CRESST Connected Learning Study

Novel Research Methodologies, and the Implications for PBS KIDS

Jeremy Roberts, September, 2018

All materials created that are funded by RTL needs to include the following funding language for this grant cycle: The contents of this document were developed under a cooperative agreement #PRU295A150003, from the U.S. Department of Education. However, these contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government. PBS KIDS® and the PBS KIDS® Logo are registered trademarks of PBS. Used with permission.

Overview

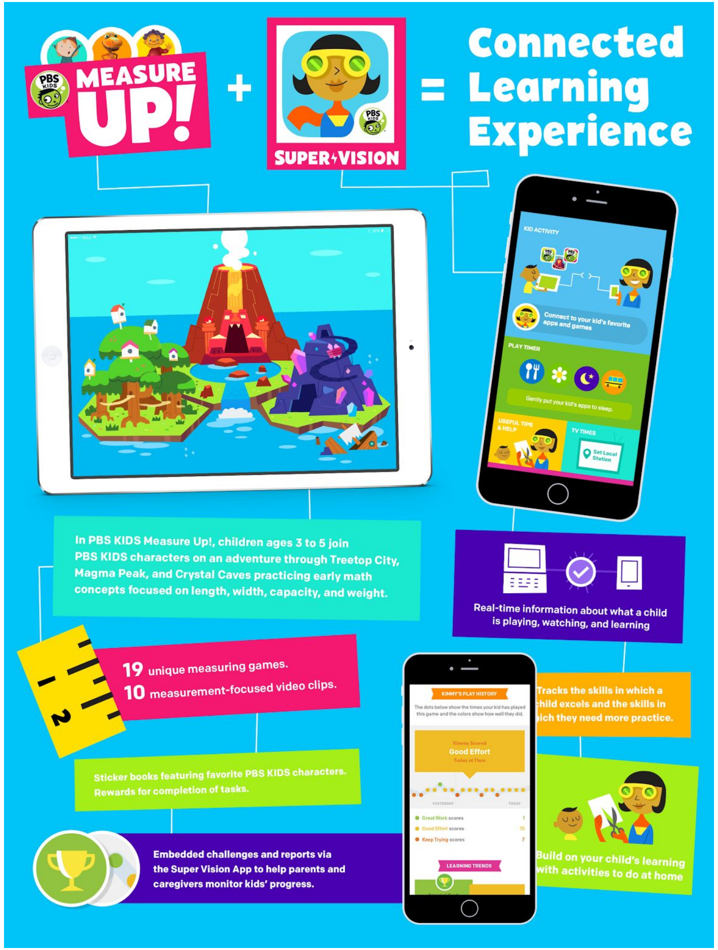
- Exploratory: Super Vision Usage
- Exploratory: Detailed Engagement Analyses
- Exploratory: Relating Specific Gameplay to Gains
- Exploratory: Detecting Misconceptions
- Implications!

Overview

- Authors: Elizabeth J. K. H. Redman, Gregory K. W. K. Chung, Katerina Schenke, Thomas Maierhofer, Charles B. Parks, Sandy M. Chang, Tianying Feng, Claudia S. Riveroll, and Joanne K. Michiuye
- PI: Gregory K. W. K. Chung

PBS KIDS Measure Up! and Super Vision

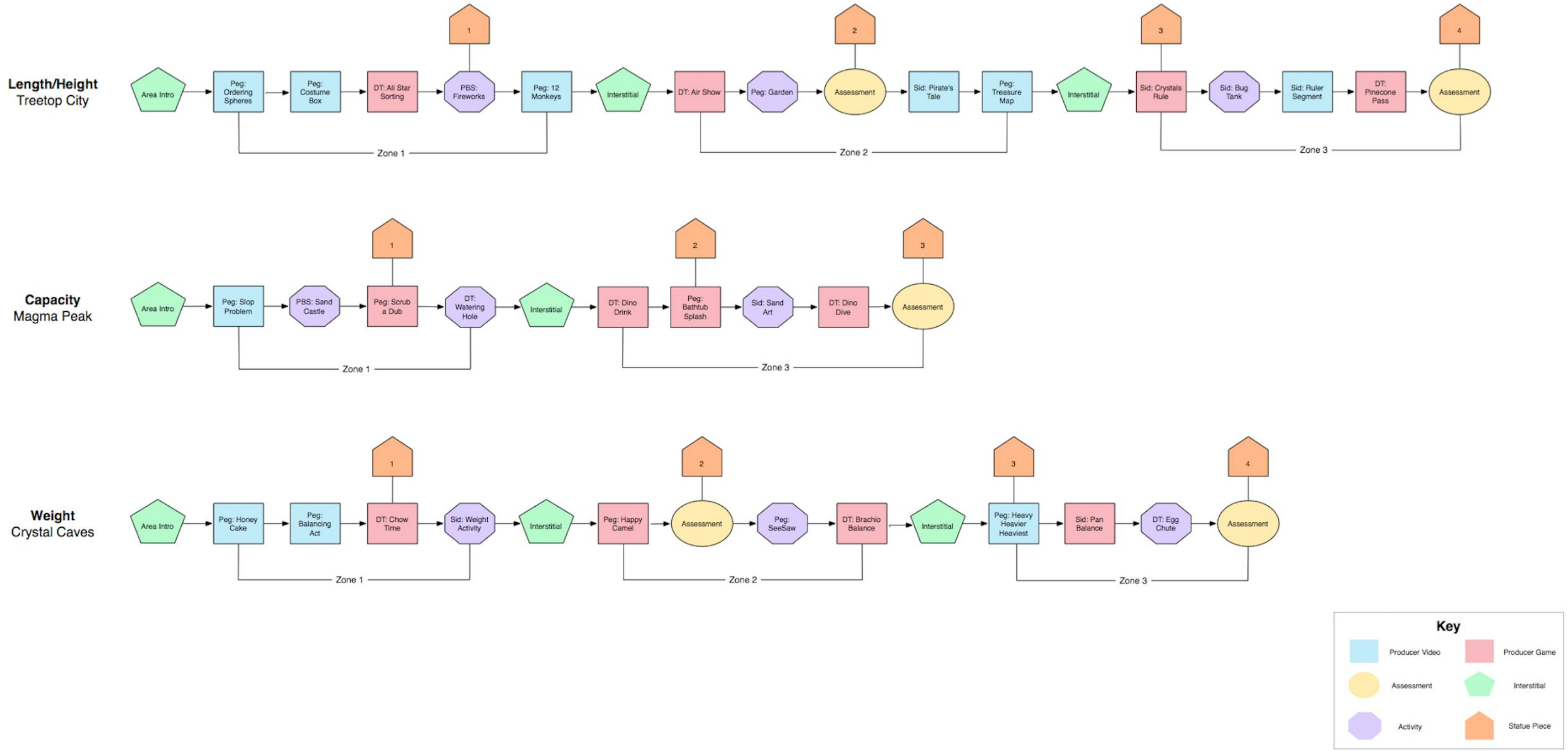
- Focus on Measurement for 3-5yo: Length + height, capacity, weight
- Adventure narrative
- Video + Sandboxes + Games + Challenges
- Modeling + Exploration + Instruction, Practice, Feedback, Assessment
- Multiple operationalizations of framework constructs
- Real-time monitoring of activity stream for parents
- Progress and performance-based reporting
- Assessment in service of learning: reporting through adults / family unit


PBS KIDS
MEASURE UP!
+
SUPER VISION
=
Connected Learning Experience
KID ACTIVITY
PLAY TIMER
USEFUL TIPS & HELP
TV TIMES
In PBS KIDS Measure Up!, children ages 3 to 5 join PBS KIDS characters on an adventure through Treetop City, Magma Peak, and Crystal Caves practicing early math concepts focused on length, width, capacity, and weight.
Real-time information about what a child is playing, watching, and learning
19 unique measuring games.
10 measurement-focused video clips.
Tracks the skills in which a child excels and the skills in which they need more practice.
Sticker books featuring favorite PBS KIDS characters. Rewards for completion of tasks.
Build on your child's learning with activities to do at home
Embedded challenges and reports via the Super Vision App to help parents and caregivers monitor kids' progress.


THE SUPER VISION APP
KID ACTIVITY
Connect to your kid's favorite apps and games
PLAY TIMER
Gently put your kid's apps to sleep.
USEFUL TIPS & HELP
TV TIMES
Set Local Station
KIMMY'S PLAY HISTORY
The dots below show the times your kid has played this game and the colors show how well they did.
Kimmy Scored
Good Effort
Today at 11am
YESTERDAY
TODAY
Great Work scores 1
Good Effort scores 15
Keep Trying scores 7
LEARNING TRENDS

PBS KIDS Measure Up! and Super Vision

PBS KIDS Measure Up! and Super Vision



Exploratory: Super Vision Usage

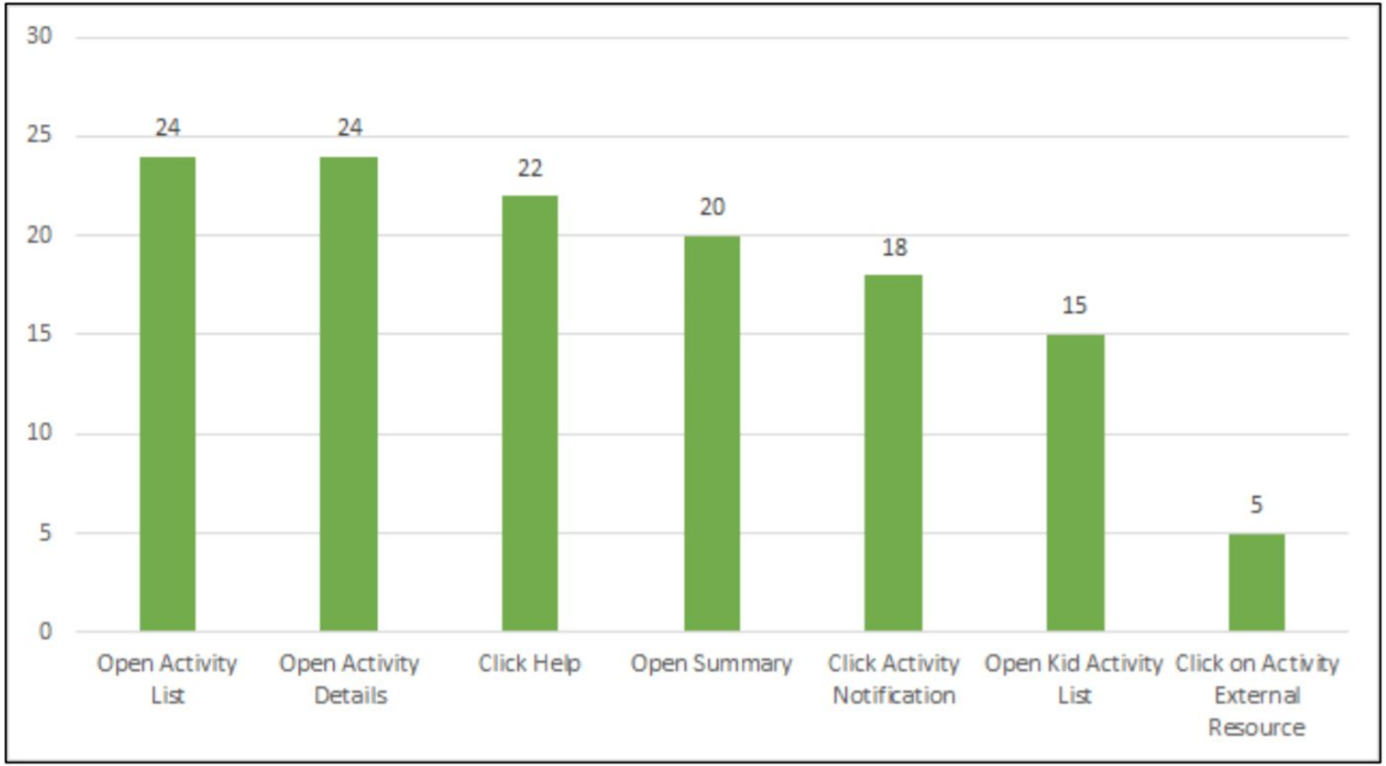


Figure 2. Parents' usage of Super Vision.

Exploratory: Super Vision Usage

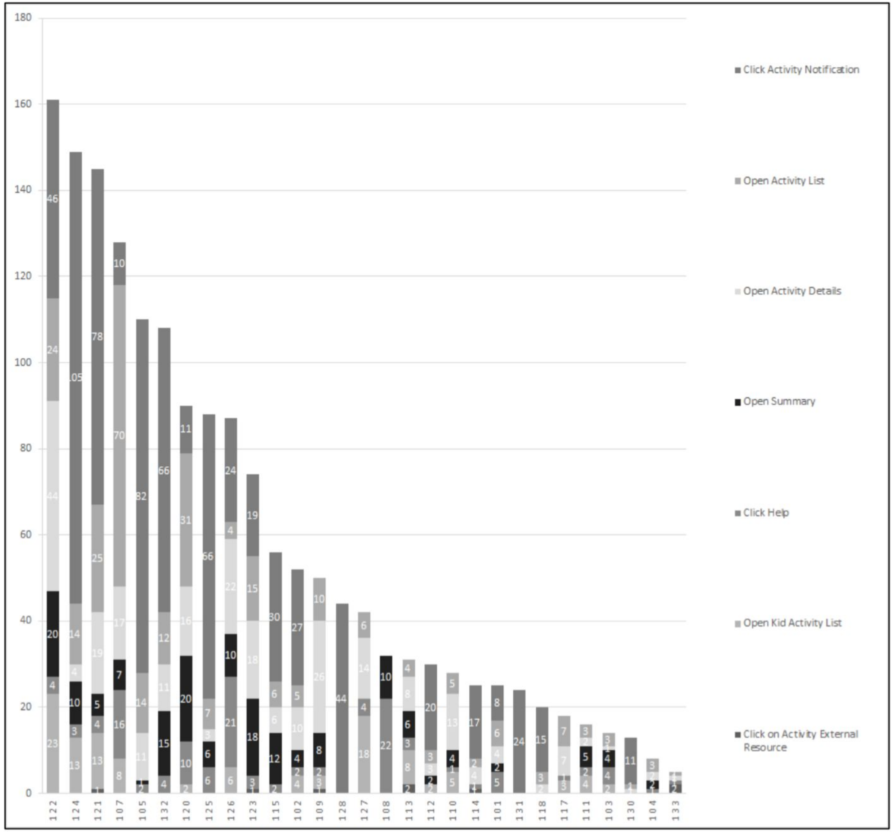


Figure 3. Individual parents' usage of Super Vision.

Super Vision Usage - Bottom Line

- Fall off observed as you look at more and more engagement with kid learning
- Wide behavior variation across parents
- About a third have more robust activity pattern
- We were able to see and analyze the details of what the parents did

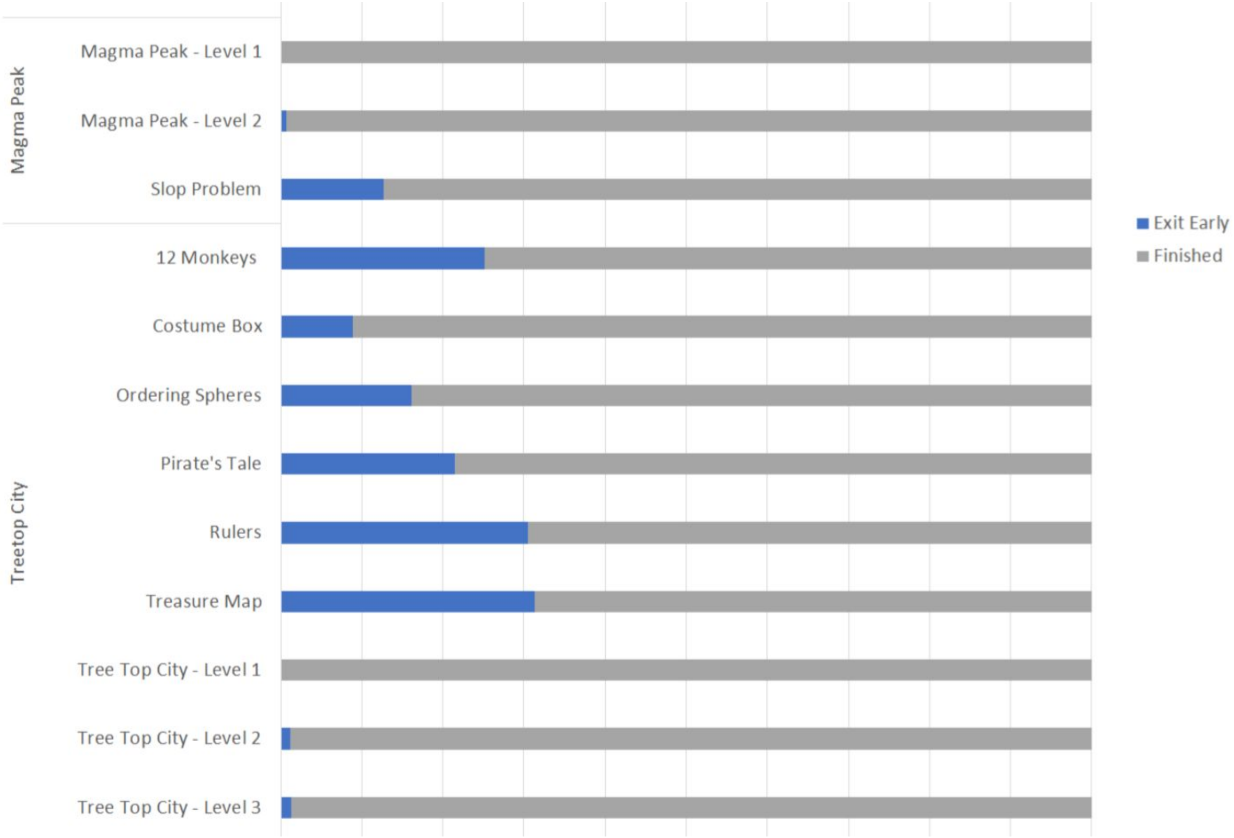
Super Vision Usage - Implications

- Help get to the bottom of the discrepancy between parent self-report and behavior
- Potential to relate specific parent activity to learner gains
- Supports data driven approaches to entire learning ecosystem (kids, parent, teacher)

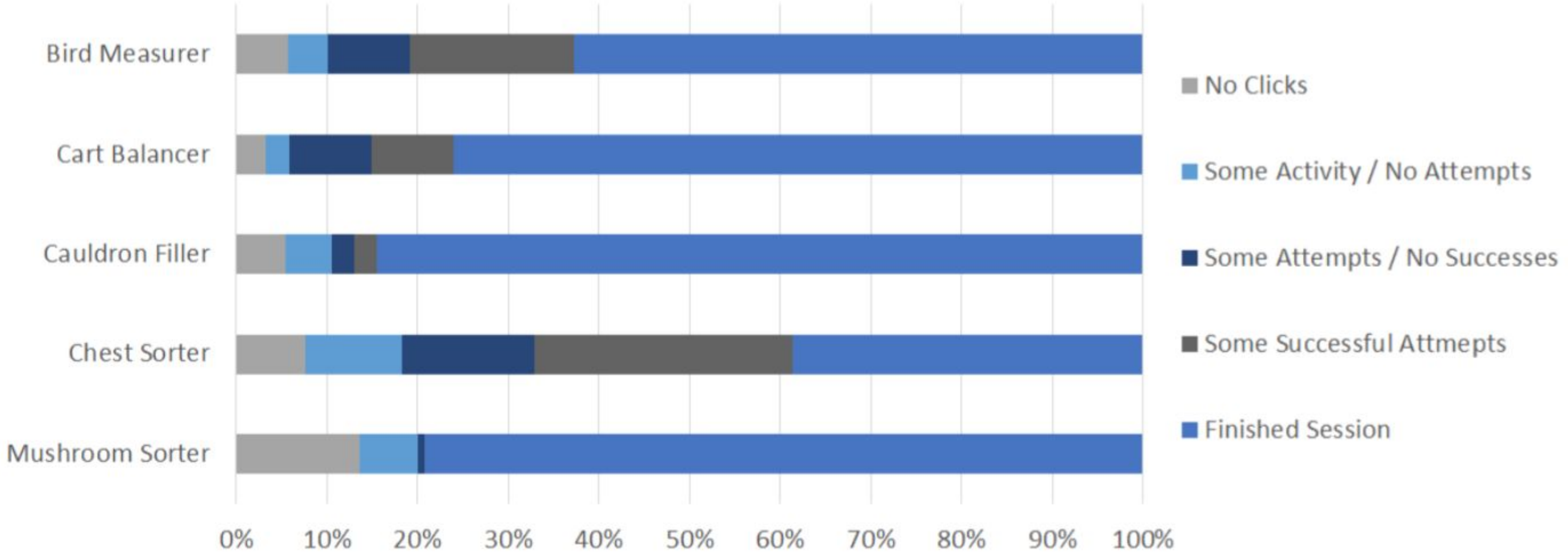
Exploratory: Detailed Engagement Analyses

- Flexibility in learner choice allows us to understand preferences
- Average and share of time
  - by world, activity, type
- Depth of engagement
- Average number of errors

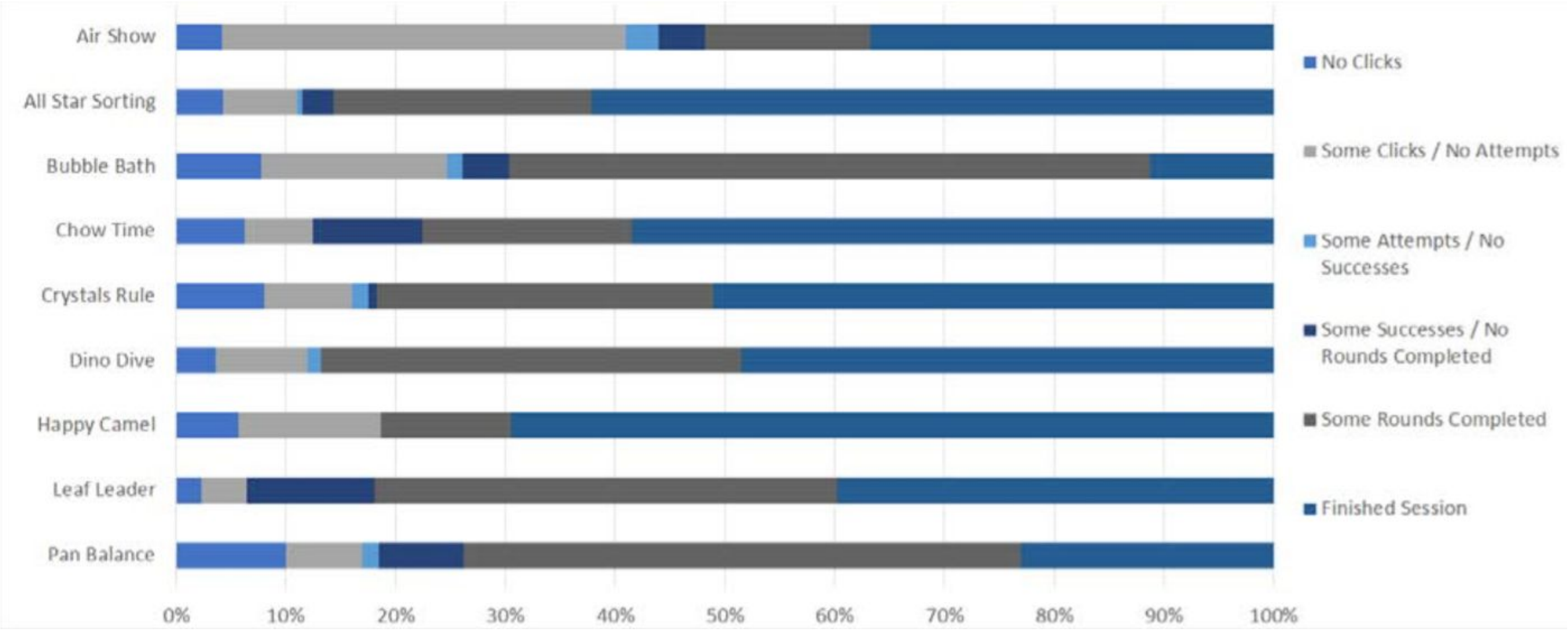
Exploratory: Detailed Engagement Analyses



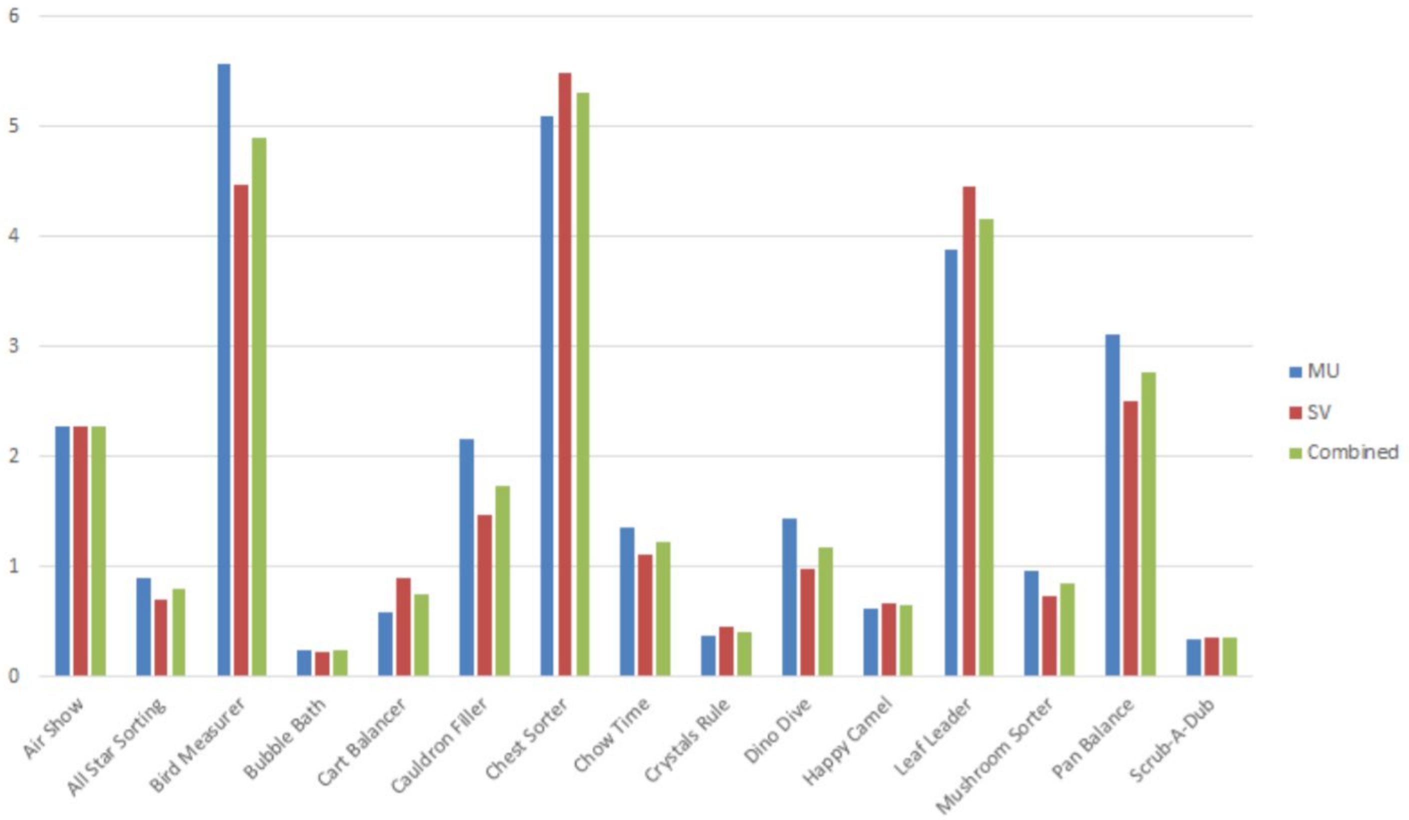
Exploratory: Detailed Engagement Analyses



Exploratory: Detailed Engagement Analyses



Exploratory: Detailed Engagement Analyses



Exploratory: Detailed Engagement Analyses

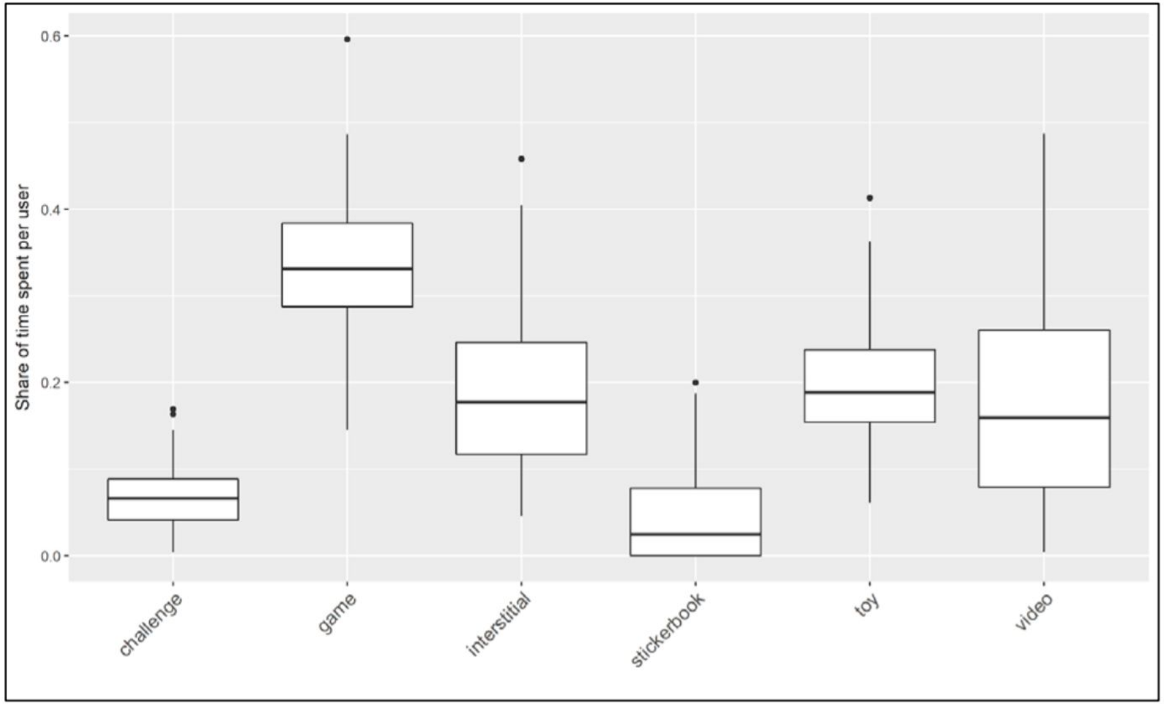


Figure J1. Share of time spent per activity type during the entire study. Every child is one observation in each boxplot.

Exploratory: Detailed Engagement Analyses

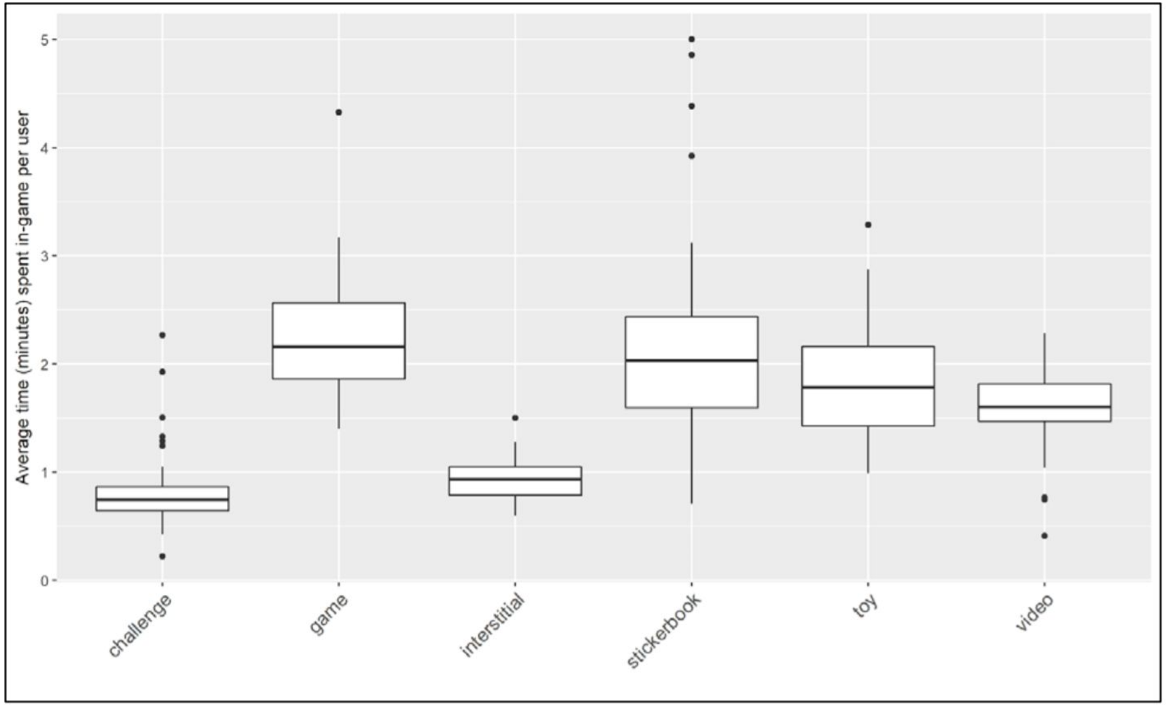


Figure J2. Average time spent in-game per activity type. If a user exits a game and comes back to it later (or immediately) the visits are counted separately. Every observation is the average time a user spends in a title of a certain activity type.

Exploratory: Detailed Engagement Analyses

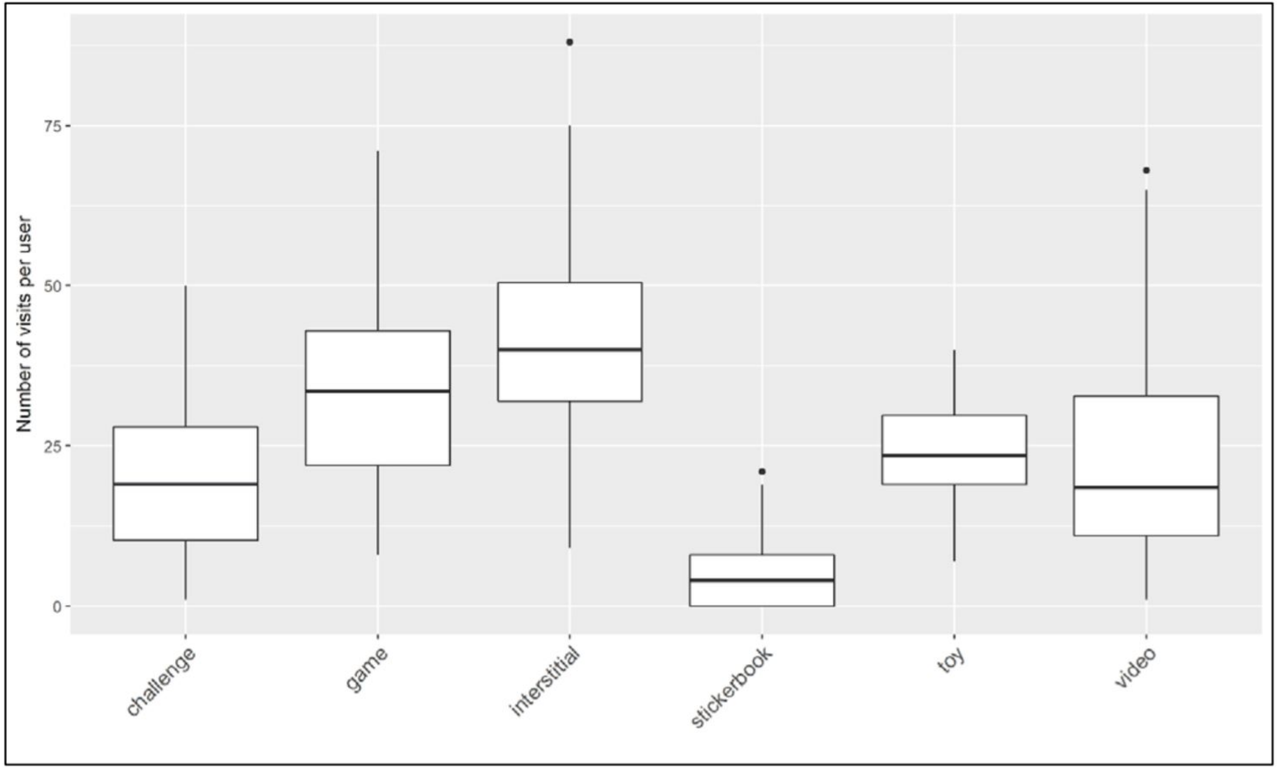


Figure J3. Total number of visits per activity type.

Exploratory: Detailed Engagement Analyses

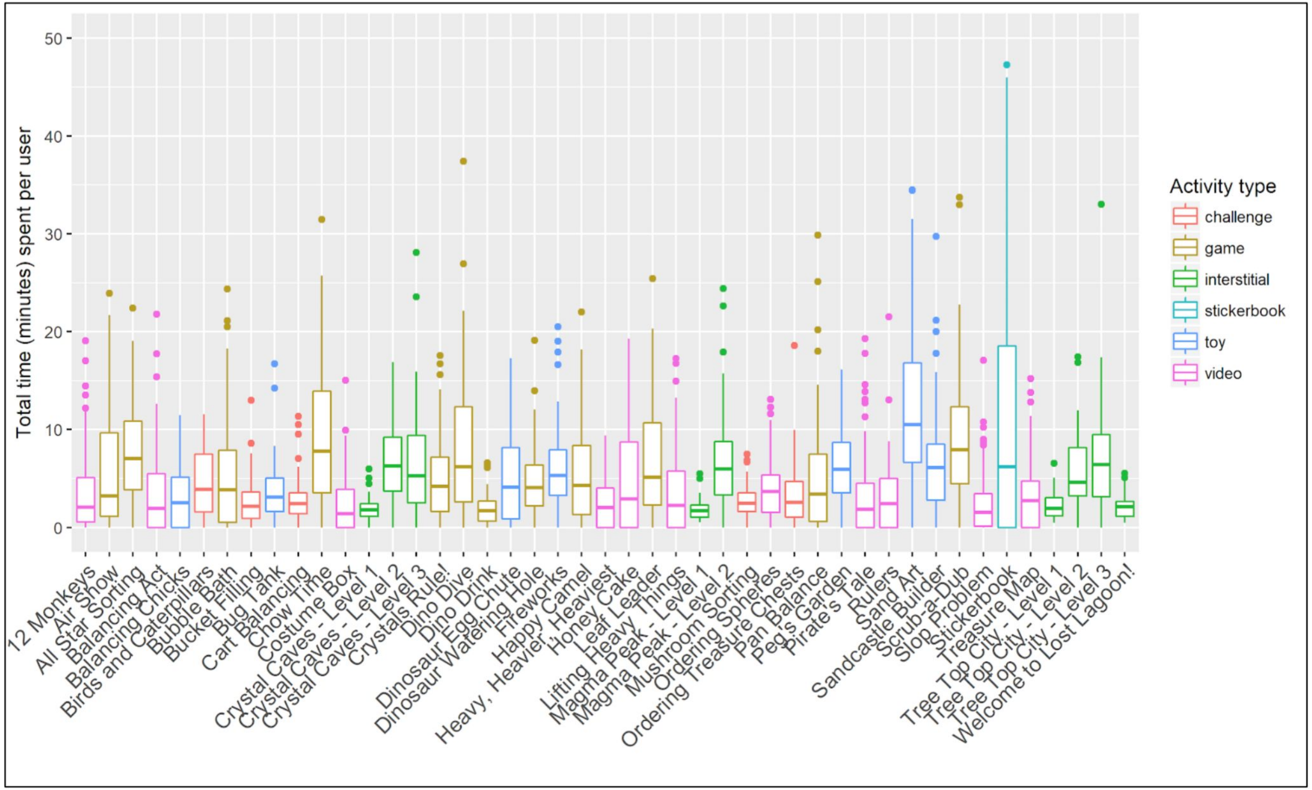


Figure J4. Total time spent per child per activity, for every activity.

Exploratory: Detailed Engagement Analyses

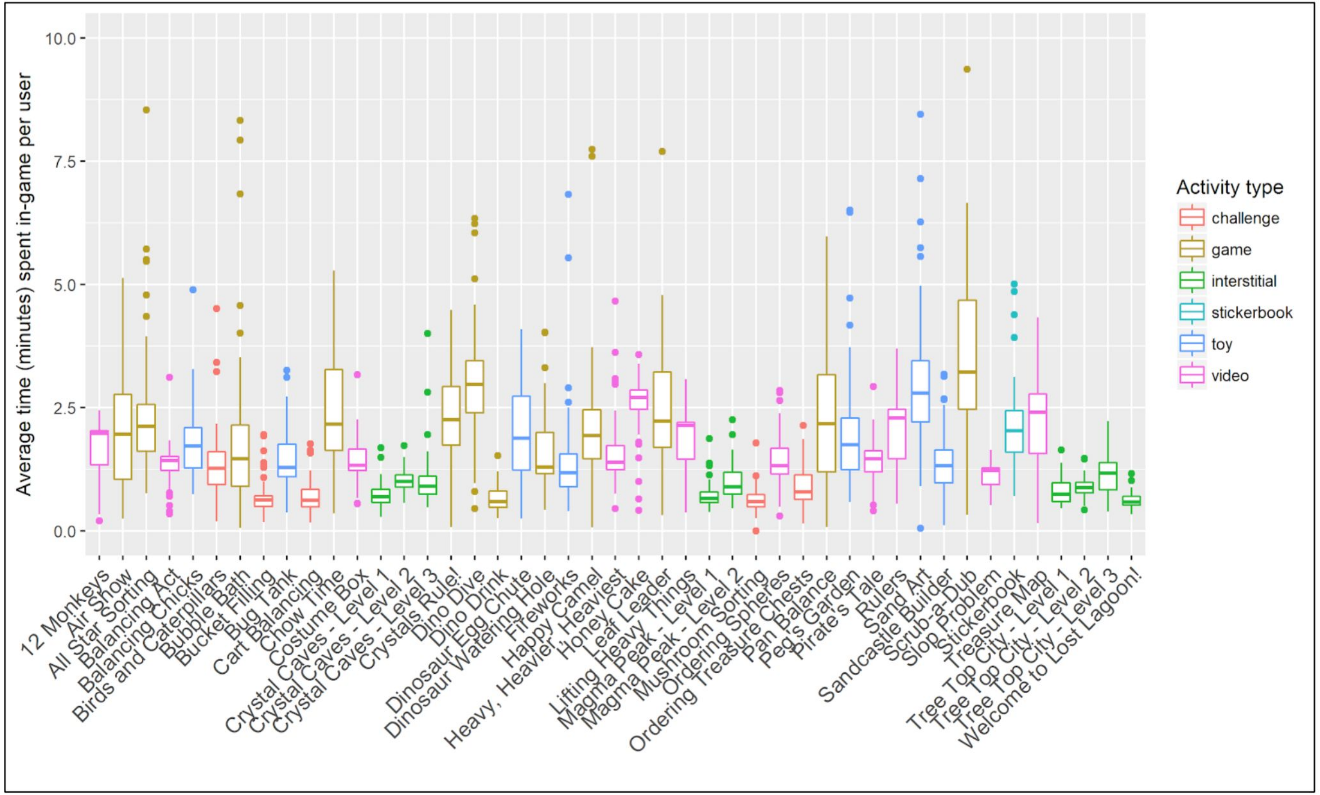


Figure J6. Average time spent in-game per activity for every activity.

Exploratory: Detailed Engagement Analyses

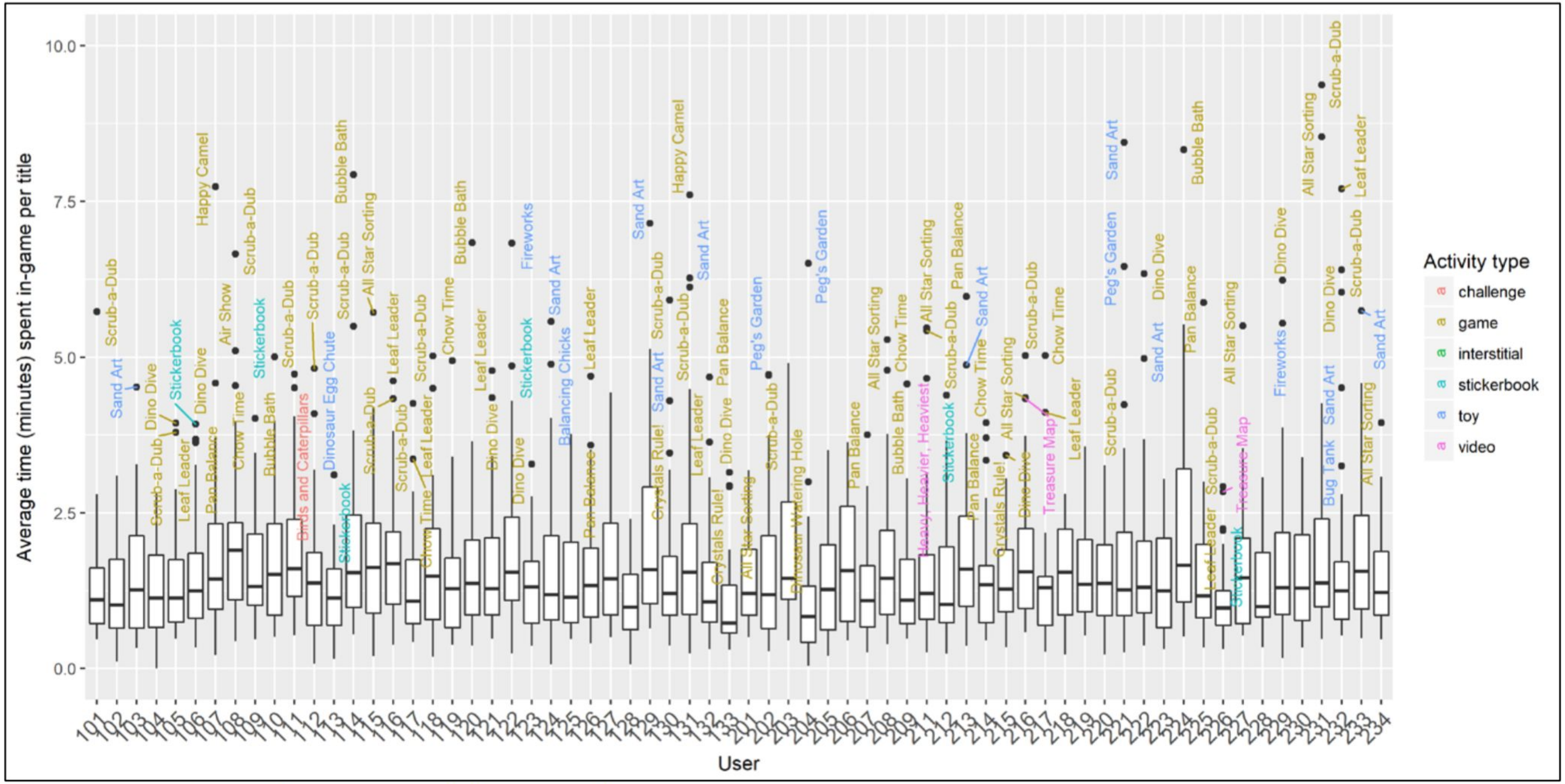


Figure J7. Average time spent in-game per activity for every child.

Exploratory: Detailed Engagement Analyses

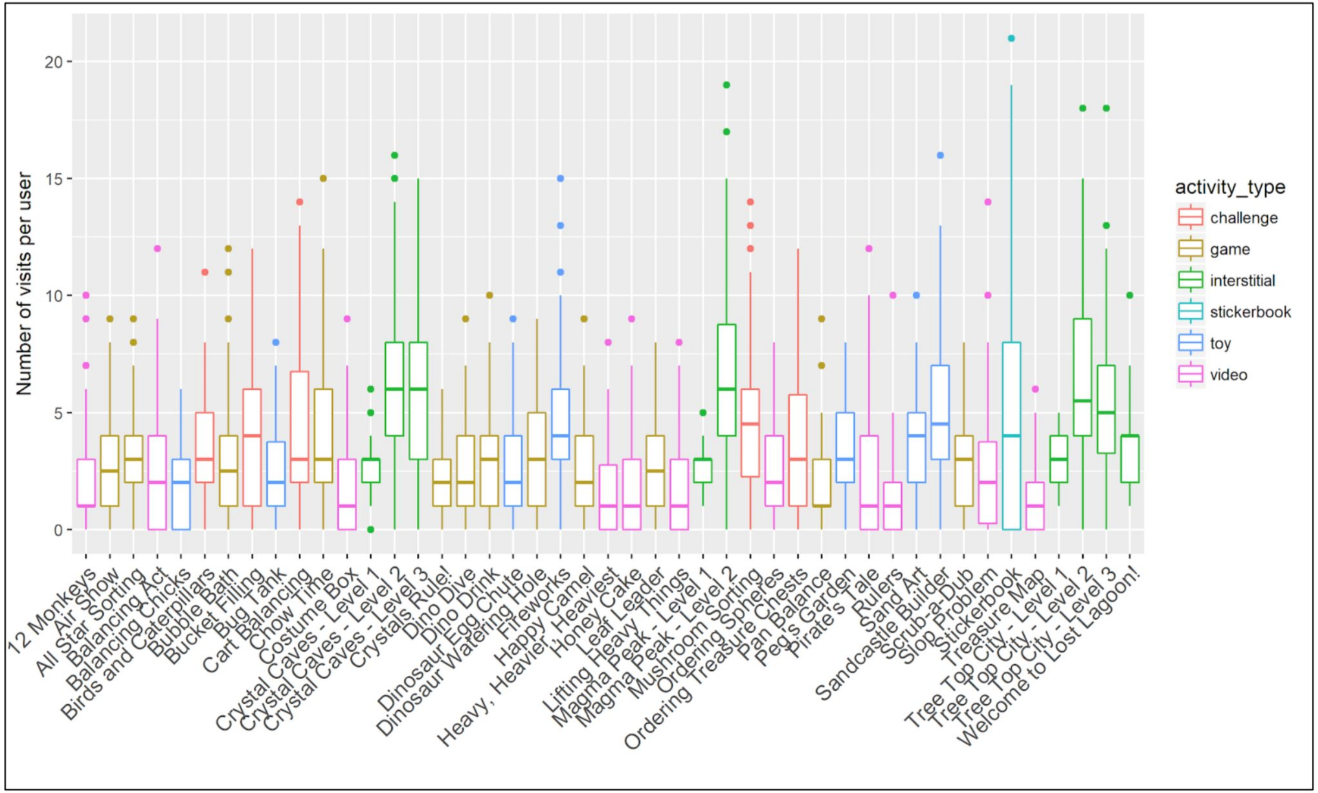


Figure J8. Number of visits per activity.

Exploratory: Detailed Engagement Analyses

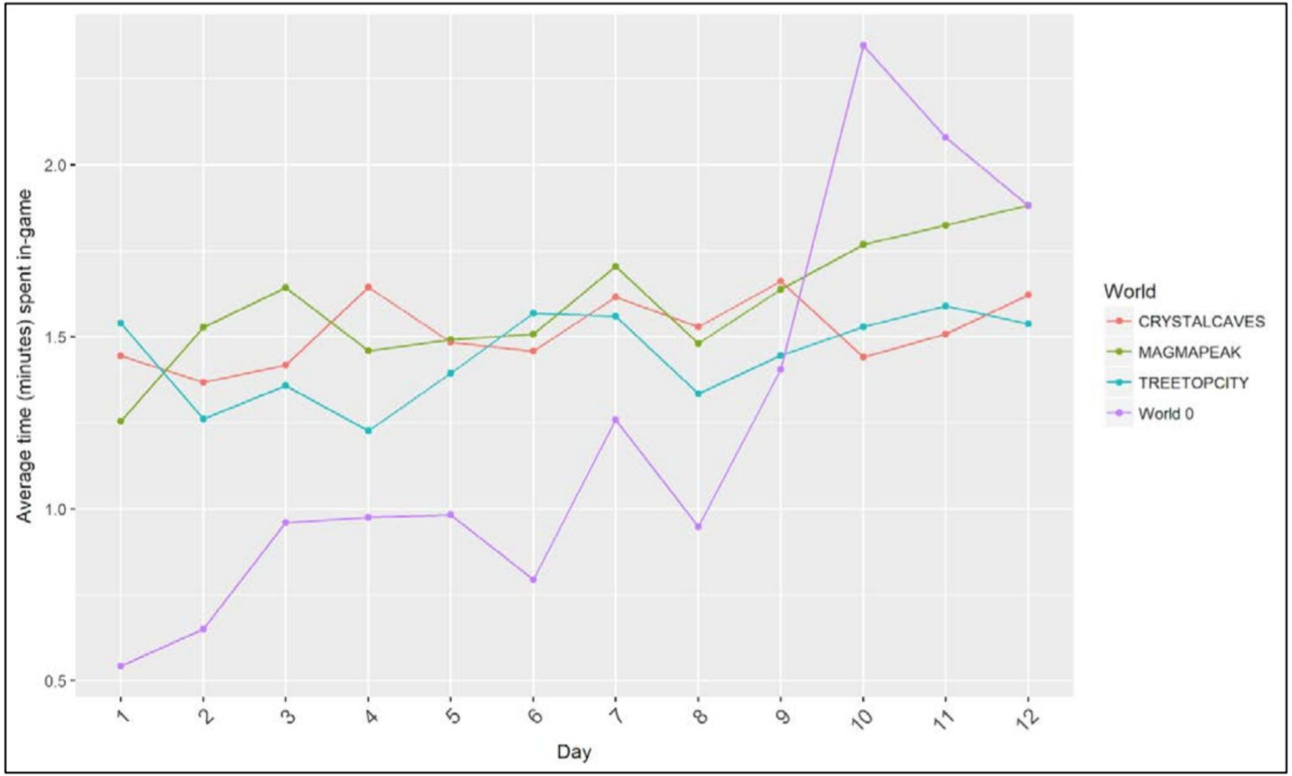


Figure J9. Average time spent in an activity by world per day (1-12).

Exploratory: Detailed Engagement Analyses

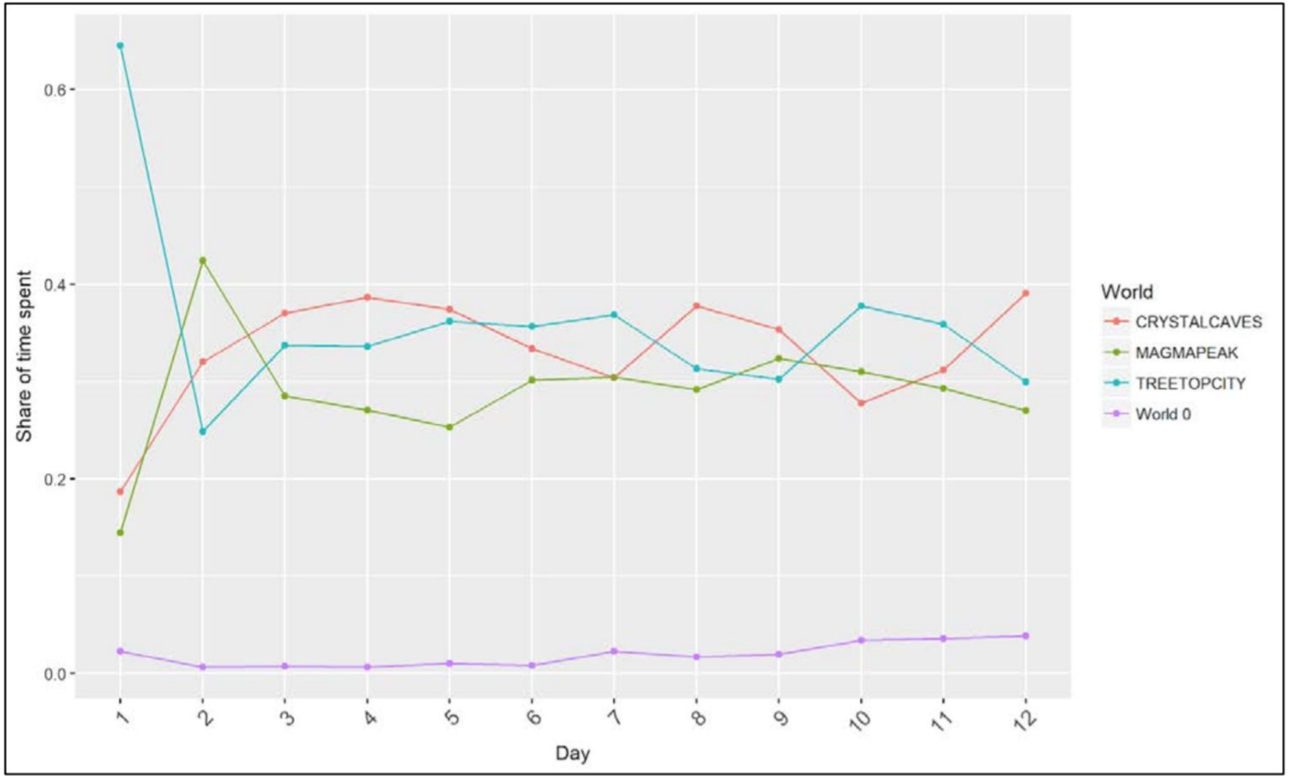


Figure J10. Share of time spent in each world per day.

Exploratory: Detailed Engagement Analyses

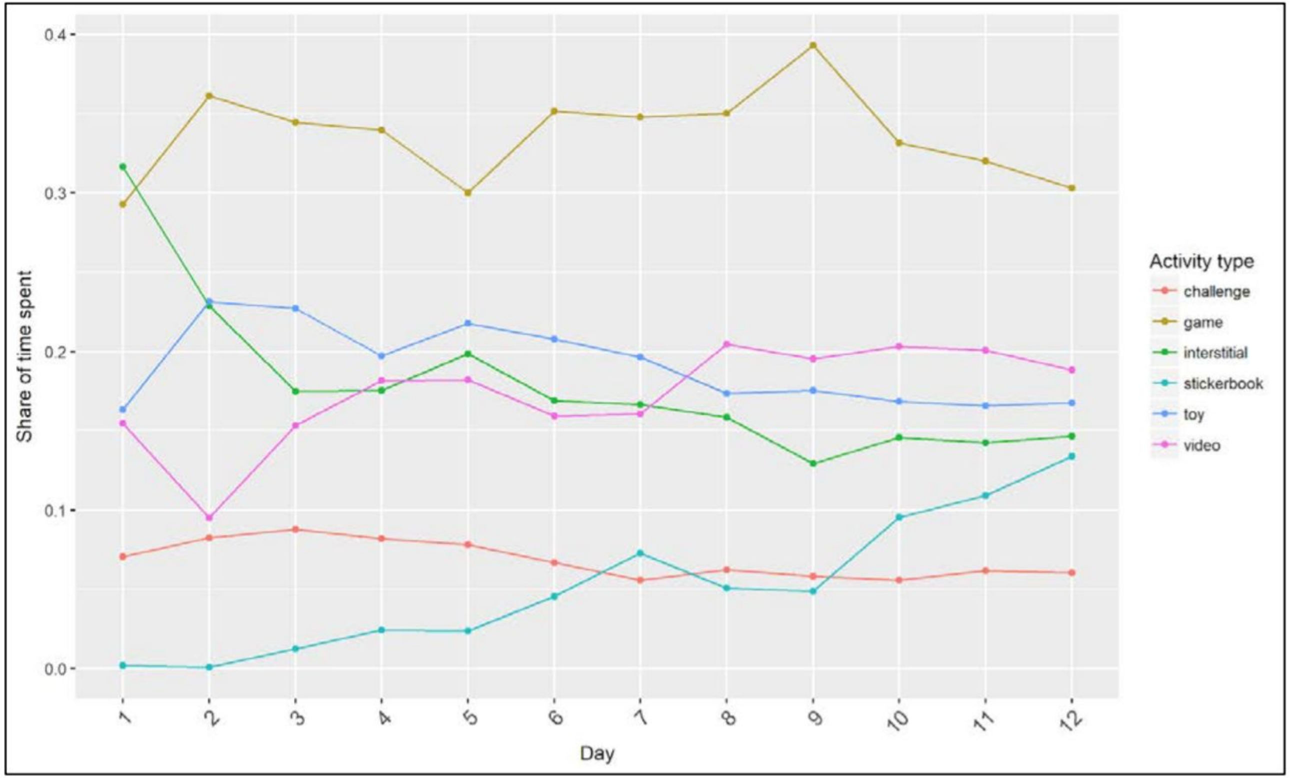


Figure J11. Share of time spent on activity types over the course of the study.

Exploratory: Detailed Engagement Analyses

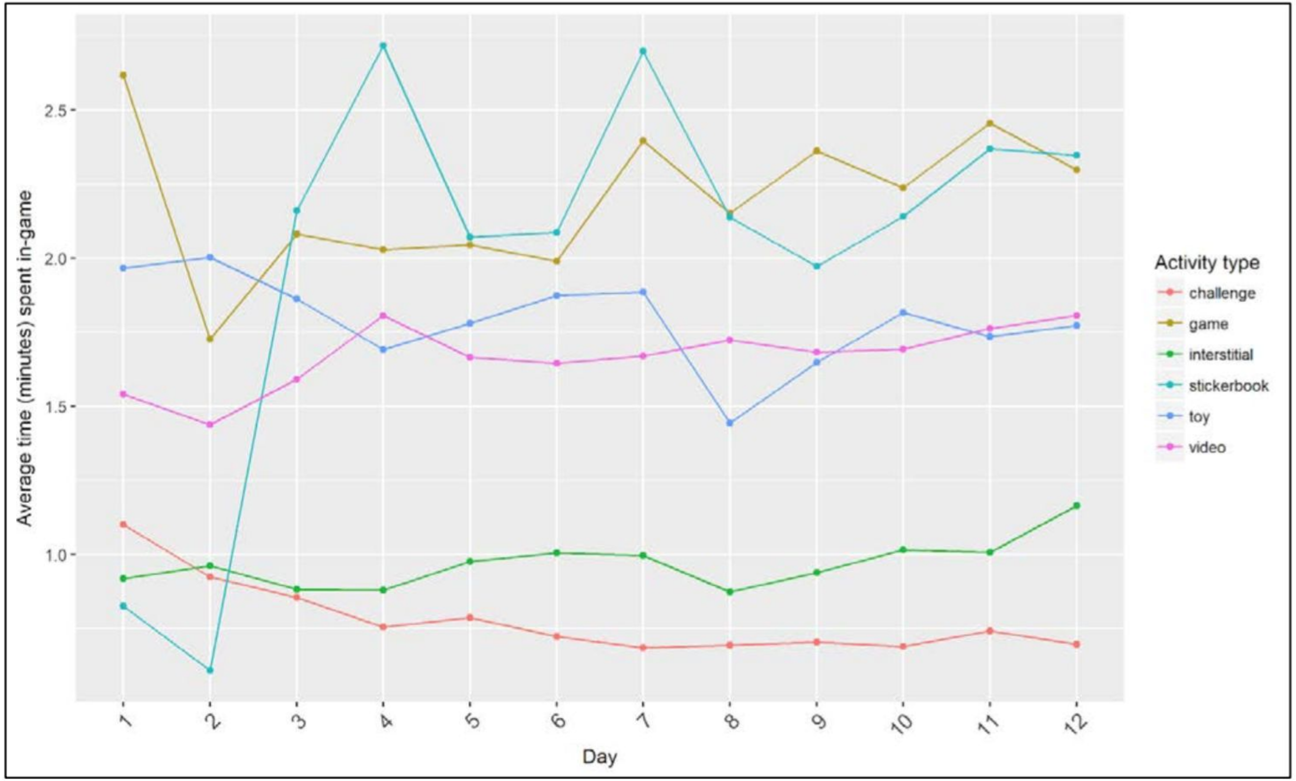


Figure J12. Average time spent in-game by activity type.

Exploratory: Detailed Engagement Analyses

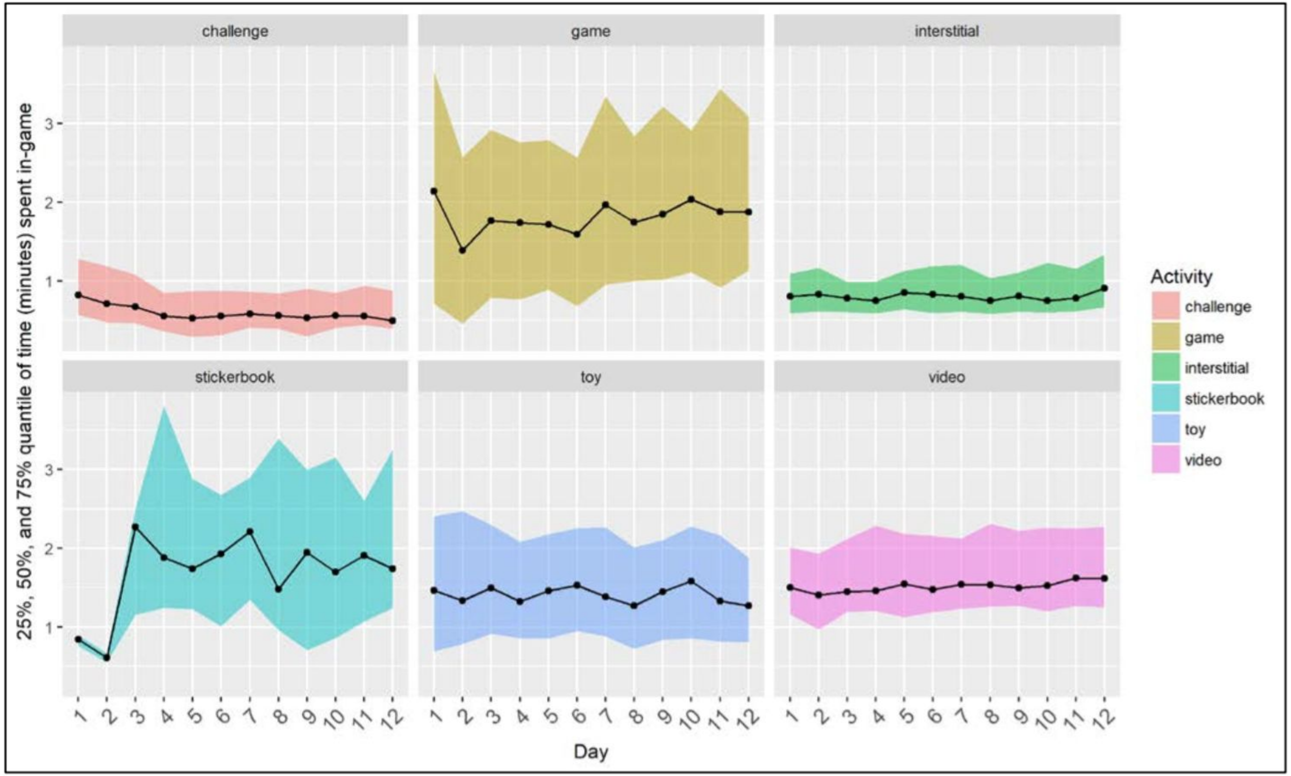


Figure J13. Time spent in-game by activity type.

Exploratory: Detailed Engagement Analyses

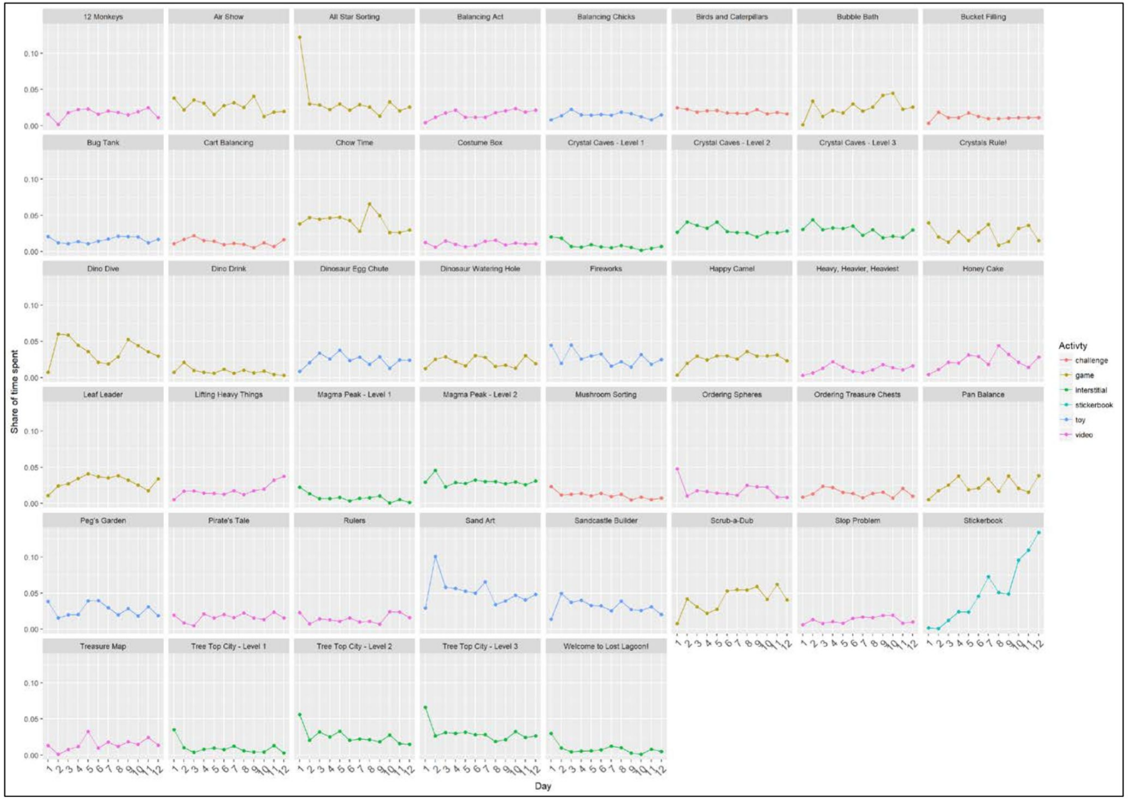


Figure J14. Share of time spent per activity over the course of the study.

Exploratory: Detailed Engagement Analyses

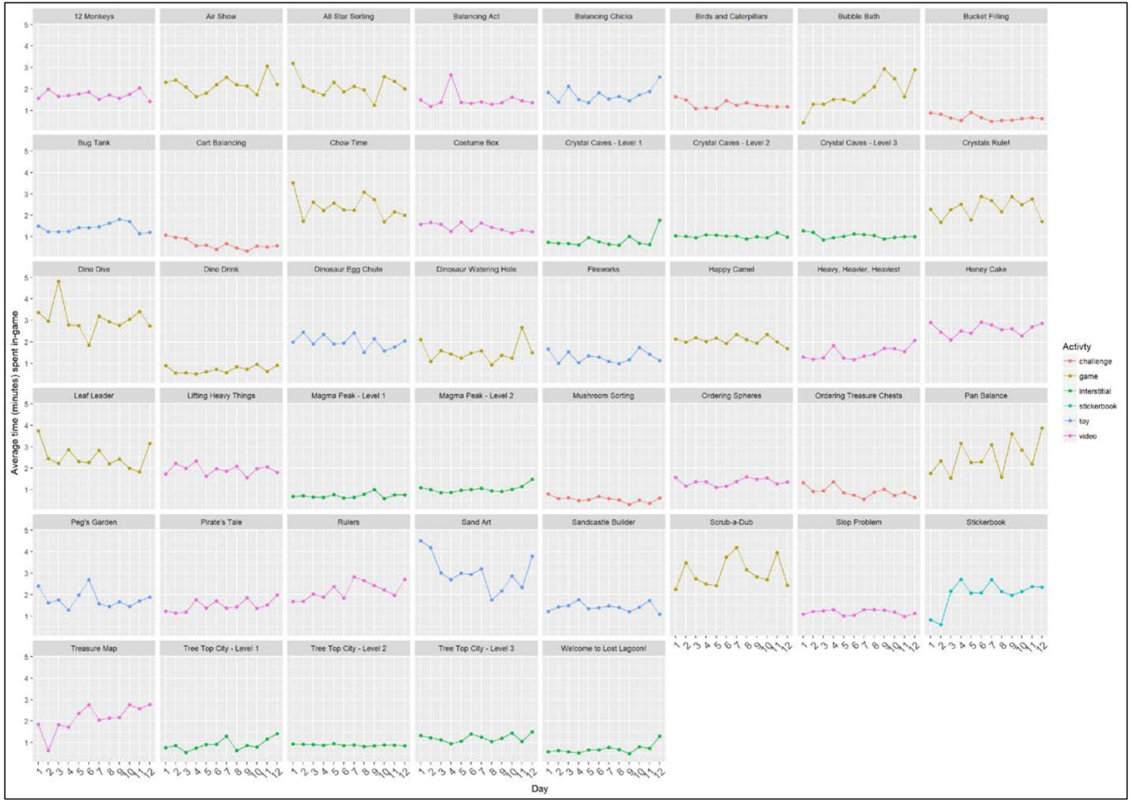


Figure J15. Average time spent in-game.

Exploratory: Detailed Engagement Analyses

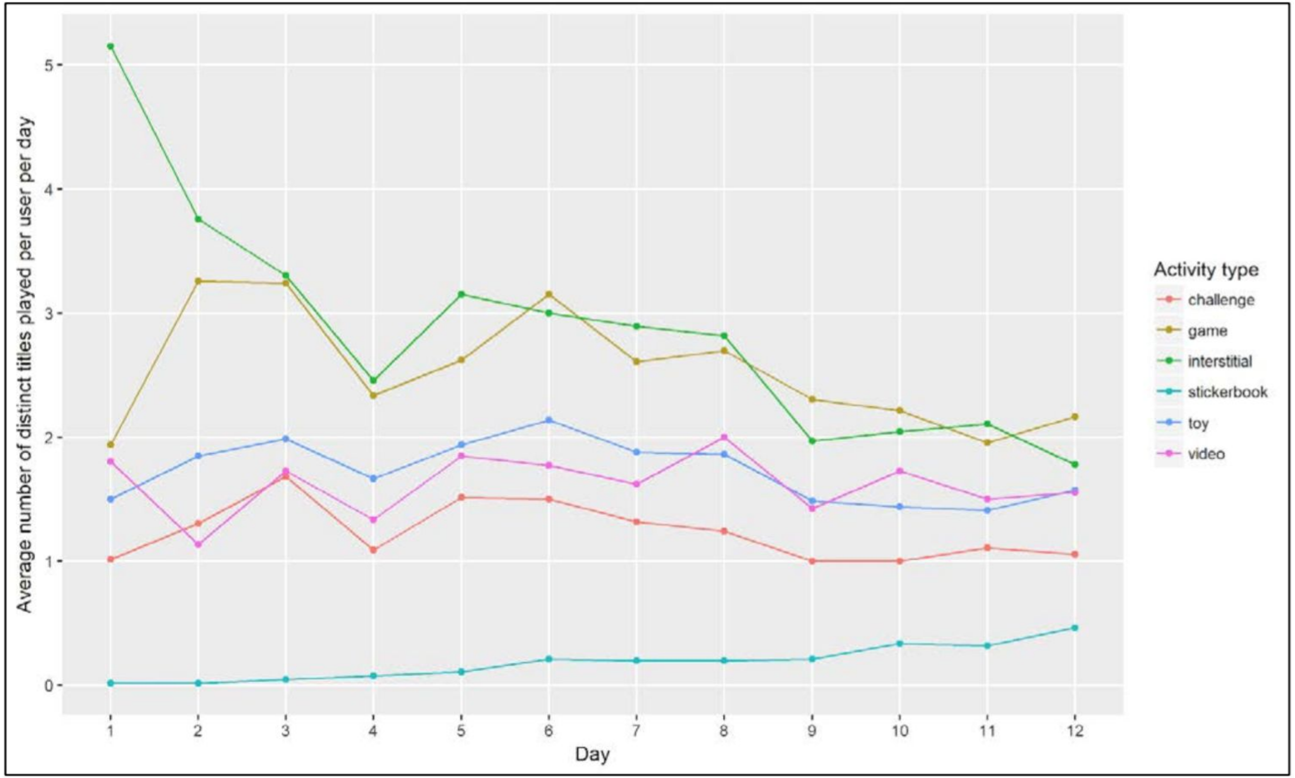


Figure J17. The average number of distinct titles entered per kid per day per activity.

Engagement Analyses: Implications

- We can use PBS KIDS LAP to help understand how content actually gets used (including patterns over time)
- We are now piloting this kind of analyses “on the fly” during product design and iteration
- There is a wide variation in favorite content
- We can see how far kids are getting, and how deeply they are engaging.
- We can identify under and out performers

Engagement Analyses: Implications

- We can understand initial appeal, return appeal, frequency, stayng power
- We can start to get at the dosage and timescales involved in learning
- We can compare and compete design strategies
- We can understand contributions made by individual components in media aggregation strategies
- Stickers rule!

Exploratory: Relating Gameplay to Gains

Table 10

Coefficients for Hierarchical Linear Regression Models Predicting Posttest Score Based on Pretest Score, Treatment Condition, and Site

Variable	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
Intercept	.41	.44	.51	.32	.35	.45	.24	.31	.50	.39	.23	.24	.29	.32	.08	.21	.19	.49	.25	.07
Pretest	.54	.31	.36	.60	.13	.40	.61	.29	.42	.33	.46	.51	.47	.40	.43	.33	.37	.48	.59	.81
MU+SV	.07	.09	.07	.05	.14	.04	-.03	.22	-.06	.03	.16	-.01	.13	.20	.14	.27	.35	.04	.05	-.05
MU	.04	.04	.02	.01	.27	.04	-.02	.11	.02	.25	.02	.01	.09	.19	.22	.19	.20	-.02	-.05	.06
Site 1		-.02	.02	.01	.00	-.05	-.01	-.03	-.06	-.03	.09	-.01	.01	.00	-.02	-.03	-.03	.00	-.02	.00
Site 2		-.03	-.08	-.01	.00	-.05	-.04	-.10	-.12	-.05	.02	-.04	-.03	.00	-.04	-.06	-.02	.00	-.06	.00
Site 3		-.04	.05	.01	.00	.00	.02	.10	.15	.02	-.07	-.03	.03	.00	.02	.08	.02	.00	.08	.00
Site 4		.09	.06	.01	.00	.18	.05	.14	.10	.13	-.01	.10	.03	.00	.02	.08	.10	.00	.11	.00
Site 5		.01	-.05	-.02	.00	-.08	-.01	-.11	-.08	-.07	-.02	-.02	-.03	.00	.02	-.08	-.06	.00	-.11	.00

Exploratory: Relating Gameplay to Gains

Table 11

Coefficients for Hierarchical Linear Regression Models Predicting Question Categories on Posttest Score Based on Pretest Score, Treatment Condition, and Site

Variable	Capacity (Q2-3)	Length (Q4-5, 8-10, 20)	Displacement (Q11)	Height (Q6-7, 18-19)	Weight (Q12-13)	WeightIPad (Q14-17)	NoIPad (Q1-13, 18-20)
Intercept	.41	.21	.23	.38	.22	.16	.54
Pretest	.43	.61	.46	.58	.60	.53	.20
MU+SV	.08	.06	.16	.03	.05	.22	.05
MU	.03	.13	.02	-.01	.05	.21	.06
Site 1	.00	-.01	.09	-.02	.00	-.01	-.05
Site 2	-.07	-.04	.02	-.02	-.04	-.04	-.07
Site 3	-.01	.04	-.07	.01	.02	.03	.04
Site 4	.09	.06	-.01	.08	.07	.03	.14
Site 5	-.01	-.05	-.02	-.05	-.05	-.01	-.06

Note. The questions falling into the specific category are in parentheses.

Exploratory: Relating Gameplay to Gains

Table 13

Child Test and Gameplay Data Related to Weight Domain

Child ID	Pretest scores on weight items	Posttest scores on weight items	Score difference on pre- and posttest	Question 12 pretest and posttest responses	Most time spent on Crystal Caves activities
116	3.0	5.0	2	*Pre*: Goes up down and side to side and spins. Swings. *Post*: They're pan balances. You use something on it to see which one is heavier. The side that tips down is heavier.	• Pan Balance (20 min) • Chow Time (19 min) • Happy Camel (17 min)
121	2.0	5.0	3	*Pre*: (no response) *Post*: This one measures stuff. How heavy is something.	• Leaf Leader (19 min) • Pan Balance (18 min)
206	2.0	6.0	4	*Pre*: I don't know. *Post*: Pan balance. Balance things. How heavy is it.	• Chow Time (24 min) • Leaf Leader (20 min)
212	0.0	5.0	5	*Pre*: They do of the same things and then they go down. *Post*: They put things so they can weigh and tell if they're balanced	• Balancing Act (22 min)
213	2.5	4.5	2	*Pre*: They measure. *Post*: Measure.	• Pan Balance (30 min)
225	0.5	4.5	4	*Pre*: They hold how many weight it takes. *Post*: They hold up stuff that's the most weight.	• Cart Balancing (11 min)

Relating Gameplay to Gains: Implications

- We can know what items showed the highest gains
- We can find the children who showed gains on those items
- We can examine those kids telemetry
- We can connect the dots!

Exploratory: Detecting Misconceptions

- Metz 1993: Preschoolers and Pan Balances

Preschoolers' Developing Knowledge of the Pan Balance: From New Representation to Transformed Problem Solving

Kathleen E. Metz
School of Education
University of California, Riverside

Although it is generally accepted that a new representation of a physics task domain frequently underlies fundamental improvements in problem solving, there is less agreement and less understanding about the process of knowledge construction from emergence of first signs of new representation to transformed problem solving. This study examines that issue in the context of preschoolers' formation of a weight approach to achieve mechanical equilibrium on the pan balance.

The study used a rich version of the balance-beam task and three levels of microgenetic analysis of the videotaped experimental sessions to elucidate this subtle transition. Cues employed included focus of visual attention, manipulation of materials and the apparatus, gestures, verbal protocol, and expressions of surprise. Forty-eight preschoolers (evenly distributed across sex and age groups) were each asked to balance the pans on a series of nine problems. Every problem consisted of a set of weights to be balanced, either identical or differing in size and/or weight. The subjects were encouraged to repair arrangements that resulted in disequilibrium.

Analyses revealed a remarkably multifaceted development from emergence of first signs of weight representation to fully elaborated weight-based problem solving. Development of weight representation entailed changes in the particular aspects of the domain represented in terms of weight and the extent to which "weight" was still bound up in sensorimotor coordinations. Development of weight-based action also involved conceptualization of the relevance of weight to the task and the gradual elaboration of weight-based diagnostic and instrumental implications. In short, the representation of weight itself was not unitary, nor was the construction of the new method of problem solving based on the representation.

Exploratory: Detecting Misconceptions

Table K1

Strategies: Substance and Organization (From Metz, 1993, Table 2)

Strategy	Indicators
One-to-one correspondence	The subject puts one element in each of the pans, by pairs or alternation of placements. Action is complete when all elements have been distributed. Focus on completion of the action (exhaustive distribution) and not developing cardinal number.
Equal number	The subject puts the same number in each pan. Concern with evolving quantity in each pan, as manifested by preliminary halving of the set or counting of set contents (as indicated by finger or verbal tagging).
Visual feedback	The subject determines next placement by apparatus's response resulting from previous placement, as indicated by train of visual attention, rhythm and sequence of actions on the apparatus, and verbal protocol.
Equal weight	The subject directly (not step-by-step by visual feedback) divides the set into two subsets of equal weight, using equal number if and only if confident the elements weigh the same and, when weights vary, by compensations between number of elements and their respective weights. Indicated by prearrangements of the elements on the table, train of visual attention, rhythm and sequence of placements in the pans, and verbal protocol.
Knob fiddling	The subject adjusts the knobs built into the beam. (Given the light weight of the knobs, knob fiddling never constitutes successful repair. In addition, directions of fiddling were also coded to differentiate from "other: torque adjustment.")
Compensation by force of hands	The subject tries to attain the goal state by forcing the apparatus into the goal state or by yanking down or thrusting up on the beam, the wire, or the pan. Indicated by actions on the apparatus and corresponding expectations regarding its conservation (as inferred from facial expression, subsequent actions, and verbal protocol).
Realign	The subject tries to attain the goal state by grasping the beam, the wire, or the pan, bringing into goal alignment, and gently releasing. Indicated by actions on the apparatus and corresponding expectations regarding its conservation (as inferred from facial expression, subsequent actions, and verbal protocol).
Exchange	The subject switches pan placements of one element from each pan or switches whole pan contents from one side to the other. Indicated by rhythm and sequence of manipulations in relation to examination of the apparatus's response and verbal protocol.
Reject	The subject excludes an element from placement in either pan. Indicated by handing the experimenter one or more elements, placing an element(s) aside, or verbal request (disallowed).

Table K2

Representation: Apparatus Focus (From Metz, 1993, Table 4)

State	Indicators
1. Yank or align	The subject yanks or aligns the apparatus for the purpose of attaining the goal, as evidenced by rhythm and sequence of the subject's apparatus manipulations, verbal protocol, and facial expressions as reflection of expectation or surprise.
2. Displace elements across pans	The subject attempts to attain the goal state either by moving elements from the up-pan to the down-pan or by moving elements in both directions, as evidenced by attention to the tilt of the apparatus after element placements, rhythm and sequence of examining tilt of apparatus and element displacements, directions of displacements, and verbal protocol. Absence of concern with heavier element per se, as evidenced by train of visual attention, manipulations, and verbal protocol.
3. Move elements from down-pan to up-pan: visual feedback	The subject attempts to attain the goal state by displacing elements from the down-pan to the up-pan, as evidenced by attention to the tilt of the apparatus after element placements, rhythms and sequence of examining tilt of apparatus and element displacements, directions of displacements, and verbal protocol. Absence of concern with heavier element per se, as evidenced by train of visual attention, manipulations, and verbal protocol.

Exploratory: Detecting Misconceptions

Table K3

Representation: Weight Focus (From Metz, 1993, Table 4)

State	Indicators
1. No representation of the elements as weights	The subject represents neither the weight differences between the elements nor the element as a weight, as evidenced by absence of differential attention in examination of the elements, absence of differential treatment in action, and verbal protocol.
2. Representation of weight without relevance	The subject represents the elements as weights but does not link the property to the task at hand. Representation is evidenced by differential examination of the elements, with the more intense examination given to the heavy element, and by verbal protocol. Lack of relevance is evidenced by verbal protocol and by absence of differential treatment in action.
3. Relevance of weight without specific implications	The subject believes that the weight differential is somehow related to the task but has not yet elaborated any specific implications. Relevance is evidenced by differential attention in examination of the elements, with the more intense examination given to the heavy element, and by verbal protocol. Absence of implications is evidenced by failure to base either diagnosis of the disequilibrium or action on the differential weights.
4. Diagnostic implication	The subject interprets the tilt of the apparatus, under conditions of equal number distributions, as signifying the placement of a heavy element in the lower pan, but weights-based thinking does not yet extend to weights-based action. Evidenced by interactions with the apparatus and verbal protocol.
5. Search for an instrumental implication	The subject searches for an instrumental implication to resolve the "one element weighs too much" problem, as evidenced by weight-related but inappropriate actions such as exchanging placements between a heavy weight and a standard weight, trying to reject a heavy weight (disallowed), or reshaping a heavy weight. Evidenced by nature of the actions carried out on different weights and verbal protocol.
6. Repair implication: visual feedback	The subject displaces weights from the lower pan to the raised pan or places the next element from the pile into the raised pan as a solution to the "one element weighs too much" problem, as evidenced by rhythm and sequence of interactions with the apparatus and different elements and by verbal protocol.
7. State implication: compensation	The subject solves the "one element weighs too much" problem by trade-offs between number of elements in the two pans and their respective weights, as evidenced by the decision of pan collections prior to any placements or feedback from the apparatus, the formation of states with number versus weights trade-offs, and verbal protocol.

Exploratory: Detecting Misconceptions

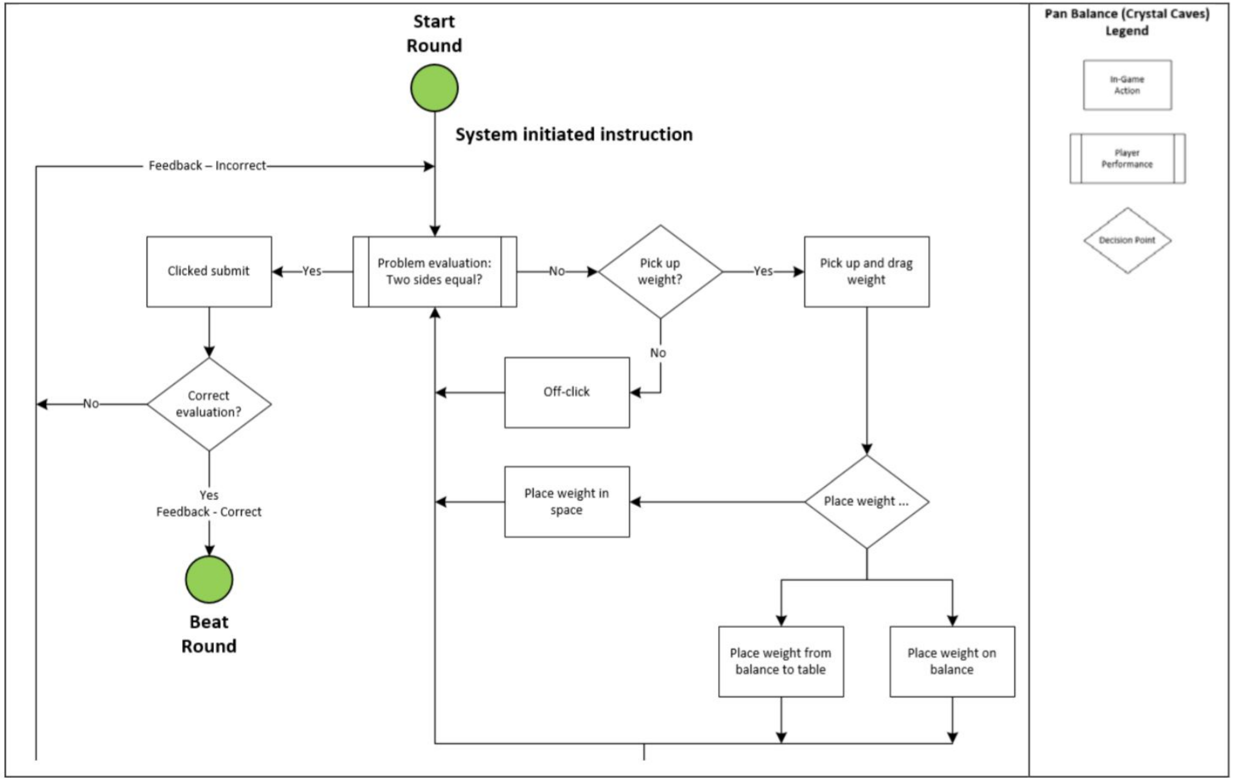


Figure K1. Interaction space mapping for “Pan Balance.”

Detecting Misconceptions

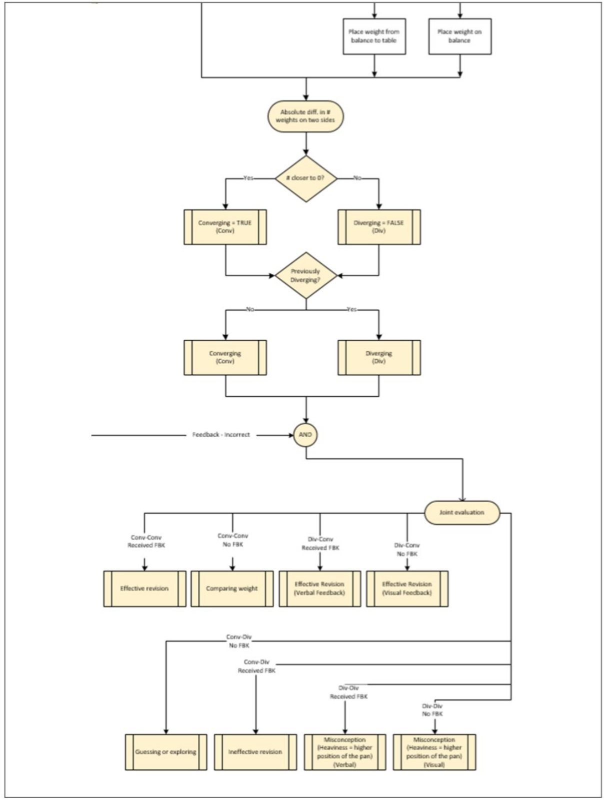


Figure K3. Analysis mapping for “Pan Balance.”

Detecting Misconceptions

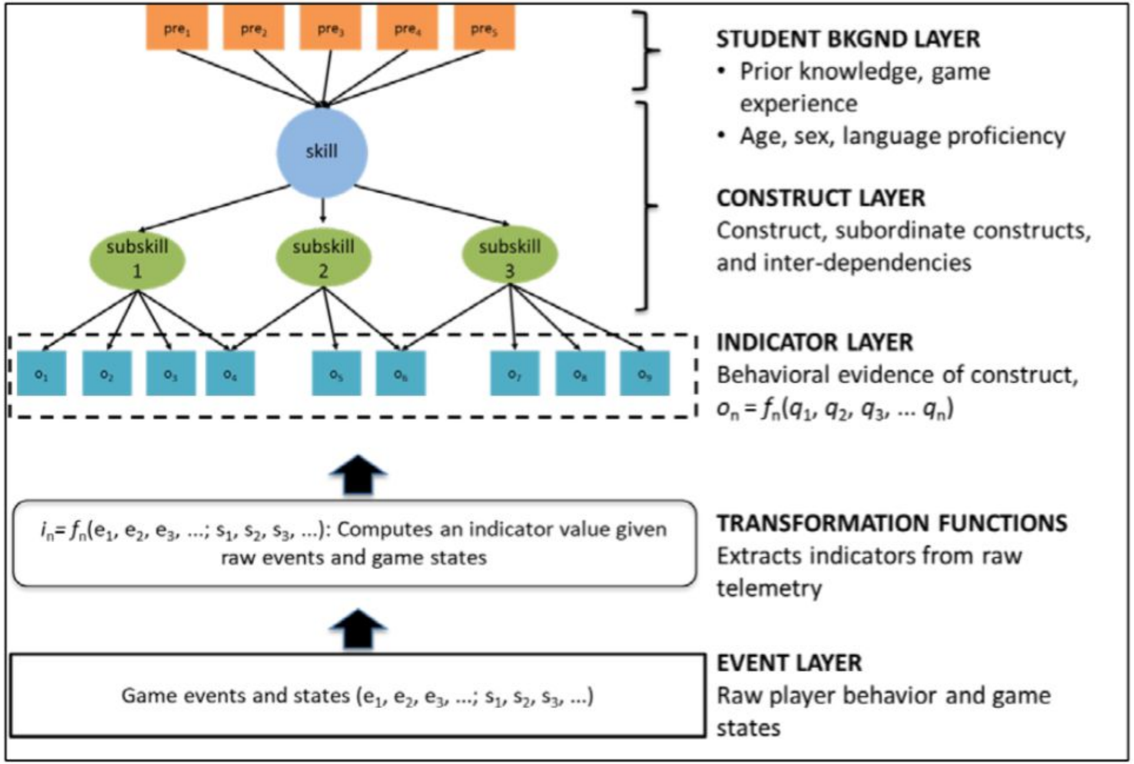


Figure K5. Computational structure.

Detecting Misconceptions

Table K4

Descriptive Statistics of Misconception and Weight Measures

Measure	*n*	*M*	*SD*	Min.	Max.
Higher is heavier misconception	53	1.64	3.13	0.0	18.00
Yank or align misconception	53	3.49	4.87	0.0	21.00
Pretest – weight items	66	2.05	1.58	0.0	6.00
Posttest – weight items	66	3.18	1.87	0.0	6.00
Gain (Posttest – Pretest) – weight items	66	1.13	1.59	-2.0	5.00
Time spent on “Pan Balance” (sec)	66	312.95	371.01	0.0	1792.17
Time spent on “Cart Balancer” (sec)	66	169.46	140.90	0.0	680.88

Detecting Misconceptions

Table K5

Intercorrelations Among Misconception Measures and External Measures of Weight

Measure	1	2	3	4	5	6
1. Higher is heavier misconception[a]	—					
2. Yank or align misconception[a]	.24	—				
3. Pretest – weight items[b]	-.04	-.06	—			
4. Posttest – weight items[b]	-.36**	-.23	.58***	—		
5. Gain (Posttest – Pretest) – weight items[b]	-.34*	-.19	-.30*	.59***	—	
6. Time spent on "Pan Balance"[b]	.21	.08	.15	.18	.07	—
7. Time spent on "Cart Balancer"[b]	.13	-.14	.15	.11	-.02	.33**

[a]$n = 53$. [b]$n = 66$.
*$p < .05$, two-tailed. **$p < .01$, two-tailed. ***$p < .001$, two-tailed.

Detecting Misconceptions: Implications

- We can consult the literature for established misconceptions
- With the right interactives, we can detect them
- We can relate having a misconception to gains, or lack thereof
- Great foundation for PAL: provide support tailored to a misconception

Overall Implications for the Future

- We now have a great methodology for evaluation
- We can use PBS KIDS LAP for unprecedented understanding of how products and services get used
- We can connect the dots between learning gains, and the specific engagement activity that correlates to it
- We can consult literature for well-defined misconceptions, and detect them
- We think we can use knowledge of misconceptions to help power Personalized and Adaptive Learning efforts

Thank You!

Thank You!

UCLA / CRESST

- National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
    - Director: Li Cai | Founding director: Eva Baker
- Expertise in the application of technology to measure knowledge and skills in low-stakes settings
    - Knowledge representation (ontologies)
    - Simulations and games (design, automated scoring)
    - Process data, evaluation of learning technologies
- Focus on validity
    - How does cognition and motivation manifest itself in digital environments, and what are behavioral indicators of both?
    - Use of process data to support measurement of complex performance

UCLA / CRESST

- Technology R&D
  - Intersection of learning, instruction, measurement, and assessment
  - Education, military, ed tech
    - PBS KIDS, EDC, NMSI, ETS
    - IES, NSF, Office of Naval Research, DARPA, USN
  - PreK to adults
  - Effectiveness studies (RCT / adhere to WWC standards)
  - Exploratory studies

Connected Learning Study

- Goal: Examine the effects of MU and SV on children’s learning of measurement concepts and parents’ awareness and support of their children’s mathematics learning

Connected Learning Study

- Pretest-posttest randomized design
- Three conditions:
  - PBS KIDS Measure Up! (MU)
  - MU + PBS KIDS Super Vision (SV)
  - Control (Literacy apps: Super WHY!)

Sample

- 4-5 years old
- Public elementary schools (Title I)
  - Prekindergarten and transitional kindergarten classrooms
- Childcare center
  - Affiliated with community college

Protocol

- Four days a week, three weeks
- 20 - 30 minutes a day
- Max 6 hours
- In-classroom play

Assessments

- 20 items, 10 minutes, 1-on-1
  - Length and height
  - Capacity and displacement
  - Weight

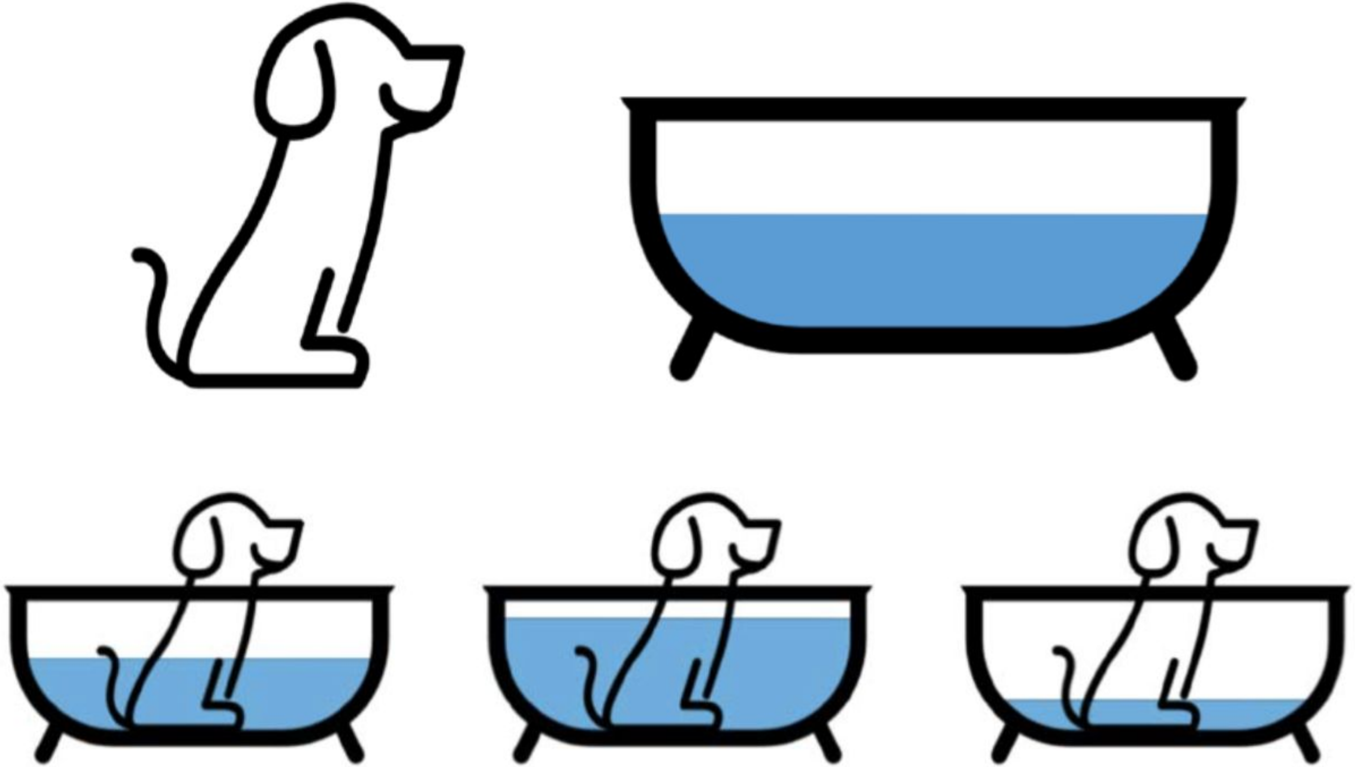
Assessments

- Adopted, adapted, and created external criterion measures
- Four CMA items
  - Extended CMA with author’s support
- Three KeyMath-3 items
- CRESST created 13 items

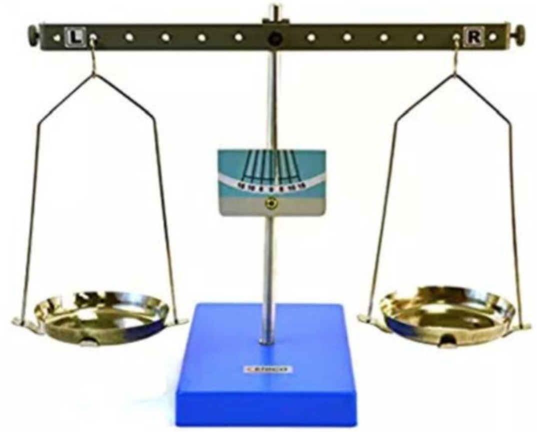
Assessments

- Eleven employed manipulatives
- Six used pictures
- Four utilized an iPad app
  - Created by CRESST to simulate using a pan balance
- Usability testing and expert review

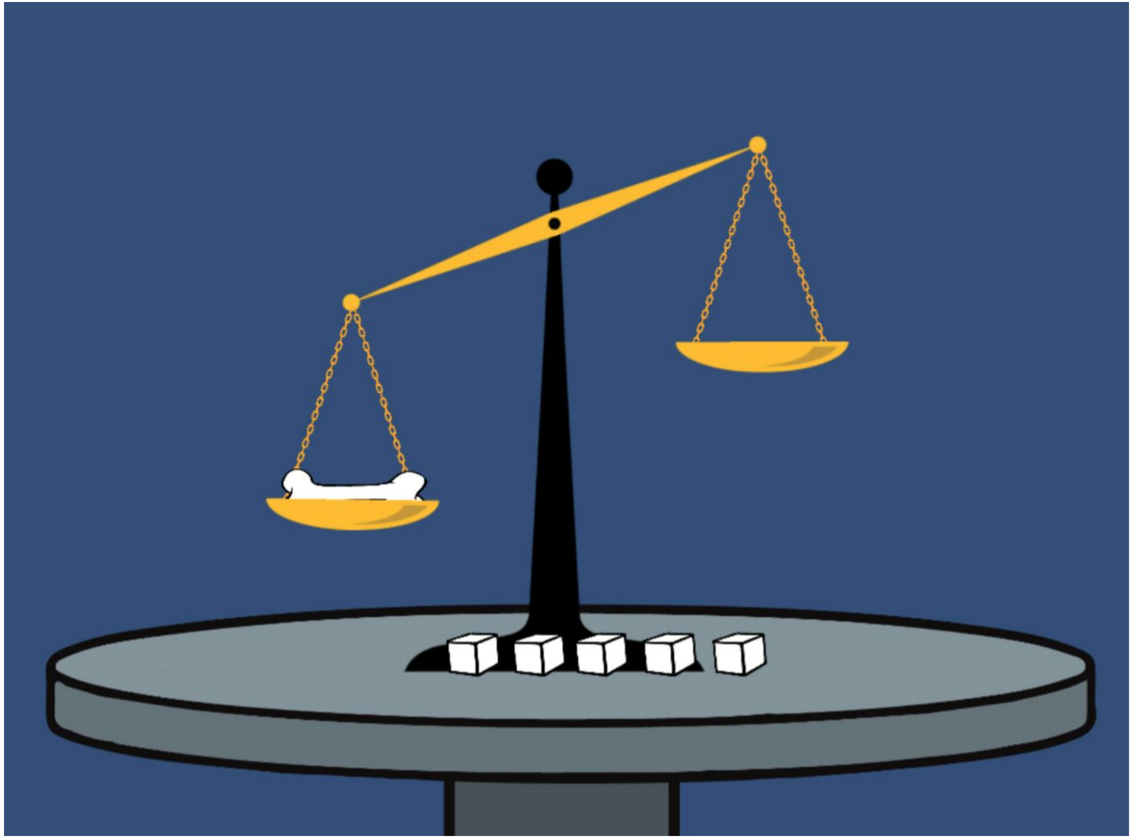
Assessments



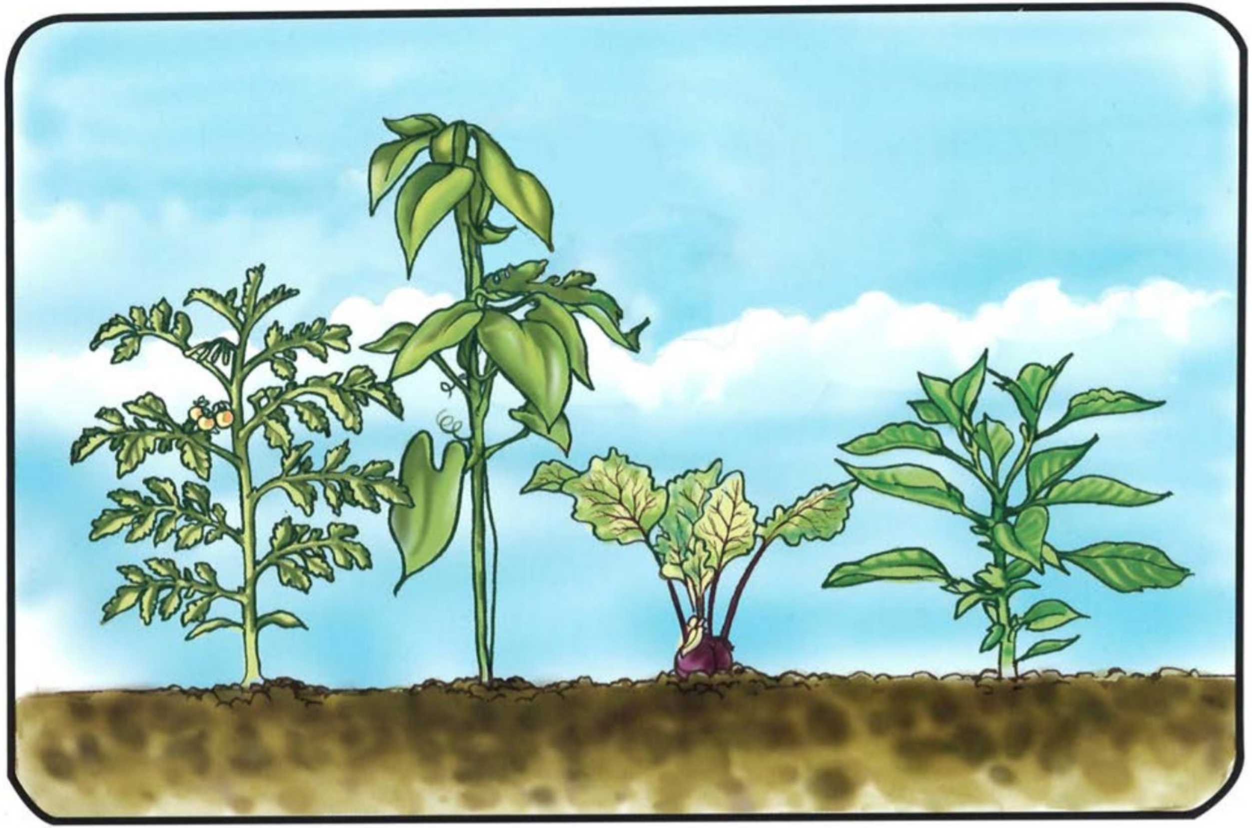
Assessments



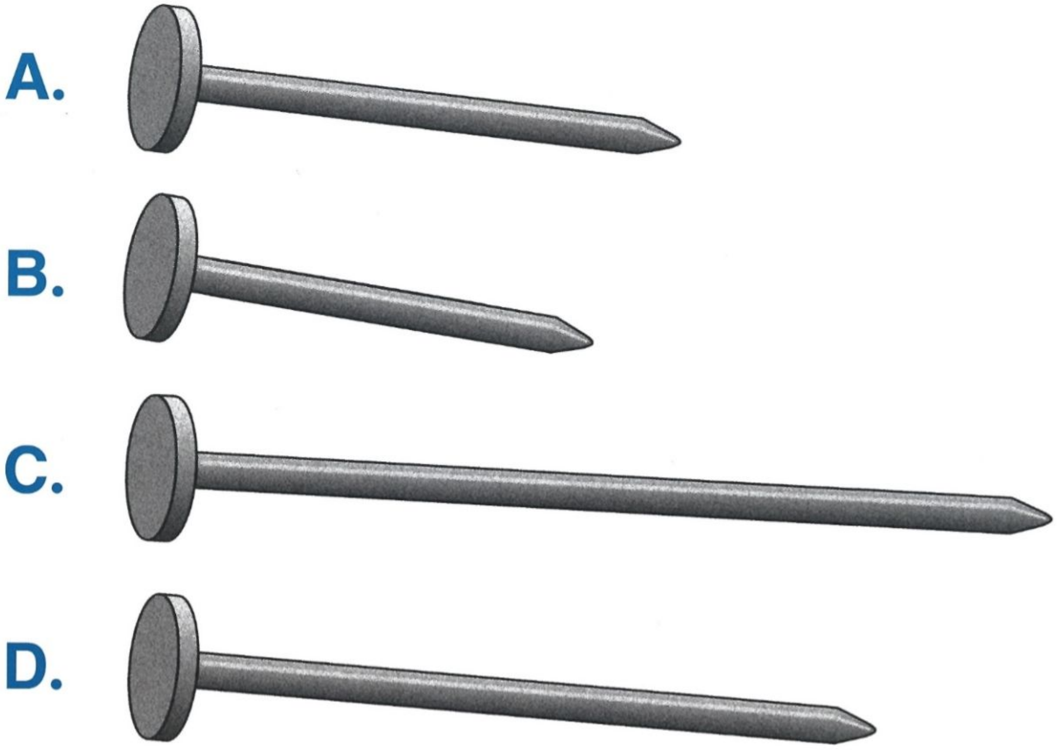
Assessments



Assessments



Assessments



And More

- Lots more available to view in report
    - All protocols, assessment items, and surveys
    - Background and demographic info
    - Descriptive statistics
    - Model definitions

Core Study - Analysis and Results

- Four models
  - Each introduces a different covariate
    - Site, age, gender, low-income
  - Robust and consistent across all 4 models
- Predictors explain a substantial amount of variance in the outcome variable

Core Study - Bottom Line

- MU condition results in **more than two** additional correct responses on posttest
- MU + SV results in **more than 1.5** additional correct responses on posttest
- Statistically significant gains compared to control (all between 8-11% improvement)
- Not statistically different from one another