Infant Behavior and Development 64 (2021) 101574

Contents lists available at ScienceDirect

Infant Behavior and Development

journal homepage: www.elsevier.com/locate/inbede

Learning to learn from video? 30-month-olds benefit from continued use of supportive scaffolding

Gabrielle A. Strouse [a,b,*], Patricia A. Ganea [c]

[a] Division of Counseling and Psychology in Education, University of South Dakota, United States
[b] Center for Brain and Behavior Research, University of South Dakota, United States
[c] Department of Applied Psychology and Human Development, Ontario Institute for Studies in Education, University of Toronto, United States

ARTICLE INFO

Keywords:
Scaffolding
Word learning
Symbolic development
Video

ABSTRACT

Young children struggle to learn new words presented on video, but adult co-viewers can support them by providing scaffolds that explicitly connect the video and real world. In this study, we asked whether scaffolding facilitates children's symbolic understanding of the video, such that they will subsequently transfer labels from video to real referents. Sixty-three 30-month-olds and 61 36-month-olds participated in a series of three word learning trials in one of three conditions. In the supportive condition, an in-person adult explicitly drew connections between each on-screen object and the corresponding real object in the room with the child. In the unsupportive condition, the in-person adult provided similar-length statements about the objects but did not draw connections between them. In the partial scaffold condition, the in-person adult provided the supportive scaffolds for the first two trials and the unsupportive version for the third trial. At 30 months, children selected the correct object on the third trial more often in the supportive than the unsupportive scaffold condition, and performance in the partial scaffold condition fell in between. At 36 months, performance on the third trial did not differ across conditions. The results showed that experiencing the scaffold twice was not enough to reliably support 30-month-olds in learning to think symbolically on the third trial; rather, they appeared to rely on the adult to connect the video image with its specific real-world referent. At 36 months, however, children did not rely on the adult scaffold to apply the video label to the real-world objects.

1. Introduction

Although there is a rich research literature on preschoolers' learning of vocabulary, numbers, and letters from educational programs like *Sesame Street*, evidence for infants and toddlers' learning from television and videos is much more mixed (Strouse, 2019). Researchers have documented that children, especially the youngest children, often learn more from in-person teaching than from the same information on video (the *video deficit*, Anderson & Pempek, 2005; Strouse & Samson, 2021). Despite this, many children's programs, and at least one entire television channel (BabyTV), are targeted at infants and toddlers, and parents of 0 to 8-year-olds report that educational or 'learning' videos are a popular category for young children to watch (Rideout, 2017). If infants and toddlers are viewing content that is intended to teach them new information, it is prudent to understand how we can best support them to overcome difficulties that the video format might present. The purpose of this study was to examine whether adults could help young

* Corresponding author at: 414 E Clark Street, Vermillion, SD, 57069, United States.
E-mail address: gabrielle.strouse@usd.edu (G.A. Strouse).

https://doi.org/10.1016/j.infbeh.2021.101574
Received 29 October 2020; Received in revised form 5 May 2021; Accepted 7 May 2021
Available online 1 June 2021
0163-6383/

children learn how to learn from video by providing scaffolding. Specifically, we were interested in whether having experienced scaffolding would contribute to children's ability to continue learning from video even after the supportive scaffolding was no longer provided.

1.1. Learning language from video

We opted to focus on children's language learning from video, specifically on children's ability to learn the label given to a novel object by an adult on video. Language learning is one of the three domains in which learning from video has been studied somewhat extensively in children age 6 years and under, and in which a consistent video deficit pattern has been observed (Strouse & Samson, 2021). It is also a domain in which several mechanisms appear to converge to make learning from video an especially difficult task for young children.

First, two-dimensional videos lack depth cues, which may make encoding memories cognitively demanding and result in less detailed memories (Carver et al., 2006; Kirkorian et al., 2016). A lack of detail could make memories hard to retrieve, and slight mismatches between memories and later real-world experiences could make transfer to later situations difficult (Barr, 2010; Simcock & Dooley, 2007; Zack et al., 2009). As children's working memory develops, they may be better able to overcome these challenges (Kirkorian, 2018). However, when children are young, these constraints may make it especially difficult for them to learn and apply information that requires transfer from the screen to the real world, such as when children are asked to apply a new label for an object presented on screen to a real-world version of that object.

Second, it can be challenging to conceptually connect video information with the real world. Even if children can retrieve the correct video memories when faced with a real-world task, they may not understand that the video information also applies to the corresponding real-world situation (Troseth, 2010). Understanding the link between the video depiction and the real world requires *representational insight,* the realization that a symbol stands for something other than itself (DeLoache & Marzolf, 1992). Once children can thus represent the dual nature of the video images, both as video images themselves as well as images that stand for and provide useful information about real-world referents (*dual representation*; Troseth & DeLoache, 1998), their transfer of information from video to the real world will improve. In a label-learning task, the ability to understand the symbolic nature of the video image and draw inferences between the image and the object it represents, will contribute to children's ability to apply the label to both the image and referent itself (DeLoache et al., 1996).

Finally, videos are relatively socially impoverished when compared to in-person learning (Krcmar, 2010; Kuhl, 2007). In some videos the speaker may not be visible, so referential cues like gaze and gesture, which children older than 9–12 months use to help them learn language (Akhtar & Tomasello, 2000; Baldwin, 1993), are completely absent. When speakers do appear on screen and use referential cues to identify which objects they are speaking about, children typically learn more words from them than when referential cues are absent (see Strouse, 2019). However, if the video is pre-recorded, the speaker cannot tailor their cues to the viewer's actions. For example, they cannot detect that the child is inattentive and call their name or point to clarify when they notice a child is misdirected. Therefore, labels may be more difficult for children to learn from video because they are unable to use joint attention and have limited referential cues to help them map the labels to objects, cues which are often used by children learning in vivo.

1.2. Co-viewer support

One way that adults can support young children in learning from video is by co-viewing it with them (Hirsh-Pasek et al., 2015). Scaffolding provided during co-viewing may support children through multiple mechanisms, including attentional, cognitive, and social support (Strouse et al., 2013). For example, parents may direct children's attention to important on-screen information. Infants as young as 6 months attend more to the screen when their mother talks with them and shares joint visual focus on it (Fidler et al., 2010). Infants between 12 and 21 months often follow their parent's gaze to the screen, shifting their gaze to it directly after their parent does (Demers et al., 2013). Infants between 12 and 25 months also attend more, vocalize more, and are more likely to transfer information to the real world when parents talk about what is on screen (Barr et al., 2008; Fender et al., 2010; Zack & Barr, 2016). For older children (ages 6–13 years), the presence of a co-viewing parent during television viewing can lead to increases in emotional arousal and allocation of cognitive resources to processing the on-screen content (Keene et al., 2019; Rasmussen et al., 2016). In some studies, co-viewers who interact actively with children support engagement more than those who are merely present. For example, 24- and 30-month-olds were more attentive and responsive when a co-viewing adult modeled responsiveness to the screen than when co-viewers did not do or say anything related to what was occurring screen (Myers et al., 2019; Strouse et al. 2018). In the case of language learning, co-viewers may encourage children to engage with videos and direct them to look to the correct referent while it is being labeled, supporting new word-object pairings. The co-viewing adult's responsiveness to the screen may also cue children that what is happening on screen is meaningful and relevant to them, and therefore worthy of learning (Strouse et al., 2018).

Co-viewing adults who interact with their children during video viewing, providing cognitive and social support, often do help them learn words. Scaffolds may include repeating information from the screen, expanding on it, or asking children recall questions. In one study, 3-year-olds whose parents were trained to ask them questions about video storybooks learned more new words from the videos than children whose parents were not trained or whose parents only made statements about what was occurring on screen (Strouse et al., 2013). In another study, 3-year-olds were more likely to learn verbs from video if a co-viewer acted out the action with dolls, than when they only saw the action performed on video (Roseberry et al., 2009). These scaffolds may help children overcome the challenges of learning from screens by cognitively supporting their encoding and retrieval of new information, as well as providing contingently responsive social cues, like joint attention, that are absent between a pre-recorded video speaker and child learner.

Furthermore, scaffolds that provide specific information about the link between two-dimensional images and the real world may provide additional conceptual support for children who do not yet consistently realize that images contain information that can be applied to real world referents. Chen (2003, p. 421) found that providing a set of scaffolds to children including that "the pictures you see may help you play the game" supported 3- and 4-year-olds in transferring a problem-solving strategy from pictures to a game played with real-world objects. In other tasks that require children to think symbolically, explicitly pointing out the correspondence between the symbol and its specific referent can be supportive (DeLoache, 1989; Kuhlmeier, 2005). For example, explicitly pointing out that furniture in a model room corresponded to furniture in a larger, real room helped 36-month-olds use information from the model to find hidden objects in the real room (DeLoache, 1989). In another study (Kuhlmeier, 2005), pointing to the location in the model room and explicitly stating it was "in the same place" in the larger room also supported 30-month-olds, although only on the first trial.

Strouse and Troseth (2014) used this type of scaffold with a video word learning task with 24-month-olds, by having parents hold up the real version of objects that appeared on screen and explicitly state their connection to the real objects in the room (i.e., "These are the same as the ones on TV"). Subsequently, these 24-month-olds were more likely to apply the new label from the video to the real-world objects compared to children whose parents provided no scaffolds. In that study, the parents did not repeat the label offered by the person on video, only highlighted the relation between the real and screen objects. Directing children's attention to the correspondence appeared to support them in associating the specific on-screen and real objects, leading them to apply the new label from the video in the real world. An open question is whether children learned only the specific scaffolded association between the label and object, or whether the scaffold helped children take a more general approach toward thinking symbolically in video contexts. If the latter, children should continue to succeed at the task even when the supportive scaffold is no longer present.

1.3. *Learning to learn*

Brown and Kane (1988) found that 3-year-olds could learn to look for correspondences to solve sets of analogous problems. Children were presented with pairs of problems. An adult helped them to come up with a solution to the first problem, then scaffolded them in transferring the solution to the next problem. On the first two pairs of problems, 3-year-olds struggled to transfer the solution. However, by the third problem set, they had learned to look for the analogy. That is, they had learned a new mindset for approaching the problems, which involved looking for similarities between the two problems in each pair.

It is possible that children, when given explicit information about the similarities between video and real objects during several video word learning tasks, could form a mindset to look for connections between videos and their referents in similar tasks. DeLoache (1995) used the term *symbolic sensitivity* to refer to a mindset in which children increasingly look for symbol-referent relations. She argued this was the primary mechanism by which children's understanding of symbols develops. She also showed that it was possible to scaffold 30- and 36-month-old children in looking for symbolic relations, e.g., by first testing them in an easy symbolic task and then asking them to complete a more difficult one (Marzolf & DeLoache, 1994). In the current study we focus on whether scaffolding of the symbolic relation between video and real-world objects could support children's development of symbolic sensitivity and result in better transfer of words from video images to real-world referents.

1.4. *The present study*

In prior studies, explicitly pointing out the relation between a symbol and its referent helped 24- to 36-month-old children solve a symbolic task (DeLoache, 1989; Kuhlmeier, 2005), including applying a label learned on video to a real-world object (Strouse & Troseth, 2014). An open question is whether experiencing this type of scaffold supports children only in drawing a specific association between symbol and referent, or whether it may also support their symbolic sensitivity, or their tendency to look for symbol-referent relations during future video tasks. We approached this problem by providing children with scaffolds like those used by Strouse and Troseth (2014), which supported 24-month-old children in a label-learning task. Children repeated the task three times, so we could test what happened when children who had experienced a supportive scaffold on the first two trials were then asked to learn a label without the support. Success on this final trial would indicate that children had learned to look for symbolic relations on their own.

In the supportive scaffold condition, children heard supportive scaffolds on all three trials. We explicitly drew connections between on-screen and real objects during a familiarization phase. The scaffolds were offered prior to labeling with the goal that they might help establish the relevance and relation of the real objects to the video, such that when children later heard the label they would understand it was meant to apply both on screen and in the real world. After the familiarization phase, children watched a video presentation in which an object was given a new label. Children were tested on their learning after the labeling presentation by asking them to identify the real-world object that corresponded to the label.

In the partial scaffold condition, we provided supportive scaffolds prior to the first two labeling presentations, but prior to the third presentation we provided unsupportive scaffolds (matched in number and timing). In the unsupportive scaffold condition, children heard the unsupportive version of the scaffolds prior to all three labeling presentations.

The primary comparison of interest was in performance on the third trial, when the supportive scaffold was replaced by an unsupportive scaffold for the partial scaffold group. If children had learned to learn from video on their own (absent the supportive scaffold), the partial scaffold group would show above-chance performance in this final trial, similar to children in the supportive scaffold group and better than children in the unsupportive scaffold group. However, if children had not learned to learn from video on their own, performance in the partial scaffold group would match that of the unsupportive group on the third trial. We also measured children's handling of the objects, as way of checking that any differences in learning were not simply due to an effect of our scaffolds increasing children's engagement with the target objects.

We adapted a word learning task previously used by O'Doherty et al. (2011). We chose to study a group of 30-month-olds because at this age children are beginning to reliably succeed on some symbolic tasks using pictures and video (DeLoache & Burns; 1994; Schmitt & Anderson, 2002; Chen & Siegler, 2013; Troseth & DeLoache, 1998). However, they do not yet display consistent video-to-real-world transfer in all tasks; 30-month-olds struggled to apply the new word from video to real objects in O'Doherty et al.'s (2011) study. Therefore, we expected that 30-month-olds would have a developing understanding of the symbolic relation between two-dimensional images and their real-world counterparts but would find our word learning task challenging in the absence of supportive scaffolding. As such, our scaffolding was not intended to teach children to begin thinking of video symbolically, but to support 30-month-olds in extending their already-developing representational insight to a new context.

As a comparison, we also recruited a group of 36-month-olds. At 36 months, children are still inconsistent in their transfer of video information to the real world (Moser et al., 2015; Zelazo et al., 1999), and benefit from adult scaffolding in video language learning studies (Roseberry et al., 2009; Strouse et al., 2013). However, they have also shown a pattern of using adult scaffolds to learn on their own in other contexts (Brown & Kane, 1988). That is, 36-month-olds successfully applied a previously scaffolded strategy to later solve a problem on their own. We expected that 30-month-olds would need the supportive scaffolds in place to do well on the third label learning trial, whereas 36-month-olds would continue to do well on the third trial in the partial scaffold condition after the supportive scaffolds were no longer offered, because they had learned to apply the appropriate strategy to solve the task.

2. Materials and methods

2.1. Participants

Sixty-three children 28.0–31.9 months (M = 29.96, SD =.89) and 61 children 34.0–37.9 months (M = 35.92, SD = 1.21) were included in the final data analyzed for this study. Participants were recruited from a database of parents who had volunteered to be contacted about research opportunities and through letters sent home at two local child care centers until we reached our pre-registered sample size of 66 per age group. To be eligible, parents needed to report that children had no known developmental delays and English accounted for at least 50 % of their language exposure. Because multilingual children were eligible for inclusion, we used the MacArthur-Bates Communicative Development Inventory (Fenson et al., 2000, 2007) to screen for children with low English proficiency. After testing, 8 children were excluded from analysis, three due to CDI scores below the 5th percentile, two due to caregiver interference, and three due to failure to make two consecutive correct selections during the pre-test.

2.2. Materials

2.2.1. Objects

Three familiar objects (a plastic turtle, fish, and duck; Fig. 1) were used in the pre-test. Three sets of three novel objects were used in the labeling videos and at test (Fig. 1). Each set of novel objects was piloted to ensure that children did not generally express a strong preference for a particular object in any set.

2.2.2. Videos

Nine stimulus videos were created, one featuring each novel object as the labeled target (https://osf.io/8fbkg/). The labeling procedure used in the stimulus videos was adapted from the procedure used by O'Doherty et al. (2011). Because each child watched three labeling videos, to reduce potential fatigue we shortened the procedure by reducing the number of objects and repetitions. We expected the increased difficulty of the shortened demonstration may be offset by our inclusion of supportive scaffolding.

2.2.2.1. Familiarization phase. Each video featured an adult seated behind a small, round child's table facing the camera. At the beginning of the video three novel objects were visible in a row on the table. The adult said, "Look at these!" then held up and rotated each object one at a time, while looking at the object she was holding. After holding up each object, she placed them into boxes on the floor next to her, out of view of the camera. This phase of the video lasted approximately 30 s.

2.2.2.2. Labeling phase. The adult then placed the three boxes with the objects onto the table. Moving from her right to her left, she made a statement, then opened the box, pulled the object out, held it up and rotated it while looking at it, placed it back in the box, and made a second statement. For non-target objects the statements were, "I'm going to show you this one. / I just showed you this one." For the target object, she stated, "I'm going to show you the fep/wug/tebu. / I just showed you the fep/wug/tebu." She then repeated the procedure a second time, resulting in children hearing the new label a total of four times. This phase of the video lasted approximately two minutes.

Fig. 1. Object sets used for pre-test (familiar objects) and test trials (wug, tebu, fep). Novel object sets were counterbalanced across the three test trials. We also counterbalanced which object within each set was labeled.

2.2.3. Questionnaires

Parents completed a demographic questionnaire and a short media questionnaire (Strouse & Ganea, 2017). Because children in this study did not fall entirely within one Level of the MacArthur Bates Communicative Development Inventory, we asked all parents to complete both the Level II B Short Form and Level III Form (Fenson et al., 2000, 2007). These forms provide a parent-report measure of children's receptive and expressive English vocabulary and were used to screen participating children for English language proficiency.

2.3. Research design

Children participated in one of three conditions: supportive scaffold, unsupportive scaffold, and partial scaffold. The first 29 children participated in the supportive scaffold condition and the remaining children were randomly assigned to condition. Pre-registration of the research questions and analysis plan was filed after approximately half of the data had been collected and none of the data had been looked at by the authors (https://osf.io/8fbkg/).

2.4. Procedure

Children participated either in a laboratory on campus (n = 117) or in a testing room at a child care center (n = 7). Procedures on campus and at the child care were identical except that parents were present on campus but not at the child care. When parents were not present, they completed written consent in advance and completed the questionnaires online. When parents were present, they completed written consent and questionnaires during the session. They were seated at the back of the room behind the child, who was seated at a child-sized table facing away from the parent. The researcher was seated next to the child at the table. A computer monitor was placed on a desk on the other side of the table so that it was directly in front of the child.

Children participated in a warm-up phase, a pre-test, and then three iterations of the video familiarization, video labeling, and test using different sets of novel objects (Fig. 1). The three conditions differed only in the type of scaffolds provided by the co-viewing researcher during the video familiarization phase. All children watched the same videos and were tested using the same procedures. The order of presentation of the object sets was counterbalanced, as was which object in the set served as the labelled target.

2.4.1. Warm up

Children were given a slide made of PVC pipe and all novel and familiar objects for 3–5 min. Children were encouraged to look at each object and put it in the slide. None of the objects were named during this time. At the end of the warm-up time the researcher collected all of the novel objects and placed them in a box out of view of the child.

2.4.2. Pre-test

The three familiar objects were placed in a blue storage bin with a lid. The researcher shook the bin, then opened the lid and asked the child to "Show me the turtle!" encouraging them to take it all the way out of the bin. If the child made a correct selection the researcher thanked them. If they made an incorrect selection, the researcher said, "That's not the turtle, let's try again!" The procedure was then repeated for the fish. If the child made an incorrect choice for either object, the researcher continued to repeat the procedure (asking for the duck, then the turtle, fish, or duck a second time, for a total of 6 possible trials) until the child successfully retrieved the requested object twice in a row. Three children who failed to successfully complete two successive pre-test trials were excluded from the analyses.

2.4.3. Video familiarization and scaffolds

Table 2 provides an outline of the in-person researchers' actions aligned with the actions that took place on video. During the video familiarization phase, the on-screen adult sequentially showed each of the three novel objects in the video to the camera. Neither the on-screen adult or the co-viewing researcher labeled any of the objects during this phase. Prior to playing the video, the researcher opened the video on the computer monitor so the first scene was visible to the child and placed the three novel objects in the video on the table in front of the child.

In the supportive scaffold condition, the supportive scaffolds were adapted from those that Strouse and Troseth (2014) had parents use to support 24-month-olds' word learning. The researcher exclaimed, "Look, these are the same as the ones on TV!" and started the video. As the adult on video held up the first object and silently rotated it, the researcher paused the video and asked the child to, "Find the one that is the same as the one she is holding on TV." After giving the child a chance to respond, the researcher, regardless of the child's response, picked up the correct object and held it up between the child and the TV screen and stated, "Look, this one is the same as that one!" She then placed the object back on the table in front of the child and unpaused the video. This process was repeated as the

Table 1
Condition difference in scaffolds children heard.

Condition	Trial 1 Scaffold	Trial 2 Scaffold	Trial 3 Scaffold
Supportive Scaffold	Supportive	Supportive	Supportive
Partial Scaffold	Supportive	Supportive	Unsupportive
Unsupportive Scaffold	Unsupportive	Unsupportive	Unsupportive

Table 2
Video actress and in-person researcher actions during video familiarization and labeling phases for each trial.

Procedure	Video actress	In-person researcher
Video familiarization	1. Three objects visible on table (on screen).	1. Video paused on screen and objects on table in room. 2. Said, “Look, these are the same as the ones on TV!” (supportive) or “Look, these are some of the toys you played with!” (unsupportive) and started the video.
	3. Said, “Look at these!”	
	4. Held up and rotated an object	4. Paused video as object held up. a Asked, “Find the one that is the same as the one she is holding on TV” (supportive) or ““Find the one that [description of object to the *right* of the one labeled]” (unsupportive) b Held object in front of screen and said, “Look, this one is the same as that one!” (supportive) or “Look, this is the one that is [description]!” (unsupportive) c Placed the object back on the table and unpaused the video.
	5. Placed object into box on the floor and moved on to next object, repeating step 4 for 2 additional objects.	5. Repeated step 4 for each object. 6. At end of video, removed real objects from table.
Video labeling		1. Said, “Now let’s listen to what she says!” 2. Played the video.
	2. Placed three boxes on table. 3. Attended to boxes one at a time a Said, “I’m going to show you this one.” (distracters) or “I’m going to show you the fep/wug/tebu.” (target) b Opened the box, held up and rotated object, then placed it back in the box. c Said, “I just showed you this one “(distracters) or “I just showed you the fep/wug/tebu.” (target) 4. Repeated attending to each of the three boxes a second time (total = 4 labels per object).	

Note. Videos are described in more detail in Section 2.2.2; scaffolds in Section 2.3. Video familiarization and labeling was followed by a test question for each trial. The duration of the recorded familiarization video was 30 s; the recorded labeling video was 2 min.

adult on video held up the other two novel objects. When the on-screen adult had placed all three objects into the boxes out of view of the camera, the co-viewing researcher paused the video and placed the three real objects into the blue storage bin and closed the lid.

In the unsupportive scaffold condition, the same procedure was followed, except the scaffolds were designed to be unsupportive for drawing a symbolic connection between the video and real objects. The initial scaffold was replaced with, "Look, these are some of the toys you played with!" The request for each object was replaced with "Find the one that [description of object that appeared on screen to the *right* of the one labeled]," where the description involved the shape and color of the object (e.g., is long and skinny with black and white stripes; is round and green and has spikes with a handle). When the researcher held up the described object, she said "Look, this is the one that is [description]!" This procedure was designed to match the researcher's actions and the number and timing of scaffolds across conditions, but to avoid highlighting the correspondence between the real and on-screen objects.

In the partial scaffold condition, the researcher followed the supportive scaffold script for the first two object sets and the unsupportive scaffold procedure for the last object set. Condition differences are displayed in Table 1.

2.4.4. Video labeling

Once the researcher had placed the three novel objects into the blue bin, she said "Now let's listen to what she says!" and unpaused the video. All children watched the video labeling phase without any scaffolds or pauses.

2.4.5. Test

The researcher shook the three objects the child had seen in the video inside the blue bin, opened the lid, tilted the bin toward the child, and said, "Now let's find the dax/wug/tebu! Show me the dax/wug/tebu!" Children were encouraged to "show" the researcher by pulling their selection out of the box. If they did not select an object, they were re-prompted. If they did not answer after the re-prompt, the objects were dumped onto the table and the child was given a final prompt to select an object. The researcher thanked the child for their selection.

2.5. Coding

2.5.1. Object selection

Children's object selections during the three tests were recorded by the researcher during the session. Children's object selections were also coded from video by a coder who was blind to the correct answer and the study condition (n = 359 coded; n = 13 were uncodable from video). There were 6 disagreements (kappa = .97), which were resolved by a third coder who reviewed the videos.

2.5.2. Object handling

The amount of time that children spent handling each object during the video familiarization phase was coded from video. With the sound on, a researcher identified the beginning and end timestamps for this phase. A different researcher (master coder) then coded these portions of each video with the sound off, so they could remain blind to condition and to which object was the target. The demonstration video was not visible in the recording of the child, so the coder was also blind to which object was being handled on screen. The master coder marked the timestamps of when the child started and stopped touching each of the objects. The duration of touching was summed for each object.

A reliability coder coded 23 percent of the sample (29 children x 3 videos x 3 objects = 261 total object handling times) using the same procedure. Inter-rater reliability was excellent (Koo & Li, 2016). That is, the intraclass correlation coefficient for a two-way single measures mixed effects model with a consistency definition was ICC = .977, 95 % CI [.971,.982]. Discrepancies were resolved in favor of the master coder.

3. Results

We first present omnibus models including both object selection and object handling as predictors of learning across all trials. We then focus specifically on testing patterns of learning on the third trial, once children in the partial scaffold condition experienced the change in scaffold, to test whether performance on the third trial differed based on prior scaffolding experience.

3.1. Condition differences in learning

To test for differences in learning, we used a binomial GEE with logit link, with test trial as a repeated measure and condition, test trial, condition x test trial, and handling proportion as fixed predictors. Because our research question was based on children's performance on the test of learning after the third video, when the scaffolding in the supportive and unsupportive conditions had not changed, whereas children in the partial scaffold condition had experienced prior support that was replaced with unsupportive statements, we also tested for condition differences on the third trial, regardless of the results of the omnibus model.

There was no significant effect of handling for either age group. For 30-month-olds, there was only a significant condition x test trial interaction, Wald χ^2 (df = 4) = 11.03, $p =$ .026. Bonferroni corrected follow-up tests indicated that there was no significant difference in learning across the three trials for children in the unsupportive or partial scaffold conditions. For the supportive scaffold condition, there was an effect of test trial number, Wald χ^2 (df = 2) = 6.88, p = .032. In this condition only, there was a significant increase in performance from trial 2 to trial 3 (p = .013 after Bonferroni correction). On the third trial, children selected the correct object more often in the supportive (71 %) than the unsupportive (33 %) scaffold condition, p = .009. Performance in the partial scaffold condition

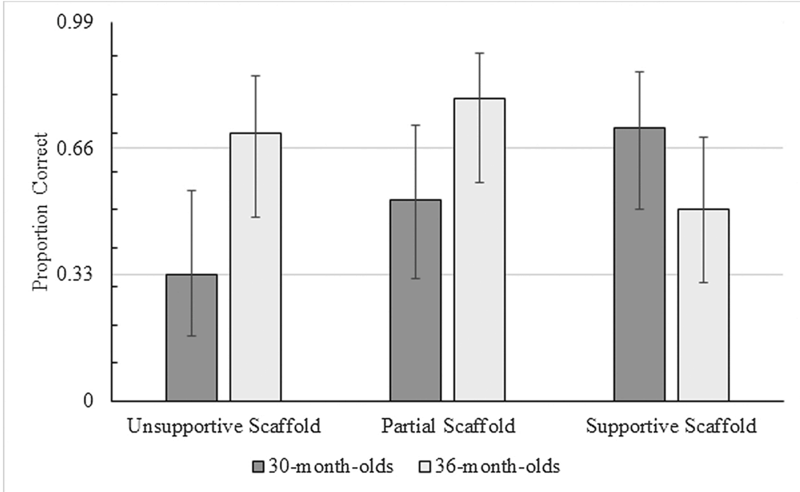

Fig. 2. Proportion of children who chose the correct target on the third learning test with Wilson confidence intervals. Chance performance = 0.33.

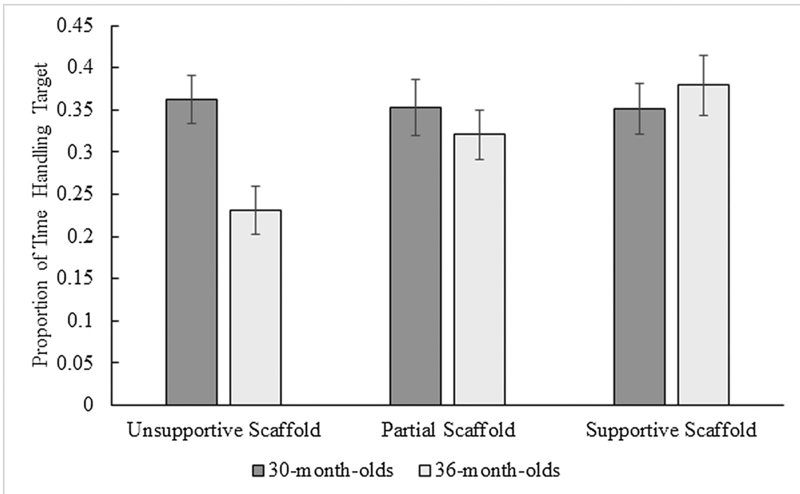

Fig. 3. Proportion of time children spent touching the target object during the video familiarization phase.

(52 %) was not significantly different from the supportive condition, $p = .398$, or the unsupportive condition, $p = .314$.

For 36-month-olds, there were no condition differences or interactions. There was only a significant main effect of trial number, Wald χ^2 (df = 2) = 9.46, $p = .009$. Follow-up comparisons with Bonferroni correction indicated a significant decline (across all conditions combined) from trial 1 to trial 2, $p = .02$ and no other significant differences. On the third trial, performance across conditions did not differ significantly (supportive 50 %, unsupportive 70 %, partially supportive 79 %).

3.2. Tests against chance

We then conducted tests against chance to determine whether children identified the target word at above chance levels after the third video (Fig. 2). We ran binomial tests against a chance value of 0.33 because three objects were available for children to choose between at test. At 30 months, only children in the full scaffold condition chose correctly more often than chance, $p < .001$. At 36 months, children in the unsupportive ($p = .001$) and partially supportive ($p < .001$) scaffold conditions chose above chance. Contrary to our prediction, children in the supportive scaffold condition did not chose significantly differently than chance ($p = .074$).

3.3. Object handling

To explore whether the supportive scaffolds we used may have simply drawn children's attention toward the target objects, rather than supporting a richer conceptual understanding, we also tested for condition differences in handling of the target objects.[1] We first ran a linear mixed model with test as a repeated measure; age group, condition, and test trial and their interactions as fixed predictors, and the proportion of the familiarization phase spent handling the target object as the dependent variable. We found only a significant age group by condition interaction, $F(2, 345.06) = 3.27$, $p = .039$. We followed-up by conducting separate mixed models for the two age groups. There were no condition differences in handling of the target objects for 30-month-olds, $F(2,165.28) = .03$, $p = .969$ (unsupportive: $M = .36$, $SD = .23$, partial: $M = .35$, $SD = .25$, supportive: $M = .35$, $SD = .24$). However, there was a significant condition difference for 36-month-olds, $F(2,165.69) = 5.53$, $p = .005$. Thirty-six-month-olds in the supportive scaffold condition handled the target object more ($M = .38$, $SD = .28$) than children in the unsupportive scaffold condition ($M = .23$, $SD = .22$), $p = .003$ after Bonferroni correction. Thirty-six-month-olds in the partial scaffold condition fell in between and did not significantly differ from the other conditions (Fig. 3).

4. Discussion

In this study, we compared 30- and 36-month-old children's performance on a label learning task when supportive scaffolds were provided across multiple trials, when supportive scaffolds had previously been present but were replaced by unsupportive scaffolds, and when unsupportive scaffolds were provided across trials. Thirty-month-old children performed above chance on the third trial only when supportive scaffolds were consistently offered across all trials. These children benefited from hearing a scaffold that drew their attention to the correspondence between the video objects and their real-world referents, like 24-month-olds in Strouse and Troseth's (2014) study. When children heard unsupportive scaffolds across all three trials, they did not choose the labeled real object at above chance rates, in line with the performance of 30-month-olds on a similar task in O'Doherty et al.'s (2011) study. Children in the partial scaffold condition, who heard the supportive scaffold for the first two trials and then the unsupportive scaffold on the third trial, also did not identify the labeled object at above chance rates on the third trial. At 30 months, two trials of scaffolding were apparently not enough to support children in learning to think symbolically about video such that they would later apply a new video label on their own. Rather, it is possible that 30-month-olds continued to depend on the presence of the supportive scaffold because it only supported them in associating the specific objects referenced in the scaffold. In fact, it appears that it took many of the 30-month-olds until the third trial to begin making the association to transfer the specific label, as the omnibus models showed no condition or trial differences except for an increase from the second to third trial for the scaffold group. In addition, 30-month-olds handled the target object for about the same amount of time in all three conditions, and their handling was not associated with learning. This suggests that at 30 months, children in the supportive scaffold condition were not simply drawn by the scaffold to pay more attention to the labeled object but were likely supported in conceptually drawing the association between the labeled video object and its specific real-world referent.

At 36 months, there were no condition differences in children's label learning on the third trial, nor across the set of three trials. At this age, many, but far from all, of the children were able to apply the video label to the real-world objects regardless of whether an adult drew their attention to the specific correspondence between the video images and their real-world referents. Although children at this age have benefited from adult scaffolding to learn language in prior studies (Roseberry et al., 2009; Strouse et al., 2013), the label-learning task we used in this study may have been more straightforward for children to solve on their own because it involved learning nouns (Childers & Tomasello, 2006) and the speaker appeared on screen and provided several cues to reference (gaze, movement; see Strouse, 2019).

Although we did not see condition differences at 36 months, only the performance of the unsupportive and partial scaffold groups was significantly above chance. However, with no condition differences on the third trial nor in the omnibus model across trials, further research is needed to determine if the lack of difference from chance in the supportive scaffold group is reliable. It is possibly a type II error. Alternatively, it may reflect a general trend that many 3-year-olds did not maintain engagement with the task. The only pattern we observed in the omnibus models was an overall drop in performance from the first to the second trial, across all groups.

However, it is also possible that performance in this group represents a type of Goldilocks effect, a pattern in which infants attend most to stimuli that are not too complex nor too simple (Kidd et al., 2012). Nussenbaum and Amso (2016) reported this type of pattern when 3- and 4-year-olds learned fewer words from video in a condition in which an on-screen actress provided the most scaffolding (high interactive condition) than in a condition where a more moderate amount of scaffolding was provided (medium interactive condition). Through eye-tracking, they observed that children in the high interactive condition maintained more attention to the labeler's face, whereas children in the medium interactive condition followed the speaker's social cues toward their referents. They argued that an optimal level of social interactivity is needed to engage children without distracting them from the target information. In our study, the scaffolds were provided by a person in the room prior to the labeling, rather than by the on-screen actress during labeling. Eye tracking may help future researchers determine how in-person scaffolding impacts children's visual attention to the on-screen speaker, and whether a moderate level of co-viewer scaffolding best supports children's attention.

[1] We also planned to code children's visual attention during the video familiarization and labeling phases, but we were unable to do so because children's eyes moved out of the camera frame too often.

4.1. Implications

At 30 months, children's learning was facilitated by continued supportive scaffolding. Children relied on the adult drawing the connection for each pair of objects and failed to learn the label when that scaffold was replaced by an unsupportive one, indicating that they had difficulty drawing a symbolic connection between the video image and its referent. This suggests it may be especially important for parents to explicitly connect video and real-world events to support children under 3 years of age in learning from video.

Learning a new word from video, holding it in mind, and later transferring that label to a real-world object is a difficult task for children ages 3 years and under (Strouse & Samson, 2021). Some of the challenges associated with this task were held constant across our three conditions, such as the perceptual details of the 2D images and the challenge of transferring from one context to another (2D to 3D). Labeling occurred when the object was not visible on screen or in the room, so children needed to hold the label in working memory to apply it to both referents. We also controlled for the social cues of the co-viewer by providing children in all conditions with the same number of scaffolds, delivered with the same timing and actions, and with similar length and content (both types of scaffolds were object-focused). The major difference between conditions was the content of the scaffold, which either emphasized the relation between the video and real objects or simply emphasized the objects. Therefore, the most likely explanation for 30-month-olds' above-chance learning in the third trial when the supportive scaffold was in place is that the scaffold helped children to conceptually understand the association between the specific scaffolded video image and paired objects. At 36 months, we did not see evidence that the supportive scaffold was effective in increasing children's success at our task. This may reflect that the major challenge for 36-month-olds was not aligning the video and real-world referents. More research is needed to identify what supports may help children in this age range succeed at this task. Because 36-month-olds' performance decreased after the first trial, it is possible that interventions that increase children's general engagement with the task would be more successful at increasing performance.

We also saw no clear evidence in either age group that adult scaffolds were effective at teaching children to shift their thinking about video in a way that would improve later learning. That is, we did not see any evidence of children increasing in symbolic sensitivity, or a mindset of looking for symbolic relations. However, the performance of the 30-month-olds in the partial scaffold group fell between that of the unsupportive and supportive scaffold groups, which may reflect that some individual children could have made this shift. Further research is needed to better understand this pattern of results.

4.2. Limitations and future directions

In order to match the adult co-viewer's behavior across conditions, we created unsupportive scaffolds matched in timing and number to the supportive scaffolds. These scaffolds involved joint attention with the child on real world objects as the co-viewer held up objects and spoke to the child about them. In other words, we chose to hold these social cues constant and vary the content of the scaffold (i.e. whether the adult explicitly drew a video-real world connection or simply described the real-world objects). It is possible that the presence of an engaged co-viewer in all conditions provided some support to children through social facilitation (Keene et al., 2019; Rasmussen et al., 2016), obscuring condition differences in the older age group. Future researchers may wish to create a condition where scaffolding and co-viewing are both entirely absent once support is withdrawn, along with a corresponding no-scaffolding/co-viewing control condition, to better understand the separate benefits of scaffolding and social facilitation.

It is also possible that instead of facilitating performance, our unsupportive scaffold was actively unsupportive, because it drew children's attention away from the screen and toward the real-world objects. Although the real objects were placed into a box out of sight before the video labeling phase occurred, and objects in both conditions were held up in front of the screen during the familiarization phase, the verbal scaffolds in the unsupportive scaffold condition only referenced the real objects. If this led children to focus more on the real objects during the video familiarization phase it could have hindered them in drawing screen-to-real-world connections. We do not, however, think this entirely explains the poor performance of the 30-month-olds in the unsupportive scaffold condition, as 30-month-olds in O'Doherty et al.'s (2011) study failed to learn words in a similar task in which no scaffolds were provided.

We also suggest that future researchers explore whether scaffolding provided over a longer period, and with more gradual withdrawal, could provide young children with the support needed to draw symbolic connections on their own. Novices may need more sustained scaffolding than more expert learners (Tawfik et al., 2018). Troseth (2003) found that 24-month-olds learned to think symbolically in later tasks after two weeks of experience with live video. Although our participants were older, 30- and 36-month-olds are still inconsistent in using symbolic thinking to solve tasks that require transfer from screens to the real world (Moser et al., 2015; Zelazo et al., 1999), so may have benefitted from more prolonged scaffolding. That is, although two trials were enough for 3-year-olds to adopt a mindset of looking for analogues in Brown and Kane's (1988) study, they may be insufficient for many 2- and 3-year-olds to learn to transfer words learned from video to real world referents. Optimal use of scaffolding should also take into account the child's experience and skill level and be faded or withdrawn based on the learner's performance (Sharma & Hannafin, 2007; van de Pol et al., 2010, Yelland & Masters, 2007). Our study did not provide that opportunity, as the withdrawal of the supportive scaffold was dictated by our procedural manipulation rather than individual children's performance. In addition, effective scaffolds in Brown and Kane (1988)'s study included not just telling children about the relation between problems, but also having children describe or teach a puppet the relation. Therefore, giving children more practice with scaffolds in place, and asking them to explain their strategies to confirm their understanding prior to withdrawing or reducing scaffolds, may provide better support for teaching young learners to learn from video on their own.

Finally, the type of scaffolding we provided in this study was based on the type of scaffolding that had been effective in prior studies using symbolic tasks (DeLoache, 1989; Kuhlmeier, 2005; Strouse & Troseth, 2014). It involved an adult explicitly stating the relation between the symbol and the referent. There is some evidence that this is not how adults typically scaffold young children's media experiences. Zack et al. (2009) asked mothers of 15- and 16-month-old infants to teach their children that a 2D object depicted on a touchscreen corresponded with a 3D version of the object. Parents scaffolded their infants' attention and demonstrated interacting with the 2D object, but only half of the mothers regularly commented on the 2D-3D correspondence. Zack et al. (2009) posited that the correspondence may have been so obvious to the mothers that they were unaware it maybe less obvious to their infants. Parents may focus their scaffolds more on the task itself (e.g., demonstrating how to interact with the new object or repeating a new label), than navigating the symbolic aspects of the medium.

5. Conclusions

Learning new words from video and applying them in the real world can be a challenging task for young learners. Multiple mechanisms likely contribute to this difficulty, one of which involves the conceptual challenge of thinking symbolically about how video images refer to real world objects. We investigated whether providing supportive scaffolding in which a co-viewer explicitly stated the connection between the video and real-world objects would teach children to later draw these connections on their own. Although the scaffolds supported 30-month-olds while they were in place, we did not see evidence that children learned to succeed at this task once the supportive scaffolding was no longer provided. This suggests that 30-month-olds may have learned to use the scaffolds to help them make specific video-object associations, rather than building a more general conceptual understanding of the symbolic relation between video and the real world. Thirty-six-month-old children did not appear to rely on the screen-to-real-world connections to be explicitly drawn in our task, as their learning did not differ with or without these scaffolds.

CRediT authorship contribution statement

Gabrielle A. Strouse: Conceptualization, Methodology, Formal analysis, Writing - original draft, Project administration, Funding acquisition. **Patricia A. Ganea:** Methodology, Writing - review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

This work was supported by an Insight Development Grant from the Social Sciences and Humanities Research Council, Government of Canada.

We would like to thank the families who participated in the research, to Salima Hackeek for data collection, Pavethra Yogeswaran for appearing in the stimulus videos, and others in the Language and Learning Lab for data coding and helpful feedback during design of the study.

References

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115–135). New York, NY: Oxford University Press.

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *The American Behavioral Scientist, 48*, 505–522. https://doi.org/10.1177/0002764204271506.

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*, 832–843. https://doi.org/10.1037/0012-1649.29.5.832.

Barr, R. (2010). Transfer of learning between 2D and 3D sources during infancy: Informing theory and practice. *Developmental Review, 30*, 128–154. https://doi.org/10.1016/j.dr.2010.03.001.

Barr, R., Zack, E., Garcia, A., & Muentener, P. (2008). Infants' attention and responsiveness to television increases with prior exposure and parental interaction. *Infancy, 13*(1), 30–56. https://doi.org/10.1080/15250000701779378.

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology, 20*, 493–523.

Carver, L. J., Meltzoff, A. N., & Dawson, G. (2006). Event-related potential (ERP) indices of infants' recognition of familiar and unfamiliar objects in two and three dimensions. *Developmental Science, 9*, 51–62. https://doi.org/10.1111/j.1467-7687.2005.00463.x.

Chen, Z. (2003). Worth one thousand words: Children's use of pictures in analogical problem solving. *Journal of Cognition and Development, 4*(4), 415–434. https://doi.org/10.1207/S15327647JCD0404_02.

Chen, Z., & Siegler, R. S. (2013). Young children's analogical problem solving: Gaining insights from video displays. *Journal of Experimental Child Psychology, 116*, 904–913. https://doi.org/10.1016/j.jecp.2013.08.009.

Childers, J. B., & Tomasello, M. (2006). Are nouns easier to learn than verbs? Three experimental studies. In K. A. Hirsh-Pasek, & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195170009.001.0001.

DeLoache, J. S. (1989). Young children's understanding of the correspondence between a scale model and a larger space. *Cognitive Development, 4*(2), 121–139. https://doi.org/10.1016/0885-2014(89)90012-9.

DeLoache, J. S. (1995). Early understanding and use of symbols: The model model. *Current Directions in Psychological Science, 4*(4), 109–113. https://doi.org/10.1111/1467-8721.ep10772408.

DeLoache, J. S., & Burns, N. M. (1994). Early understanding of the representational function of pictures. *Cognition, 52*, 83–110. https://doi.org/10.1016/0010-0277(94)90063-9.

DeLoache, J. S., & Marzolf, D. P. (1992). When a picture is not worth a thousand words: Young children's understanding of pictures and models. *Cognitive Development, 7*(3), 317–329. https://doi.org/10.1016/0885-2014(92)90019-N.

DeLoache, J. S., Pierroutsakos, S. L., & Troseth, G. L. (1996). The three 'r's of pictorial competence. *Annals of Child Development, 12*, 1–48.

Demers, L. B., Hanson, K. G., Kirkorian, H. L., Pempek, T. A., & Anderson, D. R. (2013). Infant gaze following during parent–infant coviewing of baby videos. *Child Development, 84*(2), 591–603. https://doi.org/10.1111/j.1467-8624.2012.01868.x.

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 Development, 19*(6), 613–627. https://doi.org/10.1002/icd.713.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the McArthur communicative development inventories. *Applied Psycholinguistics, 21*, 95–116. https://doi.org/10.1017/S0142716400001053.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, S., & Bates, E. (2007). *MacArthur-Bates communicative development inventories* (2nd ed). Baltimore, MD: Brookes Publishing. https://doi.org/10.1037/t11538-000.

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*(2), 176–196. https://doi.org/10.1111/j.1532-7078.2009.00013.x.

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*(1), 3–34. https://doi.org/10.1177/1529100615569721.

Keene, J. R., Rasmussen, E. E., Berke, C. K., Densley, R. L., Loof, T., Adams, R. B., Mumma, G. H., & Marshall, A. (2019). The effect of plot explicit, educational explicit, and implicit inference information and coviewing on children's internal and external cognitive processing. *Journal of Applied Communication Research, 47*(2), 153–174. https://doi.org/10.1080/00909882.2019.1581367.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS One, 7*(5), Article e36399. https://doi.org/10.1371/journal.pone.0036399.

Kirkorian, H. L. (2018). When and how do interactive digital media help children connect what they see on and off the screen? *Child Development Perspectives, 12*, 210–214. https://doi.org/10.1111/cdep.12290.

Kirkorian, H. L., Lavigne, H. J., Hanson, K. G., Troseth, G. L., Demers, L. B., & Anderson, D. R. (2016). Video deficit in toddlers' object retrieval: What eye movements reveal about online cognition. *Infancy, 21*, 37–64. https://doi.org/10.1111/infa.12102.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

Krcmar, M. (2010). Assessing the research on media, cognitive development, and infants. *Journal of Children and Media, 4*, 119–134. https://doi.org/10.1080/17482791003629586.

Kuhl, P. K. (2007). Is speech learning gated by the social brain? *Developmental Science, 10*, 110–120. https://doi.org/10.1111/j.1467-7687.2007.00572.x.

Kuhlmeier, V. (2005). Symbolic insight and inhibitory control: Two problems facing young children on symbolic retrieval tasks. *Journal of Cognition and Development, 6*(3), 365–380. https://doi.org/10.1207/s15327647jcd0603_3.

Marzolf, D. P., & DeLoache, J. S. (1994). Transfer in young children's understanding of spatial representations. *Child Development, 65*, 1–15. https://doi.org/10.2307/1131361.

Moser, A., Zimmermann, L., Dickerson, K., Grenell, A., Barr, R., & Gerhardstein, P. (2015). They can interact, but can they learn? Toddlers' transfer learning from touchscreens and television. *Journal of Experimental Child Psychology, 137*, 137–155. https://doi.org/10.1016/j.jecp.2015.04.002.

Myers, L. J., Keyser, H., & Cors, M. (2019). Co-viewers support participation in video chat interactions, but live experiences promote richer word learning for 24- to 36-month-olds in the USA. *Journal of Children and Media, 13*(4), 415–432. https://doi.org/10.1080/17482798.2019.1646294.

Nussenbaum, K., & Amso, D. (2016). An attentional Goldilocks effect: An optimal amount of social interactivity promotes word learning from video. *Journal of Cognition and Development, 17*(1), 30–40. https://doi.org/10.1080/15248372.2015.1034316.

O'Doherty, K., Troseth, G. L., Shimpi, P. M., Goldenberg, E., Akhtar, N., & Saylor, M. M. (2011). Third-party social interaction and word learning from video. *Child Development, 82*(3), 902–915. https://doi.org/10.1111/j.1467-8624.2011.01579.x.

Rasmussen, E. E., Keene, J. R., Berke, C. K., Densley, R. L., & Loof, T. (2016). Explaining parental coviewing: The role of social facilitation and arousal. *Communication Monographs, 84*(3), 365–384. https://doi.org/10.1080/03637751.2016.1259532.

Rideout, V. (2017). *The Common Sense census: Media use by kids age zero to eight*. Retrieved from. San Francisco, CA: Common Sense Media https://www.commonsensemedia.org/sites/default/files/uploads/research/csm_zerotoeight_fullreport_release_2.pdf.

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video? *Child Development, 80*(5), 1360–1375. https://doi.org/10.1111/j.1467-8624.2009.01338.x.

Schmitt, K. L., & Anderson, D. L. (2002). Television and reality: Toddlers' use of visual information from video to guide behavior. *Media Psychology, 4*(1), 51–76. https://doi.org/10.1207/S1532785XMEP0401_03.

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. *Interactive Learning Environments, 15*(1), 27–46. https://doi.org/10.1080/10494820600996972.

Simcock, G., & Dooley, M. (2007). Generalization of learning from picture books to novel test conditions by 18- and 24-month-old children. *Developmental Psychology, 43*(6), 1568–1578. https://doi.org/10.1037/0012-1649.43.6.1568.

Strouse, G. A. (2019). Using digital media to support language learning in early childhood. In J. Horst, & J. von Koss Torkildsen (Eds.), *International handbook of language acquisition* (pp. 485–503). Routledge.

Strouse, G. A., & Ganea, P. A. (2017). A print book preference: Caregivers report higher child enjoyment and more adult–Child interactions when reading print than electronic books. *International Journal of Child-computer Interaction, 12*, 8–15. https://doi.org/10.1016/j.ijcci.2017.02.001.

Strouse, G. A., & Samson, J. E. (2021). Learning from video: A meta-analysis of the video deficit in children ages 0 to 6 years. *Child Development, 92*(1), e20–e38. https://doi.org/10.1111/cdev.13429.

Strouse, G. A., & Troseth, G. L. (2014). Supporting toddlers' transfer of word learning from video. *Cognitive Development, 30*, 47–64. https://doi.org/10.1016/j.cogdev.2014.01.002.

Strouse, G. A., O'Doherty, K., & Troseth, G. L. (2013). Effective coviewing: Preschoolers' learning from video after a dialogic questioning intervention. *Developmental Psychology, 49*(12), 2368. https://doi.org/10.1037/a0032463.

Strouse, G. A., Troseth, G. L., O'Doherty, K. D., & Saylor, M. M. (2018). Co-viewing supports toddlers' word learning from contingent and noncontingent video. *Journal of Experimental Child Psychology, 166*, 310–326. https://doi.org/10.1016/j.jecp.2017.09.005.

Tawfik, A. A., Law, V., Ge, X., Xing, W., & Kim, K. (2018). The effect of sustained vs. faded scaffolding on students' argumentation in ill-structured problem solving. *Computers in Human Behavior, 87*, 436–449. https://doi.org/10.1016/j.chb.2018.01.035.

Troseth, G. L. (2003). TV guide: Two-year-old children learn to use video as a source of information. *Developmental Psychology, 39*(1), 140–150. https://doi.org/10.1037/0012-1649.39.1.140.

Troseth, G. L. (2010). Is it life or is it Memorex? Video as a representation of reality. *Developmental Review, 30*, 155–175. https://doi.org/10.1016/j.dr.2010.03.007.

Troseth, G. L., & DeLoache, J. S. (1998). The medium can obscure the message: Young children's understanding of video. *Child Development, 69*, 950–965. https://doi.org/10.1111/j.1467-8624.1998.tb06153.x.

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. *Educational Psychology Review, 22*, 271–296. https://doi.org/10.1007/s10648-010-9127-6.

Yelland, N., & Masters, J. (2007). Rethinking scaffolding in the information age. *Computers & Education, 48*(3), 362–382. https://doi.org/10.1016/j.compedu.2005.01.010.

Zack, E., & Barr, R. (2016). The role of interactional quality in learning from touch screens during infancy: Context matters. *Frontiers in Psychology, 7*, 1264. https://doi.org/10.3389/Ffpsyg.2016.01264.

Zack, E., Barr, R., Gerhardstein, P., Dickerson, K., & Meltzoff, A. N. (2009). Infant imitation from television using novel touch screen technology. *The British Journal of Developmental Psychology, 27*, 13–26. https://doi.org/10.1348/026151008X334700.

Zelazo, P. D., Sommerville, J. A., & Nichols, S. (1999). Age-related changes in children's use of external representation. *Developmental Psychology, 35*(4), 1059–1071. https://doi.org/10.1037/0012-1649.35.4.1059.