

Two-Way Play: Early Research Findings of Learning with Kinect Sesame Street TV

Meagan Rothschild, University of Wisconsin-Madison, mrothschild@wisc.edu

Abstract: Microsoft Studios has recently released *Kinect Sesame Street TV*, a new form of media for television that merges traditional means of watching episodes with game like physical actions via the Xbox and Kinect. This paper presents an overview and early findings of an initial study that investigated how ideas of embodied cognition and comprehension can be leveraged to understand the experiences of three and four year old participants, and explore the ways in which bidirectional television can facilitate new meaning-making. Early research took place at Microsoft Research in Redmond, WA, and analysis is continuing at [removed for blind review]. The findings inform the development and design of other interactive television products and programs for early learners. The presentation also breaks down directions for future analysis, showing how initial findings illustrated a need for deeper research and analysis in the nuanced ways young children learn and demonstrate knowledge.

Introduction

A narrative experience is more than a “played story”. It’s the series of ways that engagement play out that is personalized, and contextualized. It may include a story, a series of mechanics, activities, and practices. The narrative worlds we engage in influence the creation of our own live narratives. And through engagement in these practices, relevant skills can become developed. One of the most common entry points for mediated narratives in the lives of children is through television programming. The role of television experienced a revolution in the 1970s through the production and growth of Sesame Street. In the 1990s, shows like Dora the Explorer and Blues Clues expanded the role of educational television and active participation, shifting again the way we think about educational viewing. Now we are on the cusp of a new paradigm shift for television and learning, with motion enabled peripheral devices like the Xbox Kinect actually *enabling* active participation, and viewers engaging with a responsive system, rather than a scripted vehicle with flat user-feedback. One platform coming out of Microsoft’s playful learning initiatives is two-way television, a means of combining entertainment with learning experiences by engaging physical activity and embodied learning practices. In Fall 2012, Microsoft Studios released *Kinect Sesame Street TV* and *Kinect National Geographic TV*, both of which utilize the Kinect motion capture system to allow episode viewers to engage with the show differently than if they were participating in a traditional television show.

Designers of *Kinect Sesame Street TV* worked to extend an already successful media legacy. The traditional Sesame Street television format has been shown to produce learning gains in younger viewers over the last forty years, including a longitudinal study that further supported these findings (Ball & Bogatz, 1970; Bogatz & Ball, 1971, Fisch & Truglio, 1991). The designers at Microsoft needed to make sure that the added Kinect interactivity wouldn’t break the potential for learning gains found in the linear television format (Fisch & Truglio, 2001). Microsoft’s design goals included the incorporation of research on learning and media use. First, according to situated learning theory, people learn content better when it is in the context of relevant activity. Therefore, the activity of the Kinect episodes must be tightly connected to the learning content (A. Games, personal communication, October 10, 2012). Secondly, viewing learning through a lens of embodied cognition highlights that a core way to learn a concept is by connecting that concept to one’s own perceptions, which includes relationships between the content and themselves/their own bodies. Kinect needed to allow participants to take on the specific actions that physically connect them to content (Glenberg, Goldberg, Zhu, 2011; Barsalou, 2008). Next, the entry threshold must be very low, making it simple for a variety of participants at different ability levels to enter the playful learning space. Finally, designers wanted to transform television into a deeper dialogic experience by embedding performative assessment within the episodes. The goal was to allow the environment to absorb the negative connotations of feedback, and playfully encourage participants to go further in their experiences (A. Games, personal communication, October 10, 2012). The result of this design process was the version of *Kinect Sesame Street TV* released in the fall of 2012. While playtests during the iterative design cycles helped fine tune interactions and navigation, questions still remain regarding the impact of this new medium on meaning-making for children in the target audience.

The emergence of a new media format that blends traditional elements of television viewing with game-like activities for learning leads researchers to ask how ideas of embodied cognition and comprehension can be investigated in bidirectional television products in order to understand the experiences of participants, explore the ways in which the products facilitate meaning-making, and utilize findings to support design of future interactive products. This study began to investigate these questions by studying the experiences of young children who participated with a single *Kinect Sesame Street TV* episode. Existing research provided a foundation for the new line of

educational research.

Assumptions:

1. The Sesame Street TV model results in increased comprehension, particularly with content emphasized most heavily (Ball & Bogatz, 1970; Bogatz & Ball, 1971, Fisch & Truglio, 1991).
2. There are short and long term benefits of educational programming on both specific knowledge and general skills, including school readiness (Fisch, 2004).
3. Repeated viewing increases comprehension (Crawley, Anderson, Wilder, Williams, & Santomero, 1999).
4. Visual attention is linked to cognitive processing. Children attend to things they're thinking about, such as cognitively demanding tasks, or content that is meaningful (Anderson, Lorch, Field, & Sanders, 1981).
5. Activity influences the way meanings are constructed (Barsalou, Niedenthal, Barbey, & Ruppert, 2003; Glenberg, Jaworski, Rischal, & Levin, 2007; Glenberg, Brown, & Leven, 2007).

Based on these premises, the following research questions guided the inquiry, methods, and data analysis. (See Figure 1 for a visual representation of the study flow.)

1. What does engagement in Sesame Street Kinect TV look like for 3-year old participants? (Keyword: **VIEWING**)
2. How do participants demonstrate engagement when utilizing interactive features of Sesame Kinect TV? How do participants not using interactive features of Sesame Kinect TV demonstrate engagement? (Keyword: **INTERACTION**)
3. What concepts are participants of interactive and traditional Sesame Kinect TV episodes learning? (Keyword: **COMPREHENSION**)
4. How does repeat participation impact interaction and comprehension? Does this differ between interactive and traditional users? (Keyword: **REPLAY**)

Analysis with the data set from this preliminary study will be comprehensive and ongoing. This report covers early outcomes from the study, however it should be noted that deeper quantitative and qualitative analysis will continue to take place, and findings may become deeper and more highly contextualized. The first phase of analysis included descriptive statistics of the participant, an early analysis of pre/mid/posttests, and a review of observation notes for themes and presenting variables that need deeper analysis.

Method

A group of forty-two 3 and 4 year-olds participated in the study. The group contained a mix of girls and boys from across the Seattle area. Participant families were required have regular access to an Xbox 360 and Kinect in the home, have never viewed *Kinect Sesame Street TV* episodes, and be proficient in the English language. Data was predominantly collected at the Microsoft User Research Labs, and consisted of video footage, observation notes, pretests and posttests, and parent surveys (including demographic data). Participants were divided into two groups of twenty by a process of stratified random sampling, accounting for gender and known family annual income. One group of participants was designated as the *KINECT* group, in which Sesame Kinect TV experiences took place as designed, with all interactions ON. The other group was the *TRADITIONAL* group, in which all interactions in the episode were turned OFF, and the participant experienced the same content like a standard TV show (See Figure 1 on next page for a study diagram).

The study began at the Microsoft Studios User Research Lab. All participants were accompanied by a parent or guardian, and took part in a pretest prior to the viewing of the episode in the lab. Pretests covered key content presented in the selected episode. Participants then watched the episode. Parents or guardians in the room with the child were encouraged to engage with their children during the episode in any way that felt natural or comfortable for them. The guardian also completed the demographic and media use survey at that time. All assessment and play was audio/video recorded. Following the viewing, participants completed a midtest that matched the pretest (with the exception of switching from predictive story telling to a retell based on image prompts).

The next phase of the study was the independent viewing of the episode. Following the first viewing in the research lab, parents/caregivers were given a disc with the same episode viewed in the lab (*KINECT* or *TRADITIONAL*) for

use on their home Xbox/Kinect. Parents were encouraged to allow the child to watch the episode at least 3 more times, and were asked to record when and how much of the episode was watched each time. They were also given the option of providing observational feedback, which many parents took advantage of.

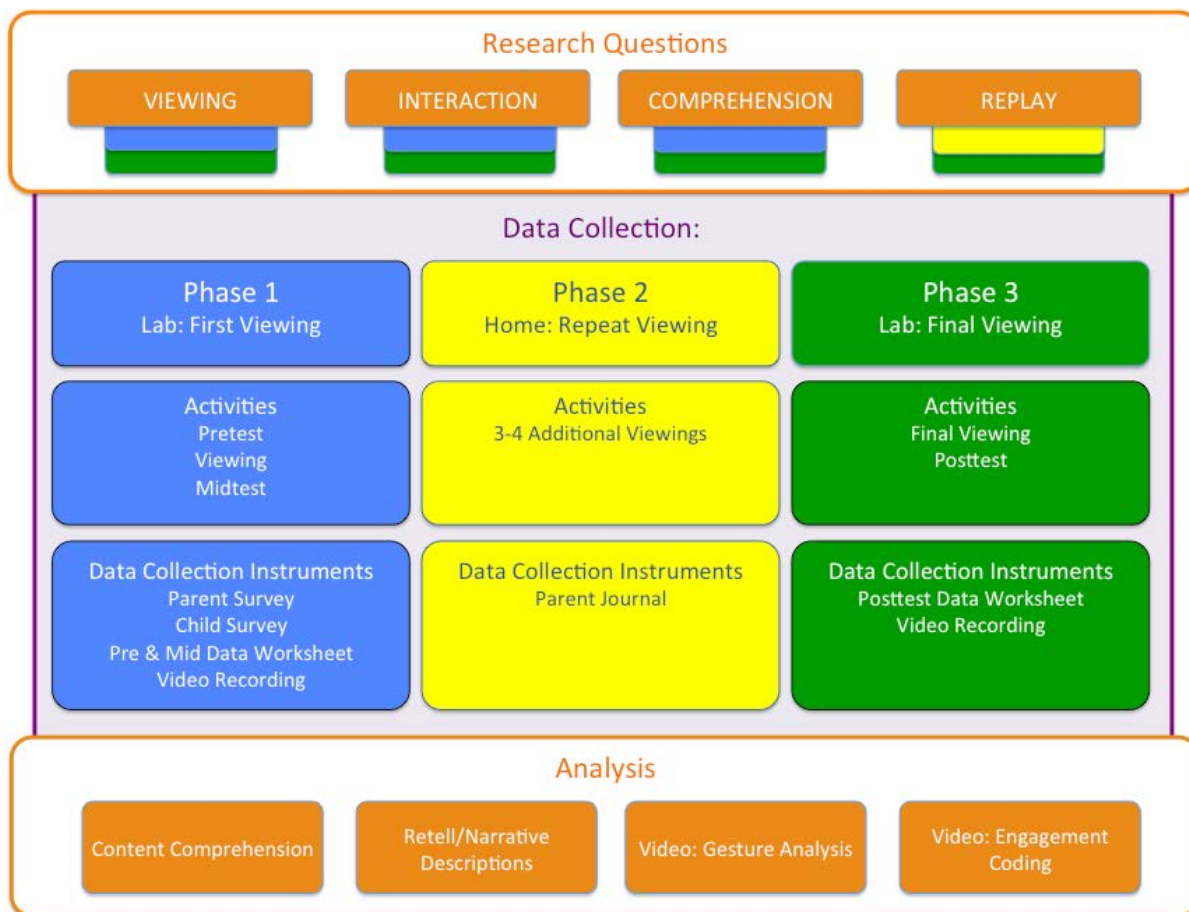


Figure 1: Visual diagram of Kinect Sesame Street TV Study.

The final phase brought participants back for a replay of the episode and post test. Participants returned to the lab approximately two weeks after their first visit. A replay of the same version of the Kinect Sesame TV episode was presented based on their status as *TRADITIONAL* or *KINECT* participants. Play was again be audio/video recorded, and parents/caregivers in the room were encouraged to engage as was natural and comfortable. All participants will took part in the same post-test immediately following their final in-lab play time, which matched the midtest.

Learning Objectives and Assessment Measures

The learning objectives assessed were culled from a content analysis of the episode selected (see Table 1 on the next page for a breakdown of assessment activities). The episode content was analyzed for core content frequency, and content tied to interactivity. The primary content identified as most targeted/emphasized in the episode were identified as core learning objectives that were assessed in the pretest, midtest, and posttest. Questions were supported with images for cued recall whenever appropriate. Questions also included performance-based activities, such as working with manipulatables, retelling, action-based content applications, etc., in order to maintain an “active participation” measure, linking to the elicited behaviors from the episode. Performance-based activities utilized non-digital manipulatables, so assessment activities in the pretest included a modeling component in an attempt to reduce potential confounds from switching activity modalities from digital to analog.

Content	Assessment Activity
Letter “H” Recognition	Identifying the letter “H” in a paper-based letter “H” show. Targeted response – clap for the letter H, don’t clap for other letters.
Vocabulary – Relational Concepts (Over, Across, Between)	Scene is created with manipulatable objects (hills, pond, Cooper puppet, Cooper’s camera). Child must navigate Cooper to the camera using the targeted relational concept.
Counting and Enumeration – 5	Child is tasked with counting five apples and putting them in Cookie Monster’s bucket.
Street Story Retell	Child is shown three key images from the street story “Failure to Launch” and is asked to retell what happened in the story.
Word understanding: “measure”	Prompted w/ image of Baby Bear and ruler, asked what it means to measure something.
Word understanding: “launch”	Prompted with images from street story, and given recap sentence that <i>Hubert the Human Cannon Ball needs to LAUNCH from the cannon to the bucket of blue gelatin that is precisely five feet away</i> . Child is asked what “launch” means.
STEM value: Curiosity & Wondering	Prompted with an image of Ernie singing his song about wondering. Child is asked if they get curious about things or if they wonder about things. (This question may be fairly abstract, but is left as an open question to see what kind of concepts the child attributes to the song. Is not evaluated for correctness/incorrectness.)
Open-ended feedback on episode experiences.	Child is prompted with key images from each segment of the episode. Child will be asked which were favorite parts, least favorite parts, what kinds of activities he/she did, etc.

Table 1: Assessment topics of Kinect Sesame Street TV pilot study.

Discussion of Findings

The study included a total of nineteen males and twenty-three females, all between the ages of 3 years old, and four years and eleven months old. Because recruitment for a group of that size that met the criteria for participation was a challenge, the study was opened up to employees of Microsoft and partner contract companies. In the study population, seventeen children were from families external to Microsoft, and twenty-five were from Microsoft/MS contractor families. The range of annual income was between \$0-15,000 and \$150,000+. The *KINECT* and *TRADITIONAL* groups were broken down evenly, with 21 children beginning in each group. The attrition between first and second visits to Microsoft was two participants; so twenty children from each group completed the viewings.

Early Analysis of Pre and Posttests

Early comprehension analysis included responses from the letter H game (four items), the relational concept manipulation activities (three items), and knowledge of the number five (two items) for a total of nine items. Preliminary analysis shows that there is a statistically significant increase for both *TRADITIONAL* and *KINECT* groups in the pre to mid to post assessments (See Figure 2 on the next page for a diagram of assessment gains). There is not a statistically significant difference between the two groups, nor is there a statistically significant difference between gender. These are preliminary findings, and require further investigation to consider possible influences of additional variables in player comprehension results.

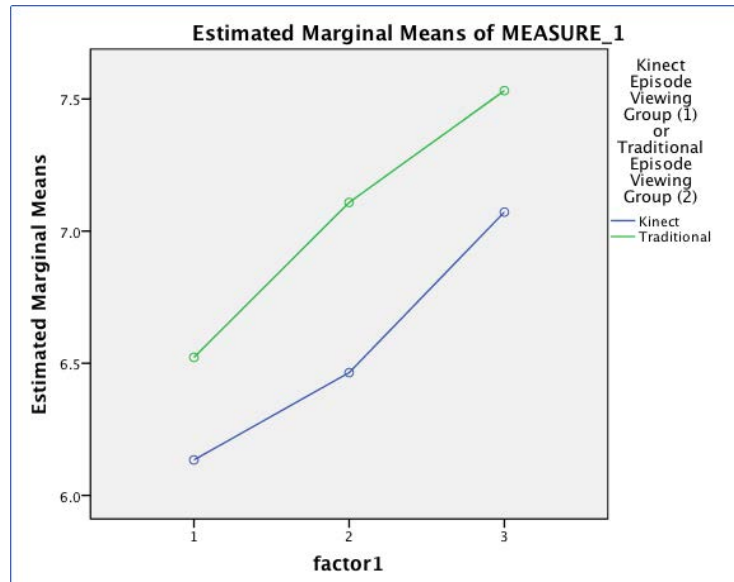


Figure 2. Estimated marginal means of KINECT and TRADITIONAL groups from pre (1) to mid (2) to post (3) test.

Emerging Themes and Areas For Deeper Analysis

Initial evaluation and notes from the viewing sessions illustrate emerging themes that require deeper evaluation. While early analysis of the comprehension data shows no significant difference between the *KINECT* and *TRADITIONAL* viewings and that both groups increased in targeted concept knowledge significantly, other observations indicate a difference in the ways participants engage with the program. For example, the *TRADITIONAL* viewers had more instances of kids watching from the floor on their tummy (lying down), and engaging physically with the prompts with less energy/movement size than those in the *KINECT* version. Also, less interactive episode segments seem to have higher rates of distractibility. The experiences of the participants are greatly nuanced, and require an analysis across multiple dimensions of viewing experience before broader and more declarative statements can be made.

One specific example is the need for deeper gesture analysis of the way participants demonstrated knowledge of the relational concept “between”. Things were initially coded with a correct/incorrect (1/0), however, the coding does not take into account the variety of ways children demonstrate their concept awareness. The assessment for the concepts took place in an activity that was designed to feel game-like. The researcher placed two green bowls upside down on the floor to represent hills or mountains, and placed a flat cutout of a lake near the middle of the two hills/mountains. A figure of Cooper Monster, one of the characters in the episode, was placed on the far side of one hill/mountain, and a figure of a camera was placed on the opposite side (See Figure 3, below).

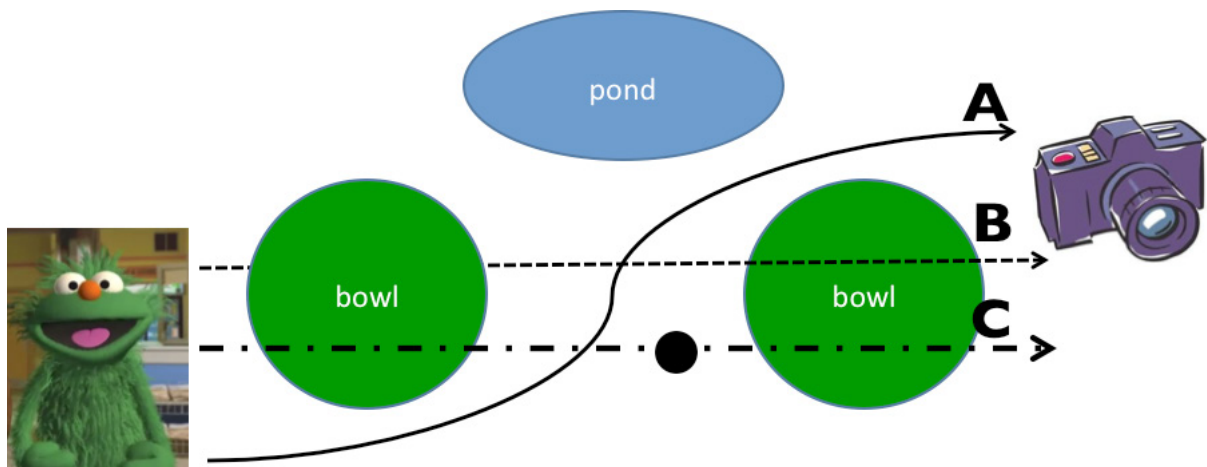


Figure 3. Demonstrations of the relational concept “between”.

In the relational concept assessment, children were told that Cooper lost his friend Flash (the camera), and that they were going to play a game to help Cooper find him. The researcher would flip a card with the relational concept (between, across, and over), and prompt the child to go between the mountains/hills, across the lake, or over the mountains/hills. Demonstration for “over” and “across” remained fairly straightforward. However, for the concept “between”, participants demonstrated it in a number of ways. The most seemingly logical response to “Help Cooper find Flash by going between the two mountains” would be to follow path A in the figure above. However, some children moved Cooper over the mountains with a solid landing in between the two (path C). Yet a couple others picked up the first bowl and slid Cooper between the bowl and the floor before continuing him on to Flash (path B). Still another child tried to push Cooper into the side of the bowl, as if wanting the bowl to split into two parts so Cooper could move between them (not pictured in Figure 3). Deconstruction of the ways children are demonstrating knowledge of the relational concept “between” will at the least require a new coding scheme and deeper gesture analysis.

Another example of the need for deeper gesture analysis of video data emerged with the way one participant showed her knowledge of the meaning of “measure”. In the pretest, she was asked what Baby Bear was doing in an image with him holding a ruler. She responded that he was measuring. When asked what a person is trying to find out when they measure something, she responded “how tall.” In the episode, she watched Baby Bear measure the distance that Hubert the Human Cannon Ball flew across Sesame Street. After watching the episode the researcher engaged her in the same line of questioning. She responded again with “how tall”, except this time, she held her hands out to indicate a horizontal distance. This shows that she is changing how she demonstrates her concept knowledge to align with the way it is framed in the episode.

More analysis will also take place with questions regarding gender. Early observations indicate that there may be differences in the ways girls demonstrate connections with narrative elements, as compared to participant boys. Finding the more nuanced differences will require an analysis of the coded engagement data that is weighed against the results of the discourse and gesture analysis.

Plans for Future Quantitative and Qualitative Analysis

The next phases in analysis includes detailed coding of participants’ open-ended answers. Coding will utilize a priori schemes, and may be adapted as patterns emerge. Engagement measures will be coded for every minute and a half of episode and will follow a Sesame Street engagement metric that has been modified for the Kinect activity and specific study. This includes noting whether the child has eyes on screen (EOS) or if they are not watching (NW). Specific behaviors will be noted, including whether a child is sitting, standing, or lying down; different types of expressive pleasure such as laughing, smiling, or clapping; responses to cues and prompts; unprompted actions or imitations; labeling, classifying, or predicting based on the actions on screen; and comments, judgements, or life connections in response to on-screen activity.

In addition, discourse and gesture analysis will take place on the open ended responses, as early observations indicate that the depth of answers that take place between first and second viewing may illustrate how participants express movement from shallower content knowledge to deeper and more situated understandings, that they then demonstrate through both discourse and gesture. More extensive findings from qualitative analysis will be presented for this presentation at the time of the conference.

Conclusion

We live in world where our experiences are increasingly technologically mediated. And these experiences can serve as the catalysts for experimentation, play, discovery, creation, and meaning-making. This means different things for different kinds of development. Childhood learning and development is a highly nuanced progression of experiences. The way they learn, know, and show is more complicated than “less learned grownups”, their processes of learning and experiences are complex, and deeply embedded in individual experiences and development trajectories. Their meaning making processes are highly socialized. Their play spaces and narrative interaction push boundaries of meaning. Meanings and identities get experimented with and worked out in play spaces. It’s an important place for children.

For both designers and researchers, we need to not only understand, but *respect* the complexities and nuances in our work. More research is needed to understand how physically interactive experiences like bidirectional television can be a tool for moving content from mediated experience to discourse and application. This research has been an exploratory pilot to begin to identify themes and connections between bidirectional television viewing and meaning-making, and this analysis an early first pass of the data corpus the pilot provided. The results from this study can be used to develop broader studies that include a larger participant base, multiple episodes, a more

targeted series of research questions, and even may include additional bidirectional properties. In addition, the outcomes from this and future studies can continue to inform design, from specific recommendations on interactions to broader design for learning questions, such as how activities can move a participant towards meaning and pleasure, how activities and practices empower the participant to make larger connections beyond the product, and given what we know and are continuing to learn about the ways children engage with their mediated narrative experiences, how products can serve as a catalyst for deeper engagement and creative expression.

References

- Anderson, D. Lorch, E. , Field, D. , & Sanders, J. (1981). The effects of TV program comprehensibility on preschool children's visual attention to television. *Child Development*, 52(1), 151-157.
- Ball S., & Bogatz, G.A., (1970). *The first year of Sesame Street: an evaluation*. Princeton, N.J.: Educational Testing Service.
- Bogatz, G.A., & Ball, S., (1971). *The second year of Sesame Street: a continuing evaluation*. Princeton, N.J.: Educational Testing Service.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-45.
- Barsalou, L.W., Niedenthal, P.M., Barbey, A., & Ruppert, J. (2003). [Social embodiment](#). In B. Ross (Ed.), *The Psychology of Learning and Motivation*, Vol. 43 (pp. 43-92). San Diego: Academic Press.
- Crawley, A. , Anderson, D. , Santomero, A. , Wilder, A. , Williams, M. , et al. (2002). Do children learn how to watch television? The impact of extensive experience with blue's clues on preschool children's television viewing behavior. *Journal of Communication*, 52(2), 264.
- Crawley, A. , Anderson, D. , Wilder, A. , Williams, M. , & Santomero, A. (1999). Effects of repeated exposures to a single episode of the television program blue's clues on the viewing behaviors and comprehension of preschool children. *Journal of Educational Psychology*, 91(4), 630-637.
- Fisch, S.M. (2004). *Children's learning from educational television: Sesame Street and beyond*. Mahwah, N.J.: Erlbaum.
- Fisch, S. M., & Truglio, R. T. (2001). *"G" is for growing : thirty years of research on children and Sesame Street*. Mahwah, N.J.: Erlbaum.
- Glenberg, A. M., Goldberg, A. B., & Zhu, X. (2011). Improving early reading comprehension using embodied CAI. *Instructional Science*, 39(1), 27-39.
- Glenberg, A. M., Jaworski, B., Rischal, M., & Levin, J.R. (2007). What brains are for: Action, meaning, and reading comprehension. In D McNamara (Ed), *Reading Comprehension Strategies: Theories, Interventions, and Technologies* (pp. 221-240). Mahwah, NJ: Lawrence Erlbaum Publishers.
- Glenberg, A. M., Brown, M., & Levin, J. R. (2007). Enhancing comprehension in small reading groups using a manipulation strategy. *Contemporary Educational Psychology*, 32, 389-399
- Goldin-Meadow, S. & Beilock, S. L. (2010). Action's influence on thought: the case of gesture. *Perspectives in Psychological Science*, 5, 664-674.
- Kinect Sesame Street TV [computer software]. (2012). Redmond, WA: Microsoft Studios
- Leslie, L., & Caldwell, J. (2011). *Qualitative reading inventory-5*. Boston, MA: Pearson.

Acknowledgments

We would like to thank Alex Games of Microsoft Studios and Rane Johnson of Microsoft Research for supporting the research that led to these findings.