UCLA | CRESST NATIONAL CENTER FOR RESEARCH ON EVALUATION, STANDARDS, AND STUDENT TESTING

A QUALITATIVE COMPARISON OF YOUNG CHILDREN'S PERFORMANCE ON ANALOGOUS DIGITAL AND HANDS-ON TASKS: ASSESSMENT IMPLICATIONS

In S. Caitlin Burns (Chair), *Assessment and Evaluation in Early Childhood Settings*. 2021 American Educational Research Association Virtual Meeting.

April 8–12, 2021

Gregory K. W. K. Chung, Ziyue Ruan, Elizabeth J. K. H. Redman

Principal Investigator: Gregory K. W. K. Chung

APRIL 2021

The work reported here was supported by the PBS KIDS Digital, through contract 20154615 to the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

The contents of this document were developed under a grant from the U.S. Department of Education. However, those contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the federal government. [PR/Award No. U295A 150003, CFDA No. 84.295A]

Table of Contents

Abstract 3
Challenges and Issues in Assessing Young Children 3
  Challenge 1: Young Children as Test Takers 4
  Challenge 2: Individualized Testing 4
  Challenge 3: Adequate Construct Representation 4
  Parallel Digital and Hands-on Tasks 6
    Digital Task 6
    Hands-on Task 6
    Research and Engineering Objectives 7
    COVID-19 Impact 8
Method 8
  Data Sources 8
  Tasks 8
    Slide Digital Game 8
    Hands-on Slide Contraption 10
  Measures 10
  Procedure 10
Results 11
  Usability of the Slide Contraption 11
  Task Performance 11
  Qualitative Observations 0
    Task Engagement 0
    Magnifying Glass Usage (Slidea-ma-zoo) 0
    Preference for Height and Speed (Slidea-ma-zoo, Slide Contraption) 0
    Multi-step Moves 1
Discussion 1
    Limitations 1
    Research Objectives 1
    Engineering Objectives 2
    Implications for Assessment 2
    Implications for Instruction 3
    Implications for Linking Digital and Physical Play 3
References 5

Appendix A Requirements for the Slide Hands-on Task (Redman et al., 2020, p. 185) 10

A Qualitative Comparison of Young Children's Performance on Analogous Digital and Hands-on Tasks: Assessment Implications[1]

Gregory K. W. K. Chung, Ziyue Ruan, and Elizabeth J. K. H. Redman

CRESST/University of California, Los Angeles

Abstract

This study conducted an exploratory study to compare children's performance on a video game-based task to their performance on an analogous hands-on physical task. The game involved a slide and concepts of height and friction. For both formats, the number of rounds it took to beat a level and the number of steps used within each round were measured.

The number of rounds to beat a level between the two modes were not drastically different. However, the number of steps (per round) to beat a round was higher in the game than the hands-on task, suggesting a mode effect. The digital format appeared to be interfering with children's ability to "sense" friction, which was not observed in the hands-on task.

Challenges and Issues in Assessing Young Children

How can the knowledge and skills learned by young children—preK through first grade—be measured reliably and validly, when many children at this age have yet to learn to read or write, have yet to develop the ability to articulate their thinking, have limited attention spans, and have little experience with standardized testing and being tested on demand? One promising approach is to use tablet-based games to assess young children's mastery of skills (e.g., Lieberman et al., 2009; Redman et al., 2019, 2020; Schenke et al., 2020). However, tasks presented in a digital format—games or otherwise—may impose artificial constraints on children's ability to express their understanding.

Thus, we conducted an exploratory study to compare children's performance on a slide task in a digital game with their performance on a hands-on physical slide task that mimicked the game slide task. For both formats, we examined the number of rounds it took to beat a level and the number of steps used within each round.

[1] We would like to thank three anonymous AERA reviewers for their helpful comments. Reviewer comments have been incorporated into this paper. We would also like to thank Jeremy Roberts of PBS KIDS Digital for supporting this work, Zhuoda (Kenny) Wu for his help with data collection, and Charlie Parks for his help with alpha version of the microcontroller.

The role of assessment is becoming increasingly important in early childhood education (Greenfield, 2015; Meloy & Schachner, 2019; Snow & Van Hemel, 2008; Wechsler et al., 2016). The surge in public and private investments in school readiness has increased the demand for assessments to fulfill both accountability requirements and program evaluations.

Challenge 1: Young Children as Test Takers

Assessing young children requires attention to many factors that are less likely to be problematic with older children and adults. For example, young children may not have developed the skills to read or write, may have short attention spans, and may become taxed and fatigued with unfamiliar assessment tasks. Young children often learn through play and the manipulation of concrete objects and other hands-on activities. Children are rarely asked to express their understanding in an abstract way or with a traditional paper-and-pencil test format. Preschool-aged and kindergarten children are more apt to *demonstrate* what they know rather than *articulate* what they know. They may have difficulties responding to cues, prompts, and directions that are not contextualized. From an assessment perspective, these factors impose unique design challenges and if unaddressed, introduce measurement error and bring into question the credibility of the assessment results (Atkins-Burnett, 2007; Brennen, 2011; Clements et al., 2015; National Research Council [NRC], 2001; Riley et al., 2016; Scott-Little & Niemeyer, 2001; Snow & Van Hemel, 2008).

Challenge 2: Individualized Testing

One-on-one testing is extremely time and labor intensive. Individually testing 20 first graders with a 20min task will take one assessor at least 6.7 hours. In addition, the assessor almost always needs to be trained to follow the protocol and to administer the task.

Challenge 3: Adequate Construct Representation

A consequence of the difficulty assessing young children, assessments rely mainly on selected-response formats. The assessor poses a question to the child, and the child provides an answer orally or through a gesture (e.g., by pointing to the option). The assessor is there to read aloud the question and answer options, record the child's response, and help the child maintain their focus on the task. A range of techniques have been used to make the assessment task more reliable: Deliver the task on a computer with the prompts and options read aloud by the computer to the child (e.g., Ruzek et al., 2020), cast the assessment task in a story or fantasy context (e.g., Klein & Starkey, 2012; or the use of puppets in Samarapungavan et al., 2009), use a highly structured interactive protocol (e.g., Grindal et al., 2019), gamify the assessment task (e.g., Howard & Melhuish, 2017; McKown et al., 2016), or use the game itself as a way to measure what a child knows and can do (Chung et al., 2016; Redman et al., 2018, 2019, 2020, in-press).

On technology increasingly used with young children are digital tablets (e.g., iPads or Android tablets). Tablets are attractive because young children can quickly learn to use tablets (Samarakoon et al., 2019). Tablet learning apps have been found to promote learning in young children (Griffith et al., 2020; Herodotou, 2018; Hirsh-Pasek et al., 2015; Huber et al., 2016; Xie et al., 2018) and tablets are now being used to deliver assessments. Digital assessments offer standardized administration, individualization (e.g., a child can easily replay prompts and instructions), automated logging of responses, and automated scoring of responses (e.g., Chung et al., 2016; Howard & Melhuish, 2017; McKown et al., 2016; Redman et al., 2018, 2020; Ruzek et al. 2020).

However, the mode or format of the task (i.e., digital or hands-on) may affect the measurement of young children's knowledge and skills especially when the content is abstract or unobservable--as is common in STEM topics. When a task situation is understood by the child (i.e., the child can perceive all the relevant information and understands what is being asked of him or her) and the child can respond in a natural way (e.g., by pointing, touching, dragging, or moving objects), then the format—digital or hands-on—has little effect on performance (Zacharia et al., 2012). However, when tasks involve concepts such as mass or friction—phenomena that is not directly observable or may require tactile sensation in order to respond appropriately (e.g., Lazonder & Ehrenhard, 2013)—format matters. Format differences can also lead to strategy differences as shown in Manches et al. (2009); children solved a numerical problem-solving task using physical blocks or digital blocks. Children using physical blocks explored many more potential solutions because they could quickly arrange the blocks in many different ways whereas children using digital blocks were constrained by the software design.

In all cases, the desire for standardization and repeatability has resulted in constraints on what children are asked to do, thus limiting what we can know about children’s capabilities and the effects of various educational interventions. Even with games—a very interactive format—children are constrained by the game design: What the child can interact with, how they can express themselves, and the degree to which the gameplay requires the child to apply their knowledge. Challenge 3 inevitably results, because of practical issues, in the use of assessment formats that tend to be measures of children’s mastery of facts and simple procedures. Yet this kind of information is unlikely to be what policymakers, teachers, and parents really want to know about children: the extent to which children can apply what they learned to new situations and the strategies and processes used in solving a problem.

Parallel Digital and Hands-on Tasks

To address Challenge 1 to Challenge 3, we developed a task format and task apparatus to enable comparison between children's performance on digital and hands-on versions of similar tasks.

Digital Task

The digital task was adopted from prior work (Redman et al., 2020). The Redman et al. study examined the extent to which a game, *Slidea-ma-zoo*, could be used for measurement purposes. *Slidea-ma-zoo*[2] is part of the PBS KIDS Digital's *Cat in the Hat Knows a Lot About That!*[3] suite of games designed to promote science and engineering concepts in young children. The *Slidea-ma-zoo* game puts the player in race situation where the player's goal is to helps one set of characters (Harry and Sally) win a slide race against another set of game characters (Thing One and Thing Two). For each level, both the opponent slide and the player's slide are set to a pre-determined height. The player then manipulates the slide height or surface friction to adjust how fast Harry and Sally travel down the slide as shown in Figure 1. Redman et al. found that children enjoyed the game and their performance in the game, particularly their manipulation of the game slide's height and friction, was related to their knowledge of height and friction concepts with slides as measured by an external measure.

Figure 1

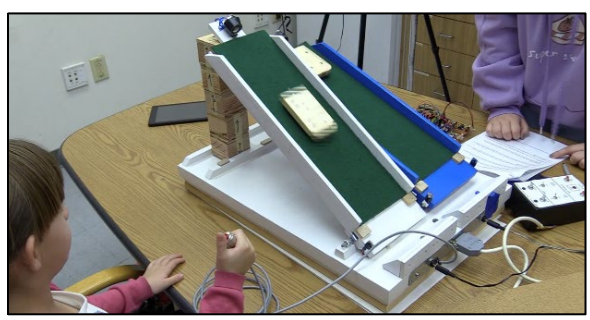
Screenshot of Slidea-ma-zoo. The player adjusts Harry and Sally's slide height via the up/down buttons and slide friction via the substances on the bottom of the screen.

Hands-on Task

The hands-on task mimicked the digital task in two ways. First, the general scenario of the game was adopted. In the hands-on task, the child races against the researcher—thus preserving the game-like feature. Both the researcher's slide and child's slide was set to a pre-determined height and surface. Second, a physical version was constructed to mimic the slide in Figure 1. The three critical features of the

Figure 2

Picture of the hands-on slide. The child races against the researcher. The child can adjust the height of the slide with blocks and change the surface friction by swapping out rames with different surfaces.



[2] https://pbskids.org/catinthehat/games/slidea-ma-zoo

[3] *Cat in the Hat Knows a Lot About That! and Slidea-ma-zoo* was produced by GBH in Boston

slides were (a) the ability to adjust the incline of the slide with the blocks such that a steeper incline (higher slide) resulted in the slide object going faster compared to a shallower incline; (b) the ability to adjust the friction of slide such that more rough surfaces resulted in the slide object going slower compared to smoother surfaces; and (c) the slide objects move slow enough to enable children to observe the winner of the race and also to observe which slide object is moving faster (e.g., a marble travels much too fast). Figure 2 shows the physical contraption.

Research and Engineering Objectives

Research Objectives. Our research objective was to examine if a mode effect exists between the digital and hands-on tasks in terms of children's task performance and task processes. While we are unaware of any research investigating such mode differences in assessments of young children, the findings of Lazonder and Ehrenhard (2013), Zacharia et al. (2012), and Manches et al. (2009) already suggest potential mode effects. If mode effects exist, then it would be important to understand the conditions under which a hands-on task (or a digital task) is more effective at eliciting evidence of young children's knowledge and skills. Thus, our research objectives were:

1. To develop an authentic, play-based task that involves a hands-on contraption that children can manipulate to demonstrate their understanding. We assumed that posing an assessment task as a game-like challenge would better maintain children's attention and interest in the task compared to more conventional methods such as being asked to predict the outcome of a scenario or explain what they would manipulate in the slide to win the race. Instead, they are simply asked to perform the task.
2. ***To examine whether and to what degree mode effects exists between the game-based task and the hands-on task*,** which is our main research question. Addressing this question would identify potential mode effects that would need to be considered when designing assessment tasks in general and digital and hands-on assessments in particular. In addition, if we find no mode effect, then digital tasks would be more practical and scalable than hands-on tasks in general.

Engineering Objectives. To support addressing the research question, and to address Challenge 1 to Challenge 3, our engineering objectives was to explore the feasibility and technical challenges of designing and constructing a proof-of-concept hands-on manipulative that could (after much further development and validation) address the Challenge 1 to Challenge 3. The key technical capability underlying a potential solution to Challenges 1 to 3 is the embedding of sensors into the contraption components the children manipulate. The sensors can provide fine-grained, moment-to-moment data about what each child is handling. Algorithms could then be developed to derive indicators of a child's degree of knowledge and skill and their use of problem-solving strategies from the fine-grained moment-to-moment data. Appendix A contains the requirements we imposed for the design of the contraption. Thus, our engineering design objectives were:

1. To design a hands-on version of the slide in *Slidea-ma-zoo* that required little instruction to operate and where the slide height and surface friction could be easily manipulated by children with little or no intervention by the researcher.
2. To embed the hands-on contraption with sensors to unobtrusively record what component in the slide is being manipulated and how it is being manipulated. This capability enables

automated observation and collection of telemetry (behavioral data), which supports scaling. Slide height, winner, and surface material would be logged automatically.

3. Use of algorithms to process the sensor data to yield measures of learning processes and learning outcomes. This capability would enable automated scoring of children's responses, which supports scaling. The measures include performance (did the child reach the goal of the task) and process (is the child using a systematic problem-solving strategy).

COVID-19 Impact.

The onset of COVID19 resulted in a halt to this study. Thus, while the physical contraptions was built, sensors embedded into the contraption, and the digital and analog circuits completed, the full telemetry data collection system was not completed. Thus, we were unable to collect telemetry from the hands-on contraption and could not address Engineering Objective 3.

Method

Data Sources

Due to COVID19, data were collected from only four participants. Participants were between 4 and 6 years old. The participants were recruited from postings on the UCLA campus. All participants were girls. Participants' manipulation of the physical slide contraption was video recorded and their performance coded. Participants' gameplay on the tablet was screen recorded and their performance coded.

Tasks

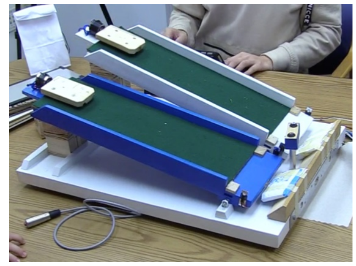
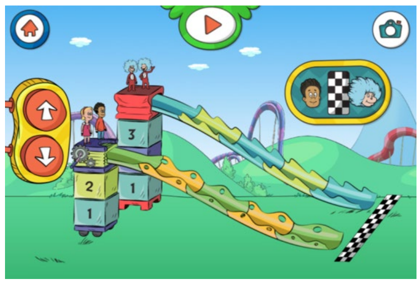
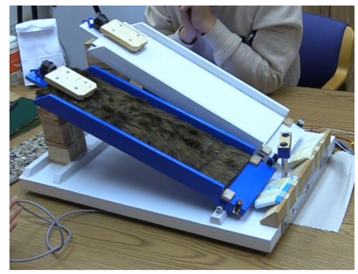
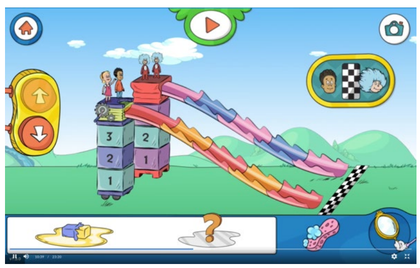
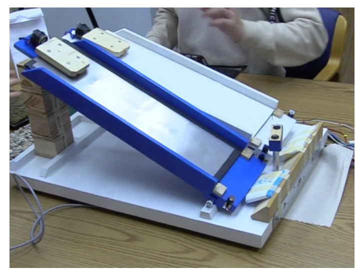
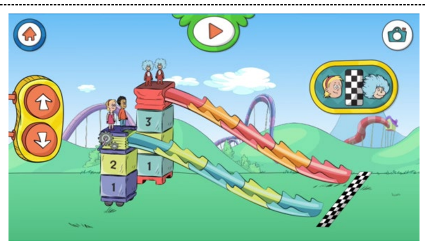
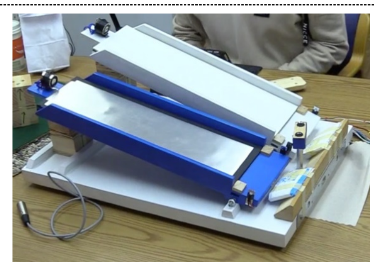
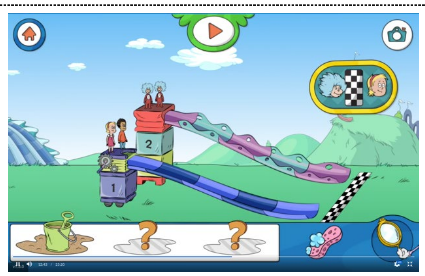
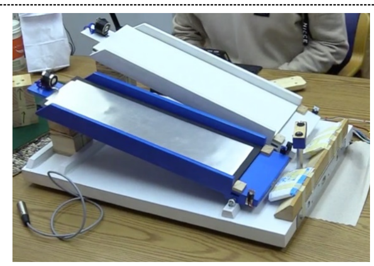
The main task was for children to manipulate a slide's height or surface friction to adjust the speed of the object on the slide. The slide task in the game, *Slidea-ma-zoo* (PBS KIDS Digital, 2020), was used for this study. A physical contraption was created to replicate the slide task.

Slide Digital Game

Slidea-ma-zoo presented the player with two slides: one that can be manipulated by the player and a reference slide, the height of which changes with the task. The game asks players to manipulate the slide (in terms of height and incline surface) to achieve a specified goal. Possible goals are the character sliding down the player's slide (a) reaching the bottom before the character on the reference slide, (b) reaching the bottom at the same time as the character on the reference slide, or (c) reaching the bottom after the character on the reference slide. The player can use a magnifying glass to examine the texture of different available incline surfaces, which have different frictional properties. The height of the slides is denoted by numbered blocks. Table 1 (left column) shows screen shots of the slide game.

Table 1

Slidea-ma-zoo *Game Levels and Corresponding Slide Contraption Initial Setup*

Level goal	Slidea-ma-zoo level	Slide contraption setup
Matched Level 1: Win		
Matched Level 2: Lose		
Matched Level 3: Lose		
Matched Level 4: Lose		
Matched Level 5: Win		

Hands-on Slide Contraption

A physical contraption was built to replicate the game slide as much as possible. The key game features that were replicated was to allow the child to manipulate the height of the slide and the friction of the slide surface. In the contraption stackable blocks were used to adjust the height, and removable surfaces with different friction (e.g., felt, cork, aluminum) were used to change the friction of the slide. The researcher's slide was used as the referent slide and the researcher posed as the opponent for the child to race against. Table 1 (right column) shows screen shots of the physical slide contraption.

Measures

We defined two measures that could be used to compare children's performance in the game and with the slide contraption (Huber et al., 2016).

Number of Steps

A step was defined as an adjustment, within a level, that a participant made to the slide before starting the race. For instance, in Slidea-ma-zoo, clicking the up-arrow button to raise the slide was considered a step. Correspondingly, with the slide contraption, a player's complete action of picking up one block and adding it to the slide contraption base was considered a step.

Number of Rounds

A round was defined as the activity between start of the race. For example, in Slidea-ma-zoo clicking on the play button marked a round. In the slide contraption, pressing on the start race switch marked a round.

Procedure

Participants played the *Slidea-ma-zoo* game for about 20 minutes. Participants then transitioned to the slide contraption. The participant was told they were racing against the researcher. Participants interacted with the slide contraption for about 20 minutes. The order of tasks varied. Participant 1, 2, and 4 used *Slidea-ma-zoo* first and then the slide contraption and Participant 3's order was the opposite.

For the slide contraption task, participants were given racing tasks that matched levels in *Slidea-ma-zoo*. The slide contraption tasks had the same goal state, same initial starting height, and the same type of allowable adjustments as the corresponding level in *Slidea-ma-zoo*. The starting height of the player's slide could be higher, lower, or the same as the competitor's slide. Allowable adjustments refer to whether a player could adjust the height, change surfaces, or both. Table 2 shows the *Slidea-ma-zoo* levels and corresponding slide contraption setup.

Results

Usability of the Slide Contraption

We consistently found that children could use the contraption with little or no instruction. We think the key design feature was to use objects that were familiar to them and operated in a way that they were familiar with. We designed the contraption to operate consistent with their prior exposure to *Slidea-ma-zoo* and presumably blocks and slides.

Task Performance

Data are presented in Table 2. In general, except for matched level 3, the slide contraption required the same or fewer rounds to beat the level. Similarly, the slide contraption required the same or fewer steps to complete a round. Matched levels 3 and 5 appeared to be more difficult for participants, regardless of mode. These levels required use of different surfaces. Interestingly, for level 3 the slide contraption required more rounds than *Slidea-ma-zoo* for all participants.

Table 2

Data for Participants 1, 2, 3 and 4 By Level

Level	Participant 1		Participant 2		Participant 3		Participant 4	
	No. of rounds	No. of steps in each round per level	No. of rounds	No. of steps in each round per level	No. of rounds	No. of steps in each round per level	No. of rounds	No. of steps in each round per level
Matched Level 1								
Slidea-ma-zoo	3	2, 2, 3	1	3	1	2	2	1, 19
Slide contraption	1	2	1	3	1	3	1	4
Matched Level 2								
Slidea-ma-zoo	1	2	2	2, 3	2	2, 2	1	6
Slide contraption	1	2	1	2	2	1, 3	1	5
Matched Level 3								
Slidea-ma-zoo	3	4, 4, 4	2	4, 2	1	6	1	4
Slide contraption	4	2, 2, 2, 2	3	4, 2, 2	8	2, 4, 3, 1, 1, 1, 1, 1	6	2, 1, 1, 2, 1, 2
Matched Level 4	Did not reach these levels				Did not reach these levels			
Slidea-ma-zoo			1	2			1	1
Slide contraption			1	2			1	5
Matched Level 5								
Slidea-ma-zoo			6	2, 3, 4, 4, 6, 8			1	4
Slide contraption			1	3			4	2, 2, 2, 2

Note. Each round is an attempt at a solution (i.e., a race). A step is the number of major actions within a round (e.g., increasing the height of the slide).

Qualitative Observations

Task Engagement

Our impressions of participants' engagement with both modes are puzzling. Participants displayed more overt emotions and curiosity when they were working with the slide contraption compared to when they were playing the game. For example, participants did not display much overt emotion when they were playing the game but they did appear to smile more, talk more, and gesture more when working with the slide contraption. However, when participants were asked at the end of the session which mode they preferred, all participants said they liked the tablet more.

Magnifying Glass Usage (Slidea-ma-zoo)

Participants 1 and 2 appeared to have difficulties using the magnifying glass (known as the Micro-ma-boodle in the game) in *Slidea-ma-zoo*. Both participants applied the magnifying glass to the slides instead of the texture items at the bottom of the stage. Participant 1 watched the entire magnifying glass instructions yet only applied the magnifying glass to the textures on the slides (and not the textures at the bottom of the stage). Participant 2 did not finish watching the magnifying glass instructions and tended to apply the magnifying glass to empty slides. These observations are consistent with the performance results suggesting difficulty on the levels involving textures.

Participant 3 and Participant 4 did not show difficulties with the magnifying glass function. However, like the other participants, Participants 3 and 4 appeared to have difficulties when the new tools related to textures were introduced (i.e., textures, sponge tool, magnifying glass, question mark icon to indicate hidden texture materials). For example, Participant 3 clicked on the butter texture multiple times to try to apply it to the slide, but they did not know they needed to hold onto the texture and drag it to the slide. Another example was when Participants 3 and 4 tried to drag the question mark icon to the slide prior to applying the magnifying glass (to reveal the texture covered by the question mark). Participant 4 also applied the sponge tool to the texture icons before realizing that the sponge tool should be applied to the texture on the slide.

*Preference for Height and Speed (*Slidea-ma-zoo*, Slide Contraption)*

During the slide contraption task, all participants showed a preference for raising the slide higher than needed to make the car go faster, and in general to make their car go faster regardless of the goal. For instance, during one level where the initial state of the participant's slide was 4 blocks high and the initial state of the researcher's slide was 2 blocks high, and the goal was to beat the researcher, the participant added one more block to her slide. This preference was also observed in the game, with Participants 3 and 4 clicking the up arrows (or down arrows) multiple times more than what was needed to beat the level.

Multi-step Moves

An interesting strategy was observed with two participants' use of the contraption. Participants 1 and 4 raised their slides two to three blocks at a time (vs. the more typical one at a time move by other participants, and the only strategy allowed in the game). Similar behavior was found in Manches et al.'s (2010) study where participants tended to move stacks of blocks together.

Discussion

Limitations

This study is limited by the low number of participants. This is primarily the result of the onset of COVID19, which shut down UCLA and all research activities. Thus, the results are very exploratory. The first limitation is the low number of participants. Thus, we do not know how representative the data in Table 2 or the qualitative analysis to the general population.

Second, we did not counter-balance the order of task administration. It is unclear how much the gameplay influences participants' performance on the hands-on task. Our sample was a usability sample where were examining timing of the digital and hands-on tasks, clarify of instructions, degree of engagement, and mode differences.

Third, the tasks were not identical. For example, in *Slidea-ma-zoo* the player drags a substance on to the slide to change the friction whereas the contraption required the player to swap surfaces. In addition, the game substances that could not be used in the hands-on version (e.g., butter, ice, sand, honey). The contraption was very much a prototype, bulky, did not use any of the game art or characters, and the hands-on task itself was not situated in a fantasy world setting.

Research Objectives

The research objective of this study was to explore to what extent children performed differently depending on mode. Were there mode differences that affected how young children demonstrated their knowledge and skills? Given the limitation, no firm conclusions can be drawn about mode differences, the following speculations may serve to guide future work: (a) the number of rounds to beat a level between the two modes were not drastically different. Participants could complete levels in both modes; (b) the number of steps to beat a round, however, appear to be higher in the game than the slide contraption; and (c) the levels that require consideration of slide friction may be more difficult than levels that only require consideration of height.

We think that the digital format may be interfering with children's acquiring a sense of friction. In the game, friction is conveyed through images of materials (e.g., ice, honey), whereas in the contraption children could readily feel the roughness of the different surfaces. When to use physical manipulatives as an assessment format may largely depend on if sensory feedback is important. That is, when students have little (physical) experience with abstract concepts such as weight, magnetic fields, and gears—and friction—physical contraptions may be a more appropriate format than a digital format. These concepts are abstract and cannot be easily perceived through visual modalities (Lazonder & Ehrenhard, 2014; Rau, 2020; Zacharia, 2015; Zacharia et al., 2012).

Engineering Objectives

Our major engineering objectives were met with the exception of the developing algorithms to derive indicators of knowledge and skill from the telemetry (i.e., telemetry was unavailable). However, we are confident that this is achievable because we have done such analysis in other areas (e.g., analysis of telemetry from digital games, analysis of telemetry from sensors; Chung 2005a, 2005b, 2007a, 2007b, 2007c, 2015; Chung et al., 2014; Chung & Parks, 2019; Nagashima, 2009a, 2009b; Parks et al., 2008)

Slide Contraption Construction. One of the most surprising outcomes of this proof-of-concept effort was that a contraption could be built that largely hid the implementation details (e.g., sensors, wires) and allowed the child to interact with the contraption with little instruction. The construction of the slide contraption, embedded sensing, event capture, and data logging were all clearly feasible. Everything was done with non-specialized machinery. With specialized machinery (e.g., a CNC router) and different materials, we are confident that a smaller, lighter, more elegant, and more intuitive contraption can be constructed.

Usability of the Slide Contraption. We consistently found that children could use the contraption with little or no instruction. We think the key design feature was to use objects that were familiar to them and operated in a way that they were familiar with.

Implications for Assessment

Our main goal was to examine the idea of using a physical contraption for measurement purposes primarily to address the challenges of testing young children. We think this approach is consistent with calls for authentic assessment in early childhood (e.g., Snow & Van Hemel, 2008). One long-term implication is that such performance tasks could be administered by local caregivers or staff at community events, in formative evaluation studies, or even usability studies to gather limited performance information.

The second implication is if format differences do exist, then we can expect continuing work in this area particularly because of the emphasis of integrating STEM into the early grades and the continuing demand for accountability; thus, this work will inform assessment designers about the limitations, advantages, and trade-offs between digital and hands-on task designs.

Third, using sensors in hands-on assessments is no different from logging data in interactive systems such as games. Capturing children's moment-to-moment interactions, whether in a game or with physical manipulatives, is conceptually identical and the analytics used to process gameplay data can be used with sensor data.

Fourth, a demonstration of new observational capability using embedded sensing in hands-on tasks may spur new research and development in early childhood (e.g., learning analytics for play and other informal activities), potentially leading to wider adoption of hands-on performance tasks and its integration with different educational media.

Implications for Instruction

Recent research suggests that children can transfer what they learned using virtual manipulatives to the use of physical manipulatives. Huber et al. (2016) found that children (4 to 6 years old) who learned a problem-solving task (Tower of Hanoi) on a tablet could transfer their learning to a physical version of the same task (Huber et al., 2016; Tarasuik et al., 2017). When the virtual and physical tasks are isomorphic—identical tasks differing only in mode—then there does not appear to be a performance difference between the modes. This finding suggests that children will be able to transition between virtual tasks (e.g., games) and physical tasks if they are structurally similar.

Including physical manipulatives as part of an intervention may be beneficial when tactile feedback is important to learning and understanding the target concept. Touch sensory feedback is particularly important when students have little (physical) experience with force-related concepts such as weight, magnetic fields, and gears. These concepts are abstract and cannot be easily perceived through visual modalities. Furthermore, when children have misconceptions related to object characteristics that can be perceived by touch, being able to physically manipulate the materials appears to be key to altering children's beliefs (Lazonder & Ehrenhard, 2014; Rau, 2020; Zacharia, 2015; Zacharia et al., 2012).

Finally, Zacharia and Michael (2016) tested the effects of combined or single modalities with sixth-grade students learning about electrical circuits. Although the sample was much older than the sample in this study, Zacharia and Michael's study sample is the closest in age we could find examining the relative benefits of physical and virtual manipulatives in science-related concepts. The study helps clarify the relative benefits of physical and virtual manipulatives. Students were assigned to a hybrid condition where they had access to both physical components and a circuit simulator, a physical components-only condition, or a circuit simulator-only condition. The students in the hybrid condition performed highest on a test of conceptual understanding that involved circuit reasoning as well as building actual circuits, compared to the other conditions; there was no difference between the other two conditions. The authors attribute the superior performance of the hybrid condition to students profiting from the unique affordances of both modalities. With the physical components, students learned how to set up a circuit and with the circuit simulator, students could rapidly test and receive feedback about their circuits.

Implications for Linking Digital and Physical Play

Once information can be gathered from physical contraptions, in principle a analytics platform could use this additional stream to coordinate gameplay and physical play to increase interaction with the child. For example, children could be prompted within the game to engage with the physical contraption. Based on what is occurring with the physical contraption, the game could scaffold to the child with additional gameplay or challenge the child to attempt a slightly more difficult physical task.

With a sufficiently powerful microcontroller and input-output components, the physical contraption could also relay information from the analytics platform or generate its own feedback. In all cases, the key capability is synchronized data streams that accurately represent player behavior and the state of the both the game and physical contraption. An example of integrating virtual and physical worlds is the Intelligent Science Station (ISS) developed by Yannier and colleagues (Yannier et al., 2016; Yannier et al., 2020). ISS is a mixed-reality system that uses vision processing to do scene detection of children's interaction with a tabletop contraption.

References

Atkins-Burnett, S. (2007). *Measuring children's progress from preschool through third grade. Mathematica Policy Research*. https://www.mathematica-mpr.com/-/media/publications/pdfs/measchildprogress.pdf

Brennen, K. (2011). Assessment for preschool science learning and learning environments. *Early Childhood Research and Practice, 13*(1). https://ecrp.illinois.edu/v13n1/brenneman.html

Chung, G. K. W. K. (2005a). *Wobble sensor data acquisition* [Computer software (.NET Visual Basic)]. Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Chung, G. K. W. K. (2005b). *Trigger sensor data acquisition software* [Computer software (.NET Visual Basic)]. Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Chung, G. K. W. K. (2007a). *Netmedia BX-24 Microcontroller-based sensor data acquisition software*. [Computer software (Netmedia Basic)]. Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Chung, G. K. W. K. (2007b). *Wireless sensor data acquisition prototype unit for measuring rifle marksmanship skill (wobble, trigger, breathing)* [Electronic hardware]. Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Chung, G. K. W. K. (2007c). *Windows-based sensor data acquisition and display software*. [Computer software (.NET Visual Basic)]. Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Chung, G. K. W. K. (2015). Guidelines for the design, implementation, and analysis of game telemetry (pp. 59–79). In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious games analytics: Methodologies for performance measurement, assessment, and improvement*. New York: Springer.

Chung, G. K. W. K., & Parks, C. B. (2019). *Pre-prototype instrumented hands-on ramp*. [Electronic hardware]. Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Chung, G. K. W. K., Choi, K-C., Baker, E. L., & Cai, L. (2014). *The effects of math video games on learning: A randomized evaluation study with innovative impact estimation techniques* (CRESST Report 841). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Chung, G. K. W. K., Parks, C. B., Redman, E. J. K. H., Choi, K., Kim, J., Madni, A., & Baker, E. L. (2016). *PBS KIDS final report* (Deliverable to PBS KIDS). University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Clements, D., Greenfield, D. B., Landry, S. H., & Sarama, J. (2015). Assessment using technology: Formative assessment with young children. In O. Saracho (Ed.). *Contemporary perspectives on research in assessment and evaluation in early childhood education* (pp. 339–372). Information Age Publishers.

Greenfield, D. B. (2015). Assessment in early childhood science education. In K. Cabe Trundle, M. Sackes (Eds.), *Research in early childhood science education* (pp. 353–380). Springer.

Griffith, S. F., Hagan, M. B., Heymann, P., Heflin, B. H., & Bagner, D. M. (2020). Apps As Learning Tools: A Systematic Review. *Pediatrics*, *145*(1). https://doi.org/10.1542/peds.2019-1579

Grindal, T., Silander, M., Gerard, S., Maxon, T., Garcia, E., Hupert, N., Vahey, P., Pasnik, S. (2019). *Early science and engineering: The impact of The Cat in the Hat Knows a Lot About That! on learning*. Education Development Center, Inc., & SRI International.

Herodotou, C. (2018). Young children and tablets: A systematic review of effects on learning and development. *Journal of Computer Assisted Learning*, *34*, 1–9. https://doi.org/10.1111/jcal.12220

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting Education in "Educational" Apps: Lessons From the Science of Learning. *Psychological Science in the Public Interest*, *16*, 3–34. https://doi.org/10.1177/1529100615569721

Howard, S. J., & Melhuish, E. (2017). An early years toolbox for assessing early executive function, language, self-regulation, and social development: Validity, reliability, and preliminary norms. *Journal of Psychoeducational Assessment*, *35*, 255–275.

Huber, B., Tarasuik, J., Antoniou, M. N., Garrett, C., Bowe, S. J., Kaufman, J., & Team, S. B. (2016). Young children's transfer of learning from a touchscreen device. *Computers in Human Behavior*, *56*, 56–64.

Klein, A., & Starkey, P. (2012). *Psychometric study of the Child Math Assessment. Final report on scaling up the implementation of a pre-kindergarten mathematics curriculum in public preschool programs* (IES Grant R305K050004).

Lazonder, A. W., & Ehrenhard, S. (2014). Relative effectiveness of physical and virtual manipulatives for conceptual change in science: How falling objects fall. *Journal of Computer Assisted Learning, 30*, 110–120.

Lieberman, D. A., Fisk, M. C., & Biely, E. (2009). Digital games for young children ages three to six: From research to design. *Computers in the Schools, 26*, 299-313.

Manches, A., O'Malley, C., & Benford, S. (2009). Physical manipulation: Evaluating the potential for tangible designs. *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction - TEI '09*, 77. https://doi.org/10.1145/1517664.1517688

Manches, A., O'Malley, C., & Benford, S. (2010). The role of physical representations in solving number problems: A comparison of young children's use of physical and virtual materials. *Computers & Education, 54*, 622–640.

McKown, C., Russo-Ponsaran, N. M., Allen, A., Johnson, J. K., & Warren-Khot, H. K. (2016). Social–emotional factors and academic outcomes among elementary-aged children. *Infant and Child Development*, *25*, 119–136.

Meloy, B., & Schachner, A. (2019). *Early childhood essentials: A framework for aligning child skills and educator competencies.* Learning Policy Institute.

Metz, K. E. (1993). Preschoolers' developing knowledge of the pan balance: From new representation to transformed problem solving. *Cognition and Instruction, 11,* 31–93. https://www.tandfonline.com/doi/abs/10.1207/s1532690xci1101_2

Nagashima, S. O., Chung, G. K. W. K., Espinosa, P. D., & Berka, C. (2009a). Sensor-based assessment of basic rifle marksmanship. *Proceedings of the I/ITSEC*, Orlando, FL.

Nagashima, S. O., Chung, G. K. W. K., Espinosa, P. D., & Berka, C. (2009b). Validity evidence for a model of rifle marksmanship skill performance using sensor-based measures. *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society*, San Antonio, TX.

National Research Council. (2001). *Eager to learn: Educating our preschoolers*. National Academy Press.

Parks, D., & Chung, G. K. W. K. (2008). *Wireless sensor data acquisition unit for measuring rifle marksmanship skill (wobble, trigger, breathing)* [Electronic hardware]. Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

PBS KIDS Digital. (2020, August, 1). *Slidea-ma-zoo*. https://pbskids.org/catinthehat/games/slidea-ma-zoo

Rau, M. A. (2020). Comparing multiple theories about learning with physical and virtual representations: Conflicting or complementary effects? *Educational Psychology Review,* 1–29. https://doi.org/10.1007/s10648-020-09517-1

Redman, E. J. K. H., Chung, G. K. W. K., Parks, C. B., Feng, T., Schenke, K., & Michiuye, J. K. (2020). *Cat in the Hat Builds That Analytics Validation Study—v2* (Deliverable to PBS KIDS). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Redman, E. J. K. H., Chung, G. K. W. K., Schenke, K., Maierhofer, T., Parks, C. B., Chang, S. M., Feng, T., Riveroll, C. S., & Michiuye, J. K. (2018). *Connected learning final report* (Deliverable to PBS KIDS). University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Redman, E. J. K. H., Schenke, K., Chung, G. K. W. K., Parks, C. B., Michiuye, J. K., Feng, T., & Chang, S. M. (2019). *Analytics validation final report* (Deliverable to PBS KIDS). University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Redman, E. J. K. H., Schenke, K., Chung, G. K. W. K., Parks, C. B., Michiuye, J. K., Feng, T., Chang, S. M., & Roberts, J. (in-press). To adapt or not to adapt: Evidence from a digital physics game. In H. F. O'Neil, E. L. Baker, R. S. Perez, & S. E. Watson (Eds.), *Using cognitive and affective metrics in education-based simulations and games.* Routledge/Taylor & Francis.

Riley, K., Miller, G. E., & Sorenson, C. (2016). Early childhood authentic and performance-based assessment. In A. Garro (Ed.), *Early childhood assessment in school and clinical child psychology* (pp. 95–117). Springer.

Roberts, J. D., Chung, G. K. W. K., & Parks, C. B. (2016). Supporting children's progress through the PBS KIDS learning analytics platform. *Journal of Children and Media, 10*, 257–266.

Roberts, J. D., Parks, C. B., Chung, G. K. W. K., Redman, E. J. K. H., Schenke, K., & Felline, C. (2019). Innovations in evidence and analysis: The PBS KIDS Learning Analytics Platform and the research it supports. In S. Pasnik (Ed.), *Getting ready to learn: Creating effective, educational children's media* (pp. 231–248). Routledge.

Ruzek, E., Jirout, J., Schenke, K., Vitiello, V., Whittaker, J. V., & Pianta, R. (2020). Children's academic orientation scale. Databrary. http://doi.org/10.17910/b7.1070

Samarapungavan, A., Mantzicopoulos, P., Patrick, H., & French, B. (2009). The development and validation of the science learning assessment (SLA): A measure of kindergarten science learning. *Journal of Advanced Academics, 20*, 502–535.

Schenke, K., Redman, E. J. K. H., Chung, G. K. W. K., Chang, S. M., Feng, T., Parks, C. B., & Roberts, J. D. (2020). Does "Measure Up!" measure up? Evaluation of an iPad app to teach preschoolers measurement concepts. *Computers and Education, 146.* https://doi.org/10.1016/j.compedu.2019.103749

Scott-Little, C., & Niemeyer, J. (2001). *Assessing kindergarten children: What school systems need to know.* SERVE. http://www.serve.org/_downloads/publications/rdakcg.pdf

Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early childhood assessment: Why, what, and how*. National Academies Press.

Tarasuik, J., Demaria, A., & Kaufman, J. (2017). Transfer of problem solving skills from touchscreen to 3D model by 3-to 6-year-olds. *Frontiers in Psychology, 8*, 1586.

Wechsler, M., Melnick, H., Maier, A., & Bishop, J. (2016). *The building blocks of high-quality early childhood education programs (policy brief)*. Learning Policy Institute.

Xie, H., Peng, J., Qin, M., Huang, X., Tian, F., & Zhou, Z. (2018). Can touchscreen devices be used to facilitate young children's learning? A meta-analysis of touchscreen learning effect. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02580

Yannier, N., Hudson, S. E., Wiese, E. S., & Koedinger, K. R. (2016). Adding physical objects to an interactive game improves learning and enjoyment: Evidence from Earthshake. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *23*(4), 1–31.

Yannier, N., Hudson, S. E., & Koedinger, K. R. (2020). Active learning is about more than hands-on: A mixed-reality AI system to support STEM education. *International Journal of Artificial Intelligence in Educatio*n, *30*, 74–96.

Zacharia, Z. C. (2015). Examining whether touch sensory feedback is necessary for science learning through experimentation: A literature review of two different lines of research across K-16. *Educational Research Review, 16*, 116–137.

Zacharia, Z. C., & Michael, M. (2016). Using physical and virtual manipulatives to improve primary school students' understanding of concepts of electric circuits. In M. Riopel & Z. Smyrnaiou (Eds.) *New developments in science and technology education: Innovations in science education and technology* (pp. 125–140). Springer.

Zacharia, Z. C., Loizou, E., & Papaevripidou, M. (2012). Is physicality an important aspect of learning through science experimentation among kindergarten students? *Early Childhood Research Quarterly, 27*, 447–457.

Appendix A
Requirements for the Slide Hands-on Task (Redman et al., 2020, p. 185)

Table 33

Requirement 1 (Adjustable Slide Height and Surface Friction): Features of the Slide Contraption, Implementation, and Rationale

Slide contraption feature	Implementation	Design rationale
Features related to the assessment target		
Stackable blocks to adjust the height of a slide	Blocks were fabricated with magnetic attachment points. See Figure 37 (rear view) and Figure 38.	• Blocks are used in Slidea-ma-zoo. • Blocks are familiar to young children. • Magnets provide a secure connection while also being easy to disconnect. See Figure 38.
Slide surfaces with different friction properties	Various slide surfaces were fabricated to have different surfaces, allowing the surface of each slide to be quickly changed. See Figure 39.	• Replacing a surface is easy for children to understand. • Each surface can have a unique ID associated with it. • Easy to fabricate.
Features unrelated to the assessment target		
Platform	A base was constructed to provide a fixed connection point for the slides and blocks. See Figure 37 (rear view).	• A base simplifies administration (kit format). • A base minimizes the number of loose parts that could be lost or broken, or cause the child to be distracted.
Slides-to-base connection	Both slides were connected to the platform with a single rod. The rod provided a common axis of rotation for both slides. The slides could only move up or down (i.e., pivot about the rod). See Figure 37 (front view).	• Ensure consistency and accuracy of the race condition. Because the angle of the slide influences the speed of the object, having the slides pivot about the same axis is critical.
Block-to-base connection	A channel was created to house the base block of each slide. The block was connected to the platform using a magnetic connection. See Figure 37 (rear view).	• A fixed location for the base block ensured consistency of the slide incline by fixing the distance from the pivot point of the slide to the base of the blocks.
Car object	A "car" object was fabricated and used for racing. The object had two parallel sleds made of aluminum. The object had a magnetized screw that was used to attach the object to an electromagnetic latch. See Figure 40.	• A sled design was used to facilitate the object traveling straight down the slide. • Aluminum was used for the sleds because of the use of magnets in the object itself and the use of magnets in the surfaces.

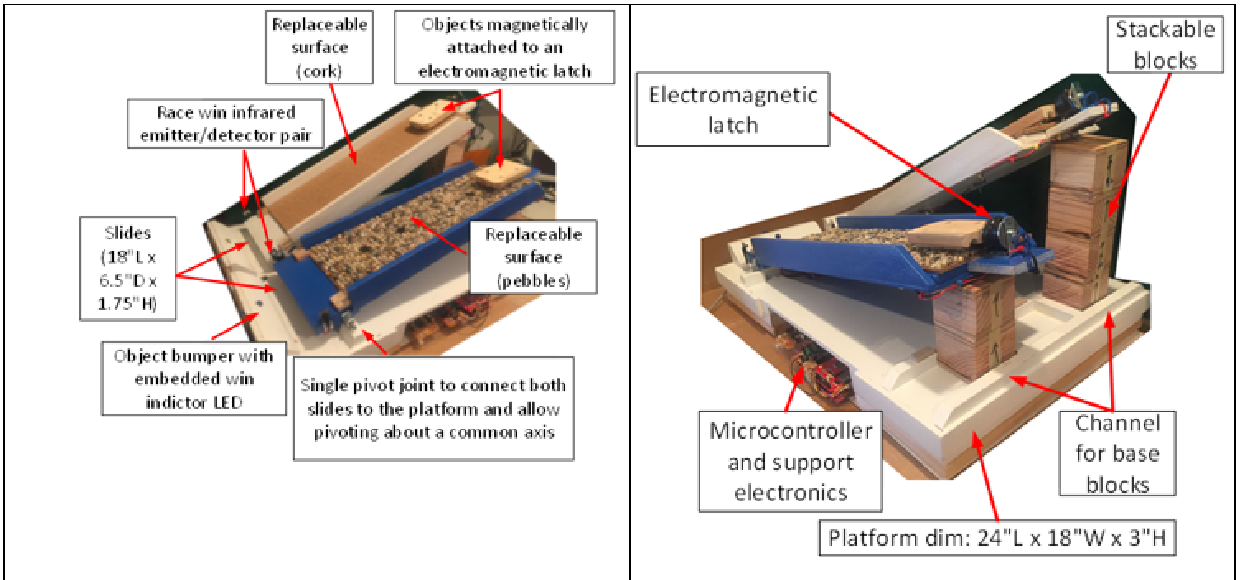
Table 34

Requirement 2 (Unobtrusive Observation): Features of the Slide Contraption, Implementation, and Rationale

Slide contraption feature	Implementation	Design rationale
Sensing slide height	Each block contained a resistor that was connected to screws and magnets in each block (see Figure 42) such that the stacking of blocks formed a ladder circuit (i.e., resistors in parallel).	• The total resistance of each stack of blocks represented the height of the slide. The higher the slide, the more resistors were in parallel, and the lower the total resistance. • The total resistance of the blocks could be sensed to indicate the height of the slide. • Event generation to indicate a change in the total resistance of the block was not implemented due to time constraints. • A passive design was used to avoid power.
Sensing slide friction	Each slide surface had a unique ID associated with it. The ID was set by embedding magnets into the underside of the slide surface. The magnetic field was sensed using reed switches embedded in the slide base. Three magnets were used to form a binary value between 1 and 7. One magnet was used to serve as an event trigger. See Figure 43.	• The combination of magnets and reed switches allowed us to easily sense which surface was present on each slide. • The state of the reed switches (open or closed) could be sensed at any time to indicate which surface was on the slide. • An event was generated when the slide surface was removed from or placed on the slide base. • A passive design was used to avoid power.
Sensing the start of the race	The start of the race was initiated by the child pressing a handheld push-button switch. The closing of the switch energized the electromagnets on each slide. The energized electromagnet released the attached car. See Figure 44.	• Use of an electromagnet allowed both car objects to be released simultaneously. • The electromagnet also has a ferrite core, which allowed for the car object to be magnetically attached to the electromagnet. See Figure 44. • An event was generated when the child pressed the push-button switch. • The electromagnet required power only when it was energized.
Sensing when the object reached the end of the slide	An infrared LED emitter/detector pair was used to sense when the object reached the end of the slide. The detector was attached to the slide and not the platform. See Figure 37 (front view).	• An infrared emitter/detector was chosen because of its simplicity. • An event was generated when a break in the infrared beam was detected. • The detector pair was attached to the slide to eliminate any obstruction a platform-mounted detector might create.

Figure 37

Front and Rear Views of the Slide Contraption



Front view of slide contraption

Rear view of contraption

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)

School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522

(310) 206-1532
www.cresst.org