Title

Visual engagement is not synonymous with learning in young children

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Shepherd, Sarah Stolp

Kidd, Celeste

Publication Date

2024

Peer reviewed

Visual engagement is not synonymous with learning in young children

Sarah Shepherd, and Celeste Kidd
Department of Psychology, University of California, Berkeley, USA

Abstract

Creators and consumers of popular media for kids tend to equate children's sustained attention with learning (Gahan, 2022; Segal, 2022). Here, we demonstrate that greater sustained visual attention does not necessarily translate to better learning—and in fact may predict learning deficits in some cases. We present the results of an empirical eye tracking study in which we demonstrate that attentionally captivating material can lead to worse learning with greater attentional capture, likely due to either distraction or overstimulation. Children who engaged most during a word-learning task learned the fewest word-object associations when they were presented on a colorful, moving background. These results support theories that suggest attentional capture due to perceptual attractors (e.g., things that are "bright, shiny") can disrupt learning. This work underscores the importance of the quality of screen-based media when considering the potential harms of children's screen time.

Keywords: attention; learning; eye tracking

Introduction

Equating sustained visual attention with learning is a popular trend among creators of children's media content. However, decades of research has demonstrated that learning value is not the sole driver of attention—or even the primary one (Haith, 1980). Attention may be guided by informational utility or highly salient perceptual features (Berlyne, 1954). Despite this literature, visual attention in popular media settings is often assumed to result from the material's learning value. For example, visual engagement is used in testing children's entertainment materials for suitability in promoting language development and literacy (Albahiri & Alhaj, 2020; Anggraini et al., 2022; Toleuzhan et al., 2022).

The trend of children's edutainment content creators conflating high visual engagement with high learning value is not new (Ryan, 2014); however, the impact and reach of the conflation has increased substantially due to recent changes in the quantity, availability, and financial incentives driving the creation of content for young children (Neumann & Herodotou, 2020).

Understanding the distinction between visual engagement and learning is pressing because more young people are watching this content more frequently for longer durations. This content is more readily available around the world via smartphones, and it's pushed harder by marketing tactics due to increased financial incentives in the new attention economy. Here, we apply classic methods from developmental psychology to test the relationship between visual engagement and learning in young children. We find no evidence of increased attentional engagement leading to greater learning in 3- to 6-year-old children, and in fact that higher rates of visual engagement for visually complex displays lead to *less* learning in participants. We draw insights and inspiration from relevant developmental psychology, connect it to recent changes in why and how content is created for children, and then present empirical evidence from a novel eye-tracking study we designed that suggests how the conflation of attentional engagement and learning in children is flawed.

The psychology behind children's attention

Attention is necessary for learning, and previous work demonstrates that it can reflect a readiness to learn (e.g. Maria Montessori's concept of "readiness" or Vygotsky's "zone of proximal development", (Montessori, 1917; Vygotsky, 1978)). Supporting empirical evidence demontrates children and non-human primates selectively attend to events of intermediate surprisal, which are theorized to hold greater informational utility (Cubit et al., 2021; Kidd et al., 2012, 2014; Oudeyer & Smith, 2016; S. Wu et al., 2022).

At the same time, attention does not *only* reflect learning value. Attention also selects perceptually salient visual material such as motion onset, high visual contrast, and hue saturation (Aslin, 2007; Cohen, 1972; Haith, 1980; Salapatek & Kessen, 1966).

While it's possible that attentional selection due to perceptual salience could lead to greater learning gains in young children—as edutainment producers commonly suggest—this has remained an untested claim. We investigate here.

Attention in economic context

The widespread adoption of the internet as a primary source of information, communication, and entertainment has monetized attention. In the modern economy attention is valuable. By its very definition attention is a finite resource. The intrinsic limits of our information processing abilities drive the attention economy: as the world becomes increasingly information rich, it must necessarily become attention poor (Simon et al., 1971). This dynamic creates competition for engagement in many information-rich but attention-scarce environments, and is evident on many popular streaming platforms.

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.*

The competition for attention is also driven by substantial financial incentives. In the US alone, digital video ad spending grew by 21% to 47.1B in 2022 (IAB, 2023). Not only is digital media advertising revenue growing rapidly, approximately half of digital video ad purchasers are quantifying viewers' visual engagement with biometric measures of attention, such as eye-tracking, demonstrating the value placed on consumer attention (IAB, 2023). The emphasis on most-watched content encourages a practice of creating content aimed at attaining high visual engagement. In line with this significant growth in online advertising revenue, many popular platforms employ recommending algorithms that prioritize content expected to maximize sustained visual attention (Lewandowsky et al., 2023; T. Wu, 2016). This includes many popular children's media platforms, like YouTube Kids, Roblox, and Twitch. The content itself, beyond the advertising within the platforms, also emphasizes attention as an important gauge of success. Creators of children's media frequently use engagement metrics to guide content production. For example, CoComelon Nursery Rhymes is the most popular children's YouTube channel having over 167 billion views as of August 2023, and is known to test how effectively their content visually engages children as young as two years old when placed near another screen showing naturalistic scenes (Segal, 2022). This practice is not necessarily unique to the streaming era; a review of children's DVDs noted the high concentration of low-level perceptual features being used in educational children's content nearly 15 years ago (Goodrich et al., 2009).

Methods

Participants

Fifty-five 3- to 6-year-old children completed the study across two locations. twenty-five participants volunteered at a local science museum near the UC Berkeley campus and thirty participated in the Kidd Lab at UC Berkeley. Four additional participants were excluded for parents not complying with experimenter instructions, and a further three additional participants were excluded due to technical issues. Testing location had no significant impact on results. All participants were compensated for their time either with $10 or a small prize.

Stimuli

Six novel objects (toys) were pseudo-randomly paired with novel words. The novel objects were small figurines of colorful monsters (see Figure 2). All toys were photographed in the same lighting conditions and backgrounds were removed. These photos of the toys were used for the screen-based portion of the task. Six familiar objects were selected from the MacArthur-Bates Communicative Development Inventory of earliest learned words (Dale & Fenson, 1996). The familiar objects were only used in the screen-based portion of the task, so simple, recognizable illustrations of these objects were selected.

Videos were used in the complex condition only. Videos were selected to be of colorful, age-appropriate naturalistic scenes with varied movement that did not include a clear singular subject. All audio was removed from videos.

Procedure

Participants were seated on their parents lap for the duration of the experiment. Exceptions were made for children who requested to sit on a booster seat or when parents were unable to hold the child on their lap. In these cases, parents sat behind their child or in the waiting room if requested by the child. To avoid parents influencing their child's behavior, all parents wore blindfolds or closed their eyes and listened to music through headphones while children participated in the study. In cases where parents closed their eyes experimenters confirmed that they complied throughout the study. Visual stimuli were presented on a monitor connected to a Gazepoint GP3 eye-tracker approximately 60cm away from the participant. The stimuli were presented from a Windows laptop using Psychopy software (Peirce et al., 2019). The eye-tracker was calibrated using a nine-point display of a shrinking dot that the participant fixated as it moved around the screen. An identical five-point validation sequence followed, allowing the experimenter to verify the calibration had been successful. If necessary, the experimenter repeated the calibration to achieve a satisfactory validation.

Learning Trials

The experiment consisted of both a screen-based eye tracking portion and a short test using the novel objects on the table in front of the participant. The screen-based portion consisted of both a learning and a test phase. The learning phase consisted of 12 trials during which the novel objects were presented with their corresponding novel word pairings. For each trial the novel object was presented with two carrier phrases "Do you see the X?" and "What a nice X!". Each novel object was presented twice in a random order during the learning trials.

Test Trials

After the learning trials the experimenter introduced the test trials as a game with toys that participants "might know the names of". The first set of test trials were the screen-based implicit test trails. Each implicit test trial presented two objects, a target and a distractor, on screen with an auditory prompt "Can you find the X?". The first 12 test trials asked participants to find a named novel object. Test trials 13-24 asked participants to find the known objects using the same format. These known-object test trials were included to check that participants were completing the task as expected. The known-object trials always occurred after the novel-object test trials. All test trials were pseudo-randomized with each object used twice as the target and twice as the distractor. Each trial displayed a unique pairing. Participant's test trial was deemed correct if their looking time to the correct object was greater during a two-second window following a 200ms delay from the time of the target word onset to account for any saccade latency. This window of analysis and criteria

(a) Simple condition (b) Complex condition

Figure 1: The learning trials in the simple condition presented the toys on a static gray background. The learning trials in the complex condition presented the toys in front of colorful videos of naturalistic scenes. Audio of child-directed speech labeled the toys with their novel names and was the same in both conditions.

Figure 2: We asked participants to identify names of novel toys. Participants learned these novel word-object associations via screens. We tested their learning via screen and with physical objects.

were the same for both novel-and known-object implicit test trials.

The final portion of the experiment consisted of explicit test trials. Children were asked to identify physical versions of the novel objects, providing an explicit measure of their learning. Physical stimuli were kept out of view from the participant until the screen-based eye-tracking portion of the study was completed. The experimenter then presented all six novel objects on the table and asked the participant to hand them each toy by name, one at a time. Between trials the experimenter replaced the selected toy and shuffled all six toys to prevent participants from selecting toys in serial order, and to make it more difficult to track which toys had already been selected. Experimenters told participants it was okay to guess if they were not sure of their answer and gave no feedback between physical trials. Participant responses were recorded on video for later coding.

Conditions

Participants were randomly assigned to either the simple or complex condition. In the complex condition, a unique video played in the background of each learning trial. Videos did not repeat such that participants should not associate a particular referent in a video with a novel word label. Each participant in the complex condition saw 12 videos in a different random order across all 12 learning trials. Stimuli in the simple condition were presented on a static gray background. Test trials and physical object trials were the same for both conditions.

Data processing

Eye tracking data was excluded from all trials with greater than 90% trackloss. After removing these trials with extreme trackloss, we averaged each participant's looking time across their validly tracked trials to calculate their average looking time for the entire screen and areas of interest. This allowed us to compare a participant's looking behavior in the learning trials to their performance on the test trials. We further excluded participants who were missing data from either all learning trials or all test trials. This resulted in excluding a further 4 trials. After processing the data in this manner, we were left with complete observations from 49 participants. 26 in the simple condition with a mean age of 4.955, and 23 in the complex condition with mean age 4.996.

Results

Increased visual engagement in complex condition

Participants looked to the screen during learning trials significantly longer in the complex condition (Figure 3). This result was expected as the complex condition was designed to increase visual engagement through the use of low-level attentional attractors. On average, participants looked to the learning trials 87.41% of the time in the complex condition, compared to 81.04% in the simple condition. This difference was confirmed to be significant by a Wilcoxon rank sum exact test (W = 463, $p < 0.001$)

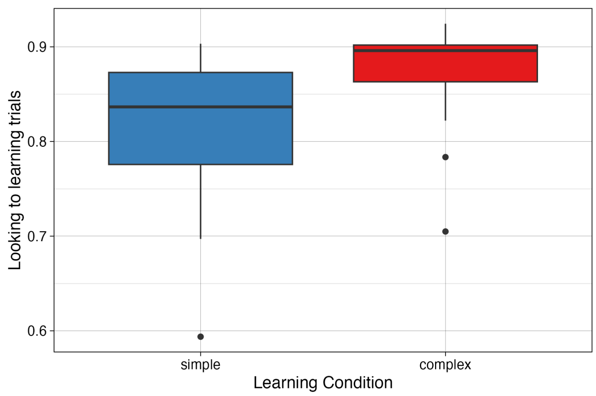
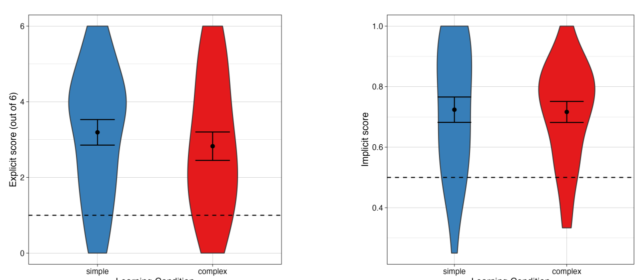


Figure 3: Participants attended to the screen more in the complex condition, where increased low-level perceptual features were present during learning trials.

Visual salience did not predict learning outcomes

As we expected, learning was not improved by increased visual engagement. Overall, participants demonstrated that they learned the names of the toys above chance in both the explicit and implicit test trials (see Figure 4). There was no significant difference in learning between conditions on either implicit or explicit word learning. On average, participants in the simple condition correctly identified the correct toy on 53% of explicit trials, compared to 47% in the complex condition. This difference was not significant in a Wilcoxon rank sum exact test (W = 260.5, $p = 0.44$). Implicit learning was nearly identical across conditions. Participants looked to the correctly named item in 72.39% of trials in the simple condition and 71.64% of implicit trials in the complex condition.



(a) Explicit scores

(b) Implicit scores

Figure 4: Participants learned the toy names similarly in both conditions, despite increased attention to the learning trials in the complex condition.

Increased visual engagement predicted worse learning when distractors were present

A linear regression controlling for age revealed that the interaction between a participant's looking time to the screen during learning trials and the condition they were in was a significant predictor of their performance on explicit learning trials ($\beta = 1.823$, $p = 0.0017$). For participants in the complex condition, increased looking to the screen during learning trials predicted worse outcomes on the explicit trials. In the simple condition looking time to the learning trials was associated with better explicit learning scores. The difference in this interaction between conditions is shown in Figure 5. This interaction was not observed in the implicit learning trials. The implicit learning trials were forced choice, and participants were largely successful in identifying the correct labeled toy.

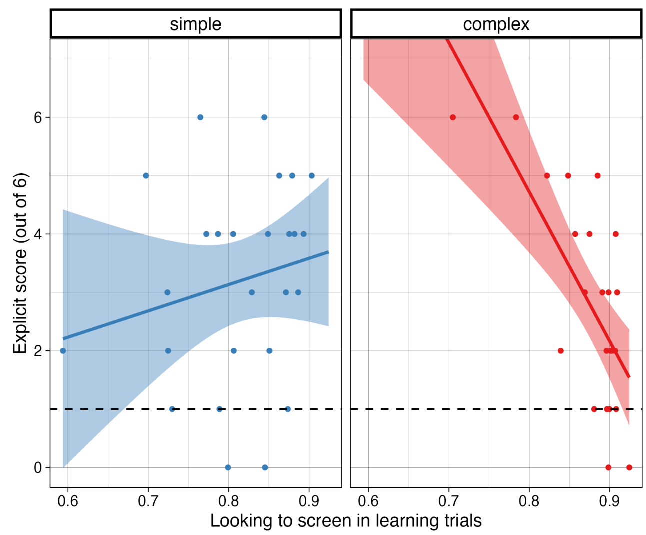


Figure 5: Increased visual attention during learning trials in the complex condition was associated with worse learning in the explicit trials. Participants who engaged the most learned the fewest toy names compared to less-attentive participants in the same condition.

Increased focus to target object did not explain learning trends

We found no evidence that participants who had better focus to the target object during the learning trials learned more word-object associations. Both overall looking time to the target and looking to the target as a proportion of all looking to the screen did not predict explicit or implicit learning in either condition.

Discussion

We found that increasing children's visual attention through heightened salient, low-level perceptual attentional attractors resulted in worse learning outcomes on a word learning task. In fact, these low-level cues elicited greater attention while hindering word learning for some children within the same condition. These highly engaged children may have learned something about the backgrounds that is not captured in our learning measures. However, their poor performance on a relatively simple and clearly defined learning task is still notable considering less-attentive children in the same condition performed better on the task. If anything, highly engaged children should have learned the names of the toys as well as something about the background videos. While we cannot speak to what they may have learned about the ambient background scenes, their increased attention did not benefit them on the word learning task and demonstrates the risk of highly engaging content that creates competition for learning, especially when this content aims to be educational.

The high-stakes competition to capture and sustain children's attention in online environments is driving the production of highly visually engaging content that may undermine well-intentioned educational messages. Using highly salient perceptual features to captivate young viewers not only interferes with their potential learning but may also increase the risk for adverse cognitive and behavioral outcomes. Our findings align with theories of overstimulation that link the adverse effects of screen time to the excessively salient audiovisual properties common in highly engaging children's content.

The potential harms of screen-time during early childhood are well documented, with known associations between excessive early screen use, developmental delays, and worse academic performance (Madigan et al., 2019). Specifically, children's screen use has been tied to disordered attention, impaired vision, sleep disruption, and poor mental health outcomes (Lissak, 2018). However, the research on children's screen usage and its potential harms is largely observational, making it difficult to determine the causal structure of this relationship despite the well-established associations. One theory suggests that overstimulating audio-visual content, like screen-based media, may be responsible for the negative cognitive and behavioral outcomes associated with children's screen time (Christakis et al., 2018). This theory began in part with observations of a relationship between fast-paced television content and disordered attention in young children (Anderson et al., 1977; Geist & Gibson, 2000). Further observational evidence has shown that fast-paced and fantastical children's television reduces young children's executive functioning for at least a short time after exposure (Essex et al., 2022; Lillard & Peterson, 2011). The theory that developing brains are susceptible to potential negative effects of overstimulation is also supported by animal models: mice exposed to heightened audio-visual stimulation early in life display both behavioral and cognitive deficits (Christakis et al., 2012; Christakis et al., 2018).

Our findings support this theory of overstimulation. Within the complex condition alone, maximized attention may be a sign of overstimulation, explaining the poor learning outcomes for children that barely looked away from the screen. These findings suggest that overstimulating children's media that maximizes visual engagement through low-level salient visual features can be detrimental to children's learning and development. It's important to consider that while our experimental design consisted of colorful, engaging videos, they were all of naturalistic scenes which we expect to be less stimulating than much of the content young children are engaging with on popular media platforms. This suggests that the potential negative impact of highly stimulating children's content may be even more severe in common online environments.

This study cannot directly speak to the level at which the presence of distractors would detract from learning. However, this experimental design provided participants with a better chance to learn simple associations than most children's content that is designed to be highly engaging. There was no interfering audio, and all background distractor videos were of ambient scenes that did not repeat with a corresponding toy to avoid confusion about which object was being labeled. Moreover, children in both conditions performed at the same level, further demonstrating that visual engagement, not the presence of distractors, was the moderating factor in the relationship between condition and explicit learning outcomes.

Recommendations for children's screen time remains general and often unclear to parents, in part due to the lack of evidence demonstrating a causal link between screen time and its potential detrimental effects. In the streaming era, where anyone can easily create and upload a video to YouTube or TikTok, placing the burden on caretakers to assess the quality of the media their children consume is overwhelming. Highly stimulating children's content poses a serious risk to the learning and development of young children that can be hard for adult caretakers to uncover because it encourages sustained visual engagement, which may appear as learning or enjoyment from the viewer.

Children's content, especially that which claims to be educational, needs to be created with the potential detrimental impacts of overstimulation in mind. We should require more careful and rigorous evaluation of children's media. In the absence of formal guidelines, caretakers and content creators should be cautious of media that may be highly engaging and perceptually salient.

References

Albahiri, M., & Alhaj, A. (2020). Role of visual element in spoken english discourse: Implications for YouTube technology in EFL classrooms.

Anderson, D. R., Levin, S. R., & Lorch, E. P. (1977). The effects of TV program pacing on the behavior of preschool children. *Educational Communication & Technology*, *25*(2), 159–166.

Anggraini, P., Apriliani, N., Supeni, I., & Handrianto, C. (2022). The use of the cocomelon YouTube channel as a medium for introducing children's english vocabulary. *SAGA Journal of English Language Teaching and Applied Linguistics*, *3*, 81–90.

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53.

Berlyne, D. E. (1954). A theory of human curiosity [Place: United Kingdom Publisher: British Psychological Society]. *British Journal of Psychology*, *45*, 180–191.

Christakis, D. A., Ramirez, J. S. B., & Ramirez, J. M. (2012). Overstimulation of newborn mice leads to behavioral differences and deficits in cognitive performance. *Scientific Reports*, *2*, 546.

Christakis, D. A., Ramirez, J. S. B., Ferguson, S. M., Ravinder, S., & Ramirez, J.-M. (2018). How early media exposure may affect cognitive function: A review of results from observations in humans and experiments in mice [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *115*(40), 9851–9858.

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences [Publisher: [Wiley, Society for Research in Child Development]]. *Child Development*, *43*(3), 869–879.

Cubit, L. S., Canale, R., Handsman, R., Kidd, C., & Bennetto, L. (2021). Visual attention preference for intermediate predictability in young children. *Child Development*, *92*(2), 691–703.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments & Computers*, *28*(1), 125–127.

Essex, C., Gliga, T., Singh, M., & Smith, T. J. (2022). Understanding the differential impact of children's TV on executive functions: A narrative-processing analysis. *Infant Behavior & Development*, *66*, 101661.

Gahan, M. B. (2022). Why children love the slow, strange world of 'CoComelon'. *Washington Post*.

Geist, E. A., & Gibson, M. (2000). The effect of network and public television programs on four and five year olds ability to attend to educational tasks [Place: US Publisher: Journal of Instructional Psychology]. *Journal of Instructional Psychology*, *27*(4), 250–261.

Goodrich, S. A., Pempek, T. A., & Calvert, S. L. (2009). Formal production features of infant and toddler DVDs. *Archives of Pediatrics & Adolescent Medicine*, *163*(12), 1151–1156.

Haith, M. M. (1980). *Rules that babies look by: The organization of newborn visual activity*. L. Erlbaum Associates.

IAB. (2023, May 3). *2022 video ad spend & 2023 outlook: Defining the next generation*. Interactive Advertising Bureau.

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 (A. Rodriguez-Fornells, Ed.). *PLoS ONE*, *7*(5), e36399.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804.

Lewandowsky, S., Robertson, R. E., & DiResta, R. (2023). Challenges in understanding human-algorithm entanglement during online information consumption. *Perspectives on Psychological Science*, 17456916231180809.

Lillard, A. S., & Peterson, J. (2011). The immediate impact of different types of television on young children's executive function. *Pediatrics*, *128*(4), 644–649.

Lissak, G. (2018). Adverse physiological and psychological effects of screen time on children and adolescents: Literature review and case study. *Environmental Research*, *164*, 149–157.

Madigan, S., Browne, D., Racine, N., Mori, C., & Tough, S. (2019). Association between screen time and children's performance on a developmental screening test. *JAMA Pediatrics*, *173*(3), 244–250.

Montessori, M. (1917). *The advanced montessori method*. Frederick A. Stokes Company.

Neumann, M. M., & Herodotou, C. (2020). Young children and YouTube: A global phenomenon [Publisher: Routledge _eprint: https://doi.org/10.1080/00094056.2020.1796459]. *Childhood Education*, *96*(4), 72–77.

Oudeyer, P.-Y., & Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, *8*(2), 492–502.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Ryan, E. (2014). Toward a redefinition of "educational" in babies' and toddlers' media:overt visual attention, verbalization, and other measures of engagement as indicators of learning. *Advances in Social Sciences Research Journal*, *1*, 105–114.

Salapatek, P., & Kessen, W. (1966). Visual scanning of triangles by the human newborn. *Journal of Experimental Child Psychology*, *3*(2), 155–167.

Segal, D. (2022). A kid's show juggernaut that leaves nothing to chance. *The New York Times*.

Simon, H. A., Deutsch, K. W., & Shubik, M. (1971). Designing organizations for an information-rich world. *Computers, communications, and the public interest*.

Toleuzhan, A., Sarzhanova, G., Romanenko, S., Uteubayeva, E., & Karbozova, G. (2022). The educational use of YouTube videos in communication fluency development in english: Digital learning and oral skills in secondary education. *International Journal of Education in Mathematics, Science and Technology*, *11*(1), 198–221.

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Wu, S., Blanchard, T., Meschke, E., Aslin, R. N., Hayden, B. Y., & Kidd, C. (2022). Macaques preferentially attend to intermediately surprising information. *Biology Letters*, *18*(7), 20220144.

Wu, T. (2016). *The attention merchants: The epic scramble to get inside our heads*. Knopf Doubleday Publishing Group.