Journal of Applied Research in Memory and Cognition

Manuscript version of

Attending Less and Forgetting More: Dynamics of Simultaneous, Massed, and Spaced Presentations in Science Concept Learning

Haley A. Vlach, Megan Kaul, Alexis Hosch, Emma Lazaroff

Funded by:

- National Science Foundation

AMERICAN PSYCHOLOGICAL ASSOCIATION SARMAC Journal of Applied Research in Memory and Cognition

ISSN: 2211-3681
https://doi.org/10.1016/j.jarmac.2021.10.007

EMPIRICAL ARTICLE

Attending Less and Forgetting More: Dynamics of Simultaneous, Massed, and Spaced Presentations in Science Concept Learning

Haley A. Vlach, Megan Kaul, Alexis Hosch, and Emma Lazaroff
University of Wisconsin

Research on children's categorization presents seemingly paradoxical results: Presenting exemplars at the same time (simultaneously) and presenting exemplars apart in time (spaced) have both been argued to support learning. This research was designed to explain these results by examining the visual attention and forgetting dynamics underlying various presentation schedules. Across three experiments, preschool-aged children ($N = 292$) were presented with science category exemplars on simultaneous, massed, and spaced schedules. The first experiment revealed that children had the strongest generalization performance in the spaced condition at the delayed post-test. In subsequent experiments, children visually attended less and forgot more during spaced learning. These results are discussed in the context of several theoretical accounts in cognitive science and applied implications for science education.

General Audience Summary
Generalization, the ability to transfer knowledge from one context to another context, is a critical cognitive process that drives learning and development. Educators are encouraged by federal guidelines, such as the Next Generation Science Standards, to present students with multiple examples of a concept so that learners can engage in the cognitive process of generalization. However, federal guidelines do not provide specific recommendations of when and how to present examples of concepts to students. The extant literature presents a seemingly paradoxical recommendation with regards to presentation timing: Examples should be presented at the same time, and examples should be presented apart in time. The research in this paper clarifies these seemingly conflicting results by directly comparing children's acquisition and generalization of science concepts on simultaneous, massed, and spaced learning schedules. We found that the timescale in which children generalize matters: At an immediate test, there was no difference between the schedules. At the delayed test, the spaced learning schedule promoted children's generalization to a greater degree than the other schedules. We also examined how the learning schedules affected children's visual attention and forgetting and found that the spaced actually led to less visual attention and more forgetting during learning. These results clarify theoretical accounts of generalization in cognitive science and have implications for science education. Science educators should use spaced practice and not get discouraged if they notice students visually attending less or forgetting during learning.

Keywords: forgetting-as-abstraction, spaced learning, comparison, categorization, conceptual development, science learning

The generalization of knowledge is a fundamental process in cognitive development and education. One domain in which categorization is particularly important is informal science learning and formal science education. Beginning early in life, children learn categories of animals, people, objects, and so on, generating a conceptual framework for understanding the natural world. When learning these categories, children must abstract the similarities and differences across examples to later generalize knowledge. For instance, when children learn the category of "dog," children must abstract across the examples of "dog" that they have

Many thanks the undergraduate research assistants of the UW-Madison Learning, Cognition, & Development Lab for their help with this project. Funding for this work was provided by NSF 1561531 awarded to Haley A. Vlach.

The authors declare that they have no conflict of interest.

Haley A. Vlach conceived of the theoretical background and research idea. All authors worked on creating the experimental design and stimuli. Megan Kaul, Alexis Hosch, and Emma Lazaroff collected and processed the data. All authors analyzed portions of the data. Haley A. Vlach prepared the manuscript, with the assistance of Megan Kaul, Alexis Hosch, and Emma Lazaroff

Correspondence concerning this article should be addressed to Haley A. Vlach, Department of Educational Psychology, 859 Educational Sciences, Madison, WI, United States. Email: hvlach@wisc.edu

Template Version: 22 February 2022 ■ 3:11 pm IST mac_j.jarmac.2021.10.007_format_proof ■ 25 February 2022 ■ 10:28 pm IST

experienced to notice the similarities (e.g., body shape) and differences (e.g., hair color, height, etc.). This knowledge will allow children to classify and generalize characteristics of "dogs" to new dogs they encounter. Indeed, the process of generalization, which includes encoding and storing information, abstracting similarities and differences across experiences, and using abstracted knowledge to categorize new exemplars, is critical to learning in and outside of the classroom.

Researchers have long sought to elucidate learning environments that promote children's ability to generalize knowledge. This body of work has shown that the timing at which exemplars are presented to children affects children's categorization and generalization. For example, one line of work has demonstrated that *viewing multiple exemplars of a category simultaneously* facilitates category acquisition and generalization (e.g., Ankowski, Vlach, & Sandhofer, 2013; Boroditsky, 2007; Gentner, Anggoro, & Klibanoff, 2011; Gick & Holyoak, 1983; Namy & Gentner, 2002; Oakes & Ribar, 2005; Thompson & Opfer, 2010; Trippas & Pachur, 2019; Vlach, Ankowski, & Sandhofer, 2012). In a typical novel noun generalization paradigm, children are presented with multiple novel objects that are labeled with the same novel linguistic label (e.g., "This is a fep!"). In these studies, the novel objects are presented on two schedules: simultaneous or massed. In simultaneous presentations, novel objects are presented at the same time so that children can visually inspect all of the exemplars during learning. That is, all exemplars are able to be viewed during the learning period. In massed presentations, novel objects are presented one at a time so that children can only visually inspect one exemplar. At an immediate test, learners are asked to generalize to a novel exemplar of the category (e.g., identify a new "fep"). This work has consistently shown that visually comparing multiple exemplars of the same category promotes generalization to a greater degree than viewing the same number of exemplars presented in immediate succession (i.e., massed presentations).

Why do simultaneous presentations promote categorization and generalization? Researchers have proposed that simultaneous presentations guide visual attention toward the commonalities among exemplars and minimize the short-term memory demands of mentally comparing multiple category exemplars. For example, according to the structure mapping theory (for a review, see Gentner, 1983; Namy & Gentner, 2002; Thompson & Opfer, 2010), simultaneous presentations attract attention to categorical similarities, inviting the mental process of comparison to occur, which promotes the abstraction of similar features between exemplars of a category. Because multiple exemplars are viewable during abstraction, learners do not need to recollect information from past learning events, thus reducing memory demands. The abstraction of similar features supports generalization to new category exemplars. In brief, presenting exemplars simultaneously has been proposed to promote children's generalization by guiding visual attention to relevant information and minimizing short-term memory demands of the task.

A separate line of research has come to a strikingly different conclusion. Many studies on spaced learning have demonstrated that *viewing multiple exemplars of a category across time* facilitates children's category acquisition and generalization (e.g., for a review, see Vlach, 2014; Vlach et al., 2008, 2012). In these studies, researchers have also used a novel noun generalization paradigm. Novel objects and words are presented to children on two schedules: a spaced schedule or a massed schedule. In both schedules, novel objects are presented one at a time so that children can only visually inspect one exemplar. In massed presentations, the novel object presentations are presented in immediate succession, just like in studies of simultaneous presentations described above. In spaced presentations, the novel object presentations are separated by intervals of time, such as 30 s (Vlach et al., 2008, 2012). After a retention interval, children are asked to generalize to a novel exemplar of the category. This work has consistently found that presenting children with category exemplars on a spaced schedule promotes generalization to a greater degree than presenting the same number of exemplars on a massed schedule.

Why do spaced presentations promote categorization and generalization? Researchers have proposed that spaced presentations support the long-term retention of categorical knowledge, supporting generalization across time. For instance, according to the forgetting-as-abstraction theory (for a review, see Vlach, 2014; Vlach et al., 2008, 2012), an extension of study-phase retrieval theories (Thios & D'Agostino, 1976), the time intervals between category exemplar presentations causes learners to forget information. The forgetting then leads learners to engage in more difficult retrieval of knowledge at each subsequent event. That is, they must engage in more cognitive effort in retrieving learned information than children in a massed condition, as these children can simply recollect information from their short-term memory. This effortful retrieval solidifies the memory trace and slows down the forgetting rate of common elements across category exemplars. We will examine these processes in the current research. Forgetting-as-abstraction theory extends study-phase retrieval theories by also proposing that irrelevant information is unlikely to be present in all category exemplars, it continues to be forgotten at a faster rate than relevant information, which likely gets retrieved from memory across most/all exemplars. In brief, presenting exemplars across time has been proposed to promote children's generalization by allowing them time to forget, thus engaging them in effortful retrieval and supporting their long-term memory for relevant categorical knowledge. Indeed, spacing effects in children's conceptual development are a desirable difficulty in learning (e.g., Bjork & Bjork, 2011, 2020; Roediger & Karpicke, 2006).

These two bodies of work present a seemingly paradoxical set of results: How is it that comparison, the presentation of exemplars at the same time, and spaced learning, the presentation of exemplars apart in time, both facilitate categorization and generalization? Two experiments (Vlach et al., 2012, Experiments 1 & 2) attempted to resolve these findings by directly comparing simultaneous, massed, and spaced presentations in a novel noun generalization task. Children were presented with four novel object exemplars with a corresponding novel word ("This is a fep!") at the same time (simultaneously), in immediate succession (massed), or distributed in time across 30 s intervals (spaced). Children's generalization was tested at an immediate test, consistent with comparison paradigms, and after delay, consistent with spaced learning paradigms. The results revealed that the testing timescale led to differing results. Children had stronger performance in the simultaneous condition at the immediate test. In contrast, children had stronger performance in the spaced condition at the delayed test. In a follow-up experiment, researchers tested children's ability to retrieve and generalize the labels during learning and found that children in the spaced condition experienced the most retrieval difficulty during learning. However, more retrieval difficulty was linked to stronger performance at the delayed post-test. Thus, researchers concluded that the timescale at which children are required to generalize is a determining factor in the efficacy of comparison and spaced learning, and forgetting and retrieval effort during learning was the mechanism that led to performance differences.

Because there have only been two experiments that have directly compared simultaneous and spaced presentations in children's learning, there are still many open questions. The current research builds on this work in two key ways to further resolve this paradox. First, we examined children's categorization and generalization in the context of science categories. By examining children's generalization of science knowledge, we can determine whether the results observed in previous research using novel objects (Vlach et al., 2012) generalize to a new set of stimuli. Moreover, this work bridges typical laboratory-based paradigms, such as the novel noun generalization task, with the types of materials educators use in early education settings, such as examples of science concepts. Indeed, both comparison and spaced learning have been argued to be efficacious for science education (Gluckman, Vlach, & Sandhofer, 2014; Jee et al., 2013; Vendetti, Matlen, Richland, & Bunge, 2015; Vlach & Sandhofer, 2012), and thus this work will help resolve the same paradox observed in studies of science education.

Second, we examined multiple mechanisms that could be contributing to the findings in this research. To date, only one mechanism has been proposed to explain why timescale may affect the efficacy of simultaneous and spaced presentations: forgetting and retrieval effort. According to the forgetting-as-abstraction theory (Vlach, 2014), simultaneous presentations prevent forgetting and thus lead to high performance at an immediate test when learners do not have an opportunity to forget between learning and test. In contrast, spaced presentations allow forgetting to happen during learning, which is beneficial at a delayed test (as described above). Indeed, these processes are foundational to study-phase retrieval theories (Thios & D'Agostino, 1976). Thus, we tested whether simultaneous and spaced presentations lead to differing levels of forgetting and retrieval difficulty during learning to test these theoretical accounts.

However, other viable mechanisms have yet to be tested. For instance, theoretical accounts of comparison and spaced learning have proposed that these presentation methods change patterns of visual attention. According to deficient processing theories of spaced learning, massed presentations cause learners to pay less attention because learners habituate to the stimulus (e.g., Hintzman, 1974; Rose, 1980). That is, learners pay more attention in spaced presentations because the time between presentations allows learners to dishabituate. Moreover, recent sub-accounts of deficient processing theories, such as the attention attenuation hypothesis (Kornell et al., 2010; Wahlheim et al., 2011), have suggested that the magnitude of attentional differences increase across learning events. That is, learners pay less and less attention to massed presentations with each subsequent learning event because they feel like they have already learned the concept. In sum, according to these theories, we expected to observe less visual attention to massed presentations relative to the other conditions and changes in the magnitude of attention differences across learning events.

We tested these possibilities in the current research. Across three experiments, preschool-aged children were presented with exemplars of science categories according to one of three schedules: simultaneous, massed, or spaced presentations. Children's generalization was tested immediately after learning and after a 5-min delay (Experiment 1). These timescales were chosen to mirror the timescales used in previous studies of comparison and spaced learning (e.g., Namy & Gentner, 2002; Oakes & Ribar, 2005; Vlach, Ankowski, & Sandhofer, 2012; Vlach, Sandhofer, & Kornell, 2008). In two follow-up experiments, children's forgetting (Experiment 2) and visual attention (Experiment 3) were measured during the learning phase. Taken together, these experiments afforded a direct test of simultaneous, massed and spaced presentations in children's science learning, and an examination of the underlying cognitive mechanisms engendered by these presentation methods.

Experiment 1

Method

We first examined how simultaneous and spaced presentations impact the categorization and generalization in the context of science categories. That is, in the first experiment, we asked: Do the results of previous research on simultaneous versus spaced presentation with novel objects generalize to a new type of stimuli? We chose to use science categories to connect this body of research with naturalistic learning contexts, such as science education.

Participants

The participants were 158 preschool-aged children (M_{age} = 50.77 months, SD = 10.27, Range: 27–75 months, 80 girls). To determine the sample size for this experiment, effect sizes were gathered from published studies on the spacing effect with this age group, which had consistently large effect sizes (e.g., Vlach et al., 2008, 2012). We used a smaller effect size in the large effect category, d = 0.9, to be conservative in determining a sample size. A power analysis for a two-tailed t-test, with α = .05, revealed that we would need at least 20 participants per condition to have 80% power to observe an effect. Thus, we decided to collect data until we reached at least 20 participants in every condition.

Children were recruited from local preschools in a mid-size Midwestern city and the surrounding metropolitan area. Parents provided written consent, and children provided developmentally appropriate assent to participate in the study. Basic demographic data was collected from parents, and children were predominately White (76.6% of children) and from middle- to upper-SES families (70.5% of children). Children received a storybook for their participation in the study. An additional 12 children participated in the experiment but were excluded from the final sample because of inability to follow instructions and/or complete the experiment. These 12 children did not differ in age from the final sample, $p > .10$.

Design

Children were randomly assigned to one of three between-subjects learning conditions (simultaneous, massed, or spaced) and one of two between-subjects testing conditions (immediate or 5 min delay). After random assignment, the immediate condition consisted of 77 participants (ns = 26 simultaneous, 26 massed, and 25 spaced; M_{age} = 54.94 months, SD = 8.47, Range: 34–75 months) and the 5-min delay testing condition consisted of 81 participants (ns = 24 simultaneous, 29 massed, and 28 spaced; M_{age} = 46.81 months, SD = 10.32, Range: 27–69 months). A series of t-tests revealed that there were no significant age differences among the six between-subjects conditions, *ps* > .10.

Apparatus and Stimuli

All children viewed the same 16 science concepts in 16 learning trials, presented in the same order, on an iPad. The science concepts used during the learning and testing phases included: camouflage, predator, herbivore, life cycle, molting, hibernation, cirrus clouds, buoyancy, erosion, fossil, molecules, precipitation, reptiles, pollution, reflection, and tundra. These stimuli were tested with a separate set of children to ensure that preschool-aged children were largely unfamiliar with these concepts and that there were no items effects. During each learning and testing phase, children were presented with an image that served as a distractor object (listed in corresponding order to science concepts): art supplies, racecar, clothing rack, dog, apple, house, airplane, soccer field, potatoes, crayons, cookie, cove, dog, rock concert, flower, goat. During each testing phase, children were presented with a new image that shared some perceptual features with target pictures from the learning phase but did not depict the target concept. These images depicted (listed in corresponding order to science concepts): rainbow snake, dog and cat cuddling, hamburger with lettuce, four unrelated objects with arrows between them, man shaving, elephant dancing with eyes closed, cumulus clouds, water waves, flooded lawn, rock formation, gumball machine, spilled salt, fish, marine fog, puddle jumping, beach. During each testing phase, children were presented with a new image that was completely unrelated to the target concept. These images depicted (listed in corresponding order to science concepts): dog panting, beaver dam, mountain, cupcake, dog laying, laughing monkeys, ocean, girl holding umbrella, cactus, car, cow, tennis shoes, playground, bird, ocean wave, cathedral.

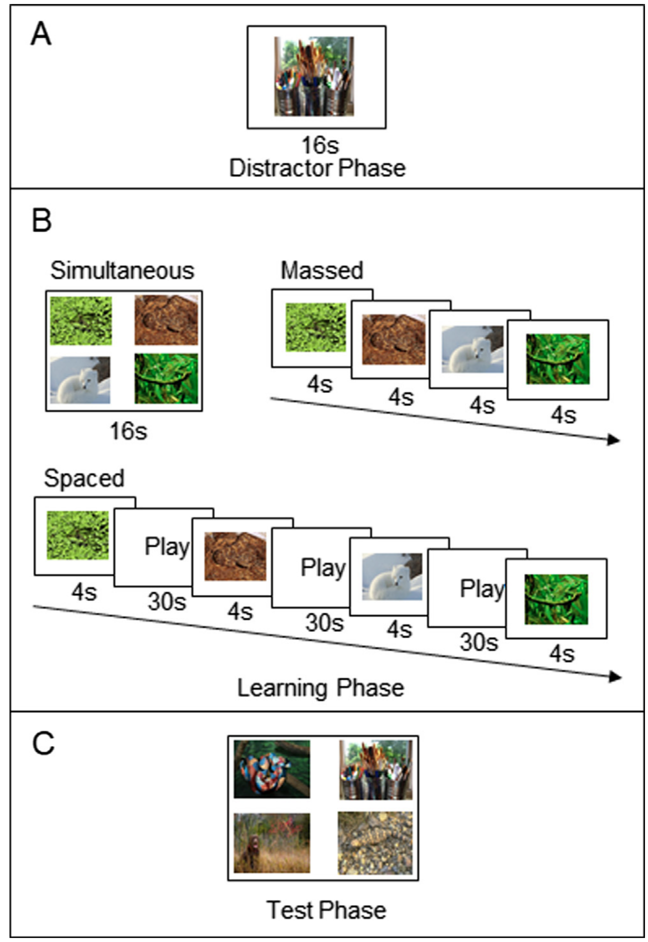
Procedure

Children completed the experiment in a quiet area of their school or at a university lab. At the beginning of the experiment, the experimenter told children they would be learning new information, but they were not told that they would be tested. During the experiment, children participated in 16 trials (Figure 1). Each trial was presented in three phases: a distractor phase, a learning phase (simultaneous, massed, and spaced), and a testing phase.

Distractor Phase. The distractor phase was the first phase of each trial. The purpose of the distractor phase was to introduce an item that would later serve as a familiarity control in the testing phase. The iPad presented the distractor image for 16 s, and the experiment prompted children to look at the image without labeling the science concept (Figure 1, Panel A; e.g., "Look at this!"). Sixteen seconds was chosen to equate learning time to the exemplars (4 exemplars × 4 s = 16 s). The 16 distractor images were unrelated to the target category.

The learning phase occurred immediately after the distractor phase. Children were presented with four exemplars of each science concept. The only difference between the three learning conditions was the timing in which the science concept exemplars were presented to children. In the simultaneous condition, children viewed all four science concept exemplars simultaneously on the same screen (Figure 1, Panel B). In the massed condition, children viewed four exemplars individually, presented in immediate succession. In the spaced condition, children viewed four exemplars individually, with 30-s intervals between each exemplar. Children participated in an unrelated activity (e.g., coloring, stickers, puzzles, etc.) during the spacing intervals.

Figure 1
Experimental Procedure of Each of the 16 Trials



Note. (A) Distractor phase. A picture was presented without a label (e.g., "this" or "it"). (B) Learning phase. Four pictures of a science concept were presented and given a label (e.g., ''camouflage'') in simultaneous, massed, or spaced presentations. (C) Testing phase. Four pictures were presented and children were asked to identify the target (e.g., ''Can you point to camouflage?''). For children in the immediate condition, testing occurred directly after the learning phase. For children in the delayed testing condition, testing occurred 5 min after the learning phase. See the online article for the color version of this figure.

In all conditions, each image of the exemplar was allotted 4 s of viewing time. In the simultaneous condition, all images were simultaneously presented for 16 s. In the massed and spaced presentations, each of the four images was presented for 4 s (for a total of 16 s). Each picture was labeled once (e.g., "This is camouflage!"). In the simultaneous condition, the experimenter would provide the label for each image on 4 s intervals (at 0 s for the first exemplar, at 4 s for the second exemplar, etc.) to mirror the other two presentation conditions. Thus, total learning time and the number of times the pictures were labeled were equated across conditions.

Testing Phase. The testing phase consisted of one forced-choice generalization test (Figure 1, Panel C). Children were presented with four pictures in random placement order and were asked to pick out the target science exemplar ("Can you point to camouflage?"). One of the four pictures, the science concept exemplar (e.g., "camouflage"), was a new exemplar of the concept. A second picture was the distractor item that had been presented during the distractor phase. A third picture was an item that shared some perceptual features with previous target pictures (e.g., a similar animal, but one not using camouflage) but did not depict the target concept. The fourth object was a new picture of an unrelated object that had not been presented during the experiment (e.g., a cactus). Children were not given feedback after making their selection.

In the immediate condition, testing immediately followed the distractor and learning phases. In the 5-min delay condition, learning and distractor phases were interleaved. For example, after the distractor and learning phases for the first trial were complete, the distractor and learning phases for the second trial immediately followed, and so on until children had completed all learning and distractor phases. Testing for each trial began exactly 5 min following the end of the last learning phase. A 5-min delay was chosen because (a) it was along enough delay to require children to access information from long-term memory, and (b) it was short enough to allow children to pay attention for the entire experiment.

Results

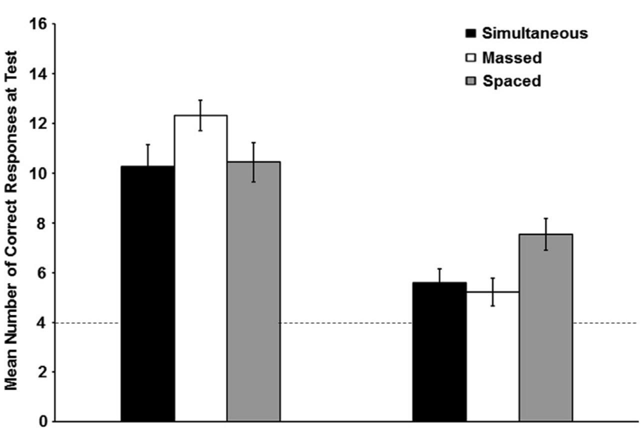
We were interested in whether there would be differences in children's generalization in relation to the presentation timing of learning and the timing of testing. Thus, we first plotted children's performance by presentation timing and testing delay, which can be seen in Figure 2. One-sample *t*-tests revealed that performance in each of the six between-subjects conditions was significantly higher or marginally higher than chance, *p*s < .10. We then computed a 3 (simultaneous, massed, or spaced) × 2 (immediate or 5-min delay) between-subjects ANOVA with the total number of correct responses at test as the outcome measure. This test revealed a main effect of testing delay, $F(1, 152) = 76.66$, $p < .001$, $\eta_p^2 = .335$, and an interaction between presentation timing and testing delay, $F(2, 152) = 8.86$, $p = .009$, $\eta_p^2 = .060$. We computed the same ANOVA with children's age entered as a co-variate, and we observed a main effect of children's age, $F(1, 151) = 79.74$, $p < .001$, $\eta_p^2 = .346$. Older children had higher overall performance than younger children. However, children's age did not affect the overall pattern of results observed in the first ANOVA; there was a main effect of testing delay, $F(1, 151) = 39.41$, $p < .001$, $\eta_p^2 = .207$, and an interaction between presentation timing and testing delay, $F(2, 151) = 3.17$, $p = .045$, $\eta_p^2 = .040$.

Because there was an interaction between presentation timing and testing delay, we conducted planned post-hoc analyses within each testing delay. Post-hoc analyses were conducted using one way ANOVAs, with the total number of correct responses at test as the outcome measure, followed-up with *t*-tests with Tukey's HSD correction. At the immediate testing condition, there was no main effect of presentation timing and there were no significant differences between the conditions, *p*s > .10. However, at the 5-min delayed test, there was a main effect of presentation timing, $F(2, 78) = 4.50$, $p = .014$, $\eta_p^2 = .103$. *T*-tests revealed that performance in the spaced condition ($M = 7.54$, $SD = 3.38$) was significantly higher than in the massed condition ($M = 5.21$, $SD = 3.05$), $t(55) = 2.73$, $p = .008$, $d = 0.724$, and marginally higher than the simultaneous condition ($M = 5.58$, $SD = 2.84$), $t(50) = 2.23$, $p = .030$, $d = 0.623$. There was no significant difference between the massed and simultaneous conditions.

The pattern of generalization observed in the 5-min delay condition replicates the one study to date that has directly compared simultaneous, massed, and spaced presentations (Vlach et al., 2012). Thus, this works extends previous research by demonstrating that, even when examining different learning materials or categories, timing protocols, and developmental periods, we still observed the strongest performance on a spaced schedule at a delayed test. However, previous research would suggest performance in the simultaneous condition should have been higher than the massed and spaced conditions at the immediate test. In Experiments 2 and 3, we examined attention and memory processes underlying the three presentation conditions to seek a potential explanation for this finding, and the spacing effect commonly observed across prior research (e.g., Vlach et al., 2012).

Figure 2

Mean Number of Correct Responses at Test, by Presentation Timing (Simultaneous, Massed, or Spaced) and Testing Delay (Immediate or 5 Minute Delay), in Experiment 1



Note. Error bars represent one standard error. Dashed line represents chance performance; performance in all conditions was significantly above chance.

Experiment 2

Method

In previous research on spaced learning in children's categorization and generalization, the mechanism that has been argued to cause the spacing effect is forgetting and the subsequent retrieval difficulty from that forgetting (Vlach et al., 2012; Vlach, 2014, 2019). We conducted Experiment 2 to determine whether we observe these forgetting and retrieval dynamics in the current experiment. We presented children with the same science categories on simultaneous, massed, and spaced schedules and asked children to retrieve and generalize during learning. We hypothesized that children would experience the most retrieval difficulty in the spaced condition and the least retrieval difficulty in the simultaneous condition.

Participants

The participants were 70 preschool-aged children (M_{age} = 49.37 months, *SD* = 9.02, Range: 36–71 months, 35 girls). These children had not participated in Experiment 1. We used the same sample size and recruitment plan procedure as Experiment 1: We collected data until we reached at least 20 participants in every condition. Children were recruited from local preschools in a mid-size Midwestern city and the surrounding metropolitan area. Parents provided written consent, and children provided developmentally appropriate assent to participate in the study. Basic demographic data was collected from parents, and children were predominately White (75.7% of children) and from middle- to upper-SES families (81.4% of children). Children received a storybook for their participation in the study. An additional 10 children participated in the study but were not included in the final sample due to inadequate testing conditions (e.g., loud classroom) and/or inability to follow directions (e.g., not responding to questions).

Design

Children were randomly assigned to one of three between-subjects learning conditions (simultaneous, massed, or spaced). Random assignment resulted in the following number of participants per condition: *n*s = 26 simultaneous (M_{age} = 47.73 months, *SD* = 7.75, Range: 36–61 months), 22 massed (M_{age} = 51.50 months, *SD* = 9.31, Range: 38–71 months), and 22 spaced (M_{age} = 49.18 months, *SD* = 10.05, Range: 38–70 months). A series of *t*-tests revealed that there were no significant age differences among the three between-subjects conditions, $ps > .10$.

Apparatus and Stimuli

Same as Experiment 1.

Procedure

The same as Experiment 1 with two critical differences: (1) children were asked to retrieve and generalize words during the learning phase, and (2) there was no testing phase, as we were interested in solely the forgetting and retrieval dynamics during learning.

Learning Phase. Across all conditions, the experimenter pointed to the first exemplar and told children (e.g., "This is camouflage!"), then paused for 4 s. For the following three exemplars, the experimenter first pointed to the picture and asked children, "What is this one called?". The experimenter gave children 4 s to respond to the query. If children responded correctly, incorrectly, or not at all within the 4 s, the experimenter then repeated the labeling phrase (e.g., "This is camouflage."). This procedure ensured the children heard the word four times during the learning phase, the same number of times as Experiments 1 and 3. The experimenter marked whether each response was correct or incorrect on a sheet of paper. To allow time for the queries, the children viewed each science concept for 28 s total; 4 s on the first science exemplar and 8 s on each of the three subsequent exemplars. Overall, this procedure mirrored the protocol for testing forgetting and retrieval in Vlach et al. (2012).

Results

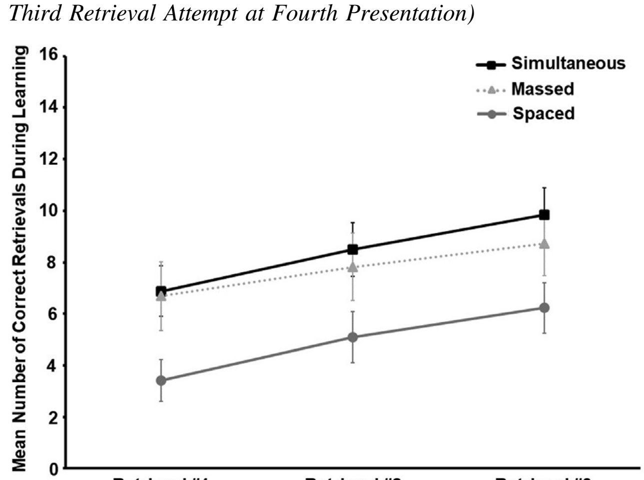
We were interested in whether children's ability to retrieve and generalize during learning would vary by presentation schedule. We started our analysis by examining differences in retrieval success across presentation schedules. Thus, we conducted a one way ANOVA, with presentation schedule (simultaneous, massed, and spaced) as the predictor variable, and the total number of correct retrievals during the learning phase as the outcome measure. We found a marginal main effect of presentation schedule, $F(2, 67) = 3.04$, $p = .055$, $\eta_p^2 = .083$. We conducted the same model with age as a co-variate and found a significant main effect of presentation schedule, $F(2, 66) = 3.36$, $p = .041$, $\eta_p^2 = .092$. Children's age was a significant co-variate, $F(1, 66) = 4.50$, $p = .038$, $\eta_p^2 = .064$. Older children had higher overall performance than younger children.

To further examine the main effect of presentation schedule, we examined differences among the three presentation conditions. A series of planned *t*-tests with Bonferroni corrections revealed that children in the spaced condition (*M* = 14.73 out of 54, *SD* = 12.68) had significantly less successful retrievals than children in the simultaneous condition (*M* = 25.23 out of 54, *SD* = 15.07), $t(46) = 2.59$, $p = .013$, $d = 0.749$, and marginally less than the massed condition (*M* = 23.23 out of 54, *SD* = 17.95), $t(42) = 1.81$, $p = .077$, $d = 0.547$. There were no significant differences in the number of successful retrievals between the simultaneous and massed conditions, $p > .10$. Thus, children in the spaced presentation condition had significantly more difficulty in retrieving and generalizing science categories during the learning phase.

We next examined children's pattern of retrieval successes across the exemplar presentations. We calculated the mean number of retrieval successes, at each retrieval event, during the learning phase (once at the second presentation, once at the third presentation, and once at the fourth presentation). As can be seen in Figure 3, there appeared to be differences in the patterns of retrieval successes

Figure 3

Mean Number of Correct Retrievals During the Learning Phase of Experiment 2, by Presentation Timing (Simultaneous, Massed, or Spaced) and Retrieval Attempt (First Retrieval Attempt at Second Presentation, Second Retrieval Attempt at Third Presentation, or Third Retrieval Attempt at Fourth Presentation)



Note. Error bars represent one standard error.

across the learning phase. A mixed ANOVA, with presentation schedule (simultaneous, massed, and spaced) as the between-factor predictor variable and mean number of retrieval success at exemplar presentations (first, second, and third retrieval event), confirmed a significant main effect of presentation schedule, $F(2, 67) = 3.04$, $p = .050$, $\eta_p^2 = .083$, and a main effect of retrieval event, $F(1, 67) = 84.08$, $p < .001$, $\eta_p^2 = .557$. There was no significant interaction between presentation schedule and retrieval event, $p > .10$.

To further understand the main effects of presentation schedule and retrieval event, we conducted a series of planned *t*-tests with Bonferroni corrections within each retrieval event. At the first retrieval event, children in the spaced condition ($M = 3.41$, $SD = 3.85$) had lower performance than children in the simultaneous condition ($M = 6.88$, $SD = 5.02$), $t(46) = 2.65$, $p = .011$, $d = 0.758$, and massed condition ($M = 6.68$, $SD = 6.24$), $t(42) = 2.09$, $p = .042$, $d = 0.630$. At the second retrieval event, children in the spaced condition ($M = 5.09$, $SD = 4.66$) had lower performance than children in the simultaneous condition ($M = 8.50$, $SD = 5.34$), $t(46) = 2.36$, $p = .024$, $d = 0.677$, but not the massed condition ($M = 7.82$, $SD = 6.09$), $p > .10$. At the third retrieval event, children in the spaced condition ($M = 6.23$, $SD = 4.58$) had lower performance than children in the simultaneous condition ($M = 9.85$, $SD = 5.33$), $t(46) = 2.50$, $p = .016$, $d = 0.724$, but not the massed condition ($M = 8.73$, $SD = 5.85$), $p > .10$. Thus, children in the spaced condition experienced the most retrieval difficulty across the retrieval events.

We also examined whether children experienced improvements in their retrieval performance across the learning phase. To test this possibility, we calculated a difference score between the first and third retrieval attempts. We then computed an ANOVA with presentation schedule (simultaneous, massed, and spaced) as the between-factor predictor variable and difference scores between the third and first retrieval event as the outcome variable, which revealed no significant differences, $p > .10$. Within each presentation condition, we calculated one sample *t*-tests comparing the difference scores to zero, as a measurement of whether retrieval improvement occurred. The results of these tests revealed that differences scores in the simultaneous ($M = 2.96$, $SD = 2.76$), $t(25) = 5.46$, $p < .001$, $d = 1.071$, massed ($M = 2.05$, $SD = 1.94$), $t(21) = 4.94$, $p < .001$, $d = 1.057$, and spaced conditions ($M = 2.82$, $SD = 2.26$), $t(21) = 5.85$, $p < .001$, $d = 1.248$, were all significantly different than zero. These findings suggest that children in all conditions experienced improvements in the retrieval performance, and to a similar degree across conditions.

These results confirm our hypotheses. Children in the spaced condition experienced the most retrieval difficulty, and children in the simultaneous and massed conditions experienced less retrieval difficulty. In brief, this study provides additional evidence for existing theories of the spacing effect in inductive learning, such as the forgetting-as-abstraction account (Vlach, 2014), which is discussed further in the General Discussion.

Experiment 3

Method

Previous research has focused on memory processes as explanations for the spacing effect observed in children's categorization and generalization (for a review, see Vlach, 2014). However, visual attention precedes encoding and retrieval processes in information processing theory (Klahr & Wallace, 1976), and thus visual attention processes may contribute to the spacing effect as well. Indeed, theories of the spacing effect in adults' learning, such as deficient processing theories (Hintzman, 1974), have proposed that learners pay less attention to massed presentations because they habituate to them faster than spaced presentations. Moreover, the attention attenuation hypothesis (Kornell et al., 2010; Wahlheim et al., 2011) proposes that learners pay less and less attention to massed presentations with each subsequent learning event because they feel like they have already learned the concept. Thus, in Experiment 3, we examined children's visual attention by using eye-tracking during the distractor and learning phases of Experiment 1. We tested the hypothesis that children visually inspect massed items to a lesser degree than spaced presentations, both overall (overall visit duration) and across learning events (exemplar visit duration), as proposed by deficient processing theories and the attention attenuation hypothesis. We also conducted a series of exploratory analyses to further characterize how simultaneous, massed, and spaced presentations engender different patterns of visual attention in children.

Participants

The participants were 64 preschool-aged children ($M_{age} = 54.17$ months, $SD = 8.45$, Range: 37–71 months, 32 girls). These children had not participated in Experiments 1 and 2. We used the same sample size and recruitment plan procedure as Experiment 1: We collected data until we reached at least 20 participants in every condition. Children were recruited from local preschools in a mid-size Midwestern city and the surrounding metropolitan area. Parents provided written consent, and children provided developmentally appropriate assent to participate in the study. Basic demographic data was collected from parents, and children were predominately White (75.0% of children) and from middle- to upper-SES families (59.4% of children). Children received a storybook for their participation in the study. An additional 10 children participated in the study but were not included in the final sample due to inadequate testing conditions (e.g., loud classroom) and/or inability to follow directions (e.g., not responding to questions).

Design

Children were randomly assigned to one of three between-subjects learning conditions (simultaneous, massed, or spaced). Random assignment resulted in the following number of participants per condition: *n*s = 21 simultaneous ($M_{\text{age}} = 51.76$ months, $SD = 8.25$, Range: 40–69 months), 23 massed ($M_{\text{age}} = 55.04$ months $SD = 9.32$, Range: 37–71 months), and 20 spaced ($M_{\text{age}} = 55.70$ months, $SD = 7.41$, Range: 41–68 months). A series of *t*-tests revealed that there were no significant age differences among the three between-subjects conditions, *p*s > .10.

Apparatus and Stimuli

Same as Experiment 1, with the addition of an eye-tracker. Children's eye gaze was measured by a Tobii Pro X3-120 eye tracker attached to a Dell Precision 7510 laptop with a screen size of 15.6 inches and a resolution of 1920 × 1080 pixels. This eye-tracker uses a corneal reflection tracking technique, in which an infrared light source attached to the laptop is directed at the eyes. The reflection of the light on the cornea relative to the center of the pupil is measured, which is used to estimate where the gaze is fixated on the computer screen. The eye-tracking system recorded gaze data at a sampling frequency of 120 Hz (accuracy = 0.4° and spatial resolution = 0.24°).

Procedure

The same as Experiment 1 with two critical differences: (1) children's visual attention was measured using an eyetracker, and thus there was a calibration procedure before the experiment began; and (2) there was no testing phase, as we were interested solely in the visual attention dynamics during learning.

A calibration procedure was carried out prior to the experiment. Children sat on a chair approximately two feet from the computer screen used to present the stimuli. The experimenter had the child adjust their seating position to make sure that the corneal reflections of both eyes were centered in the eye-tracking camera's field of view. Five calibration points were used using Tobii Pro Lab's standard calibration procedure, in which children were instructed to focus their gaze on each dot and follow the point with their eyes as it moved across the screen. The total duration of the calibration procedure was approximately 1 min. The eye tracker was able to capture $M = 64.59\%$ ($SD = 20.43\%$) of children's gaze during the experiment, not including the time intervals between spaced presentations when children were allowed to look away from the screen.

Results

Planned Analyses

All descriptive statistics for the visual attention measures are provided in Table 1. We started our analyses by examining the hypothesis that children would visually attend to simultaneous and massed presentations to a lesser degree than spaced presentations. To do so, we first created areas of interest (AOIs) on each image presented to children during the learning phase of each trial. We calculated the total visit duration to each AOI using Tobii ProLab, and calculated a mean total duration score from these values. We then conducted a one way ANOVA, with presentation schedule (simultaneous, massed, and spaced) as the predictor variable, and the mean visit duration as the outcome measure. We found a significant main effect of presentation schedule, $F(2, 61) = 16.32$, $p < .001$, $\eta_p^2 = .349$. We conducted the same model with age entered as a co-variate and found a significant main effect of presentation schedule, $F(2, 60) = 15.93$, $p < .001$, $\eta_p^2 = .347$. Children's age was not a significant co-variate.

To examine the nature of the differences across conditions, we then conducted a series of planned *t*-tests with Bonferroni corrections. These tests revealed that children in the spaced condition ($M = 1.81$ s out of 4 s, $SD = 0.55$) looked significantly less at exemplar presentations than children in the simultaneous condition ($M = 2.52$ s out of 4 s, $SD = 0.38$), $t(39) = 4.81$, $p < .001$, $d = 1.504$, and massed condition ($M = 2.81$ s out of 4 s, $SD = 0.74$), $t(41) = 4.96$, $p < .001$, $d = 1.516$. There was no significant difference in the mean visit duration between the simultaneous and massed conditions, $p > .10$. Thus, children in the spaced presentation condition looked less at the science category exemplars during the learning phase.

We also examined whether there were differences in looking duration to individual exemplars across the learning phase. That is, we analyzed whether there were significant differences in mean visit duration to the first, second, third, and fourth presentations. In the simultaneous condition, the first exemplar was defined as the top left image, the second exemplar was defined as the top right image, the third exemplar was defined as the bottom left image, and the fourth image was defined as the bottom right image. In the massed and spaced conditions, the first exemplar was defined as the first item

Table 1
Descriptive Statistics for Visual Attention Measures in Experiment 3

	Visit Duration (in seconds)				
	M (SD) All	*M (SD)* Exemplar1	*M (SD)* Exemplar2	*M (SD)* Exemplar3	*M (SD)* Exemplar4
Simultaneous	2.524 (0.383)	2.699 (0.612)	2.142 (0.427)	2.572 (0.551)	2.681 (0.541)
Massed	2.815 (0.742)	2.665 (0.740)	3.002 (0.753)	2.818 (0.772)	2.775 (0.811)
Spaced	1.815 (0.549)	2.069 (0.614)	1.695 (0.650)	1.721 (0.481)	1.774 (0.627)
	Visit Count (in number of visits)				
	M (SD) All	*M (SD)* Exemplar1	*M (SD)* Exemplar2	*M (SD)* Exemplar3	*M (SD)* Exemplar4
Simultaneous	2.843 (0.437)	2.884 (0.493)	2.905 (0.576)	2.899 (0.570)	2.685 (0.394)
Massed	1.060 (0.211)	1.060 (0.194)	1.060 (0.235)	1.057 (0.237)	1.063 (0.240)
Spaced	1.003 (0.147)	1.022 (0.148)	0.947 (0.211)	1.000 (0.154)	1.044 (0.183)
	Time to First Fixation (in seconds)				
	M (SD) All	*M (SD)* Exemplar1	*M (SD)* Exemplar2	*M (SD)* Exemplar3	*M (SD)* Exemplar4
Simultaneous	2.985 (0.588)	2.171 (0.724)	2.397 (0.693)	2.921 (0.873)	4.450 (0.990)
Massed	0.371 (0.133)	0.658 (0.301)	0.211 (0.173)	0.333 (0.178)	0.284 (0.169)
Spaced	1.232 (0.286)	1.010 (0.340)	1.267 (0.320)	1.362 (0.303)	1.290 (0.384)

presented, the second exemplar as the second item presented, and so on. The image across all three conditions was the same; for instance, the third exemplar in the simultaneous condition was the same exemplar image as the third exemplar in the massed and spaced conditions. We then conducted a mixed ANOVA, with presentation schedule (simultaneous, massed, and spaced) as the between-subjects factor and exemplar (first, second, third, and fourth) as the within-subjects outcome variable. The results revealed that there was no main effect of exemplar and a significant interaction between presentation condition and exemplar, $F(2, 61) = 3.24$, $p = .046$, $\eta_p^2 = .096$.

To follow-up on the interaction, we examined whether there were differences in mean visit duration across the exemplars within each presentation condition using a series of paired-samples *t*-tests with Bonferroni corrections. In the simultaneous condition, children demonstrated significantly less looking to the second exemplar than to the first exemplar, $t(20) = 4.67$, $p < .001$, $d = 1.015$, third exemplar, $t(20) = 3.31$, $p = .003$, $d = 0.723$, and fourth exemplar, $t(20) = 4.55$, $p < .001$, $d = 0.995$. In the massed condition, children demonstrated significantly more looking to the second exemplar than the first exemplar, $t(22) = 6.84$, $p < .001$, $d = 1.425$, and third exemplar, $t(22) = 3.16$, $p = .005$, $d = 0.658$. In the spaced condition, children demonstrated significantly more looking to the first exemplar than the second exemplar, $t(19) = 4.29$, $p < .001$, $d = 0.959$, third exemplar, $t(19) = 3.87$, $p = .001$, $d = 0.865$, and fourth exemplar, $t(19) = 3.33$, $p = .004$, $d = 0.743$. Thus, each of the presentation conditions engendered differential looking to individual exemplars.

We also calculated a difference score between the mean visit duration to the first and fourth exemplars, as a measure of mean visit duration change across the learning phase. We then computed an ANOVA with presentation condition (simultaneous, massed, and spaced) as the between-subjects variable and difference score (mean visit duration first exemplar – mean visit duration fourth exemplar) as the outcome measure. The results revealed a significant main effect of presentation condition, $F(2, 61) = 3.20$, $p = .048$, $\eta_p^2 = .095$. To examine the nature of the differences across conditions, we conducted a series of planned *t*-tests with Bonferroni corrections. These tests revealed that children in the spaced condition had a significantly larger decrease in looking time over the learning phase ($M = 0.29$ s, $SD = 0.40$) compared to children in the massed condition ($M = -0.11$ s, $SD = 0.41$), $t(41) = 3.27$, $p = .002$, $d = 1.000$. There was no significant differences between the simultaneous condition ($M = 0.02$ s, $SD = 0.72$) and the other conditions, *p*s > .10.

These results are the opposite of what is predicted by deficient processing theories (Hintzman, 1974) and the attention attenuation hypothesis (Kornell et al., 2010; Wahlheim et al., 2011). Children attended more (not less) to the massed and simultaneous presentations than the spaced presentations. Moreover, children in the spaced condition, not massed condition, looked significantly less to exemplars across the learning phase. Taken together, these results suggest that massed presentations do not always engender less looking, and thus deficient processing theories and the attention attenuation hypothesis are unlikely to be viable explanations for all observed spacing effects in prior research.

Exploratory Analyses

Because this was the first study to examine children's visual attention during simultaneous, massed, and spaced learning, we did not have planned analyses beyond the analyses described above. However, we conducted a series of exploratory analyses to (a) further characterize how these different presentations styles engender different visual attention dynamics to serve as a basis for hypotheses in future research, and (b) elucidate possible explanations for why we did not observe higher performance in the simultaneous condition at the immediate test of Experiment 1.

Visit Count. Our first exploratory analysis looked at whether there were differences in the mean visit count to exemplars. One difference between the three conditions is the number of exemplars on the screen; in the simultaneous condition, there are four images on the screen, and in the massed and spaced conditions, there is one image on the screen. Thus, this could have caused differences in whether the children looked at/away from exemplars, as children have more opportunities to look at other stimuli in the simultaneous condition. We conducted a one way ANOVA, with presentation schedule (simultaneous, massed, and spaced) as the predictor variable and the mean visit count as the outcome measure. We found a significant main effect of presentation schedule, $F(2, 61) = 270.99$, $p < .001$, $\eta_p^2 = .899$. We conducted the same model with age entered as a co-variate and found a significant main effect of presentation schedule, $F(2, 60) = 273.09$, $p < .001$, $\eta_p^2 = .901$. Children's age was not a significant co-variate.

To examine the nature of the differences across conditions, we then conducted a series of *t*-tests with Bonferroni corrections. These tests revealed that children in the spaced condition ($M = 1.00$ looks, $SD = 0.15$), $t(39) = 17.90$, $p < .001$, $d = 5.587$, and massed condition ($M = 1.06$ looks, $SD = 0.21$), $t(42) = 17.48$, $p < .001$, $d = 5.275$, looked significantly fewer times at exemplar presentations than children in the simultaneous condition ($M = 2.84$ looks, $SD = 0.44$). There was no significant difference in the mean visit count between the massed and spaced conditions, $p > .10$.

We also examined whether there were differences in visit count to individual exemplars across the learning phase. That is, we analyzed whether there were significant differences in mean visit count to the first, second, third, and fourth presentations. To test this possibility, we conducted a mixed ANOVA, with presentation schedule (simultaneous, massed, and spaced) as the between-subjects factor and exemplar (first, second, third, and fourth) as the within-subjects outcome variable. The results revealed that there was no main effect of exemplar and a significant interaction between presentation condition and exemplar, $F(2, 61) = 4.25$, $p = .019$, $\eta_p^2 = .122$. To follow-up on these analyses, we examined whether there were differences in mean visit duration across the exemplars within each presentation condition using a series of paired-samples *t*-tests with Bonferroni corrections. However, none of these *t*-tests resulted in significant results after correcting for multiple comparisons.

We also calculated a difference score between the mean visit count to the first and fourth exemplars, as a measure of mean visit count change across the learning phase. We then computed an ANOVA with presentation condition (simultaneous, massed, and spaced) as the between-subjects variable and difference score (mean visit count first exemplar – mean visit count fourth exemplar) as the outcome measure. The results revealed a significant main effect of presentation condition, $F(2, 61) = 4.80$, $p = .012$, $\eta_p^2 = .136$. To examine the nature of the differences across conditions, we conducted a series of planned *t*-tests with Bonferroni corrections. These tests revealed that children in the simultaneous condition had a significantly larger decrease in visit count to the first versus fourth exemplars (M = 0.199 looks, SD = 0.397) compared to children in the massed condition (M = −0.003 looks, SD = 0.142), $t(42) = 2.29$, $p = .027$, $d = 0.691$, and spaced condition (M = −0.022 looks, SD = 0.149), $t(39) = 2.34$, $p = .025$, $d = 0.731$. There was no significant difference between the massed and spaced condition, $p > .10$.

Taken together, these results show that the simultaneous condition engenders more looks to exemplars and more variation in the number of looks to exemplars across the learning phase. These results provide a potential explanation for why we did not observe higher performance in the simultaneous condition at the immediate test, as is often observed in prior research (e.g., Namy & Gentner, 2002; Vlach, Ankowski, & Sandhofer, 2012). There were no significant differences between the simultaneous and massed conditions in mean visit duration, but there was a significantly greater number of looks to exemplars in the simultaneous condition than the massed condition. This suggests that simultaneous presentations divide children's attention. That is, children in the simultaneous condition engaged in more, but shorter, looks than children in the massed condition. Thus, it may be that the benefits of simultaneous presentations for abstraction are diminished when children's visual attention is too divided. This possibility is further outlined in the General Discussion.

Time to First Fixation. We examined whether there were differences in the mean time to first fixation across the conditions. Time to first fixation was calculated as the duration of time between the beginning of the learning trial to the time children's eyes first visited the exemplar AOI. As noted above, one difference between the three conditions is the number of exemplars on the screen; in the simultaneous condition, there are four images on the screen, and in the massed and spaced conditions, there is one image on the screen. Thus, this could have caused differences in how long it took children to first look at exemplars; indeed, children in the simultaneous condition may have took more time to look at each exemplar than children in the massed and spaced conditions. To test this possibility, we conducted a one way ANOVA, with presentation schedule (simultaneous, massed, and spaced) as the predictor variable, and the mean time to first fixation as the outcome measure. We found a significant main effect of presentation schedule, $F(2, 61) = 264.96$, $p < .001$, $\eta_p^2 = .897$. We conducted the same model with age entered as a co-variate and found a significant main effect of presentation schedule, $F(2, 60) = 251.49$, $p < .001$, $\eta_p^2 = .893$. Children's age was not a significant co-variate.

We then conducted a series of t-tests with Bonferroni corrections to examine differences across the conditions. These tests revealed that children in the simultaneous condition (M = 2.98 s to fixate, SD = 0.59) took significantly longer to look at each exemplar than children in the massed condition (M = 0.37 s to fixate, SD = 0.13), $t(42) = 20.75$, $p < .001$, $d = 6.266$, and spaced condition (M = 1.23 s to fixate, SD = 0.29), $t(39) = 12.03$, $p < .001$, $d = 3.762$. Moreover, children in the spaced condition took significantly longer to look at each exemplar than children in the massed condition, $t(41) = 12.94$, $p < .001$, $d = 3.966$. Thus, children in the massed condition were faster to look at the exemplars than children in the other two conditions.

We also examined whether there were differences in mean time to first fixation on individual exemplars across the learning phase. That is, we analyzed whether there were significant differences in mean time to fixation to the first, second, third, and fourth presentations. To test this possibility, we conducted a mixed ANOVA, with presentation schedule (simultaneous, massed, and spaced) as the between-subjects factor and exemplar (first, second, third, and fourth) as the within-subjects outcome variable. The results revealed that there was a main effect of exemplar, $F(1, 61) = 76.96$, $p < .001$, $\eta_p^2 = .558$, and a significant interaction between presentation condition and exemplar, $F(2, 61) = 84.14$, $p < .001$, $\eta_p^2 = .734$.

To follow-up on these analyses, we examined whether there were differences in mean time to first fixation across the exemplars within each presentation condition using a series of paired-samples t-tests with Bonferroni corrections. In the simultaneous condition, all t-tests revealed a significant difference except the comparison between the first and second exemplar, $p > .10$. Children took significantly longer to look at the fourth exemplar compared to the first exemplar, $t(20) = 8.73$, $p < .001$, $d = 1.905$, second exemplar, $t(20) = 11.51$, $p < .001$, $d = 2.510$, and third exemplar, $t(20) = 6.48$, $p < .001$, $d = 1.415$. Children also took significantly longer to look at the third exemplar compared to the first exemplar, $t(20) = 3.67$, $p = .002$, $d = 0.802$, and second exemplar, $t(20) = 2.90$, $p = .009$, $d = 0.642$. In the massed condition, children took significantly longer to look at the first exemplar compared to the second exemplar, $t(22) = 6.84$, $p < .001$, $d = 1.429$, third exemplar, $t(22) = 5.90$, $p < .001$, $d = 1.235$, and fourth exemplar, $t(22) = 5.28$, $p < .001$, $d = 1.104$. In the spaced condition, we observed the opposite pattern of performance to the massed condition: Children took significantly less time to look at the first exemplar compared to the second exemplar, $t(19) = 2.96$, $p = .008$, $d = 0.671$, third exemplar, $t(19) = 4.32$, $p < .001$, $d = 0.962$, and fourth exemplar, $t(19) = 3.84$, $p = .001$, $d = 0.860$.

We also calculated a difference score between the mean time to the first fixation to the first and fourth exemplars, as a measure of mean visit count change across the learning phase. We then computed an ANOVA with presentation condition (simultaneous, massed, and spaced) as the between-subjects variable and difference score (mean visit count first exemplar – mean visit count fourth exemplar) as the outcome measure. The results revealed a significant main effect of presentation condition, $F(2, 61) = 75.77$, $p < .001$, $\eta_p^2 = .713$. To examine the nature of the differences across conditions, we conducted a series of planned t-tests with Bonferroni corrections. These tests revealed differences between all three conditions; children in the simultaneous condition had a significantly larger increase in time to first fixation across first to fourth exemplars (M = −2.28 s, SD = 1.20) compared to children in the massed condition (M = 0.37 s, SD = 0.34), $t(42) = 10.20$, $p < .001$, $d = 3.079$, and spaced condition (M = −0.28 s, SD = 0.33, $t(39) = 7.22$, $p < .001$, $d = 2.256$. Children in the spaced condition had a significantly larger increase in time to first fixation across first to fourth exemplars compared to children in the massed condition, $t(41) = 6.15$, $p < .001$, $d = 1.964$.

In sum, children in the massed condition took significantly less time to look at exemplars than children in the other conditions. These results suggest the massed condition fostered faster looking to exemplars after the first exemplar, whereas the spaced condition engendered slower looking to exemplars after the first exemplar. Finally, the simultaneous condition led children to view the images in the order consistent with how an adult would when reading English: top left to top right and bottom left to bottom right. The vast majority of preschool-aged children do not know how to read, but their experience being read to by caregivers may influence their lower-level visual attention (Ferretti, Mazzotti, & Brizzolara, 2008).

The implications of these findings are discussed in the General Discussion.

General Discussion

Previous research presents seemingly paradoxical results: Both presenting exemplars at the same time and presenting exemplars apart in time promote children's learning. To date, only one study has addressed these paradoxical results (Vlach et al., 2012); this work suggests that the timescale of generalization may determine which presentation method supports learning. That is, simultaneous presentations support immediate learning, but spaced presentations support long-term learning. We built upon this work by empirically testing whether the learning conditions lead to differences in forgetting and retrieval dynamics. We also tested alternative explanations, such as whether processes of visual attention could explain these results, as proposed by deficient processing theory and the attention attenuation hypothesis.

In Experiment 1, we found that children had the highest performance in the spaced condition at the 5-min delayed test. Thus, this result replicates previous research on the spacing effect in children's categorization and generalization (e.g., Vlach et al., 2008, 2012). However, we did not observe higher performance at the immediate test in the simultaneous condition. Why? The results of Experiment 3 provide insights into why we did not observe a benefit of simultaneous presentations. Although the overall looking time in the simultaneous condition did not differ from the massed condition, the number of looks to each exemplar did vary. That is, children were making shorter, more frequent looks to exemplars in the simultaneous condition, suggesting that children were experiencing more divided attention. The science category exemplars in this study were much more complex than the simple objects used in prior research (e.g., Namy & Gentner, 2002; Oakes & Ribar, 2005; Vlach, Ankowski, & Sandhofer, 2012), and thus children may not have been able to visually inspect each of the exemplars sufficiently. Indeed, this divided attention may have diminished the abstraction benefit hypothesized to underlie simultaneous presentations. Moreover, studies of comparison often provide learners with two exemplars (e.g., Namy & Gentner, 2002; Oakes & Ribar, 2005), rather than four exemplars, and thus much of the research on simultaneous presentations has placed smaller demands on children's visual attention. In brief, we propose that this is a constraint on the efficacy of simultaneous presentations; if the exemplars are so numerous or complex that children's visual attention becomes divided, then there is unlikely to be a benefit of simultaneous presentations at an immediate test.

In Experiments 2 and 3, we focused on the learning phase of the experiment to examine the nature of the visual attention and retrieval dynamics underlying learning. In Experiment 2, we tested predictions of forgetting-as-abstraction theory (Vlach, 2014) and study-phase retrieval theories (Thios & D'Agostino, 1976), which proposes that forgetting and retrieval dynamics are the mechanisms that lead to spacing effects. According to this account, children forget in between learning events in spaced presentations, causing retrieval of prior learning to be more difficult (compared to massed and simultaneous schedules). This retrieval practice strengthens the memory trace, in turn facilitating future retrieval and generalization. This is exactly what we found in Experiment 2: Children had less retrieval success in the spaced condition than the simultaneous and massed conditions. Thus, this work provides further support for forgetting-as-abstraction theory, and study-phase retrieval theories more broadly.

In Experiment 3, we examined children's visual attention during the three learning schedules. We had two key hypotheses: First, based on deficient processing theories (Hintzman, 1974), we predicted children would pay less attention to massed presentations because they habituate to them faster than spaced presentations. We found the opposite: Children looked more to simultaneous and massed presentations than spaced presentations. Thus, this work provides evidence against deficient processing theories. Second, based on the attention attenuation hypothesis (Kornell et al., 2010; Wahlheim et al., 2011), we predicted that children pay less attention to massed presentations, but not spaced presentations, with each subsequent learning event. We found the opposite: Children looked less at the spaced presentations with each subsequent learning event. We also conducted a series of exploratory analyses in Experiment 3, in which we observed several ways in which the three presentation conditions engender different patterns of visual attention. For instance, the simultaneous condition engendered more unique looks to exemplars but longer first fixation times to exemplars. These exploratory analyses introduce the question of whether the observed visual attention dynamics contribute to learning outcomes, both individually and in interaction with each other, or whether they represent tradeoffs between other cognitive processes, such as memory processes.

Taken together, this work yields several important contributions to our understanding of children's categorization and generalization. First, this work provides evidence that supports the mechanisms proposed by the forgetting-as-abstraction theory (Vlach, 2014); spaced presentations led to more forgetting and retrieval difficulty than the other conditions. Second, this work provided evidence against deficient processing theories (Hintzman, 1974) and the attention attenuation hypothesis (Kornell et al., 2010; Wahlheim et al., 2011), suggesting that visual attention is not an explanation for why we observe spacing effects in prior research. Third, because this is the first study to directly compare children's visual attention in simultaneous, massed, and spaced presentations, this work provides several important empirical findings that should be pursued by future research.

There are many exciting directions for future research; we highlight just three areas here. First, given that there is a small number of studies in this line of work, researchers should continue to expand the scope of these phenomena, such as by determining causal relations between attention, memory, and presentation timing; examining learning across longer timescales, and the cognitive processes that may occur with more time (e.g., consolidation); and widening the demographics (age, SES, culture, etc.) of participants in our research. Second, future work should identify the constraints on the benefits of simultaneous presentations. That is, how can we, a priori, predict when simultaneous presentations are not efficacious for learning? Based on the current research, we predict that the number and complexity of the exemplars will likely be important, but there may be other constraints as well. For example, working memory abilities may constrain children's ability to process multiple exemplars simultaneously. The third area of future research is to examine why less visual attention was observed in the spaced condition. Our hypothesis is that less visual attention could be facilitating retrieval. For instance, children may need to look away from stimuli in order to dedicate cognitive resources to successfully retrieving and generalizing prior learning. Indeed, research with adult learners has observed that looking at nothing (i.e., empty spaces) during learning is often related to stronger memory performance (for a review, see Ferreira et al., 2008), and thus researchers have argued that looking at nothing supports retrieval processes. Moreover, looking away and/or in new locations may indicate the type of categorization processes children are engaging in, such as similarity-based or rule-based categorization (Scholz et al., 2015).

On a final note, we believe that this work also has important implications for the field of science education. The Next Generation Science Standards (National Research Council, 2015), a set of national standards for science education, encourage teachers to present students with multiple examples of a science concept to promote transfer. However, the standards do not provide clear guidance on when and how to present these examples. We believe that this work, in conjunction with prior research on spaced learning in science education curricula (e.g., Custers, 2010; Dresner, de Rivera, Fuccillo, & Chang, 2014; Gluckman, Vlach, & Sandhofer, 2014; Kapler, Weston, & Wiseheart, 2015; Reynolds & Glaser, 1964; Vlach & Sandhofer, 2012) gives a clear recommendation: if the educational goal is to support long-term generalization of science concepts, present students with examples that are distributed across time. If students seem like they are visually attending less and forgetting more, that is to be expected. Indeed, this research serves as a powerful example of how theories and research in the field of cognitive development can be used to inform educational practices.

References

Ankowski, A. A., Vlach, H. A., & Sandhofer, C. M. (2013). Comparison vs. contrast: Task specifics affect category acquisition. *Infant and Child Development*, *22*(1), 1–23. https://doi.org/10.1002/icd.1746.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, J. R. Pomerantz (Eds.) & FABBS Foundation, *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, *9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003.

Boroditsky, L. (2007). Comparison and the development of knowledge. *Cognition*, *102*(1), 118–128. https://doi.org/10.1016/j.cognition.2002.08.001.

Custers, E. J. (2010). Long-term retention of basic science knowledge: A review study. *Advances in Health Sciences Education*, *15*(1), 109–128. https://doi.org/10.1007/s10459-008-9101-y.

Dresner, M., de Rivera, C., Fuccillo, K. K., & Chang, H. (2014). Improving higher-order thinking and knowledge retention in environmental science teaching. *BioScience*, *64*(1), 40–48. https://doi.org/10.1093/biosci/bit005.

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11), 405–410. https://doi.org/10.1016/j.tics.2008.07.007.

Ferretti, G., Mazzotti, S., & Brizzolara, D. (2008). Visual scanning and reading ability in normal and dyslexic children. *Behavioural Neurology*, *19*(1–2), 87–92. https://doi.org/10.1155/2008/564561.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170. https://doi.org/10.1016/B978-1-4832-1446-7.50026-1.

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure mapping and relational language support children's learning of relational categories. *Child Development*, *82*(4), 1173–1188. https://doi.org/10.1111/j.1467-8624.2011.01599.x.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38. https://doi.org/10.1016/0010-0285(83)90002-6.

Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum. *Applied Cognitive Psychology*, *28*(2), 266–273. https://doi.org/10.1002/acp.2997.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology; The Loyola symposium*, (pp. 77–97). Erlbaum.

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., & Sageman, B. (2013). Finding faults: Analogical comparison supports spatial concept learning in geoscience. *Cognitive Processing*, *14*(2), 175–187. https://doi.org/10.1007/s10339-013-0551-7.

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction*, *36*, 38–45. https://doi.org/10.1016/j.learninstruc.2014.11.001.

Klahr, D., & Wallace, J. G. (1976). *Cognitive development: An information-processing view*. Lawrence Erlbaum.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, *25*(2), 498–503. https://doi.org/10.1037/a0017807.

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, *131*(1), 5–15. https://doi.org/10.1037/0096-3445.131.1.5.

National Research Council. (2015). *Guide to implementing the next generation science standards*. National Academies Press, https://doi.org/10.17226/18802.

Oakes, L. M., & Ribar, R. J. (2005). A comparison of infant's categorization in paired and successive presentation familiarization tasks. *Infancy*, *7*(1), 85–98. https://doi.org/10.1207/s15327078in0701_7.

Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology*, *55*(5), 297–308. https://doi.org/10.1037/h0040734.

Roediger,, III, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x.

Rose, R. J. (1980). Encoding variability, levels of processing, and the effects of spacing of repetitions upon judgments of frequency. *Memory & Cognition*, *8*(1), 84–93. https://doi.org/10.3758/BF03197555.

Scholz, A., von Helversen, B., & Rieskamp, J. (2015). Eye movements reveal memory processes during similarity-and rule-based decision making. *Cognition*, *136*, 228–246. https://doi.org/10.1016/j.cognition.2014.11.019.

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*(5), 529–536. https://doi.org/10.1016/0022-5371(76)90047-5.

Thompson, C. A., & Opfer, J. E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development*, *81*(6), 1768–1786. https://doi.org/10.1111/j.1467-8624.2010.01509.x.

Trippas, D., & Pachur, T. (2019). Nothing compares: Unraveling learning task effects in judgment and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45* (12), 2239–2266. https://doi.org/10.1037/xlm0000696.

Vendetti, M. S., Matlen, B. J., Richland, L. E., & Bunge, S. A. (2015). Analogical reasoning in the classroom: Insights from cognitive science. *Mind, Brain, and Education*, *9*(2), 100–106. https://doi.org/10.1111/mbe.12080.

Vlach, H. A. (2014). The spacing effect in children's generalization of knowledge: Allowing children time to forget promotes their ability to learn. *Child Development Perspectives*, *8*(3), 163–168. https://doi.org/10.1111/cdep.12079.

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *38*(1), 246–254. https://doi.org/10.1037/a0025260.

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*(4), 1137–1144, 10.111/j.1467-8624.2012.01781.x.

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109* (1), 163–167. https://doi.org/10.1016/j.cognition.2008.07.013.

Vlach, H. A. (2019). Learning to remember words: Memory constraints as double-edged sword mechanisms of language development. *Child Development Perspectives*, *13*, 159–165. https://doi.org/10.1111/cdep.12337.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39* (5), 750–763. https://doi.org/10.3758/s13421-010-0063-y.

Received March 25, 2021
Revision received October 16, 2021
Accepted October 21, 2021 ■