Parent-Child Screen Media Co-Viewing: Influences on Toddlers' Word Learning and Retention

Clare E. Sims (Clare.Holtpatrick@Colorado.Edu)
Department of Psychology & Neuroscience, 345 UCB
Boulder, CO 80309-0345 USA

Eliana Colunga (Eliana.Colunga@Colorado.Edu)
Department of Psychology & Neuroscience, 345 UCB
Boulder, CO 80309-0345 USA

Abstract

Screen media, such as television and videos, are a common part of young children's lives. Yet infants and toddlers have been shown to learn less effectively from screens than from interactions with another person. Using a quasi-experimental design we explored how social factors of screen media co-viewing impact children's learning outcomes. We observed parents co-viewing a novel word training video with their children, then tested children for immediate and delayed word learning. We then investigated the links between parental speech during co-viewing and children's subsequent word learning. Parental speech that encouraged children to produce the novel words predicted better retention of word learning, whereas speech that focused more on the video itself rather than the content was negatively associated with learning.

Keywords: Screen media; co-viewing; word learning.

Introduction

Screen media, such as television, videos, computers, and hand-held devices like smartphones and tablets, are an increasingly common part of young children's lives. Although the American Academy of Pediatrics recommends that children under the age of two not watch any screen media (AAP, 2010), survey data reveal that 43% of children in this age group watch TV every day (Rideout & Hamel, 2006). This may be partly due to the recent proliferation of screen media content aimed specifically at infants and toddlers (e.g., *Baby Einstein*). Much of this content is presented with educational claims, including in the domain of language development (Fenstermacher et al., 2010). However, research that actually compares children's learning from screen media to their learning from face-to-face interactions with another person reveals a video deficit effect (Anderson & Pempek, 2005). That is, across varied experimental paradigms, infants and toddlers learn less effectively from a screen than from a person.

Is it possible to effectively facilitate infants' and toddlers' learning from screen media? Factors that may impact this learning include the complexity of the information on screen, whether key information is highlighted or repeated, or the presence of social cues that help children direct attention to what is important. In the work presented here we employed a quasi-experimental approach to investigate the role of a common social context for screen mediated learning: co-viewing a video with a parent. We explored the relationship between parental behavior during active co-viewing and children's learning outcomes. The results suggest meaningful links between characteristics of the screen media viewing context and children's learning of video content. Exploring the attributes and effects of co-viewing is vital for guiding further research on children's screen mediated learning and applying this work in practice. In the remainder of the introduction we will briefly review the evidence for the video deficit effect in the domain of language as well as the existing research on co-viewing with young children. Many questions on this topic remain, and our approach of linking co-viewing to word learning in the moment contributes to this emerging literature.

The Video Deficit in Language Learning

From birth, infants are constantly exposed to language, both from live and mediated sources. By manipulating whether information to be learned is presented through screen media or through face-to-face interactions, researchers have identified a video deficit effect in the domain of language development—infants and toddlers more effectively learn language from a person than from screen media. The video deficit has been demonstrated in two types of tasks in this domain: phoneme distinction and word learning.

One study took advantage of a change in phoneme discrimination that occurs in infancy. At six months of age infants can discriminate phonetic speech contrasts from their native language as well as from a non-native language to which they have never been exposed (e.g., Werker, Gilbert, Humphrey, & Tees, 1981). However, over the first year of life, infants become attuned to distinctions between speech sounds in their native language, leading to an inability to recognize speech contrasts that do not exist in that language. Kuhl, Tsao, and Liu (2003) tested what kinds of exposure to non-native speech distinctions can prolong 9-month-old infants' ability to perceive those distinctions. These researchers found a video deficit in maintaining phonetic distinctions: infants exposed to non-native speech on video lost their sensitivity to non-native speech distinctions, whereas those exposed to live speech were still able to discriminate between non-native phonemes. In fact, infants in the video exposure group performed equivalently to a control group with no exposure; screen mediated speech led to the same outcome as no speech at all.

Another major developmental step in language acquisition is learning words. Researchers have looked at how young children learn unfamiliar or novel words from various live and screen mediated sources. In a typical word-learning task, children are shown an object and told a label for that object. To test for learning, the child is presented with an array of objects, including the trained object and distractor objects, and is asked to identify the object corresponding to the trained label. One study found a U-shaped developmental progression of the video deficit in word learning: while children between 13 and 20 months of age showed a significant deficit, younger children (aged 6 to 12 months) and older children (aged 21 to 24 months) did not show a deficit (Krcmar, 2010). Another word learning study found evidence for a video deficit among children between 15 and 24 months of age, which was actually stronger after children were exposed to a commercial video compared to a lab-created video (Krcmar, Grela, & Lin, 2007). While the studies mentioned so far were conducted in a lab setting, another study confirmed the video deficit in word learning among 12- to 18-month-olds exposed to a commercial video in their own homes (DeLoache et al., 2010). One study using a slightly older age group did not find any evidence for a video deficit in a word learning task among 30-month-olds (O'Doherty et al., 2011). However, recent work in our lab suggests that the video deficit may persist depending on how word learning is assessed. We found that 30- to 36-month-olds learned and retained one-to-one word-referent mappings from a screen, but the same children showed a video deficit in their retention of lexical categories (Sims & Colunga, in preparation). When toddlers must generalize what they previously learned from a screen, inferring categories based on the words they learned, screen mediated learning still seems to be at a disadvantage.

In sum, studies that compare language learning from screen media to learning from a person show an early emerging video deficit effect that diminishes with age and depending on the type of task used. But why do children struggle with screen mediated learning when they can easily learn the same information in person? Some have proposed that social factors of screen media, or lack thereof, drive the video deficit effect (Richert, Robb, & Smith, 2011). The screen mediated environment typically lacks the kind of rich social interactions and contingencies that children get when learning directly from a person. The video deficit may be rooted in the socially impoverished nature of screen media itself, particularly in the fundamentally social domain of language learning. Therefore, the social context of screen mediated learning is the focus of a related area of research, including work on co-viewing.

Screen Media Co-Viewing

Most parents watch TV with their child either all or most of the time (Rideout & Hamel, 2006). Parental co-viewing may provide a social context for children's screen media use. By actively co-viewing with their children, parents have opportunities to scaffold children's learning from the screen. Most of the work linking parent-child co-viewing and children's learning from a screen has looked at slightly older age groups than studies of the video deficit effect. However, more recent research is beginning to explore how co-viewing impacts infants' and toddlers' attention to and learning from screen media.

Some studies of preschoolers have experimentally tested how different types of co-viewing interactions influence children's subsequent learning. For example, one set of studies tested co-viewing in a context familiar to many young children: *Sesame Street* episodes. In one study, Reiser, Tessmer, and Phelps (1984) manipulated whether or not adults asked content-specific questions and provided feedback and encouragement while co-viewing segments teaching letters and numbers. Children learned the content of the video more effectively in the experimental condition compared to the control condition, in which adults did not provide any commentary. Reiser, Williamson, and Suzuki (1988) added to this result by showing that asking questions, with or without providing feedback, resulted in better learning than simply directing children's attention to the screen when educational content was being shown. Together these studies show that adult commentary and questions during co-viewing can directly facilitate children's learning from real screen media content.

Fewer studies have linked co-viewing interactions to outcome measures among infants and toddlers. Some studies have linked qualities of parental behavior or parent-child interactions to a precursor for children's screen mediated learning: attention to the screen. Results show that parents' eye-gaze to a screen modulated 12- to 21-month-olds' attention to screen media (Demers, Hanson, Kirkorian, Pempek, & Anderson, 2012). Further, sensitive and reciprocal parent-child interactions during co-viewing predicted 6- to 18-month-olds' looking time to an infant-directed video (Fidler, Zack, & Barr, 2010). Another study classified parental interaction styles based on co-viewing behaviors (Barr, Zack, Garcia, & Muentener, 2008). Parents were classified into different levels of scaffolding, and these clusters of co-viewing behaviors predicted looking time and responsiveness to infant-directed videos among 12-, 15-, and 18-month-old infants. The high-scaffolding parents tended to use verbalizations that oriented their children to the video and focused on the content therein. Together these studies show that parents who used eye gaze, high-quality, responsive interactions, and content-focused verbalizations were most effective in establishing joint attention to the screen and getting their children actively involved in co-viewing.

Only one study that we are aware of has started to link observations of co-viewing to measures of learning specifically among this younger age group. In this study, parents co-viewed a video with their 12- to 25-month-old children that was intended to teach words (Fender, Richert, Robb, & Wartella, 2010). The authors observed parental co-viewing behaviors as well as child verbalizations, and, importantly, also asked parents which of the words in the

video their child was unfamiliar with. In this way, the authors measured learning by observing how often children produced words that they had been unfamiliar with prior to seeing the video. Parents tended to cluster into different groups depending on how much their co-viewing behavior was focused on teaching the words in the video to their children. Children produced more words that they were previously unfamiliar with when their parents had a higher teaching focus during co-viewing. Further analyses showed that these parents tended to focus specifically on the words that they knew their children were unfamiliar with. This result is particularly interesting in light of the literature on the video deficit effect in word learning. This study shows that sensitive parental scaffolding during co-viewing with infants and toddlers may be able to reduce the video deficit.

Rationale and Predictions

In the current study, we investigated the link between parental co-viewing behavior and toddlers' word learning. Parents and their 2½- to 3-year-old children watched a video of a person teaching novel words for novel objects. Children were subsequently tested on their word learning, both immediately after watching the video and after a week-long delay. Different kinds of parental speech during co-viewing were coded, analyzed using principal component analysis, and examined as predictors of children's word learning.

This approach offers several contributions to the literature. First, by training and testing children on novel words for novel objects, we controlled for any prior knowledge or exposure children may have had to the content of the task. Second, by testing children on the content of the video, both immediately and after a delay, we were able to make direct links between parental speech during co-viewing and children's word learning and retention. Third, we included an older age group compared to most studies of the video deficit in word learning to see what behaviors impact learning once children are becoming better able to learn from a screen. Fourth, this work uses an interdisciplinary approach, drawing on methodologies from the fields of linguistics and psychology. The results will provide a first step in identifying specific co-viewing behaviors that facilitate, or possibly inhibit, word learning and retention in young children.

Based on the work reviewed above, the extent to which parents focus their speech on the key information to be learned on screen should predict children's word learning performance. Parents who are responsive and help their children focus on the content of the video should promote their children's attention to and learning from the screen. In this study that means talking about specific objects shown and novel words presented in the video. Further, parents who focus more on the novel labels being taught in the task should help children better learn the correct word-object mappings. An emphasis on the novel labels concurrent with the presentation of objects in the video should help children establish and retain these mappings.

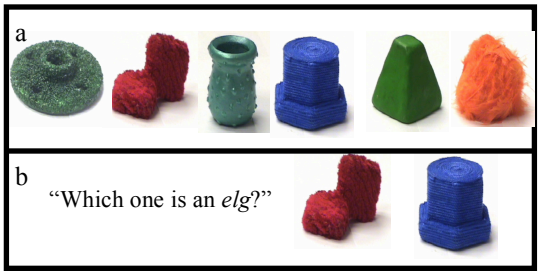

Figure 1. a. Novel target objects taught to children in the word learning task. b. Example word learning trial.

Method

Participants

Fifty children were recruited for participation from the Boulder, CO area. Six subjects were excluded from analyses due to missing or inadequate co-viewing data (two children) or co-viewing speech being primarily in a language other than English (four children). Therefore, the final sample included here consisted of 44 children (M_{age} = 32.1 mo., *SD* = 1.3 mo., 25 girls).

Materials

Children were taught six novel words (*elg*, *ife*, *nork*, *gub*, *zeb*, and *lug*) for six novel objects (see Figure 1a). The novel objects and words were presented in a lab-made training video of a research assistant whom the children did not meet during the study. In the video, the assistant presents one novel object at a time, placing it on a table in front of her and rotating it as she speaks. Addressing the camera directly, she labels the object in three ways: "This is a/an __. Do you see the __? This is my __." After an object is labeled, the video cuts to a 3 second still close-up image of the object. Each object is presented two times each, and thus labeled six times total, for a total video duration of 2 minutes and 50 seconds.

Two standardized vocabulary measures were used to assess language development. Children were tested on the Peabody Picture Vocabulary Test 4 (PPVT; Dunn & Dunn, 2007), a test of receptive vocabulary. Parents were given the MacArthur-Bates Communicative Development Inventory (CDI-III; Fenson et al., 2007), a checklist on which they indicated words their children knew. Parents also completed a survey on their children's screen media use at home.

Parent-child interactions during training video co-viewing were recorded and later transcribed in ELAN (Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands: http://tla.mpi.nl/tools/tla-tools/elan/; Brugman & Russel, 2004). Parental verbalizations were subsequently coded using a scheme adapted from Barr et al. (2008). We used codes from this study and also developed codes specifically related to labeling in our task (see Results section for detailed description of the included codes).

Table 1: Factor loading values for the four components resulting from the factor analysis.

	Describing Objects	Label Elicitation & Feedback	Narrating	Open-Ended Questions vs. Explicit Labeling
Tag questions	**.664**	.389	-.057	.125
Descriptions	**.853**	-.341	-.060	-.040
Label elicitation questions	.189	**.777**	-.024	.012
Confirmations	-.270	**.739**	-.112	-.066
Evaluations	.141	-.146	**.831**	.007
Interactive verbalizations	-.245	.020	**.846**	-.061
Wh- questions	-.411	-.199	-.183	**-.774**
Labels	-.179	-.190	-.198	**.903**

Procedure

Training At the beginning of their first visit to the lab, parents and children watched the novel word training video. Parents were encouraged to actively co-view the video with their children. The experimenter explained that the video was meant to teach some new words and that the parent could teach their child about the words in the video as they would while watching at home. The experimenter left the room for the remainder of training, and parent-child co-viewing was videotaped for later analysis. When the training video had ended, the experimenter re-entered the room.

Testing Children were tested both immediately after watching the training video and again a week later. To test word learning, children were presented with pairs of trained novel objects and asked to identify one by name (e.g., "Which one is an *elg*?"; see Figure 1b). Each object was asked for once, for a total of six testing trials. At the end of their first visit to the lab children were given the PPVT vocabulary test.

Children were given the same word learning test at their second visit. Importantly, children were not re-trained on any of the novel words or objects at this time. Therefore, the delayed testing session captured retention of the novel words.

Results

The first question to assess was how well children learned and retained the novel word-object mappings. Children's proportions of correct target object choices at each testing session were first compared to chance performance. Children were accurate at above-chance levels both at immediate (M = .57, SD = .22, $t(43)$ = 2.20, p = .03) and delayed testing (M = .58, SD = .20, $t(43)$ = 2.58, p = .01). Next, a paired t-test showed that accuracy did not differ between visits (t < 1, p > .05), confirming that children retained the word-object mappings they had learned initially. It is worth noting that accuracy performance at the group level was not particularly high, and yet varied a fair amount across individual subjects. This may suggest that something about the learning context of individual children influenced their accuracy in the task.

The next question was how parent speech during co-viewing related to children's word learning performance. We began with two common measures of the quantity of parental speech: total word count and mean length of utterance (MLU). These variables were entered as predictors of children's word learning outcomes at each visit in two multiple regression analyses. These measures of parental speech quantity did not explain a significant proportion of the variance in either immediate or delayed child word learning performance (R^2 < .18, $F(2, 43)$ < 1, p > .05 for both models). Further, neither total word count nor MLU were significant predictors of word learning outcome at either visit ($t(41)$ < 1.50, p > .05 for all coefficients). Because these parent speech quantity variables did not predict learning outcome, we next explored the quality of parental speech during co-viewing.

First, to reduce the dimensionality of the qualitatively coded parent speech data, we conducted a principal component analysis (PCA). After removing several codes that were used by very few parents in the sample, eight coding categories were entered into the PCA and resulted in four components with Eigenvalues above 1.0. The first component explained 19.47% of the variance, the second 18.94%, the third 18.74%, and the fourth 18.01% for a total explained variance of 75.15%. An orthogonal Varimax rotation was used to facilitate interpretation of the components.

The four components are shown in Table 1 with factor loadings on each included coding category. The first component included tag questions and descriptions of the items shown in the video. Tag questions, which are statements with a question appended at the end, were typically used by parents to talk about the items on screen. Many of the observed instances of tag questions were also descriptions (e.g., "it's a green lug, huh?"). This component will be referred to as describing objects because it captures parents' focus on the individual objects depicted on the screen. The next component includes label elicitation questions and confirmations, which also appeared together often in co-viewing speech. This component will be referred to as label elicitation and feedback because it captures how often parents explicitly asked children to produce labels, including giving positive feedback for doing so. The third component includes evaluations and interactive verbalizations. Parents often used evaluations to make general comments about the video or item shown on screen (e.g., "that's a cool one"), and interactive verbalizations were comments about the video itself (e.g., "it says let's see those again"). This component will be called narrating because it captures parental speech about the video and

Table 2: Multiple regression output for the novel word learning task at immediate and delayed testing.

	Immediate Word Learning			Delayed Word Learning		
	Unstandardized Coefficients		Standardized Coefficients	Unstandardized Coefficients		Standardized Coefficients
Predictors	B	SE	Beta	B	SE	Beta
Age	-.017	.030	-.102	.001	.025	.010
Gender	-.091	.075	-.209	.051	.063	.130
Vocabulary Percentile	-.001	.001	-.076	.001	.001	.169
Screen Time	.000	.001	.077	.000	.001	-.104
PCA Components						
Describing Objects	.021	.036	.096	-.020	.031	-.101
Label Elicitation & Feedback	.025	.038	.114	.069	.032	.335*
Narrating	.005	.036	.022	-.052	.031	-.269†
O-E Questions v Explicit Labeling	.046	.035	.211	.015	.030	.079

Note. †$p < .10$. *$p \leq .05$.

screen viewing context more broadly. The fourth and final component includes wh- questions and explicit labeling. These codes loaded in opposite directions onto this component, so the component will be referred to as open-ended questions vs. explicit labeling. In the positive direction this component captures the extent to which parents provided the novel labels being taught on screen, and in the negative direction it captures the extent to which parents asked their children open-ended questions about the video (e.g., "what is that?").

Hierarchical multiple regression analyses were conducted on children's word learning outcomes at each visit. In each analysis, the independent variables were entered in two blocks. Demographic and standardized test variables were entered in the first block. These included child age in months, child gender (dummy coded), vocabulary percentile score (averaged over the CDI-III and PPVT), and average screen use per day in minutes. The second block included the four co-viewing components from the factor analysis. This method of analysis allowed for evaluating the predictive value of the co-viewing components over and above the other included variables.

Table 2 displays unstandardized and standardized regression coefficients after entry of both independent variable blocks for immediate and delayed testing. None of the included variables significantly predicted immediate word learning accuracy ($R^2 = .37$, $F(4, 43) = .68$, $p > .05$ after entry of all independent variable blocks). Although the full model did not reach significance ($R^2 = .48$, $F(4, 43) = 1.31$, $p = .27$), delayed word learning accuracy was predicted by two of the independent variables entered into the regression.

Parental label elicitation and feedback was a positive predictor of learning, indicating that the more parents asked children to produce novel labels and provided positive feedback, the better children retained word-object mappings. Specifically, the model predicts that for every standard deviation increase in parents' use of label elicitations and feedback, children's delayed word learning performance should increase by 0.34 standard deviations. On the other hand, narrating was a marginally negative predictor of retained novel word learning. This suggests that parental speech that was focused on the video itself and the content only in a general way actually inhibited children's correct retention of novel word-object mappings in this task. Specifically, the model predicts that for every standard deviation increase in parents' use of narrating speech, children should be 0.27 standard deviations worse at retaining the novel word-object mappings.

Discussion

The results of the current study show that although toddlers as a group learned novel words from a video, certain aspects of parental speech were associated with differences in this learning. An exploratory factor analysis revealed several variables of parent speech quality that characterized the co-viewing linguistic environment and that predicted children's learning outcomes. Of note, the quality of parent speech only predicted children's retention of learning. This suggests that toddlers' immediate word learning from screen media may be relatively robust, and less influenced by the co-viewing environment. Yet the retention of this information may be sensitive to parental intervention.

The first key result showed that the extent to which parents elicited labels from their children and provided feedback while watching the training video predicted children's retention of word-object mappings. Although the finding that labels during co-viewing facilitated word learning is in line with prior work and with our predictions, the specific form of this labeling is informative. Children's retention was predicted not by hearing parents label the items on screen, but by parents cuing children to produce the labels themselves and providing responsive feedback.

Another key result was that the extent of parents' narrating during co-viewing was negatively associated with children's retention of word learning. This shows that parent speech about the video itself or general, non-specific speech about what is shown on screen is not conducive to novel word retention. This result suggests a negative impact on learning due to focusing on the form rather than the content of screen media. Further, this also suggests that general evaluative speech about the content on screen is not much more informative than talking about the video itself, and both of these together may actually impede learning outcomes.

Together these results are consistent with prior work indicating that responsive behavior during co-viewing promotes children's attention to and learning from screen media. The results build on prior work by demonstrating specific co-viewing behaviors that are responsive and thus scaffold children's learning, as well as behaviors that are

linked to detrimental learning outcomes. Together they provide new insight into the role of co-viewing in learning.

The results of this study represent a first step in linking specific co-viewing behaviors to children's learning outcomes in the context of screen media. There are various ways to refine and build on this work. One future direction would be to incorporate measures of child behavior and parent-child interaction quality in the kinds of analyses presented here. The relative timing of utterances and responses between a child and parent may be particularly predictive of word learning. For example, children may learn most effectively when parents respond promptly and provide information about the item that is the focus of the child's attention in that moment. The co-viewing data collected for this study could be coded for contingencies in interactions between parents and children. This kind of analysis would also resonate with research on the social aspects of screen mediated learning (e.g., Richert et al., 2011). Although social information was not manipulated directly in video training in this study, it could be informative to test how different extents of social contingency in co-viewing link to learning outcomes.

Another future direction for this work would be to guide experimental investigations of co-viewing. The kinds of analyses presented here can be used to develop experimental manipulations of the linguistic environment surrounding screen media co-viewing. For example, the current results suggest that the type and extent of labeling during co-viewing may impact learning in different ways. This could be tested by manipulating whether labels are provided to children or elicited from them and how many labels are used during co-viewing. This would allow for greater control of other characteristics of the co-viewing context, randomized assignment of children to conditions, and causal conclusions about the role of labels in screen mediated word and category learning. Similar experiments could be designed to test the effects of specific, content-focused speech compared to broad, screen-focused speech during co-viewing. Future work could also investigate co-viewing and learning outcomes from real, professionally produced child-directed media. In conclusion, the current study makes novel contributions to the emerging literature on screen mediated language learning in young children and highlights directions for future research on facilitating this learning.

References

American Academy of Pediatrics, Council on Communications and Media (2010). Media education. *Pediatrics, 126*(5), 1012-1017.

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *Am. Behav. Sci., 48*(5), 505-522.

Barr, R., Zack, E., Garcia, A., & Muentener, P. (2008). Infants' attention and responsiveness to television increases with prior exposure and parental interaction. *Infancy, 13,* 30-56.

Brugman, H., & Russel, A. (2004). Annotating multimedia/ multi-modal resources with ELAN. *Proc. of LREC 2004.*

DeLoache, J. D., Chiong, C., Sherman, K., Islam, N., Vanderborght, M., Troseth, G. L., Strouse, G. A., & O'Doherty, K. (2010). Do babies learn from baby media? *Psych. Science, 21*(11), 1570-1574.

Demers, L. B., Hanson, K. G., Kirkorian, H. L., Pempek, T. A., & Anderson, D. R. (2012). Infant gaze following during parent-infant coviewing of baby videos. *Child Dev.*

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test.* Minneapolis, MN: Pearson Assessments.

Fender, J. G., Richert, R. A., Robb, M. B., & Wartella, E. (2010). Parent teaching focus and toddlers' learning from an infant DVD. *Infant and Child Dev., 19,* 613-627.

Fenson, L. et al. (2007) *MacArthur-Bates communicative development inventories: User's guide and technical manual.* Baltimore, MD: Paul H. Brookes Pub. Co.

Fenstermacher, S. K., Barr, R. R., Salerno, K., Garcia, A., Shwery, C. E., Calvert, S. L., & Linebarger, D. L. (2010). Infant-directed media: An analysis of product information and claims. *Infant and Child Dev., 19*, 557-576.

Fidler, A.E., Zack, E., & Barr, R. (2010). Television viewing patterns in 6- to 18-month-olds: The role of caregiver-infant interactional quality. *Infancy, 15,* 176-196.

Krcmar, M. (2010). Can social meaningfulness and repeat exposure help infants and toddlers overcome the video deficit? *Media Psychology, 13*(1), 31-53.

Krcmar, M., Grela, B., & Lin, K. (2007). Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology, 10,* 41-63.

Kuhl, P. K., Tsao, F., & Liu, H. (2003). Foreign-language experience in infancy: Effect of short-term exposure and social interaction on phonetic learning. *PNAS, 100*(15), 9096-9101.

O'Doherty, K., Troseth, G. L., Goldenberg, E., Akhtar, N., Shimpi, P. M., & Saylor, M. M. (2011). Third-party social interaction and word learning from video. *Child Dev., 82*(3), 902-915.

Reiser, R.A., Tessmer, M.A., & Phelps, P.C. (1984). Adult-child interaction in children's learning from "Sesame Street." *Edu. Comm. & Tech. Journal, 32,* 217-233.

Reiser, R.A., Williamson, N., & Suzuki, K. (1988). Using "Sesame Street" to facilitate children's recognition of letters and numbers. *Edu. Tech. Res. and Dev., 36,* 15-21.

Richert, R. A., Robb, M. B., & Smith, E. I. (2011). Media as social partners: The social nature of young children's learning from screen media. *Child Dev., 82*(1), 82-95.

Rideout, V. J., & Hamel, E. (2006). *The media family: Electronic media in the lives of infants, toddlers, preschoolers, and their parents.* Henry J. Kaiser Family Foundation.

Sims, C. E., & Colunga, E. (In preparation). Toddlers' word and concept learning from screen media.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Dev., 52*(1), 349-355