The Effect of Modality on Children's Higher-Order Concept Learning

Emily M. Neer (emilyneer@ucla.edu)
Raymond Shao (rshao@ucla.edu)
Catherine M. Sandhofer (sandhofer@psych.ucla.edu)
Department of Psychology, University of California, Los Angeles
502 Portola Plaza
Los Angeles, CA 90095 USA

Abstract

Podcasts are unique forms of unimodal modality because they include features like conversation, description, and sound effects to encourage audio engagement. Research shows that learners benefit from learning in two modalities (audio + visual) when information is complementary, not redundant. However, these previous studies used audio narration of text as auditory stimuli which differs from podcast formats. Do children learn from podcasts, and does providing supporting visual information affect learning? Children listened (or listened and viewed related images) to an 11-minute science podcast and answered recall and transfer questions. There was no effect of modality on children's learning, and children in both conditions performed above chance on transfer questions. Using a semantic textual similarity analysis, we show that children in the audiovisual condition do not incorporate visual information in their description of concepts. These results highlight the uniqueness of podcasts as a unimodal context that could benefit higher-order concept learning.

Keywords: modality, podcasts, learning, higher-order concepts

Introduction

Audio podcasts are digital media delivered in a single modality (i.e., the auditory modality). Audio podcasts have gained popularity in families with young children as an alternative to screen media (Kids Listen, 2021). Children's science podcasts are of particular interest because they attempt to engage children with higher-order, complex concepts (e.g., components of an atom, the evolutionary history of dinosaurs). Higher-order concepts involve making inferences, drawing connections, and building representations beyond the provided material (Resnick, 1987).

Children's podcasts include multiple features to encourage audio engagement. One feature of children's podcasts is that they are conversational. In an interview-style format, a host or a couple of hosts interview an expert and discuss a science topic, typically by the host asking questions and clarifying the expert's explanations. In another format, narrative-style science podcasts, characters act out a storyline that encompasses the science topic (e.g., traveling to Mars to learn about NASA's Martian research). Often, hosts will add elements of humor to these conversations and storylines. In addition, podcast hosts often speak directly to the child listener by asking questions to help children make connections to the topic being discussed, or they suggest related learning activities (i.e., experiments) that children can complete at home. Furthermore, children's podcasts use sound effects and music effects throughout the podcast episode to further engage listeners. Combined, these features attempt to attract and maintain children's attention to information presented in the podcast.

Little research exists on the degree to which children learn information presented in podcasts. Studies have found podcasts positively affect children's engagement with material (Grack Nelson et al., 2021), but have not examined whether children learn and remember information presented in podcasts created for children. Studies on adult learning from podcasts show mixed results. Some studies report positive learning outcomes when college students have access to podcasts as a resource with course material (e.g., Lonn & Teasley, 2009; Kennedy et al., 2016). On the other hand, other studies show that podcasts do not support learners' retention of information (e.g., Daniel & Woody, 2009). However, a key feature of the podcasts in these adult studies is that they are adaptations of text-based information (e.g., reading the textbook aloud). This differs substantively from the structures of podcasts produced for child audiences (e.g., conversational, interview style).

This misalignment of perceptual information (i.e., text-based information delivered in the auditory modality) used in adult podcast learning studies may contribute to the mixed findings of learning from podcasts. Would learning outcomes differ if learners were provided with podcasts in which perceptual information aligned with one another? In the current study, we compared children's higher-order concept learning from a unimodal context formatted like a child's podcast and a multimodal context (child's podcast + related static images) to examine the effect of modality on children's learning in this unique unimodal format.

Learning in Unimodal Contexts

There are multiple reasons to expect unimodal contexts to be effective learning contexts. One approach is to compare the cues, or information, available to learners when engaging with different forms of media and how these cues affect learners' attention and processing of information.

The cognitive theory of multimedia learning (Mayer, 1997) proposes that learners simultaneously use verbally- and visually- based models to select, organize, and integrate to-be-learned information. However, learning can be impaired in dual modal contexts when information like on-screen text and narration conveys redundant information (Mayer, 2005;

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Kalyuga & Sweller, 2014). This is known as the redundancy principle. Receiving redundant information in a dual modality context could have negative effects because redundancy splits a learner's attention between the two streams of information (Sweller et al., 1998). Split-attention introduces cognitive load as the learner is overwhelmed with information, leading to decreased learning performance (Sweller & Chandler, 1994). Redundant information like audio narration and on-screen text overwhelms the learner's auditory channel. So, if the narration of a concept can be understood on its own, then adding additional information, like on-screen text, would not provide additional relevant information, placing a burden on the learner. Information presented in a unimodal context would mitigate the issue of redundancy in learning materials.

However, redundancy is not a detriment to learning when the visual stimulus presented with an audio narration is an image (i.e., a picture). According to the Integrated Model of Text and Picture Comprehension (ITPC) and the redundancy principle, presenting auditory text narration with visual images is optimal for learning because the visual information and auditory information are processed in the visual and auditory channels, respectively, and do not overwhelm one channel (Mayer et al., 1996; Schnotz, 2005).

Learning in Dual Modal Contexts

Prior research shows that humans are more likely to learn information presented in more than one modality compared to a single modality (e.g., Bahrick et al., 2002; Mayer, 1997; Seger et al., 2019). According to the cognitive theory of multimedia learning, it is important to have both verbal and visual representations in order to make connections across modalities. However, how the verbal and visual representations are presented is important. Cues from the verbally- and visually-based models aid learning in situations where the cues are complementary and not redundant. Prior research with adults shows that learners demonstrate greater retention of information when short captions are presented with images and narration versus on-screen text that is identical to narration (Adesope & Nesbit, 2012; Yue et al., 2013; de Koning et al., 2017). Similarly, learners presented with short on-screen captions and narrations outperform learners who receive just narration on retention tests (Mayer et al., 1996; Mayer & Johnson, 2008). Furthermore, receiving information from audio narration and picture visuals would benefit learning because these cues are processed in two different channels. Therefore, learning in dual modalities can benefit learning when information is not redundant.

Children also demonstrate greater memory and transfer of information when text-based information is presented in an audiovisual format (with static or dynamic visual representations) compared to an auditory-only format (Knoop-van Campen et al., 2018; Seger et al., 2019). One reason we may expect a positive effect of audiovisual contexts on children's learning is that audiovisual cues provide redundant information (audio + visual) that could be helpful for children learning a concept for the first time. In other words, visuals could provide children with concrete representations of new, unfamiliar concepts. For example, elementary school-aged children learning a second language performed better on a retention test when provided with narration, visuals, and on-screen text, indicating that the redundancy principle may not extend to all learning contexts and ages (Jeu & Mohamad, 2014).

Relatedly, research with children has shown that correlated, redundant cues support children's learning of information in word learning and categorization domains (e.g., co-occurrence of features that matter for specific categories like using shape to learn the categories of solid objects; Sloutsky & Fisher, 2004; Yoshida & Smith, 2005; Sloutsky & Robinson, 2013; Luna & Sandhofer, 2021). Young children are sensitive to co-occurrence in their environments, and redundant cues can strengthen connections between concepts (Thiessen & Saffran, 2003; Yoshida & Smith, 2005; Luna & Sandhofer, 2021). Therefore, correlated cues provide children with the necessary support to learn a concept.

Current Study

Children's podcasts are a unique medium that include features for audio engagement with complex topics. In the current study, we asked if children learn from podcasts and if providing supporting visual information affects their learning. Based on the cognitive theory of multimedia learning, co-occurring information from the audio and visual modalities may be useful for children in learning a challenging science concept. However, due to the auditorily engaging nature of podcasts, it is possible that visual information in this context may not benefit children's learning because the visuals do not contribute additional, relevant information to the information conveyed through audio.

To examine how a unimodal context formatted like a podcast affects learning, we compared the learning outcomes of children who listened to a (unimodal) or listened to a podcast and viewed relevant images (dual modality). Children answered ten questions that assessed a) recall of information presented in the podcast and b) transfer to new contexts. Our design diverges from previous studies examining the effects of podcasted media on learning. Instead of taking a predominately visual modality context (i.e., text) and adapting it into an auditory only format, we took information designed for an auditory context (e.g., sound effects, conversation between individuals, vivid description) and added visual cues.

Method

Participants

Participants were 69 elementary school aged children (M_{age} = 8.1 years, range = 7.0 – 8.9 years, 34 females) recruited through social media platforms, email listservs, and a birth records database. Parents reported children's racial and ethnic identity with majority of children identified as White

(N = 39). The majority of parents (N = 65) held a college degree. Parents also reported children's podcast listening habits. Twenty-five of the participants reported that their child frequently (1.5 hours/week) listened to podcasts. Four participants were excluded due to parental interference during testing, technical difficulties, and missing data. This study was approved by the authors' institutional review board and families received a $5 gift card via email after the study concluded.

Design and Materials

This study was a 2x2 mixed-subjects design in which condition (i.e., audio-only and audiovisual) was between-subjects, and test question type was within-subjects (i.e., memory and transfer). The water cycle topic was selected as the science topic for this podcast to mitigate prior knowledge concerns as it is typically incorporated into science curriculums in 5th or 6th grade (National Research Council, 2012).

In both conditions, children listened to an 11-minute podcast created specifically for the study. The podcast included two female speakers discussing the water cycle and its various processes (i.e., precipitation, evaporation, condensation, and transpiration). The number of times the term of a process (i.e., precipitation, evaporation, condensation, and transpiration) was mentioned within the podcast was standardized, with each term mentioned eight times. Sound effects were also added to the podcast to maintain attention and model the typical podcast format produced for child audiences. Sound effects included intro and outro music and highlighted main processes (e.g., boiling water sound effect when discussing evaporation examples) and tangential information (e.g., cricket and frog sound effects when ponds are mentioned).

Audio-only and Audiovisual Media

In the audio-only condition, children viewed a blank white screen while listening to the podcast. In the audiovisual condition, children viewed a PowerPoint of relevant images synced with the audio (Figure 1). Each image appeared on the screen for 8 seconds. Two images were associated with each concept (e.g., evaporation). Each image depicted an example of a process mentioned within the podcast. However, details like arrows, dots representing water molecules, and animals were included in the images to provide children with rich visual representations of the processes. The images were loosely modeled after how the water cycle is depicted in textbooks (e.g., arrows showing the direction of water movement). These types of details were not explicitly mentioned in the podcast. When an image was not present on the screen, children viewed a blank white screen while listening to the audio.

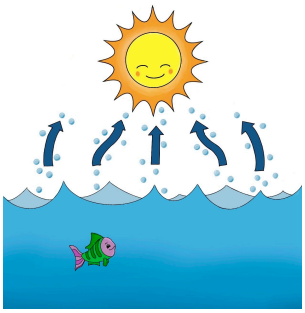

Figure 1: One of the eight images presented to children in the audiovisual condition. This image depicts the process of evaporation.

Engagement Questionnaire

Children's engagement with the media was measured using an adapted version of the Engagement in Science Learning Activities instrument (Chung et al., 2016). Children were asked eight questions that assessed their affective, behavioral, and cognitive engagement with the media (e.g., "During the podcast: I felt bored."). Children were shown a green bidirectional arrow with incremental magnitudes of yes and no along the arrow. The experimenter read a statement, and children were asked to verbally indicate how much they agreed or disagreed with the statement. Each item was based on a 4-point Likert scale with four items reverse coded. Children's responses were scored and averaged across eight items, resulting in an overall engagement score.

Test Questions

A 10-item test assessed children's recall and transfer of information presented in the study. Five recall questions assessed children's memory of information by asking children to recall information directly stated in the media (e.g., What is it called when water vapor gets cold, turns into a liquid, and then turns into a cloud?). Five transfer questions required children to transfer the information learned from the podcast to a novel scenario (e.g., "When it is hot outside, and we are drinking a cold drink from a glass, sometimes water droplets form on the outside of the glass. Is this an example of evaporation, condensation, or precipitation?). The order of test questions was randomized within the recall and transfer blocks. However, each child received all five recall questions first and then the five transfer questions. Responses to questions were scored using the rubric described below.

Procedure

Study sessions took place over Zoom and were recorded. Participants were randomly assigned to the audio-only or audiovisual condition. In the audio-only condition, parents were walked through how to minimize the Zoom participant and self-view windows to reduce distraction. The experimenter shared a blank, white PowerPoint slide through Zoom screen sharing for children to look at while listening to

the podcast. Parents and children were sent a Box link to the podcast using the Zoom chat function and asked to play the audio when ready. Once the podcast started playing, they were asked to switch back to the Zoom screen so that they viewed the blank white screen. Participants were asked to keep their video and audio on while the experimenter muted themselves and turned off their video camera.

In the audiovisual condition, children were sent a Box link to a video that included the audio synced with visual images using the Zoom chat function. Each image was on the screen for 10 seconds. Parents and children were asked to keep their attention on the web browser and not switch to Zoom when they pressed play on the video. In both conditions, the experimenter asked parents not to comment on the media material or help their children throughout the study.

After children finished listening to or listening to and watching the media, the experimenter assessed their engagement with the podcast using the Engagement in Science Learning Activities instrument (Chung et al., 2016). Children were told that the podcast creators were looking for feedback to make the podcast the best it could be for other children to enjoy; therefore, they should share their honest opinions.

To assess children's recall and transfer of information from the media, the experimenter read the ten questions and noted the child's answers on the test answer sheet. Children were not provided any feedback, but experimenters responded with neutral statements of encouragement (e.g., "You are doing great!"). After answering the ten test questions, we asked children, "In your own words, can you define [evaporation/precipitation]?"

Scoring and Reliability Coding of Test Questions

Responses to test questions were scored using a rubric designed by the first author. Each question was worth one point except for the first recall question, which had three correct responses, liquid, solid, and gas. Partial points were possible for the majority of questions.

The experimenter scored each child's responses after the study session. To assess scoring reliability, a second research assistant viewed each study session recording and scored children's responses to test questions. We obtained intra-class correlation coefficients (ICC) using the *irr* package (0.84.1; Gamer et al., 2022) in R version 4.3.1 (R Core Team, 2023). ICC was calculated based on a single measure (k =2), absolute agreement, 2-way mixed-effects model, ICC = .934, 95% CI [.923 – .947]. Any disagreements were resolved by a third coder and included in the final dataset.

Semantic Textual Similarity Analysis

We conducted a semantic textual similarity analysis comparing children's responses to the two open-ended evaporation and precipitation questions at the end of the study and descriptions of the evaporation and precipitation images presented to children (Figure 2). We were specifically interested in whether children in the audiovisual condition attended to and incorporated the information in the images into their understanding of evaporation and precipitation. If so, we would expect children's similarity scores in the audiovisual condition to be significantly higher than in the audio-only condition because children in the audio-only condition did not see the visuals. Therefore, the audio-only similarity scores served as a baseline measure. We chose to ask children open-ended questions about evaporation and precipitation processes because children in the piloting stage, regardless of condition, demonstrated a greater understanding of these processes compared to transpiration and condensation. This was confirmed after data collection was completed based on a test item analysis.

A trained research assistant transcribed children's open-ended responses. GPT-4 generated image descriptions based on the images and the prompt, "This is an illustration of [precipitation/evaporation] in a children's educational book. Can you write a description of this image using two sentences? Do not comment on whether the quality of the picture serves its purpose, focusing on describing the content."

To compare the similarity between children's open-ended responses and image descriptions, we used cosine similarity, a term-based similarity measure that calculates the cosine of the angle between two vectors (i.e., sentence embedding pairs; Gomaa & Fahmy, 2013). In our study, a high cosine similarity value means that a child's response to the question shared a high semantic similarity to the image description. For the analysis, we removed filler words (e.g., "uhm" and "hmm") along with stop words (e.g., "I" and "this") from children's responses. Words were then lemmatized (e.g., "goes" would become "go") and converted into embeddings with the Sentence Transformers library (version 2.2.2; Reimers & Gurevych, 2019) in Python (version 3.10.13). These embeddings were then compared using cosine similarity.

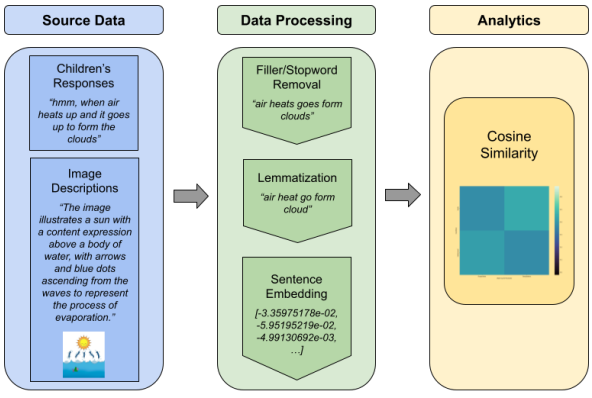

Figure 2: A schematic outlining the data processing steps in the semantic textual similarity analysis.

Results

Preliminary Results

To ensure that both conditions included children of similar ages, we examined if participants' ages differed between the audiovisual and audio-only conditions. We found no difference between the age of participants in the audio-only condition (M_{age} = 8.0 years, SD = 0.63 years) and the audiovisual condition (M_{age} = 8.2 years, SD = 0.56; $t(67)$ = -1.48, p = .144). Therefore, further analyses do not include the age of participants.

Furthermore, we conducted a linear regression analysis to assess if modality affected children's self-reported engagement. Out of the 67 children who reported engagement, we found no differences between the audio-only (M = 3.20, SD = 0.55) and audiovisual condition (M = 3.19, SD = 0.50) in children's overall engagement ratings, R^2 = -0.02, $F(1, 65)$ = .003, β = -0.007, p = .957, 95% CI[-0.26, 0.25]. In other words, modality did not influence children's engagement in the study, and thus engagement would not account for any modality differences in recall and transfer performance at test.

Modality and Learning

Our primary questions involved whether children's learning performance differed between the audio-only (n = 35) and audiovisual (n = 34) conditions. Because chance levels differed for the recall and transfer questions, we examined children's overall test performance, as well as their performance on recall and transfer questions separately. As seen in Figure 3, children's overall performance at test did not significantly differ between the audio-only and audiovisual conditions, $t(67)$ = 0.44, p = .66. We could not compare children's recall performance to chance on recall questions because chance on these questions was theoretically infinite. Recall questions provided children with a definition, and they had to generate the term in their responses. We compared children's performance on recall test questions only, and also found no significant difference in children's performance in the auditory-only condition (M = 0.44, SD = 0.25) compared to the audiovisual condition (M = 0.41, SD = 0.21) on recall test questions, $t(67)$= 0.66, p = .506.

However, because transfer questions provided children with answer choices, we could compare children's transfer performance to chance. Chance performance on transfer questions was 0.40. We conducted two, one-sample t-tests comparing children's transfer performance in the audio condition and children's transfer performance to chance. We found that children in the audio-only condition (M=0.65, SD = 0.17) performed significantly above chance on transfer questions (p < .05). Children in the audiovisual condition (M = 0.65, SD = 0.19) also performed significantly above chance on transfer questions (p < .05). Finally, we found no difference in children's performance in the auditory-only condition compared to the audiovisual condition on the transfer test questions $t(67)$ = -0.03, p = .97. In our task, children did not show any significant differences in their learning performance when presented science information in an auditory-only format like a podcast or an audiovisual format (a podcast with associated static images). However, children in both the audio-only and audiovisual conditions demonstrated evidence of learning, specifically the ability to transfer concepts presented in the podcast.

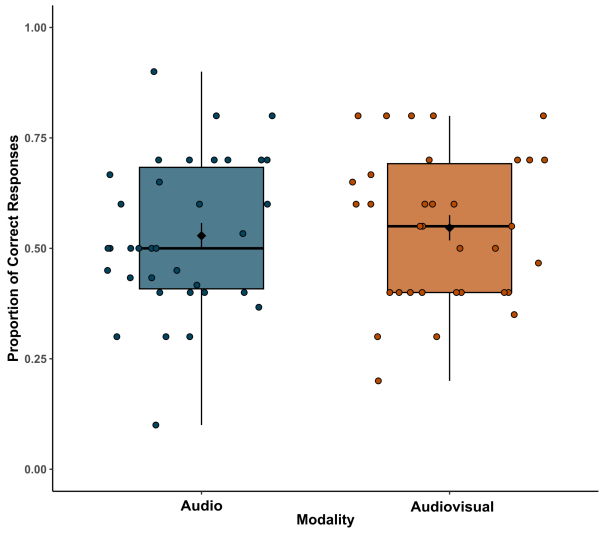

Figure 3: Children's proportion of correct responses at test by condition. Each dot represents one participant.

Analysis of Children's Open-Ended Responses

We calculated cosine similarity values for children's responses to the open-ended evaporation and precipitation questions and the evaporation and precipitation image descriptions generated by GPT-4. Each participant in the audio and audiovisual condition had four cosine similarity values – two values for the two evaporation image descriptions and two for the precipitation image descriptions.

First, we averaged the participants' two cosine values for evaporation to create an average cosine similarity value for that process. We followed the same procedure for precipitation cosine similarity values. Then, we calculated an average cosine similarity value for the audio-only and the audiovisual conditions for evaporation responses. We also calculated an average for the two conditions for precipitation. We conducted two t-tests (with Bonferroni corrections to account for multiple comparisons) to compare children's cosine values in the audio-only and audiovisual conditions for each process. We found no significant difference between the cosine similarity values of children in the audio-only condition (M = 0.38, SD = 0.10) and the audiovisual condition (M = 0.42, SD = 0.08) for evaporation, $t(51)$ = -1.63, p = .11. Similarly, we found no significant difference between the cosine similarity values of children in the audio-only condition (M = 0.45, SD = 0.13) and the audiovisual condition (M = 0.37, SD = 0.17) for precipitation, $t(38)$ = 1.53, p = .14. In other words, we did not find evidence that children in the audiovisual condition used the visual

information in their explanation of evaporation and precipitation processes.

Discussion

The aim of this study was to examine the effect of modality, audio-only versus audiovisual, on children's higher-order learning. Our study took a novel approach by using an audio-only stimulus designed for the auditory modality (e.g., a conversational interview format with sound effects) instead of narrating written text. We found no differences in learning between the audio-only and audiovisual conditions. However, we found evidence that children learned about the water cycle in both conditions as children in both conditions performed above chance on transfer questions.

Prior research on the effect of modality on learners' comprehension and retention of material is mixed. However, these studies use text narration as their audio stimulus. Studies with adults sometimes find that dual modal contexts are not beneficial for learning, especially when students read a passage and listen to its audio narration (e.g., Ari & Calandra, 2022). According to the redundancy principle in the cognitive theory of multimedia learning, this introduces cognitive load and impairs learning.

Conversely, some studies show an effect of modality on learners' performance (Mayer et al., 1996; Harskamp et al., 2007). Particularly with science information, secondary school students and college students retained and transferred information when they received narration of science lessons with illustrations (Harskamp et al., 2007) or when summaries of science lessons included narration with images (Mayer et al., 1996) providing evidence that dual modal contexts are beneficial for learning when visual (i.e., images) and audio information are complementary.

It is possible that visual information in our study was neither redundant nor complementary to the information presented auditorily. A semantic textual similarity analysis found no evidence that children were using the visual information in their explanations of evaporation and precipitation in the audiovisual condition. This could indicate that information presented in a podcast format with features that engage listeners auditorily is sufficient for children's learning. In other words, maybe they do not need the support from the visual modality to learn the information. However, we only queried open ended responses for the evaporation and precipitation concepts. Children demonstrated greater mastery of evaporation and precipitation compared to condensation and transpiration concepts. Perhaps visual information would have a greater effect on learning more challenging concepts.

Additionally, the visuals in our study were static images and were only on the screen for a brief period of time (8 seconds each). Our images may not have been salient enough to attract children's attention and affect learning. Future research should examine if changing the nature of visual information presented with podcasts, for example, using dynamic visualizations or introducing irrelevant information in the visuals, affects children's attention and learning. It is possible that different types of visual cues would disrupt learning.

According to the ITPC framework, visual information creates a mental model for learners, resulting in deeper processing (Schnotz, 2005). However, the podcast audio could create a mental model for learners through vivid, verbal description and sound effects. Future work should examine the effect of podcast audio on learning within the ITPC framework.

One cognitive process that we do not account for in the current study is selective attention. Selective attention, the ability to attend to relevant information and ignore irrelevant details, develops with age (Chong & Treisman, 2005). Prior research has found that young children attend holistically to multiple details rather than attending narrowly to the relevant information needed for learning (Deng & Sloutsky, 2016; Plebanek & Sloutsky, 2017). In contexts where multiple cues (i.e., narration and visuals) are vying for the learners' attention, it may be difficult for children to identify and attend to relevant information needed for successful learning. Our results may not capture individual differences in children's selective attention. Future research could include selective attention measures to examine how children's selective attention abilities affect learning in unimodal and dual-modal contexts.

Overall, the current study highlights the potential of podcasts to scaffold higher-order concept learning. Future research can build on this study to examine unimodal learning from an audio-only stimulus formatted like a podcast. Examining the effect of a podcast compared to a text-based audio recording on learning could extend our current understanding of learning from media and the cognitive processes involved in processing information in unimodal media contexts.

Conclusion

The current study found no effect of modality on children's learning of a science concept. However, changes in specifics of the study design may find effects. Children demonstrated learning of science information from both a single (i.e., podcast) and dual (i.e., podcast and images) modality context. Interestingly, in the dual modality context, children did not appear to use the visual information to learn the science concepts. This study calls for additional research using podcasts as an audio-only stimulus to understand modality effects on higher-order learning.

Acknowledgements

We thank the parents and children who participated in this study. Special thanks to research assistants, Urvi Shah, Stephanie Rivas-Lara, Juliana Kissinger, Zoya Egiazaryan, Kendall Schemmer, Katherine Ponce, Chalyn Valdez, Joyce Lin, Nicole Xu, Emma Chang, and Margaret Shen. We would also like to thank the Language and Cognitive Development Lab and Alice Xu for helpful comments in preparing this proceeding.

References

Adesope, O. O., & Nesbit, J. C. (2012). Verbal redundancy in multimedia learning environments: A meta-analysis. *Journal of Educational Psychology*, *104*(1), 250–263. https://doi.org/10.1037/a0026147

Ari, O., & Calandra, B. (2022). Dual modality effects on college students' comprehension of short texts during a simulated self-study session. *College Teaching*, *70*(4), 433–442. https://doi.org/10.1080/87567555.2021.1971604

Bahrick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental Psychobiology*, *41*(4), 352–363. https://doi.org/10.1002/dev.10049

Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13. https://doi.org/10.3758/BF03195009

Chung, J., Cannady, M. A., Schunn, C., Dorph, R., Bathgate, M. (2016). Measures technical brief: Engagement in science learning activities. Retrieved from: http://www.activationlab.org/wpcontent/uploads/2016/02/Engagement-Report-3.1-20160331.pdf

Daniel, D. B., & Woody, W. D. (2010). They hear, but do not listen: Retention for podcasted material in a classroom context. *Teaching of Psychology*, *37*(3), 199–203. https://doi.org/10.1080/00986283.2010.488542

De Koning, B. B., Van Hooijdonk, C. M. J., & Lagerwerf, L. (2017). Verbal redundancy in a procedural animation: On-screen labels improve retention but not behavioral performance. *Computers & Education*, *107*, 45–53. https://doi.org/10.1016/j.compedu.2016.12.013

Deng, W., & Sloutsky, V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, *91*(December), 24–62. https://doi.org/10.1016/j.cogpsych.2016.09.002

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement.* R package version 0.84.1, https://CRAN.R-project.org/package=irr

Gomaa, W. H., & Fahmy, A. A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications*, *68*(13), 13-18.

Grack Nelson, A., Her, C., Van Cleave, S., & Dominguez-Flores, J. (2021). *Children's STEM-focused podcasts as promising learning experiences: Research summary.* Science Museum of Minnesota: St Paul, MN.

Harskamp, E. G., Mayer, R. E., & Suhre, C. (2007). Does the modality principle for multimedia learning apply to science classrooms? *Learning and Instruction*, *17*(5), 465–477. https://doi.org/10.1016/j.learninstruc.2007.09.010

Jeu, N. P., & Mohamad, M. (2014). Effects of redundancy and modality principle in multimedia learning: Concentration on motivation and gender differences among primary school students. *Journal of ICT in Education, 1*, 42-53.

Kalyuga, S., & Sweller, J. (2014). The Redundancy Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 247–262). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.013

Kennedy, M. J., Hirsch, S. E., Dillon, S. E., Rabideaux, L., Alves, K. D., & Driver, M. K. (2016). Using content acquisition podcasts to increase student knowledge and to reduce perceived cognitive load. *Teaching of Psychology, 43*(2), 153-158. https://doi.org/10.1177/0098628316636295

Kids Listen. (2021). *Kids Listen. Understanding the kids and family podcast audience.* https://www.kidslisten.org/survey

Knoop-van Campen, C. A. N., Segers, E., & Verhoeven, L. (2019). Modality and redundancy effects, and their relation to executive functioning in children with dyslexia. *Research in Developmental Disabilities*, *90*, 41–50. https://doi.org/10.1016/j.ridd.2019.04.007

Lonn, S., & Teasley, S. D. (2009). Podcasting in higher education: What are the implications for teaching and learning? *The Internet and Higher Education*, *12*(2), 88–92. https://doi.org/10.1016/j.iheduc.2009.06.002

Luna, M. L., & Sandhofer, C. M. (2021). Arbitrary but predictive cues support attention to overlooked features. *Journal of Memory and Language*, *120*, 104251. https://doi.org/10.1016/j.jml.2021.104251

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*(1), 1–19. https://doi.org/10.1207/s15326985ep3201_1

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge Handbook of Multimedia Learning*, *41*, 31-48.

Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology*, *100*(2), 380–386. https://doi.org/10.1037/0022-0663.100.2.380

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, *88*(1), 64–73. https://doi.org/10.1037/0022-0663.88.1.64

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. The National Academies Press. https://doi.org/10.17226/13165.

Plebanek, D. J., & Sloutsky, V. M. (2017). Costs of selective attention: When children notice what adults miss. *Psychological Science*, *28*(6), 723-732. https://doi.org/10.1177/0956797617693005

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Processing of the 2019 Conference on Empirical Methods in Natural Language Processing.* Association for Computational Linguistics. https://arxiv.org/abs/1908.10084

Resnick, L. B. (1987). *Education and learning to think*. National Academy Press.

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), Cambridge handbook of multimedia learning (pp. 49 –70). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.005

Seger, B. T., Wannagat, W., & Nieding, G. (2019). How static and animated pictures contribute to multi-level mental representations of auditory text in seven-, nine-, and eleven-year-old children. *Journal of Cognition and Development*, *20*(4), 573–591. https://doi.org/10.1080/15248372.2019.1636799

Sloutsky, V. M., & Fisher, A. V. (2004). When development and learning decrease memory: Evidence against category-based induction in children. *Psychological Science, 15(8)*, 553-558.

Sloutsky, V. M., & Robinson, C. W. (2013). Redundancy matters: Flexible learning of multiple contingencies in infants. *Cognition*, *126*(2), 156–164. https://doi.org/10.1016/j.cognition.2012.09.016

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*(3), 185-233.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251-296.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Statistical and stress cues in infant word segmentation. *Developmental Psychology, 39*, 706–716.

Yoshida, H., & Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science, 16*(2), 90-95.

Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology*, *105*(2), 266–277. https://doi.org/10.1037/a0031971