implementation, and post-curriculum implementation. I used the mid-curriculum Teaching Strategies GOLD data for this pilot validation, as the assessment took place simultaneous to the evaluated KIBO lessons.

Four objectives from the Teaching Strategies GOLD were identified as potentially relating to PTD: Objective 1 – Regulates own emotions and behaviors; Objective 2 – Establishes and sustains positive relationships; Objective 3 – Participates cooperatively and constructively in group situations; and Objective 10 – Uses appropriate conversational and other communication skills. I examined Objective 1 (Regulates own emotions and behaviors) as potentially relating to Choices of Conduct. Research on the development of behavior disorders suggests that poor emotional regulation in preschool is related to behavioral problems both in preschool and early elementary school (Cole & Zahn-Waxier, 1996). No literature suggests that this relation would not remain true in the technology environment. I examined Objective 2 (Establishes and sustains positive relationships) as potentially relating to Community Building. The subscales of this objective include forming relationships with adults, making friends, and interacting with peers, each of which are involved in engaging with and strengthening a community (Lambert et al., 2015). I examined Objective 3 (Participates cooperatively and constructively in group situations) as potentially relating to Collaboration. The PTD rubric defines Collaboration as "the act of working together toward a shared goal," and the anchors reference different forms and levels of group work, such as sharing ideas, sharing support, or sharing space. Finally, I examined Communication as potentially relating to Objective 10 (Uses appropriate conversational and other communication skills), because the wording of this objective specifically references use of communication skills.

KIBO Project Rubric. The initial thesis proposal proposed that projects created during free exploration in Lessons 14 and 19 would be recorded and scored using the KIBO Project Rubric (Govind & Bers, 2021). This rubric evaluates projects both on project complexity and creativity. Unfortunately, project collection was disrupted due to COVID-related absences and behaviors that prevented children from completing the project components of lessons. Although the PTD allowed for documentation of a child opting to not participate in the activity, the KIBO Project Rubric does not accommodate for situations in which a child opted out of participating or had step away from the activity to go for a calming walk with a staff member, so project scores in Room A were only collected for one student for Lessons 14 and 19. Room B did not complete Lesson 14 with independent projects for students but instead with a prompted project for the whole class, so student projects were only collected for Lesson 19. In addition, there was not variation within the project scores.

Although the Project Rubric has been validated in students in early childhood classrooms and with multiple levels of KIBO knowledge, it has not been validated specifically as a measure for assessing KIBO projects in preschool classes or with four- and five-year-old children and may not have the sensitivity at its lower end to evaluate variation in projects of students in this age group. As such, due to both the limited number of projects collected and the lack of variation in Project Rubric scores, the Project Rubric was removed from the thesis as a validation measure for the PTD rubric. For this reason, I could not evaluate construct validity for Creativity or Content Creation in this thesis.

Data Analysis

In this initial pilot phase of data analysis, a few analyses were conducted to establish reliability and validity of the rubric items. All data analyses were conducted in R Studio using the tidyverse, moments, irr, and ltm packages (Gamer et al., 2019; Komsta & Novomestky, 2015; R Core Team, 2020; Rizopoulos, 2006; Wickham et al., 2019).

Reliability

I examined reliability across time (stability reliability) and across assessor (interrater reliability), as well as internal consistency reliability.

I evaluated interrater reliability by calculating Cohen's kappa coefficients to account for potential errors associated with chance interrater association. I calculated two Cohen's kappa coefficients. Because I was the only consistent rater for all sessions, I calculated one Cohen's kappa coefficient comparing my observation scores with those of the second researcher, and one Cohen's kappa coefficient comparing my observation scores to those of the classroom teacher.

I evaluated stability reliability by computing Pearson product-moment correlations of scores from two observations, both for each item and for each child's as well as for their mean score overall. I estimated these correlations twice – once with both observations recorded by me, as well as a set of correlations with one observation recorded by me and the second recorded by the second researcher.

I evaluated internal consistency by estimating two Cronbach's alpha scores. I estimated one score using ratings from each child's first observation and the other using ratings from each child's second observation.

Validity

I evaluated criterion validity by comparing student mean scores for each item to students' scores on related assessments. As I noted earlier in this section, I was not able to collect the

KIBO project rubric, which was meant to serve as a validation measure for the Content Creation and Creativity items, for most students. Therefore, I did not evaluate construct validity for those two items.

I evaluated construct validity for the remaining four items (Choices of Conduct, Communication, Community Building, and Collaboration) by computing Pearson's product-moment correlations between ratings from Lesson 12 with related scores from the Social and Emotional Learning and Language measures of the Teaching Strategies GOLD assessment (Lambert et al., 2014, 2015). I selected Lesson 12 because it had high attendance with only one student missing the lesson. I chose to use ratings from a single lesson to remove variability due to lesson differences. I compared Choices of Conduct with the Teaching Strategies GOLD Objective 1 (*Regulates own emotions and behaviors*). Collaboration was paired with the Teaching Strategies GOLD Objective 3 (*Participates cooperatively and constructively in group situations*). I paired Community Building with the Teaching Strategies GOLD Objective 2 (*Establishes and sustains positive relationships*). Communication was paired with Teaching Strategies GOLD Objective 10 (*Uses appropriate conversational and other communication skills*).

Results

Descriptive Statistics

Descriptive statistics for each item and overall mean rubric score are included in Table 2. Item scores for each PTD Rubric item as well as overall PTD Rubric mean scores were adequately normal as determined by skew and kurtosis.

Teachers did not provide any feedback on the items, phrasing, or anchors of the rubric. When asked about issues after the first observation, teachers commented that they had no issues with or feedback on the rubric.

Demographic Statistics

The mean age of the participating children was 4.79 years old. Three of the children were male and seven children were female. The group was also racially, ethnically, and linguistically diverse; although all children were fluent in English in the classroom, four primary languages were spoken by the children at home (English (n = 7), Spanish (n = 1), Amharic (n = 1), and Haitian Creole (n = 1)). One child was non-Hispanic White, six children were non-Hispanic Black/African American, one child was Hispanic Black/African American, and two children were Other Hispanic. No children were identified by their teachers as having been diagnosed with or referred for disability services at the time of observations. There were no statistically significant demographic differences between the children in the two classrooms.

PTD Correlation with Age

Children's age was moderately positively correlated with collaboration (r(7) = 0.38, p = 0.31) and community building, (r(7) = 0.39, p = 0.30) in Lesson 12. There was a strong positive association between age and choices of conduct (r(7) = 0.57, p = 0.11). There were weaker positive association between age and communication (r(7) = 0.26, p = 0.49) and between and age and overall PTD (r(7) = 0.27, p = 0.48). There was a moderately high negative association between age and content creation (r(7) = -0.43, p = 0.24) and a weaker negative association between age and creativity (r(7) = -0.19, p = 0.31).

Reliability

Interrater Reliability

I evaluated interrater reliability by computing Cohen's kappa coefficients to compare my scores to those of the second observing researcher and those of the teacher. I computed these Cohen's kappa coefficients both for the overall datasets as well as for each rubric item (Table 3). Cohen's kappa coefficients suggested moderate agreement overall between me and the second graduate student researchers (kappa = 0.48, $p < 0.0001$), and slight overall agreement between myself and the teachers (kappa = 0.20, $p < 0.0001$). There was also variation in agreement between the different rubric items. Between myself and the second researchers, the kappa scores ranged between fair agreement for choices of conduct (kappa = 0.17, $p = 0.15$) and moderate agreement for collaboration (kappa = 0.43, $p < 0.01$). Between the teachers and myself, there was no agreement on the creativity or content creation items, but fair agreement on the communication items (kappa = 0.31, $p < 0.01$). Agreement was not stronger for the same items between the second researcher and myself and between the teachers and myself. Although collaboration had been the item with strongest agreement and choices of conduct the item with weakest agreement when comparing myself with the second researcher, agreement between myself and the teacher appeared higher for the choices of conduct item (kappa = 0.31, $p < 0.01$) than for the collaboration item (kappa = 0.17, $p = 0.11$).

Stability reliability

I evaluated stability using Pearson product-moment correlations to compare scores on each item for each child's first and second attended observation. I computed two sets of these correlations, one in which both observation scores were rated by myself, and one correlating two ratings completed by myself, and one involving ratings from the second researcher and myself. Reliability varied across the different PTD items, with some items appearing to have much stronger stability reliability than others (Table 4). I was not able to assess if this variation in reliability across items was statistically significant due to lack of power.

My ratings for the Choices of Conduct items were the most strongly correlated of my scores from the two observation points across each child's two timepoints ($r(7) = 0.76$, $p = 0.02$). The ratings for Content Creation were also not significantly correlated at the two timepoints ($r(7) = 0.60$, $p = 0.09$), although there was a positive trend. The two ratings of Creativity ($r(7) = 0.33$, $p = 0.39$), Communication ($r(7) = 0.44$, $p = 0.24$), and Community Building ($r(7) = 0.25$, $p = 0.52$) items were not significant. There was no association between the two ratings for Collaboration ($r(7) = -0.06$, $p = 0.88$). Correlation coefficients between my rating scores at one timepoint and those of the second researcher at the second are included in Table 4. The scores for Collaboration and Creativity were not significantly related between the two timepoints (collaboration: $r(7) = 0.47$, $p = 0.20$; creativity: $r(7) = 0.45$, $p = 0.23$). The remaining items were also not significantly related across time.

Internal Consistency

I evaluated internal consistency using Cronbach's alpha, which provided an inconsistent view of internal consistency. I computed Cronbach's alpha twice, one time using each child's first observed visit and one time using each child's second observed visit. Cronbach's alpha when calculated using the data from each child's first observed visit was 0.80 suggesting good internal consistency. Cronbach's alpha when calculated using the data from each child's second observed visit was 0.36 suggesting poor internal consistency.

Construct Validity

Choices of Conduct was strongly and significantly correlated with the Teaching Strategies GOLD Objective 1 (*Regulates own emotions and behaviors*) ($r(7)=0.74$, $p = 0.02$).

Collaboration was not significantly correlated with the Teaching Strategies GOLD Objective 3 (*Participates cooperatively and constructively in group situations*) ($r(7)=0.48$, $p = 0.19$). Community building was strongly correlated with the Teaching Strategies GOLD Objective 2 (*Establishes and sustains positive relationships*) ($r(7)=0.70$, $p = 0.03$). Communication was not significantly correlated with the Teaching Strategies GOLD Objective 10 (*Uses appropriate conversational and other communication skills*) ($r(7)=0.32$, $p = 0.40$).

Discussion

In this pilot study, the PTD Rubric was evaluated to determine its reliability and validity as a potential tool for assessing children's Positive Technological Development behaviors in the classroom.

One of the goals in designing the PTD Rubric was to create a measure that was easily usable for teachers and that teachers could administer during typical classroom instruction. A second related goal was that the Rubric would not require significant training but would be easy to administer based on the anchors, knowledge of the PTD behaviors, and familiarity with children's normative behavior. Both design goals were evaluated through the successful administration of the rubric. Each classroom consisted of multiple students being observed during the KIBO lesson, as well as other students participating in the lesson who were not being observed. Over the course of a single KIBO lesson, teachers were leading the lesson, interacting with students (sometimes in multiple languages at once), and managing student behaviors for a class of between 12 and 18 students. The fact that teachers were also able to easily complete the rubric for five students simultaneously suggests that the rubric is a low-effort and usable material for teachers to add to their instructional routine.

In this pilot administration, teachers did not complete a training or professional development specific to the rubric. Teachers attended a general Coding as Another Language KIBO professional development as part of the CAL-KIBO PreK validation project. This training had included an introduction to the constructs of PTD and the Coding as a Playground pedagogy. For training on the rubric, teachers were given the rubric and instructions on how to complete and return the rubrics, but the rubric was not included in the professional development training. Nevertheless, teachers were able to complete the rubric, and did so with no questions. Further discussion below will discuss reliability and validity issues, which may also relate to the lack of specific training on how teachers and researchers should complete the questions. Overall, the ability of teachers to complete the rubric for multiple students simultaneously per taught lesson with no issues suggests that the general format and phrasing work for teachers to create an easily usable, short, and quick assessment.

Reliability

Interrater reliability

Interrater reliability was not strong for many items on the rubric, and particularly between researchers and teachers. Overall, there was some evidence for moderate agreement between the two researchers, but only slight agreement between the researcher and the teacher raters. There are a few possible explanations for this finding.

One possible explanation is that the teachers are having a different experience administering the rubric in the classroom than the researchers are as silent observers. Teachers in the early childhood classroom are engaging in many behaviors at once, such as leading a lesson for a class of up to 15 students, handling the technology, managing children’s engagement, and responding to behaviors. In between these roles, we were now also asking teachers to rate PTD behaviors for the identified children, either keeping track of the behaviors while teaching or reflecting on and rating the behaviors after teaching. The PTD Rubric was designed for this setting to be completed quickly, either when teaching or immediately after, quickly rating children's behaviors after managing multiple elements of a lesson.

In this context, the only role of the researchers was to sit, watch, take notes, and rate students. Between two and five children were rated at a time, depending on student attendance, so researchers were only having to attend to the behavior or two or five children, and did not have to respond to these behaviors other than to document them. As such, researchers were able to pay attention to minor details in student behavior and could spend time noticing the finer points between two scores. Whereas teachers were rating students based on a single reflection, researchers were rating students repeatedly, looking for individual behaviors. These two situations could lead to a discrepancy in scores and a discrepancy in reliability between the two researchers, versus between a researcher and a teacher.

If this difference in rating processes were the cause for the discrepancy, it might be helpful to add a time element to the rubric. For example, the anchors for Community Building are currently written as follows:

0. *Does not take part in group activities. Only works alone without talking to others.*
1. *Sometimes takes part in some group activities when prompted.*
2. *Always takes part in group activities when prompted and will sometimes take part in group activities when prompted.*
3. *Actively engages in group activities including taking steps to engage others, but only suggests ideas for group activities when prompted.*

4. *Actively engages in and suggests ideas for group activities.*

The items could be revised to be more centered around time-based anchors rather than an amalgamation of student behaviors. In this way, teachers and researchers would both have to reflect primarily on how often a behavior occurred, as compared to which of many behavior types occurred. Here is an example of how I could revise these items:

0. *Works alone without talking to others. Takes part in group activities less than half the time.*
1. *Takes part in group activities half the time.*
2. *Takes part in group activities more than half the time.*
3. *Takes part in all group activities but is actively engaged less then half the time.*
4. *Takes part in and actively engages in all group activities.*

This modification to the anchors would provide researchers with an operationalized, observable definition that aligns more strongly with the observations of teachers. With a clearer and simpler rubric, researchers could make a differentiated decision more easily and without attending to behaviors that teachers are not able to attend to due to their primary role in the classroom. As a secondary benefit, such a simplified rubric would also allow for researchers to rate more children, allowing for the next stage of validation to include researchers observing whole classrooms simultaneously as teachers rate all of their students.

Aligning the researcher rater experience to the experience of the teacher allows the focus to be on the experience of the teacher as the ultimate primary user of the assessment, and therefore the user whose needs should be the focus in the validation process. Training the researchers to rate more in line with the teachers recognizes the expertise of the teachers in understanding the behaviors of their students and the validity of the teachers' knowledge. However, the teacher raters should also be aligned with the researcher ratings, as the researchers are the ones observing multiple classrooms, and therefore generating scores across classroom.

In addition to the difference in interrater reliability between the researcher pairs and researcher-teacher pairs for the overall rubric, there were also differences in agreement between the types of rater pairs across individual items. For most items, such as Collaboration or Creativity, agreement appeared higher between researchers than between researchers and teachers. On the other hand, agreement on ratings for the Choices of Conduct item appeared higher between teachers and myself than between the other researchers and myself.

One possible explanation for some of these differences is that operationalized definitions in the rubric may not clear for the constructs, particularly when considering that many of these constructs share names with existing behaviors or constructs outside the technology setting. In most cases, this difference in prior knowledge might affect teachers, who are used to alternate definitions in the classroom and less familiar with the PTD framework definitions of these constructs. Teachers, who are used to these alternate meanings for these constructs, may be basing their ratings partially on their own understandings of the meanings of the named constructs, in comparison to the constructs as defined by the PTD and Coding as a Playground theoretical frameworks. Researchers, who have had greater experience with these frameworks, may be drawing on their existing knowledge of the framework when rating students. As pairs of researchers are operating from a shared knowledge of the framework, there may be less of a visible impairment to agreement created by poorly operationalized definitions and anchors. Revising the operationalized definitions of the PTD constructs and the operationalized anchors for scoring could help teachers make more precise ratings of behaviors within the technology environment without requiring the teachers to attend time consuming training on the PTD Framework and pedagogy. This approach could improve reliability between researchers and teachers, as well as overall reliability for these items.

Agreement on ratings for the Choices of Conduct item was higher between the teacher and myself than between other graduate researchers and myself. Again, this difference between the two pair groups may be explained by differences in a shared knowledge base, in this case a shared knowledge of the preschool setting. The appropriateness of behaviors related to Choices of Conduct are dependent on the context, rules, and expected behaviors of the setting. A rater is only able to score the appropriateness of the child's behavior if the rater is aware of and knowledgeable about the setting. Many researchers have experience working with children as babysitters or at summer camps, but there are different expectations for children's behavior in the home or at summer camp than at preschool. Many (if not all) of the graduate student researchers who came with me had little or no teaching experience, and specifically no teaching experience in the preschool setting. For a construct such as Content Creation, ratings might be made based on the process of creation of artifacts, and for a similarly created artifact and process, ratings might not change between summer camp and classroom settings. However, the two settings might differ in behavioral expectations regarding physical activity, autonomy, or interactions with peers and adults, leading to differences in ratings for Choices of Conduct (Sibthorp et al., 2020; Wilson et al., 2017).

Even for adults, the appropriate choices can change between spaces and contexts. As a hypothetical example, take three STEM classrooms within a university lecture hall. In a college biology classroom, the appropriate choice when you finish your work might be to quietly entertain yourself on your computer until your peers are finished. In the applied engineering room next door, when a group is finished with their project, the appropriate choice might be to go over to another completed group and brainstorm ideas to expand on their projects for the final. And in the computer science midterm examination down the hall, the appropriate behavior for students finishing their work might be to stand up and leave. When we speak of Choices of Conduct, we are speaking of the behavior of selecting, and therefore ability to select, the appropriate choices for the setting and context.

Along a similar line, there appeared to be similar agreement for the Communication item between the two researchers, and between the researcher and the teacher. One possible explanation for this lack of the difference seen with other items is that Communication may be easier to observe than some of the other PTD behaviors, as there is a more universal understanding of what counts as communication and what would be expected communication. Another possible explanation is Communication is physically easier to observe and more salient than some of the other rubric items. Children often communicate by talking, which attracts attention, increasing the likelihood of both the rater noticing the behavior and also accurately remembering the behavior in contrast to a less salient behavior (Mather & Sutherland, 2011). That said, the reliability was still only fair, so the next iteration of the rubric will include refinement of the anchors for this item as well.

Stability Reliability

Scores for the PTD rubric were not consistent for students across sessions. This lack of stability is not necessarily a fault of the rubric. Stability reliability assumes the existence of stability of the constructs on the rubric, with the measured constructs acting as a trait of the individuals and therefore staying consistent across time and place. These results suggest that PTD behaviors and engagement vary across lessons, which is consistent with prior research suggesting that constructs related to PTD may be context dependent, such as the Coding as a Palette of Virtues Framework.

The Coding as a Palette of Virtues framework proposes that individuals draw on and express different virtues while coding or engaging in technology, depending on their own preferences, choices, and interests for that day (Bers, 2021). According to the Coding as a Palette of Virtues Framework, individuals can select from any of the 10 identified virtues (Curiosity, Perseverance, Open-mindedness, Optimism, Honesty, Patience, Generosity, Gratitude, Forgiveness, and Fairness), a combination of these virtues, or other virtues of their choice depending on their own needs for the moment (Bers, 2021). The virtues chosen will vary depending on the individual, day, and group context. Teachers, when teaching lessons from curricula, will also select which virtues to emphasize in their lessons that day for their students.

The Coding as a Palette of Virtues framework and the Coding as a Playground framework are theoretically linked with the PTD behaviors serving as the demonstrated behaviors of the Palette of Virtues values, and the Palette of Virtues values serving as the guiding values promoting the PTD behaviors (Bers, 2021). The CAL curriculum includes the Coding as a Palette of Virtues framework as one of its pedagogical frameworks, and teachers are invited to select the virtues to promote within each lesson rather than being restricted to an individual virtue for each lesson (Bers, 2019). For this reason, even though I selected the lessons for the reasons of their similarity, teachers may have selected different virtues from the Palette of Virtues to promote when teaching the selected lessons, leading to the promotion of different PTD behaviors across lessons as well.

Although we tried to control for the classroom context with the selection of the lessons for observation, it was impossible to entirely control for the variation in the classroom learning context between days. Multiple factors had the potential to alter the classroom context between the lessons, including the presence or absence of new or different students, the absence of a teacher, the presence of an outside observer completing an observation unrelated to this research study (such as for Head Start, NAEYC, or a disability services referral), or small but known differences between the lesson activities.

Internal Consistency

Internal consistency as measured by Cronbach's alpha was not consistent across the first two points of measurement. One possible explanation for this finding may have to do with assumptions of Cronbach's alpha and what the coefficient is measuring. Cronbach's alpha assumes that all items are equally representative of the overall latent variable, or that the factors load equally onto the latent variable. We may not expect all PTD items to be equally present on all days, especially in the early lessons, wherein children are learning and exploring different components of the KIBO robotics platform on the different days. For example, in a lesson where children learn to scan, we may expect to see much higher Creativity than Content Creation as children explore different methods of scanning simple programs. In a lesson allowing for elaborate and purposeful program creation, Content Creation and Creativity might be more equally represented and more equally representative of a child's overall PTD.

Construct Validity

As mentioned above, I was not able to evaluate construct validity for two of the six items due to student absences and teacher alterations to curriculum implementation. Of the four items I was able to evaluate construct validity for, results suggest moderate construct validity compared to other assessments of children's development. That is, two of these relations were statistically significant, which was particularly meaningful with such a small sample size and therefore such low power.

Two of the four items were strongly correlated with related measures: Choices of Conduct to a measure of children's emotion and behavior regulation, and Community Building to a measure of relationship building and sustaining. Both relations had correlation coefficients above 0.70 but below 0.75, suggesting the PTD measures evaluated similar but not exactly aligned measures of behavior as the development measures from the TS GOLD. The relation of Community Building with relationship building and sustaining suggests that children's building of relationships in their general classroom environment is mirrored in the technology microworld as well. The relation of Choices of Conduct with emotion and behavior regulation aligns with prior research suggesting that healthy emotion regulation in early childhood was associated with reduced levels of negative behavior (Cole & Zahn-Waxier, 1996). It was not surprising that the correlations were significant and strong but not exactly 1.0. The PTD rubric was specifically looking at a child's behavior within the specific technology context and the TS GOLD looks at children's developmental abilities across all contexts. This finding suggests that the PTD rubric items for Choices of Conduct and Community Building are looking at constructs that are closely related to, but not identical to, the behaviors assessed by the TS GOLD.

The other two of the four evaluated rubric items, Collaboration and Communication, had only moderate associations with their related measures from the Teaching Strategies GOLD. This finding suggests that the constructs measured by the items were related to those measured in the Teaching Strategies GOLD assessment, but the alignment was less strong than for the Community Building and Choices of Conduct items described above.

A child's participation in communicative behaviors in the technology environment is not entirely dependent on their language development. Three of the children in this project spoke a primary language other than English, and none of the children had disabilities or language-related delays. We would not necessarily expect these variables to affect the Communication domain of Positive Technological Development behaviors in the same ways. Future research should examine how children whose primary language does not match that of their learning context demonstrate Communication behaviors in the technology environment, as well as all domains of PTD. In addition, as children with disabilities were not included in this sample of 10 children future research should examine the usability of the rubric to assess the behaviors of students with disabilities. Ultimately, the PTD Rubric can be to serve as a tool in assessing the potential of coding and technology playgrounds to promote PTD environments and contexts for success for children with difficulty communicating in the classroom.

What does the Rubric teach us about PTD and its measurement?

One of the key theoretical components of the Design-Based Research methodological model is that through the design-based research process, the researcher can examine and learn about the phenomenon, in addition to designing the tool. In this thesis, the phenomena studied were behaviors of Positive Technological Development and their measurement. The design research process of this rubric asked about the behaviors of PTD, as well as how they could be observed and measured in the classroom setting.

The findings from this project reinforce that PTD is a series of behaviors rather than an intrinsic trait. Although the behaviors may be associated with developmental domains, they are not a developmental domain in and of themselves. These PTD behaviors performed by a child are then state or context dependent. A child's performance of these behaviors might vary from day to day or lesson to lesson depending on the classroom or learning context. We should therefore consider PTD to be a state measure, rather than a trait measure. PTD levels may change from lesson to lesson or day to day, as different lessons encourage different levels of each PTD behavior and as students' in-school and out-of-school environments change from day to day.

Considering PTD as a state rather than a trait measure could impact how we design future research studies examining PTD. When looking to compare PTD performance data across students, the data collected should be for the same lesson, because variation due to differences between lessons will create error and mask the variation due to differences between students. When collecting PTD performance data for a lesson or to compare across lessons, researchers or teachers should select lessons carefully and intentionally. Each lesson may be promoting different instances of the PTD behaviors. If the goal is to gather a holistic view of PTD, researchers may want to collect more than one data point for each student, but for the same lessons for each. For example, each student can have data collected for Lessons 5, 10, and 15, but if the scores are averaged to create mean PTD, any student missing a datapoint should be considered as missing data and be removed from the data set.

Finally, I showed that teachers could record PTD behaviors of multiple children simultaneously in the classroom while teaching lessons. However, the features teachers notice about PTD behaviors may not be the same features that researchers observe. This difference may be due to differences in understanding of PTD behaviors, differences in familiarity with the student leading to different reflections on the students' behavior, or differences in understanding of the classroom context. However, this finding suggests that an in-classroom observational assessment of PTD is feasible as an assessment strategy for teachers in the early childhood classroom.

Limitations

As a pilot study administered in the 2021-2022 school year, this validation study had many limitations, including those associated with the COVID-19 pandemic. Some additional limitations included small sample size, poor student attendance, missing data, lack of a consistent second researcher assessor, and the absence of project rubric data.

The sample size for this study was only 10 students, five students in each of two classes. This sample size was selected partially for the proof-of-concept feature of the pilot study, as one of the research goals of the project was for a researcher and a teacher to administer a rubric to multiple children at once. However, having only 10 children selected limited the number of validation analyses that I was able to conduct because many analyses, including Pearson product-moment correlations, did not have enough power with 10 or fewer participants (Bujang & Baharum, 2016). With only 10 children, missing data due to absences was highly impactful.

Although the sample of 10 students included students of multiple races, ethnicities, and primary languages, I was not able to examine differences between student groups due to the small number of students in the sample. In a future study, having a larger number of students would allow me to examine the validity of the rubric for students who speak different languages. With the data currently available, I was only able to conclude that teachers could use the rubric to rate behaviors of students who spoke multiple languages alongside their English-speaking peers.

Another limitation of the study was poor attendance. Although teachers selected students for their likelihood of attending lessons, many students still missed observed lessons for a variety of reasons. Each of these absences was also more noticeable due to each student's significant role as 10 percent of the student sample. There was no lesson that all students attended. As mentioned above, the rubric was also not stable across lessons. Accordingly, I opted to conduct comparisons of students using data from the same lesson. Student attendance reduced the sample for each lesson to, at highest, nine students. Because this investigation was a pilot study, I was able to complete analyses with this sample size; but, for a future validation study, I would want to have a larger sample to validate the assessment.

The absence of the Project Rubric data also limited the ability of this pilot study to fully evaluate the validity of the PTD Rubric. Multiple limiting factors led to the absence of this item from the pilot study. The lesson format used by Class B for Lesson 14, in which the whole class worked on a single project as a group, resulted in a lack of individual scores for the students in that class. This change in format was made by the teacher. In the professional development training, I encouraged teachers to alter the curriculum as necessary for their classroom, including altering the group size of activities between large group, small group, and individualized activities. Teachers explained that they had altered the activity size of curricular activities due to the number of KIBOs available in the classroom, because the class was limited to whole group activities when only one KIBO was available for the whole class to share.

Whereas curriculum alterations led to a lack of scores in Class B, poor attendance led to a lack of scores overall for students in Class A. This situation included both poor class attendance (i.e., missing class), and poor attendance for the Project Rubric part of the lesson. Students who left early, either of their own accord, for pullout sessions, or for an emotion regulation break with an adult, did not create a free play project for the specified lessons. The absence of this measure meant that I was not able to evaluate construct validity for two of the rubric items, Creativity and Content Creation.

There were also missing data from multiple sessions. This absence happened for three reasons. First, in some cases, a second assessor was not present due to scheduling conflicts and so there was only teacher and primary researcher data. In another other case, teacher data were missing due to teacher confusion with paperwork during a classroom COVID shutdown. Finally, a miscommunication in the identities of the observed students led to a teacher not completing the ratings for two observed students for that session. Because data were missing for different lessons across different classrooms between teachers and second researchers, it is possible that differences in interrater reliability are due to variation within behaviors present in specific lessons.

The identity of the second researcher was also not consistent, with four individuals playing the role of second researcher. All individuals were graduate students at Tufts University and members of the DevTech research group, but their positions in the research group varied, including doctoral students, masters' students, and full-time staff. This variation likely led to a lack of consistency within the scoring of the second researchers, and there was no evaluation of this error as there were no items rated by more than one of the second assessors. The lack of a consistent second researcher also led to a violation of the assumptions of the Cohen's kappa coefficient used to assess interrater reliability. In a formal evaluation of interrater reliability, the second assessor would need to be a consistent individual for all observations to remove any potential variability and error related to assessor differences.

Future Steps for Design and Research

Although one of the goals of this project was to create a usable assessment for teachers in the early childhood classrooms, this thesis did not include usability data from teachers beyond completion data. As mentioned above, this pilot study was embedded within a broader implementation study of the CAL-KIBO curriculum at Horizons for Homeless Children. Post-implementation interviews from the CAL implementation project at Horizons will include a question about teachers' experiences administering the PTD rubric in their classroom. Resulting data should be analyzed to inform the next phase of revisions to the PTD rubric. Research in the next phase of the PTD Rubric development project should include more in-depth research questions about the user experiences of teachers administering the rubric. These questions could be both qualitative and quantitative, taking the form of interviews, asking teachers to leave comments on the rubric, or surveys.

One of the goals of the project was to develop a rubric that could be administered to multiple children. Therefore, one of the goals of the pilot study was to assess if it were possible for raters to administer the rubric to multiple children at one time. However, for an assessment to be usable in the classroom, a teacher would need to be able to administer it to all students, not only five at a time. The findings from this study suggest that PTD rubric results are influenced by classroom context including the lesson specifics and the day. Therefore, if the teacher were to administer the rubric to different students on different days, the data may not be comparable across days. For the assessment to be usable in the classroom, the teacher would need to administer the rubric to all children on the same day. The next phase of research with the PTD rubric should examine if teachers and researchers can administer the rubric to all children in a classroom at one time.

This pilot project focused only on children between the ages of four and five in a structured pre-kindergarten classroom. This setting is unique, in addition to being a specific age group. This situation was not necessarily a limitation, because it still allows for the validation of a rubric for classrooms for four- and five-year-old children. However, children's behavior in domains related to PTD, such as language development and self-regulation changes significantly across the early childhood years. For example, over the preschool and early elementary school years, children develop the ability to maintain sustained conversations and provide feedback to listeners, which could then promote greater communication and collaboration in the technology context (Becker Bryant, 2009). Results from this thesis suggest that the PTD rubric and its results are highly context dependent. Four- and five-year-old classrooms, such as the BPS UPK classrooms included in this study, are a unique learning context (Gray-Lobe et al., 2021). As is the case with younger preschool classrooms, these classrooms have multiple learning centers, large learning spaces such as carpets, short lessons, and many opportunities for free selection by children of choices. However, these classrooms also begin to introduce structured lessons, group lessons, and traditional literacy and numeracy experiences to prepare children for the transition to Kindergarten. For this reason, the rubric may not apply to either the younger preschool context, or the older early elementary context. Further research with students of those younger and older age groups would be necessary to explore the implementation of the rubric in learning settings for children of those ages.

As I have noted, the study of this rubric included data collection across lessons and locations, but with a small sample size of students, and results suggested that PTD varies across contexts. I have noted as well that the small sample size limited the ability of this thesis to examine factors that may contribute to variation in PTD. As part of the larger Horizons for Homeless Children project, we are collecting photos and video recordings of the classrooms and instruction, which we could analyze using qualitative methods alongside researcher field notes to understand variation in classroom contexts that might affect the PTD behaviors performed by a child. This variation in contexts could be due to differences in teacher behaviors, the behaviors of surrounding students, or the physical learning environment. Future analyses could compare these differences to differences in PTD scores to develop a preliminary understanding of what factors in the classroom context affect variation in PTD.

In the next phase of the design and research process, more qualitative data about the classroom environment should be collected. This addition could include requesting teacher comments about external factors that could affect student learning for the day (for example, if the student was distressed about the loss of a grandparent or excited about an upcoming birthday), or could involve taking close field notes on classroom factors that could affect the classroom dynamic (for example if the students were returning from a week of vacation or if a new student had just entered the class). Future research could also include more video and photographic data of the KIBO lessons themselves, including teacher instruction, the physical space for student learning, and the overall class dynamic at this time.

The sample of students observed in this pilot study included students whose primary language was Spanish, Haitian Creole, and Amharic in addition to students of only English-speaking backgrounds. This diversity suggests that the assessment can be used with students of each of these linguistic backgrounds. However, there was not a large enough sample of each of these students to run statistical analyses comparing scores. No students with disabilities were included in the sample. When future validation studies are conducted with a larger sample, students with disabilities and a larger sample of students whose primary language is not English should be included to determine the validity of the rubric for other groups of students.

Finally, this thesis found some opportunities for revision of the rubric. Some of the construct definitions may not have been well operationalized for the technological environment, particularly as the names of the constructs already refer to other, related constructs in the general preschool environment. As written, the operationalized definitions do not include mention of technology or the technology environment, although the PTD constructs specifically apply to the technology environment. To address this point, the operationalized definitions can be revised to include mention of the technology environment. This change may improve reliability between researcher and teacher scores.

The method used in this validation study encouraged the researcher to spend more time reflecting on the rubric while rating the child. In contrast, the teacher used the rubric in reflections on the student's behavior after teaching. The rating style of the teacher was the intended method of rubric completion for which I initially designed the rubric, as referenced in the original conjecture map (Figure 1). To promote the quicker, reflective method of rubric completion, the rubric can be revised to include more time-based anchors. The rubric currently uses time-anchors such as "does not," "sometimes," "consistently," and "always." However, future revisions of the rubric can transition these terms to "never," "less than half the time," "more than half the time," and "always." Future research can include assessment of all children in a classroom. This step will both allow for collection of more data, providing opportunities for more statistical analyses, as well as encouraging a quicker method of rubric completion for researchers, who will have less time to reflect on each rubric item for each student.

In sum, the next step for this project is to continue with the iterative design process. I will begin by revising the rubric's definitions and anchors as described above. Then, we will begin collecting data of all children in a classroom, allowing for greater data collection. At this point, we will also begin working with more classrooms to collect a larger amount of data and include greater numbers of bilingual students and students with disabilities. Alongside quantitative data, we will collect qualitative data in order to evaluate potential factors related to variation in PTD behaviors. Finally, we will begin collecting teacher usability data to evaluate the experience of teachers as the target user for this tool.

References

Alvino, F. J. (2000). *Art Improves the Quality of Life: A Look at Art in Early Childhood Settings.*

Becker Bryant, J. (2009). Pragmatic development. In E. L. Bavin (Ed.), *The Cambridge Handbook of Child Language* (pp. 339–354). Cambridge University Press.

Bell, T., Witten, I. H., & Fellows, M. (1998). *Computer Science Unplugged. . . Off-line activities and games for all ages*. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.308.6827&rep=rep1&type=pdf

Bers, M. U. (2006). The role of new technologies to foster positive youth development. *Applied Developmental Science*, *10*(4), 200–219.

Bers, M. U. (2010). Beyond computer literacy: Supporting youth's positive development through technology. *New Directions for Youth Development*, *2010*(128), 13–23. https://doi.org/10.1002/yd.371

Bers, M. U. (2012). *Designing digital experiences for positive youth development: From playpen to playground*. Oxford University Press.

Bers, M. U. (2018). Coding, playgrounds and literacy in early childhood education: The development of KIBO robotics and ScratchJr. *2018 IEEE Global Engineering Education Conference (EDUCON)*, 2094–2102. https://doi.org/10.1109/EDUCON.2018.8363498

Bers, M. U. (2019). Coding as another language: A pedagogical approach for teaching computer science in early childhood. *Journal of Computers in Education*, *6*(4), 499–528. https://doi.org/10.1007/s40692-019-00147-3

Bers, M. U. (2020a). *Coding as a playground: Programming and computational thinking in the early childhood classroom*. Routledge.

Bers, M. U. (2020b). *Norfolk Public Schools CAL-KIBO Research Findings Report* (p. 48). DevTech Research Group, Tufts University.

Bers, M. U. (2021). Coding, robotics and socio-emotional learning: Developing a palette of virtues. *Pixel-Bit, Revista de Medios y Education*(62), 309–322. https://doi.org/10.12795/pixelbit.90537

Bers, M. U., Doyle-Lynch, A., & Chau, C. (2009). Positive technological development: The multifaceted nature of youth technology use towards improving self and society. *Constructing the Self in a Digital World*, 110–136.

Bers, M. U., González-González, C., & Armas–Torres, M. B. (2019). Coding as a playground: Promoting positive learning experiences in childhood classrooms. *Computers & Education*, *138*, 130–145. https://doi.org/10.1016/j.compedu.2019.04.013

Bowers, E. P., Li, Y., Kiely, M. K., Brittian, A., Lerner, J. V., & Lerner, R. M. (2010). The Five Cs Model of Positive Youth Development: A Longitudinal Analysis of Confirmatory Factor Structure and Measurement Invariance. *Journal of Youth and Adolescence*, *39*(7), 720–735. https://doi.org/10.1007/s10964-010-9530-9

Bowers, E. P., Napolitano, C. M., Arbeit, M. R., Chase, P., Glickman, S. A., Lerner, R. M., & Lerner, J. V. (2013). On a Pathway Towards Thriving: Evaluating the Effectiveness of the “GPS to Success” Tools to Promote Positive Development and Intentional Self Regulation in Youth. *Journal of Youth Development*.

Bowne, J. B., Magnuson, K. A., Schindler, H. S., Duncan, G. J., & Yoshikawa, H. (2017). A Meta-Analysis of Class Sizes and Ratios in Early Childhood Education Programs: Are Thresholds of Quality Associated With Greater Impacts on Cognitive, Achievement, and

Socioemotional Outcomes? *Educational Evaluation and Policy Analysis*, *39*(3), 407–428. https://doi.org/10.3102/0162373716689489

Brackmann, C. P., Román-González, M., Robles, G., Moreno-León, J., Casali, A., & Barone, D. (2017). Development of Computational Thinking Skills through Unplugged Activities in Primary School. *Proceedings of the 12th Workshop on Primary and Secondary Computing Education*, 65–72. https://doi.org/10.1145/3137065.3137069

Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *The Journal of the Learning Sciences*, *2*(2), 141–178.

Bujang, M. A., & Baharum, N. (2016). Sample Size Guideline for Correlation Analysis. *World Journal of Social Science Research*, *3*(1), 37. https://doi.org/10.22158/wjssr.v3n1p37

Capp, M. J. (2017). The effectiveness of universal design for learning: A meta-analysis of literature between 2013 and 2016. *International Journal of Inclusive Education*, *21*(8), 791–807. https://doi.org/10.1080/13603116.2017.1325074

Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language*, *43*(6), 1310–1337. https://doi.org/10.1017/S0305000915000689

Caudle, L., Harper, F. K., Quinn, M., & Greene, D. (2021). *Co-Constructing and Sharing STEAM Knowledge through a Culturally Relevant Literacy-Based Early Childhood School-University Partnership*. *2*.

Chen, C.-H. (2008). Why Do Teachers Not Practice What They Believe Regarding Technology Integration? *The Journal of Educational Research*, *102*(1), 65–75. https://doi.org/10.3200/JOER.102.1.65-75

Code.org. (2021). *Hour of Code*. https://hourofcode.com/us

Code.org, CSTA, & ECEP Alliance. (2021). *2021 State of computer science education: Accelerating action through advocacy* (p. 116). https://advocacy.code.org/stateofcs

Cole, P. M., & Zahn-Waxier, C. (1996). Individual Differences in Emotion Regulation and Behavior Problems in Preschool Children. *Journal of Abnormal Psychology*, *105*(4), 518–529. https://doi.org/10.1037/0021-843X.105.4.518

Collinson, V., & Fedoruk Cook, T. (2001). "I don't have enough time" - Teachers' interpretations of time as a key to learning and school change. *Journal of Educational Administration*, *39*(3), 266–281. https://doi.org/10.1108/09578230110392884

Congress, U. (1975). *The Individuals with Disabilities Education Act–IDEA.*

Damon, W. (1990). *The Moral Child: Nurturing CHildren's Natural Moral Growth.* The Free Press.

de Ruiter, L. E., & Bers, M. U. (2021). The Coding Stages Assessment: Development and validation of an instrument for assessing young children's proficiency in the ScratchJr programming language. *Computer Science Education*, 1–30. https://doi.org/10.1080/08993408.2021.1956216

Dede, C., Nelson, B., Ketelhut, D. J., Clarke, J., & Bowman, C. (2004). *Design-based research strategies for studying situated learning in a multi-user virtual environment*. 158–165.

DevTech Research Group. (n.d.). *Positive Technological Development (PTD) Engagement Checklist: Children/Child*. https://ase.tufts.edu/devtech/PTDChild.pdf

D'ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT press.

Dixson, D. D., & Worrell, F. C. (2016). Formative and Summative Assessment in the Classroom. *Theory Into Practice*, *55*(2), 153–159. https://doi.org/10.1080/00405841.2016.1148989

Dym, B., Pasupuleti, N., Rockwood, C., & Fiesler, C. (2021). “You don’t do your hobby as a job”: Stereotypes of Computational Labor and their Implications for CS Education. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 823–829. https://doi.org/10.1145/3408877.3432396

Eisenberg, N., Cameron, E., Tryon, K., & Dodez, R. (1981). Socialization of Prosocial Behavior in the Preschool Classroom. *Developmental Psychology*, *17*(6), 773–382.

Eisenberg, N., Spinrad, T. L., & Knafo-Noam, A. (2015). Prosocial Development. In R. M. Lerner (Ed.), *Handbook of Child Psychology and Developmental Science* (7th ed., Vol. 3). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118963418.childpsy114

Ertmer, P. A., & Newby, T. J. (1993). Behaviorism, Cognitivism, Constructivism: Comparing Critical Features from an Instructional Design Perspective. *Performance Improvement Quarterly*, *6*(4), 50–72.

Ervin-Tripp, S. (1982). Ask and It Shall be Given Unto You: Children’s Requests. In H. Byrnes (Ed.), *Georgetown University Round Table on Languages and Linguistics 1982* (pp. 235–245). Georgetown University Press.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement* [R package version 0.84.1]. https://CRAN.R-project.org/package=irr

Garrigan, B., Adlam, A. L. R., & Langdon, P. E. (2018). Moral decision-making and moral development: Toward an integrative framework. *Developmental Review*, *49*, 80–100. https://doi.org/10.1016/j.dr.2018.06.001

Garvis, S., & Twigg, D. (2010). Exploring Art in Early Childhood Education. *The International Journal of the Arts in Society: Annual Review*, *5*(2), 193–204. https://doi.org/10.18848/1833-1866/CGP/v05i02/35830

Gould, S. J., & Gold, S. J. (1996). *The mismeasure of man*. WW Norton & company.

Govind, M., & Bers, M. U. (2021). Assessing Robotics Skills in Early Childhood: Development and Testing of a Tool for Evaluating Children's Projects. *Journal of Research in STEM Education*, *7*(1), 47–68.

Gray-Lobe, G., Pathak, P., & Walters, C. (2021). *The Long-Term Effects of Universal Preschool in Boston* (No. w28756; p. w28756). National Bureau of Economic Research. https://doi.org/10.3386/w28756

Greitemeyer, T., Traut-Mattausch, E., & Osswald, S. (2012). How to ameliorate negative effects of violent video games on cooperation: Play it cooperatively in a team. *Computers in Human Behavior*, *28*(4), 1465–1470. https://doi.org/10.1016/j.chb.2012.03.009

Hall, G. C. N., Yip, T., & Zárate, M. A. (2016). On becoming multicultural in a monocultural research world: A conceptual approach to studying ethnocultural diversity. *American Psychologist*, *71*(1), 40–51. https://doi.org/10.1037/a0039734

Harper, F. K., Caudle, L., Quinn, M., Rainwater, T., Larsen, J., Greene, D., Flowers, C. E., & Sadovnik, A. (2022, April). *Culturally relevant computing for young children: A conceptual analysis.* American Educational Research Association Annual Conference, San Diego, CA.

Hunt, L., & Bers, M. U. (2021). Coding, Computational Thinking, and Cultural Contexts: In M. U. Bers (Ed.), *Teaching Computational Thinking and Coding to Young Children* (pp. 201–215). IGI Global. https://doi.org/10.4018/978-1-7998-7308-2.ch010

Iacono, T., Goldbart, J., Douglas, S. N., & Garcia-Melgar, A. (2022). A Scoping Review and Appraisal of AAC Research in Inclusive School Settings. *Journal of Developmental and Physical Disabilities*. https://doi.org/10.1007/s10882-022-09835-y

ISTE Standards For Students (p. 2). (2016). International Society for Technology in Education. iste.org/standards

K–12 Computer Science Framework. (2016). http://www.k12cs.org

Komsta, L., & Novomestky, F. (2015). *moments: Moments, cumulants, skewness, kurtosis and related tests* (0.14) [Computer software]. https://CRAN.R-project.org/package=moments

Lambert, R. G., Kim, D.-H., & Burts, D. C. (2014). Using Teacher Ratings to Track the Growth and Development of Young Children Using the Teaching Strategies GOLD Assessment System. *Journal of Psychoeducational Assessment*, *32*(1), 27–39. https://doi.org/10.1177/0734282913485214

Lambert, R. G., Kim, D.-H., & Burts, D. C. (2015). The measurement properties of the Teaching Strategies GOLD® assessment system. *Early Childhood Research Quarterly*, *33*, 49–63. https://doi.org/10.1016/j.ecresq.2015.05.004

Lerner, R. M., Lerner, J. V., Murry, V. M., Smith, E. P., Bowers, E. P., Geldhof, G. J., & Buckingham, M. H. (2021). Positive Youth Development in 2020: Theory, Research, Programs, and the Promotion of Social Justice. *Journal of Research on Adolescence*, 1–21.

Lerner, R. M., Lerner, J. V., P. Bowers, E., & John Geldhof, G. (2015). Positive youth development and relational-developmental-systems. *Handbook of Child Psychology and Developmental Science*, 1–45.

Levinson, T. (2021a). *CAL Reflection* [Unpublished manuscript].

Levinson, T. (2021b). *PTD Pilot* [Unpublished manuscript].

Levinson, T., Hunt, L., & Hassenfeld, Z. R. (2021). Including Students With Disabilities in the Coding Classroom: In M. U. Bers (Ed.), *Teaching Computational Thinking and Coding to Young Children* (pp. 236–248). IGI Global. https://doi.org/10.4018/978-1-7998-7308-2.ch012

Lieber, J., Horn, E., Palmer, S., & Fleming, K. (2008). Access to the General Education Curriculum for Preschoolers with Disabilities: Children's School Success. *Exceptionality*, *16*(1), 18–32. https://doi.org/10.1080/09362830701796776

Maloney, J., Resnick, M., Rusk, N., Silverman, B., & Eastmond, E. (2010). The Scratch Programming Language and Environment. *ACM Transactions on Computing Education*, *10*(4), 1–15. https://doi.org/10.1145/1868358.1868363

Marotta, L., Tramonte, L., & Willms, J. D. (2022). *Equivalence of Testing Instruments in Canada: Studying Item Bias in a Cross-Cultural Assessment for Preschoolers*. 24.

Mather, M., & Sutherland, M. R. (2011). Arousal-Biased Competition in Perception and Memory. *Perspectives on Psychological Science*, *6*(2), 114–133. https://doi.org/10.1177/1745691611400234

Maxwell, G. S. & Queensland School Curriculum Council. (2001). *Teacher observation in student assessment*. The Council.

McClelland, M. M., John Geldhof, G., Cameron, C. E., & Wanless, S. B. (2015). Development and Self-Regulation. In R. M. Lerner (Ed.), *Handbook of Child Psychology and Developmental Science* (7th ed., Vol. 1). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118963418.childpsy114

Milgrom-Elcott, T. (2019, November 7). *When STEM Becomes STEAM, We Can Change The Game*. https://www.forbes.com/sites/taliamilgromelcott/2019/11/07/when-stem-becomes-steam-we-can-change-the-game/?sh=36b420ec6905

Napolitano, C. M., Bowers, E. P., Arbeit, M. R., Chase, P., Geldhof, G. J., Lerner, J. V., & Lerner, R. M. (2014). The *GPS to Success* Growth Grids: Measurement Properties of a Tool to Promote Intentional Self-Regulation in Mentoring Programs. *Applied Developmental Science*, *18*(1), 46–58. https://doi.org/10.1080/10888691.2014.866768

National Center for Education Statistics. (2021). *Digest of Education Statistics: 2019*. U.S. Department of Education. https://nces.ed.gov/programs/digest/d19/

Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. *Cognition*, *108*(1), 222–231. https://doi.org/10.1016/j.cognition.2007.12.003

Papert, S. (1980). *Constructionism vs. Instructionism*. Proceedings from Japanese Educators Conference. http://www. papert. org/articles/const_inst.

Pederson, P. V. (2007). What Is Measured Is Treasured: The Impact of the No Child Left Behind Act on Nonassessed Subjects. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *80*(6), 287–291. https://doi.org/10.3200/TCHS.80.6.287-291

Public School Teacher Data File (U.S. Department of Education). (2017).

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Research Foundation: The Creative Curriculum®. (2010). Teaching Strategies, Inc.

Resnick, M., Ocko, S., & Papert, S. (1988). LEGO, LOGO, AND DESIGN. *Children's Environments Quarterly*, *5*(4), 14–18.

Resnick, M., & Rusk, N. (2020). A Different Approach to Coding. *Communications of the ACM*, *63*(11), 120–127.

Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses. *Journal of Statistical Software*, *17*(5), 1–25.

Roque, R., & Rusk, N. (2019). Youth perspectives on their development in a coding community. *Information and Learning Sciences*, *120*(5/6), 327–348. https://doi.org/10.1108/ILS-05-2018-0038

Sandoval, W. (2014). Conjecture Mapping: An Approach to Systematic Educational Design Research. *Journal of the Learning Sciences*, *23*(1), 18–36. https://doi.org/10.1080/10508406.2013.778204

Sibthorp, J., Wilson, C., Povilaitis, V., & Browne, L. (2020). Active ingredients of learning at summer camp. *Journal of Outdoor and Environmental Education*, *23*(1), 21–37. https://doi.org/10.1007/s42322-019-00050-6

Slocombe, K. E., & Seed, A. M. (2019). Cooperation in children. *Current Biology*, *29*(11), R470–R473. https://doi.org/10.1016/j.cub.2019.01.066

Spivak, A. L., & Farran, D. C. (2012). First-Grade Teacher Behaviors and Children’s Prosocial Actions in Classrooms. *Early Education and Development*, *23*(5), 623–639. https://doi.org/10.1080/10409289.2011.566164

Strawhacker, A., & Bers, M. U. (2018). Promoting Positive Technological Development in a Kindergarten Makerspace: A Qualitative Case Study. *European Journal of STEM Education*, *3*(3). https://doi.org/10.20897/ejsteme/3869

Sullivan, A., & Bers, M. U. (2016). Girls, Boys, and Bots: Gender Differences in Young Children’s Performance on Robotics and Programming Tasks. *Journal of Information*

Technology Education: Innovations in Practice, *15*, 145–165. https://doi.org/10.28945/3547

Unahalekhaka, A., & Govind, M. (2021). Examining Young Children's Computational Artifacts. In *Teaching Computational Thinking and Coding to Young Children* (pp. 265–294). IGI Global.

van Aalst, J., & Chan, C. K. K. (2007). Student-Directed Assessment of Knowledge Building Using Electronic Portfolios. *Journal of the Learning Sciences*, *16*(2), 175–220. https://doi.org/10.1080/10508400701193697

Viglas, M., & Perlman, M. (2018). Effects of a Mindfulness-Based Program on Young Children's Self-Regulation, Prosocial Behavior and Hyperactivity. *Journal of Child and Family Studies*, *27*(4), 1150–1161. https://doi.org/10.1007/s10826-017-0971-6

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1683. https://doi.org/10.21105/joss.01686

Wilson, C., Sibthorp, J., & Brusseau, T. A. (2017). Increasing Physical Activity and Enjoyment Through Goal-Setting at Summer Camp. *Journal of Park and Recreation Administration*, *35*(4), 24–36. https://doi.org/10.18666/JPRA-2017-V35-I4-7930

Xu, F., Xie, J., & Correia, A.-P. (2022). How Users Seek Computational Thinking Resources on Teachers-Pay-Teachers: A Review-Based Content Analysis. *San Diego*, 8.

Zimmerman, E. (2009). Reconceptualizing the Role of Creativity in Art Education Theory and Practice. *Studies in Art Education*, *50*(4), 382–399. https://doi.org/10.1080/00393541.2009.11518783

Tables and Figures

Table 1

Student Attendance for Observed Lessons

Lessons	Room B					Room A				
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Lessons Attended	5	4	4	5	5	3	4	2	5	2
Lesson 5	Yes	Yes	Pullout	Yes	Yes	Absent	Yes	Yes	Yes	Absent
Lesson 12	Yes	Yes	Yes	Yes	Absent	Yes	Yes	Yes	Yes	Absent
Lesson 13	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Absent	Yes	Yes
Lesson 14	Yes	Yes	Yes	Yes	Yes	On Walk	Yes	Left early	Yes	Left early
Lesson 19	Yes	Absent	Yes	Yes	Yes	Yes	Absent	Absent	Yes	Absent

Table 2

Descriptive Statistics for PTD Rubric Item and Overall Scores

Measure	*M*	*SD*	*Kurtosis*	*Skewness*
Overall PTD	2.08	1.06	2.44	-0.38
Collaboration	2.42	1.55	1.59	-0.39
Communication	2.67	1.28	2.39	-0.67
Content Creation	1.22	1.48	2.30	0.87
Community Building	2.20	1.18	2.33	-0.24
Creativity	1.13	1.41	2.25	0.87
Choices of Conduct	2.74	1.46	2.12	-0.75

Table 3

Agreement Across Assessors for PTD Rubric Item and Overall Scores

Measure	Author - Assessor 2	*p*	Author - Teacher	*p*
PTD	0.48	0.000	0.20	0.001
Collaboration	0.43	0.002	0.19	N.S.
Communication	0.31	0.008	0.31	0.003
Creativity	0.27	0.041	0.04	N.S.
Community Building	0.19	N.S.	0.25	0.015
Content Creation	0.38	0.001	0.07	N.S.
Choices of Conduct	0.17	N.S.	0.31	0.008

Table 4

Stability Reliability for PTD Rubric Items for Single and Multiple Assessors

Item	Pearson's product-moment *r*, single assessor	*p*	Pearson's product-moment *r*, two assessors	*p*
Choices of Conduct	0.76	0.02	0.25	N.S.
Collaboration	-0.06	N.S.	0.47	N.S.
Creativity	0.33	N.S.	0.45	N.S.
Content Creation	0.6	0.09	0.11	N.S.
Communication	0.44	N.S.	0.29	N.S.
Community Building	0.25	N.S.	-0.08	N.S.

Figure 1

Conjecture Map for the Design Based Method

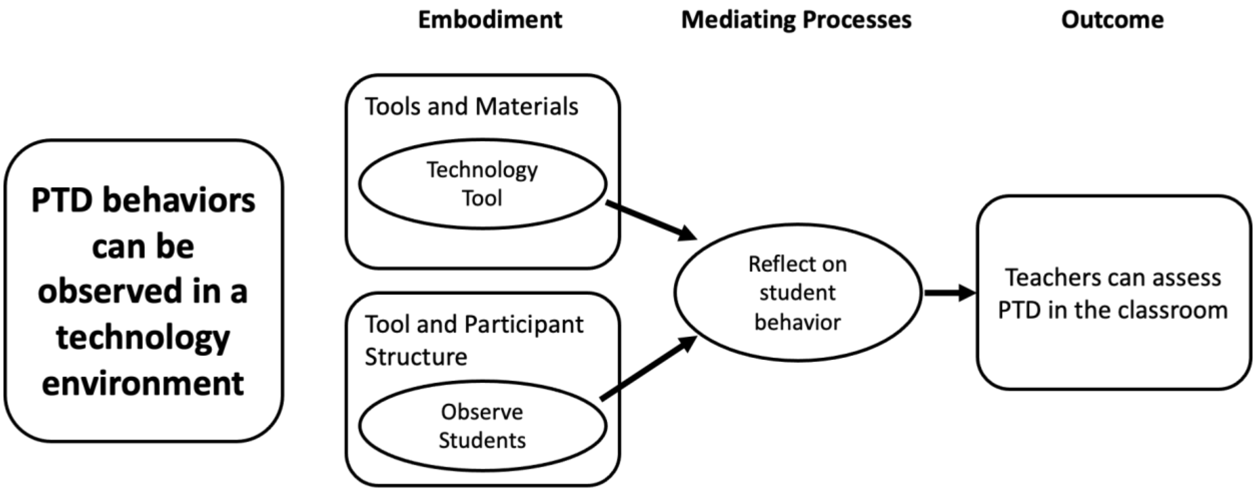
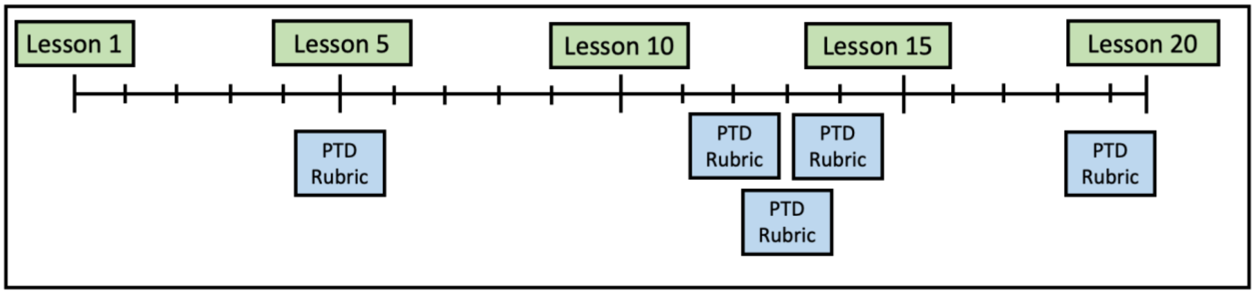


Figure 2

Project Timeline



Note. This timeline ends with the end of PTD data collection. The CAL curriculum project continues beyond this timeline, but I did not include data from that period in this thesis. Therefore, those data are not included in the timeline.

Figure 3

An example of a classroom carpet where some CAL-KIBO lessons were taught

Figure 4

The carpet in the STEM room where some CAL-KIBO lessons were taught

Appendix A: Positive Technological Development Rubric

	Definition	0	1	2	3	4
Collaboration	the act of working together towards a shared goal	Only works alone	Works on an individual activity in parallel with others	Works on individual project while asking, giving, and receiving help	Works on individual project but with shared/communal ideas	Works with another student on shared project with common goal
Communication	the act of initiating and participating in conversations	Does not participate in conversations with others and is not engaged	Listens to conversations without active engagement	Participates in conversations when directly prompted (i.e. answering yes/no questions)	Always participates in conversations when prompted and sometimes when unprompted	Always actively engages in group conversations, including speaking and listening, when unprompted
Content Creation	the act of making purposeful and original projects	Does not create projects	Creates projects by directly copying examples or other projects	Creates projects that are modified copies of examples or other projects	Creates original projects using copied elemtents from other projects	Creates original, purposeful projects
Community Building	the act of taking actions to strengthen and promote group cohesion and comradery	Does not take part in group activities. Only works alone without talking to others	Takes part in some group activities passively when prompted	Always takes part in group activities when prompted and will sometimes take part in group activities without prompting	Actively engages in group activities including taking steps to engage others, but only suggests ideas for group activities when prompted	Actively engages in and suggests ideas for group activities
Creativity	the act of participating in new and original ways	Does not create projects and/or only creates replicas of existing or example projects	Creates new projects that directly follow prompts	Creates projects with elements beyond prompts by copying other or example projects	Sometimes adds innovative elements to projects	Consistently uses tools, elements, or materials in new or unprescribed ways
Choices of Conduct	the act of behaving in ways that promote a respectful, fair, and positive environment	Acts in unsafe, disrespectful, unkind ways. Does not follow class rules.	Follows class rules when reminded and prompted	Usually follows class rules without prompting but when facing conflict, does not make safe, kind, or respectful choices without prompting or assistance	When facing conflict, sometimes makes safe, kind, and respectful choices without prompting or assistance	Always makes safe, kind, and respectful choices without adult prompting

Appendix B: Positive Technological Environment Rubric

Collaboration

Teacher definition: the extent to which the teacher's actions scaffold children working together on shared projects or ideas

T0= Teacher gives projects designed to be completed individually

T1= Teacher encourages students to ask each other questions or for help, but assignments are meant to be completed individually

T2= Teacher encourages collaborative brainstorming

T3= Teacher assigns students to work in groups

T4= Teacher puts students in group and scaffolds group work with discussion of group roles

Environment Definition: The extent to which the workspace supports children working together on shared projects or ideas

E0= Children sit in fixed seating arrangements with individual workspaces

E1= Children sit in fixed seating arrangements with individual workspaces, but children can see each other's work from their workspace

E2= Children sit in fixed seating arrangements with individual workspaces at shared tables

E3= Children sit in fixed seating arrangements with large, shared workspaces

E4= Workspaces are large enough for multiple children, and children can move between workspaces

Communication

Teacher definition: the extent to which the teacher's actions scaffold conversations between children

T0= Teacher discourages discussion and encourages silent work

T1= Teacher does not encourage or discourage either discussion or silent work.

T2= Teacher asks children questions but does not actively encourage discussion between students

T3= Teacher encourages discussion between students but does not provide scaffolding for discussions

T4= Teacher scaffolds discussion between students

Environment Definition: the extent to which the environment supports conversations between children

E0= Children are seated individually in rows facing foward

E1= Children are seated individually but can turn to face each other

E2= Children are seated facing each other

E3= Children can walk freely around the room to talk to others

E4= Children have flexible seating and can move their workspace around the room

Content Creation

Teacher definition: the extent to which the teacher scaffolds the creation of innovative projects

T0= Teacher gives assignments that do not include element of creation (i.e. puzzle solving only)

T1= Teacher gives assignments that include creating replicas of sample projects

T2= Teacher provides narrow prompts for creative projects and encourages children to follow prompt

T3= Teacher provides narrow prompts for creative projects but does not discourage children from deviating from prompt.

T4= Teacher provides open-ended prompts for creative projects

Environment definition: the extent to which the classroom supports children in creating innovative projects

E0= No materials are available to children

E1= The materials or tools required to complete the prompted activity are available to children

E2= A broad range of materials or tools are available to children upon request.

E3= The classroom is decorated with a variety of art and projects created by students

E4= A broad range of materials or tools are available and accessible to children without assistance.

Community Building

Teacher definition: the extent to which the teacher scaffolds children's taking part in whole-classroom learning activities

T0= Teacher encourages children to work individually. Lessons include only individual work with no sharing.

T1= Teacher displays children's work in classroom, but children work individually and lessons do not include a sharing component.

T2= Teacher includes group activities such as games, songs, or read-alouds in lesson

T3= Teacher prompts children to share work with class

T4= Teacher scaffolds children to engage in the process of sharing, receiving, and encorporating feedback on projects

Environment definition: the extent to which the design of the classroom supports whole-classroom learning activities and class culture

E0= Children work only at workstations. Projects do not leave workstations.

E1= Children work only at workstations, but there is a space where work can be displayed.

E2= There are decorations on the walls representing local holidays, symbols, or events.

E3= Children sit at workstations but can take part in whole-class activities such as games or project sharing from their seats.

E4= There is a learning space outside of children's typical workspaces for children to take part in whole-class activities, such as a large carpet.

Creativity

Teacher definition: the extent to which the teacher scaffolds the use of tools and materials in new and innovative ways

T0= Teacher assigns activities that should be completed in one way (i.e. a puzzle with one solution, or a project made by copying a sample).

T1= Teacher does not discourage children from exploring new approaches of using materials or solving problems, but does not actively encourage or scaffold such exploration.

T2= Teacher encourages children to try new tools or materials.

T3= Teacher prompts children to use material or tool in new way

T4= Teacher scaffolds discussion on multiple approaches (i.e. "This program is longer but this program has more characters. I love how both tell wonderful stories about our classroom.")

Environment Definition: the extent to which the environment supports the use of tools or materials in new or innovative ways

E0= Only the materials or tools required to complete the prompted activity are available to children

E1= A range of materials or tools are available to children.

E2= Tools and materials children have not been taught yet are available to children.

E3= Projects featuring multiple and varied uses of tools and materials are on display in the classroom.

E4= Tools and materials with unspecified uses are available to children

Choices of Conduct

Teacher definition: the extent to which the teacher scaffolds children is making decisions that are positive, safe, and kind when presented with scenarios involving potential conflict

T0= Teacher does not remind children of classroom rules or expectations

T1= Teacher reminds children of rules and expectations for safe, kind, and respectful behavior

T2= Teacher provides children with materials that are breakable or delicate and reminds children of expected behaviors

T3= Teacher labels positive and negative behaviors and describes impact of behavior when giving feedback (i.e. "I see you are walking slowly with your tablet. That shows respect to the classroom materials.")

T4= Teacher scaffolds discussion of process of making positive choices

Environment Definition: the extent to which the environment supports children in making positive, safe, and kind choices when presented with scenarios involving potential conflict

E0= There are no classroom rules posted for children to see.

E1= Classroom rules are posted in the classroom, but may not be accessible to the children (i.e. only written for a class of non-readers).

E2= Classroom rules are posted in the classroom in an accessible format to the children (i.e. written in accessible language or accompanied by pictures).

E3= Decision flow-charts (i.e. "options for when you are mad" anchor chart) are posted in an accessible format alongside classroom rules to support children in making positive choices.

E4= There is an area of the room dedicated to positive decision making (i.e. a calm-down corner) with materials and accessible instructions for students to make positive choices.