Cognitive Development 60 (2021) 101099

Contents lists available at ScienceDirect

Cognitive Development

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

Relating a picture and 1000 words: Self-derivation through integration within and across presentation formats

Alena G. Esposito[a,*], Katherine Lee[b], Jessica A. Dugan[b], Jillian E. Lauer[c], Patricia J. Bauer[b]

[a] *Department of Psychology, Clark University, United States*
[b] *Department of Psychology, Emory University, United States*
[c] *Department of Psychology, New York University, United States*

ARTICLE INFO

Keywords:
Graphic representation
Self-derivation through integration
Integration across text and graphic representations

ABSTRACT

To build knowledge, separate yet related learning episodes can be integrated with one another and then used to derive new knowledge. Separate episodes are often experienced through different formats, such as text passages and graphic representations. Accordingly, in the present research, we tested integration of learning episodes provided through different presentation formats with children in the laboratory (Experiment 1; $n = 24$; $M = 8.36$ years) and in classrooms (Experiment 2; $n = 85$; $M = 9.34$ years and Experiment 3; $n = 154$; $M = 10.67$ years). Children in the laboratory were successful in both same-format and different-format conditions. Children in the classroom were also successful in both conditions, but in Exp. 2 showed a cost to integration across two different presentation formats compared to the same-format condition. In Exp. 3, greater support for encoding the graphic information was added and performance no longer showed a cost between conditions.

Learning does not occur at one place and time, but rather, lessons are distributed across experiences. For the products of learning to accumulate, separate yet related learning episodes must be integrated with one another. Indeed, integration of content is central to inferential processes, including anaphoric inference (e.g., McKoon & Ratcliff, 1980), transitive inference (e.g., Bryant & Trabasso, 1971), and analogical problem solving (e.g., Gentner & Smith, 2013), to name a few. The products gained through integration are of particular interest because these products are how knowledge is built over time. As important as integration is to building knowledge, little is known about the process of memory integration and the contexts that support successful extension of knowledge through integration. Critically, to date, inferential processes have been tested in paradigms with the same presentation format between related episodes (e.g., Esposito & Bauer, 2019; Preston, Shrager, Dudukovic, & Gabrieli, 2004). Outside the laboratory, however, learning crosses modalities: information is presented not only through text, but through images, videos, podcasts, and radio to name a few. Indeed, the IES *What Works Clearinghouse* even recommends the use of graphic presentations as a mechanism for increasing learning (Pashler et al., 2007; although with conditions, see Schnotz, 2014, for discussion and guidelines). Yet the process of integration of non-redundant information across presentational formats—such as text and a graphic display—has yet to be examined. Accordingly, in the present research, we tested integration of separate, non-redundant episodes rendered in different presentational formats by children in the laboratory (Experiment 1) and in the classroom (Experiments 2 and 3).

* Corresponding author at: Department of Psychology, 950 Main Street, Clark University, Worcester, MA, 01610, United States.
E-mail address: aesposito@clarku.edu (A.G. Esposito).

https://doi.org/10.1016/j.cogdev.2021.101099
Received 12 November 2020; Received in revised form 25 June 2021; Accepted 9 August 2021
Available online 23 August 2021

Inferential processes such as analogy, deduction, and induction (e.g., Gentner, 1983, 1988; Goswami, 2011) are generally recognized as a major mechanism of cognitive development (e.g., Bauer, 2012; Brown, 1982; Siegler, 1989). In the present research, we focused on the specific inferential process of self-derivation based on integration (e.g., Bauer & San Souci, 2010). This process is especially well suited for examining the contexts that support integration. In self-derivation based on integration, the integration of related facts results in a novel product that serves as evidence of successful integration. For example, a fact learned in one episode (e.g., "dolphins talk by clicking and squeaking") can be integrated with a related fact learned in a different episode (i.e., "dolphins live in groups called *pods*"), and the combined or integrated fact representation can be recruited to answer the question "How does a pod talk?" Although never given directly, the answer that "pods talk by clicking and squeaking" can be derived through inferential processes operating over the integrated fact representation. Successfully answering the integration questions is evidence that integration has occurred.

Focus on the inferential process of self-derivation through integration is especially well suited for educationally relevant work because it has been shown to be a valid model of how new factual knowledge accumulates over time and experience (discussed in Bauer, Esposito et al., 2020). As reviewed elsewhere (e.g., Bauer et al., 2019; Esposito & Bauer, 2018), information newly self-derived through memory integration is rapidly incorporated into the knowledge base (Bauer & Jackson, 2015) and is retained over time by both children and adults (e.g., Varga & Bauer, 2013, 2017, respectively). The process occurs over a range of semantic content, including content from academic curricula. The process has been observed in the laboratory (e.g., Bauer & Larkina, 2017; Esposito & Bauer, 2018; Varga & Bauer, 2013) and in elementary classrooms (e.g., Esposito & Bauer, 2017; 2019). Moreover, self-derivation performance is related to academic achievement in both language arts (i.e., literacy and writing) and math in the elementary school years (Esposito & Bauer, 2017) and predicts grade point average (GPA) in college students (Varga, Esposito, & Bauer, 2019). These observations suggest that memory integration as measured by self-derivation is a valid model of accumulation of semantic knowledge through productive processes.

Importantly, all of the studies of self-derivation through memory integration to date have required that participants integrate information from separate episodes of the same presentation format. That is, to-be-integrated facts are conveyed either through separate brief passages of text (e.g., Bauer & San Souci, 2010; Esposito & Bauer, 2018) or through separate sentences (e.g., Bauer, Blue, Xu, & Esposito, 2016; Esposito & Bauer, 2018). Homogenous presentation mediums are not unique to the self-derivation through integration test paradigm. For example, in the associative inference test paradigm, integration of two images or abstract shapes is tested (e.g., Preston et al., 2004). Yet related experiences outside the laboratory seldom share such high levels of surface similarity. More often, a text passage will relate to an oral lecture, a museum exhibit, a video, or a graphic representation, for example. Thus, examination of integration across mediums of presentation is needed to better model ecologically valid learning. As such, the present work stands to inform potential challenges of integrating under conditions of a common method of instruction: text complemented by graphic representations. In addition, the work will address the generalizability of the process of self-derivation through memory integration.

1. Integration across presentation formats

In the present research, we tested integration of separate yet related episodes of new learning when one of the episodes is text-based and the other conveys new information graphically. Integration across textual and graphic presentations is especially relevant for at least two reasons. First, the requirement to integrate across text and graphic representations becomes increasingly frequent as children move through their formal educational curriculum, especially in science disciplines (Devetak & Vogrinc, 2013). Indeed, requirements for integration across texts and graphic representations are part of one of the dominant education standards in the United States, the Common Core State Standards (www.corestandards.org).

Second, integration across textual and graphic presentations is a test of integration under conditions of low surface similarity. Surface similarity refers to the details of a presentation that are not specific to content meaning. For example, if the goal is to teach polygon recognition, then the number of sides and angles is pertinent but the color in which the different shapes are portrayed is a surface-level feature. Numerous studies on productive processes broadly have identified low surface similarity between related stimuli as a challenge to their deployment (e.g., Gentner, 1988; Winner, Rosenstiel, & Gardner, 1976). For example, young children are dependent on surface-level features when tasked with finding analogous relational structures in an analogical problem solving task (Kotovsky & Gentner, 1996). Considering the specific process of self-derivation through memory integration, results under low surface similarity conditions have been mixed. For example, in Bauer, King, Larkina, Varga, and White (2012), 6-year-old children had lower levels of self-derivation under low surface similarity conditions relative to under high surface similarity conditions. That is, when the main characters in related passages were different from one another (e.g., a lizard and a dog: low surface similarity), children's responses to open-ended questions (i.e., questions without answer choices provided such as in forced-choice) were correct on only 37 % of trials. In contrast, under high surface similarity conditions (i.e., when the characters in the related passages were the same: e.g., a lizard in both stories) self-derivation in open-ended testing was 63 % correct. This effect of surface similarity has not always been found in the classroom, however. For example, children in 2nd grade (mean age 8.17 years) did not benefit from the high surface similarity condition, performing similarly in both low and high surface similarity conditions (36 % and 40 %, respectively; Bauer et al., 2019). The performance across conditions resembled low surface similarity performance in laboratory testing, suggesting that the high surface similarity presentation was not a facilitator in the classroom. Differences in performance between these studies are not likely to be developmental in nature, as multiple studies have shown that performance improves across development in both the laboratory and the classroom context (e.g., Bauer & Larkina, 2017; Esposito & Bauer, 2017). Thus, developmentally, the older children would have been expected to perform at a higher level across both conditions rather that the lower performance that was found. Instead, the results underscore the need for research in both the laboratory and classroom contexts, under different surface-similarity conditions.

Additional evidence that surface similarity is but one contributor to performance comes from another classroom study. Esposito and Bauer (2019) reported that children in Grades 3 and 4 integrated under a different type of low surface-similarity condition, namely when one member of a pair of related passages was presented in English and the other in Spanish. Under these conditions, children in 3rd grade English/Spanish dual-language instruction successfully integrated only in the high surface similarity (same-language) condition; their performance in the low surface similarity (different-language) condition was at chance. Children in 4th grade English/Spanish dual-language instruction, however, performed equally well when required to integrate passages under the same low and high surface similarity conditions.

2. Graphic presentations

Integration of textual and graphic presentations presents another ecologically valid test of memory integration under conditions of low surface similarity. Graphic presentations are generally considered an aid to comprehension (see Levin, Anglin, Carney, Willows, & Houghton, 1987, for a meta-analysis and Carney & Levin, 2002, for a review). Graphic presentations that accompany text are thought to draw attention to important elements of the text, aid in comprehension by depicting abstract concepts in a concrete form, and facilitate encoding. Thus, they are thought to be especially beneficial and are often used in science materials (Cook, 2006; Hannus & Hyönä, 1999). The majority of the research examining the role of illustrations in science learning has been conducted with college students. An exception is an investigation with elementary children of the role of graphic presentations in understanding scientific equipment. Mayer and Gallini (1990) found that graphic presentations were especially helpful for children, compared to college students, in a condition in which participants had low levels of prior knowledge on the topic of instruction. This indicates that children especially may benefit from graphic presentations.

Graphic presentations are generally thought to have a positive impact on comprehension. Yet there is evidence that graphic presentations can have a negligible or even negative impact on learning (e.g., Bergey, Cromley, & Newcombe, 2015; Canham & Hegarty, 2010; Coleman & Dantzler, 2016; Cromley et al., 2013; Florax & Ploetzner, 2010; Roberts, Norman, & Cocco, 2015). The main challenge is lack of comprehension of the graphic presentations themselves. Graphic presentations require the individual to sift through some ambiguity and even "clutter" to extract targeted information that may also require interpretation. For example, Arteaga, Batanero, Contreras, and Cañadas (2012) found that both elementary students and their teachers had difficulty interpreting and extracting information from graphics. Similarly, professors and designers of graphics for textbooks did not identify errors in graphics and did not accurately predict where college students might struggle with incorrect interpretations (Benson, 1997). Even when images are recognized and labeled accurately, students in third grade through college only correctly interpret the intended meaning of graphic representations approximately one-third of the time (Boling, Eccarius, Smith, & Frick, 2004). Thus, the utility of graphic presentations is limited by whether they are interpreted accurately.

Another possible challenge associated with graphic presentations is that typically, they must be integrated with accompanying text. In a review of multimedia learning, Mayer (2005) describes the integration of information learned through text with information learned through a visual medium as "extremely demanding" due to the complexity introduced to integrate across two different forms of media, and he was referring to learning taking place within one episode. Learning across both time and medium is expected to be even more challenging. In a study that induces the integration of text and graphics, only participants who generated their own graphic summaries showed evidence of integrating text with graphics (Paoletti, 2005). In addition to the complexity of integrating across media, Paoletti posits that learners do not always adequately attend to graphic presentations, resulting in shallow processing that does not lend itself to an integrated representation. Thus, although integration between text and graphic presentations becomes increasingly important as education progresses, there are many challenges to the interpretation and comprehension of graphic presentations as well as subsequent integration with text material.

3. The present study

The process of memory integration is a means by which we build new factual knowledge across separate learning episodes. The specific demand to integrate across separate learning episodes presented through text and graphic presentations is integral to science education in particular. Graphic presentations are thought to be helpful learning aids, yet interpreting graphics can be a challenge. It is also unclear whether the low surface similarity between text and graphic presentations introduces an additional challenge to integration. Thus, in the present study, we investigated whether there is a cost or benefit to presenting related information in separate episodes of learning through text and graphic presentations compared to the previously examined homogenous presentation formats.

We conducted three experiments to examine memory integration of separate episodes of learning rendered in text and graphic form. Experiment 1 is a laboratory examination of memory integration of information presented in separate episodes through text and in graphic form. We directly compared performance in a text-graphic condition to the homogenous text-text condition (e.g., Bauer & San Souci, 2010; Esposito & Bauer, 2018). In Experiments 2 and 3, we extended the inquiry into the classroom. To foreshadow the results, there was a cost to integrating across text and graphic presentations in Experiment 2. In Experiment 3, we examined whether the difficulty in the text-graphic condition was due to difficulty encoding the target graphic information or difficulty integrating across presentation mediums by providing additional support for graphic comprehension at encoding.

In all three experiments, we focused on school age children (grades 2–5) for several reasons. First, children in this age range transition from "learning to read" to "reading to learn," an important distinction that brings about greater need to interpret and integrate text with graphic presentations (e.g., Lesnick, Goerge, Smithgall, & Gwynne, 2010). Also during this age period, there are mixed findings regarding navigation of differences in surface similarity. As reported, there is evidence that children experience difficulty integrating across low surface similarity conditions (Bauer et al., 2012; Esposito & Bauer, 2019) as well as evidence that children can be successful under these conditions (Bauer et al., 2019; Esposito & Bauer, 2019). Finally, from a practical perspective, children of this age range are capable of participating in the multiple trials of the task necessary to examine the difference between text-text and text-graphic conditions (Esposito & Bauer, 2018). Children in the laboratory based experiment (Exp. 1) were younger than the classroom based experiments (Exps. 2 & 3) ($M = 8.36$ vs. 9.34 & 10.7 years, respectively) for two reasons. First, prior research indicated that the classroom adds an additional challenge to memory integration such that older children in the classroom perform more similarly to younger children in the laboratory (e.g., Esposito & Bauer, 2017; 2019). Second, the classroom based studies were in collaboration with a school system and the work is at the discretion of the school administrators; the grade levels included represent a population sample of the grade levels with which we were invited to collaborate. Importantly, the statements regarding the importance of this age range apply across all three experiments.

In summary, there is ample evidence that low surface similarity conditions pose a challenge to productive processes. It is unclear how different conditions of low surface similarity impact memory integration. Graphic presentations are of particular interest because of the prevalence in informative texts (Devetak & Vogrinc, 2013) and the requirement that information conveyed in graphic presentations be interpreted and also integrated with information conveyed through text. Though graphic presentations could be an aid to memory and comprehension, the challenge of interpreting graphic presentations and integrating across media formats could result in lower performance. We hypothesized that conditions of high surface similarity (text-text) would facilitate integration performance compared to low surface similarity conditions associated with cross-media presentation (text-graphic) in both the laboratory and the classroom, and that the challenges of cross-media integration could be mitigated by supports at encoding.

4. Experiment 1

The major purpose of Experiment 1 was to examine text-text compared to text-graphic performance in 7- to 9-year-old children in the laboratory using the newly developed text and graphic stimuli. We predicted that the text-graphic condition would be more

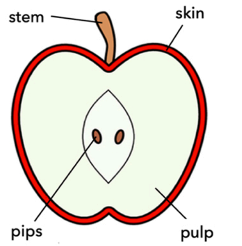
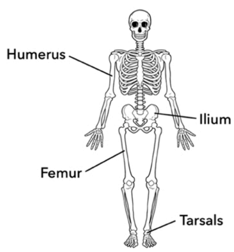
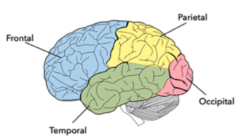
Stem Fact 1		Stem Fact 2	Stem Fact Questions	Integration Question
Text	Graphic			
Apple seeds are called pips.		Apple seeds contain cyanide.	What are apple seeds called? What do apple seeds contain?	What do pips contain?
The bone in the thigh is called the femur.		The bone in the thigh is the longest bone in the body.	What is the bone in the thigh called? Where is the longest bone in the body?	What is the longest bone in the body called?
The smallest part of the brain is called the occipital.		The smallest part of the brain is used for vision.	What is the smallest part of the brain called? What is the smallest part of the brain used for?	What is the occipital part of the brain used for?

Fig. 1. Example stem facts pairs and self-derivation (SD) questions. Children were presented with Stem Fact 1 in either text or graphic form and Stem Fact 2 in text form. Children in all experiments were tested on the same integration questions, with the exception of a few modifications to stimuli in Experiment 3 to accommodate appropriate unknown curriculum.

challenging than the text-text condition due to the challenges of interpreting graphic presentations and integrating across mediums. However, graphics could also have a facilitatory effect (see Carney & Levin, 2002 for review) that could result in similar performance across conditions or even superior text-graphic performance.

5. Method

5.1. Participants

Twenty-four children (15 girls), aged 7–9 years ($M = 8.36$, $SD = 0.64$ years) participated. Based on parental report, the sample was 4 % Asian, 29 % Black or African American, 42 % Caucasian, and 21 % multi-racial; 4 % did not report their race. Eight-percent of the children also were identified as Hispanic or Latinx. Recruitment was determined by apriori G*Power analyses to achieve 80 % power (Faul, Erdfelder, Buchner, & Lang, 2009). Prior to the start of the session, parents provided written informed consent for their children to participate; children provided verbal assent. At the end of the single laboratory session, children were given a small toy as a thank you, and parents were given a $5 gift card to a local merchant.

5.2. Stimuli

Given that the ultimate goal was to inform educational practice, we designed the stimuli and procedures with typical classroom practices in mind. First, text and graphic presentations are often used to supplement each other for better learning outcomes (e.g., Mayer, 2005). Typical instruction includes using both text and graphic presentations, but it is much less typical to present only graphic instructional items without accompanying text (we found no examples of science textbooks without text). Given the low prevalence of graphic-graphic instruction and dominance of text-text and text-graphic, we focused on the latter two conditions in the present research.

Second, we examined graphic presentations used in text books and for informational classroom décor. Although teachers and students prefer simple graphics, the majority of text books and educational materials are complex with many distracting features (McTigue & Flowers, 2011). We designed our graphics in line with teacher preferences, which meant we reduced the amount of information that could be extracted. Even so, the graphic presentations, by nature, include more elements than a text sentence. As discussed, these elements could be either beneficial to learning and memory (see Carney & Levin, 2002 for review) or could be a distraction (Cook, 2006), especially for lower performing students (Hannus & Hyönä, 1999). Thus, the inherent differences in text and graphic presentations cannot be equated without eliminating validity as an instructional tool. The differences in modality could result in differences in comprehension of each as well as difficulty in integrating across the different representations.

The stimuli were 16 true but novel facts ("stem" facts) that formed 8 pairs of related facts that when integrated with one another could be used to self-derive new factual knowledge, indicating successful integration. For each stem-fact pair, we created a graphic representation of one of the stem facts (e.g., a figure presenting an apple with a label indicating that its seeds are called pips; see Fig. 1). Graphic presentations were designed with ecological relevance, meaning they required extraction and interpretation. Each graphic presentation featured 4 single-word labels. When the graphic representation was integrated with the related text fact, the pairs supported successful response to the "integration" question (e.g., *What do pips contain?*; see Fig. 1).

Stimuli were pilot tested in a university classroom setting, in groups of two to four children, by one of the authors. In pilot testing, we examined whether children successfully answered the integration question with exposure to only one fact from the stem-fact pairs (which would suggest that children had not needed to integrate the two stem facts to be successful) and whether children's memory differed for stem facts presented in text from facts presented as graphics. Overall, children were not successful in generating integration facts when presented with only one, but not both, members of stem-fact pairs. Children were successful recalling (open-ended testing) and recognizing (forced-choice testing) the stem facts, regardless of whether they were presented in text or graphic format.

5.3. Procedure

Children were tested individually in a university laboratory room outfitted with a table, chairs, and small couch. All testing was done by one of two female experimenters (including one of the authors). The procedure involved three phases. In Phase 1, children were presented with the first member of each of the eight stem-fact pairs, followed by a 10-min buffer activity. In Phase 2, children were presented with the second member of each of the eight stem-fact pairs, followed by a 10-min buffer activity. Both phases featured pre-recorded audio presentations. For facts presented through sentences, the audio presented the sentence. For facts presented through graphics, the audio presented the labels on the graphics. Pre-recorded audio was used (rather than live reading), both to equate stimulus-presentation timing across experimenters and to make the procedure more similar to that used in Experiments 2 and 3 (explained below). Across Phases 1 and 2, children were presented with both members of each of eight stem-fact pairs: four pairs in which both members of the pair were presented in text (text-text condition) and four pairs in which one member of the pair was presented in text and the other in graphic form (text-graphic condition). The stimulus pairs were counterbalanced such that each member of each pair was presented in the first and second phases approximately equally often and each pair was presented in text-text and text-graphic presentation format equally often, across participants.

In Phase 3 of the procedure, children's integration and stem fact performance were tested in open-ended format and forced-choice format. Children first were tested on integration questions in open-ended format (i.e., no answer choices were presented). The open-ended questions were read aloud by the experimenter one at a time; the questions also were printed on answer sheets on which children recorded their responses. Following open-ended testing, the integration questions were posed again, this time with three alternative responses, one of which was correct. The experimenter read each forced-choice question aloud, along with the three alternatives. The questions and alternatives also were printed on the answer sheets on which children recorded their responses.

This was followed by open-ended testing of the stem facts (e.g., *What are apple seeds called?*; see Fig. 1), following the same protocol as used for the open-ended testing of the integration questions. Lastly, children completed forced-choice testing of the stem-fact questions. As for the integration questions, the experimenter read aloud each question and the three answer choices, when applicable. For both the integration and stem-fact questions, all forced-choice options had been experienced during the presentation phase, thus eliminating novelty as a controlling variable. The integration and stem-fact questions were presented in one of eight predetermined orders, selected at random (see Fig. 1 for example stem fact and integration questions).

5.4. Scoring

Children were given 1 point for each integration question answered correctly in each of open-ended and forced-choice testing. For each testing format, the maximum possible score was 4 in each of the text-text and text-graphic conditions. Children also were given 1 point for each stem-fact question answered correctly in both open-ended and forced-choice testing, such that the maximum possible score was 8 in each condition for each testing format.

6. Results

Table 1, Panel A provides descriptive statistics on children's performance overall and in the text-text and text-graphic conditions, as measured in open-ended and forced-choice testing. The data met assumptions of normality in measures of skew and kurtosis (range from $|.1|$ to $|1|$ for both measures for all variables).

We first report the results for the integration questions, the key measure for the research questions. For the integration questions, children's levels of memory integration did not differ in the text-text and text-graphic conditions when tested in either open-ended [$t(23) = 0.65, p = .52$] or forced-choice [$t(23) = 1.28, p = .21$] format. Children's overall levels of forced-choice selection of the answers to the integration questions exceeded the chance level of 33 %: $M = 0.74, SD = 0.18$; $t(23) = 11.43, p < .001, d = 2.33$. When examined separately for the two stimulus types, children's forced-choice performance exceeded chance in the text-text condition, $t(23) = 8.83, p < .001, d = 1.80$, and in the text-graphic condition, $t(23) = 6.55, p < .001, d = 1.34$. Thus, children showed evidence of memory integration when they were required to integrate across episodes both of which were presented in text as well as when they were required to integrate across episodes rendered in different presentation formats.

We now report the results of the stem-fact questions for the purpose of examining whether children learned both facts and learned through both formats (text and graphic). The results of analyses of recall (open-ended) and selection (forced-choice) of the stem facts paralleled those for integration facts in that there were no differences between stem facts presented as text and stem fact presented as graphics, $ts(23) = 0.00$ and -0.46, $ps > 0.99$ and $.65$, for open-ended and forced-choice (respectively). Children's overall levels of forced-choice selection of the answers to the stem-fact questions exceeded chance (33 %): $M = 0.73, SD = 0.15$; $t(23) = 12.92, p < .001, d = 2.64$. When examined separately for the two stimulus types, children's levels of forced-choice selection of the answers to the stem-fact questions exceeded chance (33 %) in both the text-text and text-graphic conditions [$ts(23) > 9.67, ps < .001, ds > 1.97$]. Thus children had overall good recall for the stem facts and remembered them equally well across both text and graphic formats. This

Table 1
Mean Proportion Correct by Condition and Test Format for Integration and Stem Facts for Experiment 1(Panel A), Integration Facts for Experiment 2 (Panel B), and Integration and Stem Facts for Experiment 3 (Panel C).

	Test Condition			Test Statistic (Text-Text vs. Text-Graphic conditions)		
Fact Type /Test Format	*M (SD)* Overall	*M (SD)* Text-Text	*M (SD)* Text-Graphic	*t*	*p*	*d*
Panel A: Experiment 1						
Integration Facts						
Open-ended	.27 (.15)	.27 (.21)	.26 (.20)	0.19	0.85	0.04
Forced-choice	.73 (.17)	.77 (.21)	.69 (.26)	1.28	0.21	0.26
Stem Facts						
Open-ended	.38 (.16)	.38 (.19)	.38 (.20)	0.00	> .99	0.00
Forced-choice	.73 (.15)	.72 (.16)	.74 (.21)	−0.46	0.65	−0.09
Panel B: Experiment 2						
Integration Facts						
Open-ended	.13 (.11)	.15 (.20)	.10 (.16)			
Forced-choice	.57 (.19)	.63 (.26)	.52 (.28)	2.61	0.01	0.29
Panel C: Experiment 3						
Integration Facts						
Open-ended	.10 (.15)	.08 (.15)	.12 (.20)			
Forced-choice	.43 (.22)	.40 (.27)	.46 (.27)	2.41	0.02	0.22
Stem Facts						
Forced-choice	.60 (.23)	.57 (.25)	.66 (.45)	2.35	0.02	0.25

indicates that children were able to extract and interpret stem facts depicted through graphic presentations.

7. Discussion

The results of Experiment 1 indicate that, as tested in the present research, 8-year-old children's memory integration performance did not differ when to-be-integrated material was presented through the same format (text-text) compared to different formats (text-graphic). This finding is noteworthy for two reasons. First, previous research has documented children's difficulties extracting information from graphic presentations (e.g., Arteaga et al., 2012; Boling et al., 2004). In the present research, this challenge may have been mitigated by the prerecorded audio highlighting each of the graphic's elements. Importantly, the information that was the target of integration was not differentially highlighted, relative to the other elements of the graphic. The second reason that findings of equivalent performance across presentation formats are noteworthy is that prior research has documented deficits in performance when the surface similarity of related episodes is low, as when the main characters differ across the paired stories, relative to when it is high, as when the same character is featured across stories (Bauer et al., 2012). The difficulty of forming integrated representations between text and graphic presentations has previously been noted (e.g., Mayer, 2005; Paoletti, 2005). In the present research, surface similarity was higher in the text-text relative to the text-graphic condition. Yet children's performance did not differ across conditions. The results of the present experiment thus extend the boundary conditions under which self-derivation through integration is observed.

One motivation for examining memory integration across text and graphics was that this challenge is common in children's classrooms. Especially in STEM disciplines, textual material is often accompanied by material presented in graphic form. The number of graphic presentations included in texts has increased with time and continues to do so (Carney & Levin, 2002). Thus, to fully appreciate their lessons, children must integrate information that is presented in text and graphic form across learning episodes. The purpose of Experiment 2 was to evaluate memory integration of separate episodes of learning in elementary classrooms, under text-text and text-graphic conditions. The extension to the classroom provided the opportunity to test the replicability of our findings and whether they generalized across the different contexts. Consider that the laboratory and classroom contexts differ in the conditions of encoding and tests of retrieval. In the laboratory, children meet one-on-one with an experimenter in a discrete episode of learning that takes place in a novel, distinctive context. In contrast, in the classroom, children participate with their classmates in a socially and cognitively "noisy" environment where they experience multiple hours of instruction in a row, increasing the likelihood of interference. As a result, the classroom context offers more distractions as well as fewer cues to the target learning episode relative to all other episodes of learning that take place in the classroom. Based on these differing conditions, it cannot be assumed that what is observed in the laboratory would also be observed in the classroom.

8. Experiment 2

In Experiment 2, we examined memory integration across separate learning episodes in the classroom, when both facts were presented in text compared to when one was in text and the other through a graphic representation. In Experiment 1, performance did not differ when children were tasked with integrating across two presentational formats (text-graphic) compared to one (text-text). Prior research has found that manipulations in surface similarity do not have the same effect in the classroom as found in the lab (see Bauer et al., 2012, and Bauer et al., 2019). As discussed above, relative to the laboratory, the conditions of the classroom are less controlled and more cognitively challenging. As such, students may be more taxed by the change in presentation format between related episodes. Thus, we predicted that the text-graphic condition would be more challenging than the text-text condition due to facing the challenges of interpreting graphic presentations and integrating across mediums while in the classroom context. Alternatively, the classroom could offer no additional difficulty to performance, resulting in a replication of Experiment 1.

9. Method

9.1. Participants

Eighty-five children (42 girls) aged 8–10 years ($M = 9.34$, $SD = 0.79$) participated. The children were tested in their 3rd (68 % of children) or 4th (32 % of children) grade classrooms in a public school system in the rural southeastern United States. Seventy-eight percent of the parents of the children returned demographic questionnaires on which they reported on the gender and race/ethnicity of their children. Based on parental report, the sample was 21 % Black or African American, 38 % non-Hispanic Caucasian, 32 % Hispanic or Latinx, and 9% multi-racial. In the year in which the data were collected, 86 % of the children in the participating school system were eligible for free or reduced lunch. As described in **Procedure**, the protocol was administered in the children's classrooms. The parents of the 85 children included in the sample had provided written informed consent for their children to participate prior to the testing session. Children in the same classroom whose parents did not provide consent (40 children) were tested in the protocol alongside their classmates, but their data were not analyzed. Given that classroom data often has greater variability and more influences on performance compared to the controlled laboratory environment, we aimed to increase the sample size from Experiment 1. The collaborating school system offered us a population sample, meaning all children in the collaborating grades were invited to participate, the results of which increased our sample size (from 24 to 85, respectfully). At the end of the testing session, all children were given a small token of appreciation for their participation (e.g., pencil, eraser). The procedures were reviewed and approved by the university Institutional Review Board and by the School Board of the participating school system.

9.2. Stimuli

The stimuli were the same as used in Experiment 1. An additional 8 fact pairs were included in the protocol as part of a larger study and are not analyzed in this report.

9.3. Procedure

Testing took place in children's classrooms during a single session. The tasks were administered by two 2-person research teams, each consisting of an experimenter and an assistant. Each team tested an approximately equal number of classrooms. As in Experiment 1, the procedure involved three phases. In the first phase, children were presented with 8 stem facts. The facts were projected on a screen in the front of the classroom and, as in Experiment 1, pre-recorded audio was used to narrate the presentations. For the facts presented as sentences, the audio read the sentence; for the facts presented as graphics, the audio read out the labels. This information was presented via pre-recorded audio to equate stimulus-presentation timing across experimenters and classrooms. The 8 facts were followed by a 10-min unrelated buffer activity. In the second phase, children were presented with 8 more stem facts projected on the screen accompanied by pre-recorded audio. The 8 facts were followed by another 10-min buffer activity. As in Experiment 1, across these phases, children were presented with both members of each of 8 stem-fact pairs. For half of the pairs, both members of the pair were presented in text (text-text condition); for the other half of the pairs, one member of the pair was presented in text and the other was presented in graphic form (text-graphic condition). The order of presentation of the stimulus pairs was counterbalanced such that each pair was presented in each condition (text-text, text-graphic) approximately equally often across participants.

In the third phase of the procedure, children were tested for memory integration through open-ended and forced-choice response to integration questions. They first were tested in open-ended format. Children were given a paper with each of the eight questions and spaces for their responses. Children read the questions silently and recorded their responses in pencil. Children were encouraged to ask for assistance if they needed help reading the questions and experimenters circulated to be available to do so. The questions were presented in one of four different random orders; children seated side-by-side were given different orders. After open-ended testing, the same integration questions were presented in forced-choice format through PowerPoint®. The questions were read aloud by the experimenter and children were asked to select the "best answer" from among three options, one of which was correct (33 % expected by chance). Children used individual response devices to provide responses, which were recorded via TurningPoint® software. The questions were presented in one of four different random orders; each order was presented approximately equally often across classrooms. Due to time constraints, we were unable to test for recall or recognition of stem facts.

9.4. Scoring

Children were given 1 point for each integration question answered correctly in both open-ended and forced-choice testing. For each phase of testing, the maximum possible score was 4 in each of the text-text and text-graphic conditions. Scores were converted to proportions.

10. Results

Descriptive statistics on children's performance in the text-text and text-graphic conditions, in open-ended and forced-choice testing, are provided in Table 1, Panel B. The forced-choice data met assumptions of normality in measures of skew ($\leq |.24|$ for all variables) and kurtosis ($\leq |.81|$ for all variables). However, in open-ended testing, there was a lack of variability in response (near floor levels). Low performance in open-ended testing for memory integration in classroom contexts is not unusual (e.g., Esposito & Bauer, 2017, 2019), but also warrants caution in interpretation of results. For this reason, only forced-choice selection data are analyzed further. In forced-choice testing, in both the text-text and text-graphic conditions, across grade levels, children selected the correct integration facts from among distracters at levels significantly greater than would be expected by chance: $ts > 3.68$, $ps < .001$, $ds > 0.67$. A 2 (Grade) x 2 (Stimulus type) mixed ANOVA revealed a main effect of stimulus type, such that performance in the text-text condition was significantly higher than in the text-graphic condition, $F(1, 81) = 6.52$, $p = .01$, $\eta^2 = .07$. There was not a main effect of grade, $F(1, 81) = 0.06$, $p = .81$, $\eta^2 = .001$, nor a significant Grade x Stimulus type interaction, $F(1, 81) = 0.12$, $p = .73$, $\eta^2 = .002$.

11. Discussion

The results of Experiment 2 indicate that, when they are tested in the classroom setting, children 8–10 years of age experience some challenge to memory integration across separate episodes of learning. In open-ended testing, they had low levels of performance regardless of whether the required integration was of two facts both presented in text or two facts, one of which was presented in text and the other in graphic form. Importantly, performance improved substantially when children were permitted to select the correct response to the integration questions from among distracters. This pattern suggests that the integrations children formed at the first prompt to integrate (i.e., the open-ended integration question) were not sufficiently robust to support open-ended performance. As discussed in Bauer, Cronin-Golomb, Porter, Jaganjac, and Miller (2020) and Bauer, Esposito, and Daly (2020), first integrations may be tentative and overt evidence that separate sources have been integrated may depend on the additional support provided through forced-choice testing. Importantly, forced-choice performance on integration questions cannot be accurately classified as recognition. All of the forced-choice options were familiar from the testing session and were viable answers. Responding above chance required that there be some integration work put in on the child's part to select the correct option. Under forced-choice conditions, there was evidence of successful memory integration in both conditions, although performance in the text-text condition exceeded the text-graphic condition.

A goal of Experiment 2 was to move the laboratory work into the classroom to assess whether the classroom context (and necessary changes in protocol to permit classroom testing) would garner a different pattern of results compared to the laboratory. The results of this classroom study support the hypothesis that integrating across a text and graphic presentation presented in separate episodes of learning would pose challenges in the classroom. This pattern suggests that under the conditions of the present experiment, children either 1) experienced difficulty extracting and encoding the target information from graphics, 2) were challenged to navigate the lower surface level similarity of the text-graphic relative to the text-text condition, or 3) both. Challenges in encoding and extracting the target information are consistent with the difficulty of interpreting graphics in children as well as adults, including teachers (e.g., Arteaga et al., 2012). Alternatively, the demand to integrate across both time and presentation type is a possible source of challenge given the difficulty integrating across these modalities (Mayer, 2005) and more broadly from work on inferential processes that indicates the difficulty in integrating when surface similarities are low (e.g., Bauer et al., 2012; Gentner, 1988). And of course, a combination of these factors also might explain the pattern. The purpose of Experiment 3 was to examine these alternative mechanisms.

12. Experiment 3

In Experiment 3, we again examined memory integration across separate learning episodes in the classroom. However, relative to Experiment 2, we provided additional support for encoding the graphic presentations to better understand whether the challenge of integrating across text and graphic presentation formats is due to difficulty extracting the target information from the graphic or due to the demand of integrating across presentation formats. Encoding of the individual facts is a necessary precursor to integration, although not sufficient. Individuals can encode information without integrating, but you cannot integrate what does not exist in memory. Unlike the text facts, which are directly provided, the graphic fact require an additional level of extraction. Thus, to examine if this is a failure at encoding or at integration, we offered support at encoding to mitigate the difficulty of the extra step required from graphics compared to text facts.

We predicted the additional support at encoding would mitigate the challenge of integrating across two presentation mediums (text and graphic). If the difficulty from Experiment 2 was in extracting and encoding, providing additional support at encoding should increase memory integration performance of text-graphic pairs relative to text-text. However, if the difficulty is stemming not from encoding, but the challenges of integrating across mediums, then additional support at encoding would not improve memory integration performance on text-graphic pairs and the deficit in comparison to text-text pairs would remain.

The collaborating school system offered us more time and a larger population from which to recruit participants in Experiment 3 compared to Experiment 2. As a result, we were able to assess memory for individual facts to assess encoding of facts presented graphically. The sample size increased due to the larger population of students involved in the recruitment. Whereas Experiment 2 was conducted with children in 3rd and 4th grade, Experiment 3 did shift to 4th and 5th grade. Analyses were also conducted with only the 4th grade groups to ensure that the pattern of results for full sample was also replicated in the overlapping grade levels.

13. Method

13.1. Participants

The participants were 154 children (81 females) ages 9–11 years ($M = 10.7$ years, SD = .68) across grade 4 (36 % of children) and grade 5 (64 % of children). All children attend the same school in a rural community in the southeastern United States and were drawn from the same population as Experiment 2. No children from Experiment 2 were included in Experiment 3. Ninety-eight percent of the parents of the children returned demographic questionnaires on which they reported on the gender and race/ethnicity of their children. Based on parental report, the sample was 34 % Black or African American, 34 % non-Hispanic Caucasian, 27 % Hispanic or Latinx, and 5% multi-racial. Consent forms were again sent out via parent communication folders and only the data from children whose parents/guardians returned signed consent forms were included in the analysis (approximately 46.71 % of the population). Children in the same classroom whose parents did not provide consent were again tested in the protocol alongside their classmates, but their data were not analyzed. Children, families, and school personnel were compensated as they were in Experiment 2. The university institutional review board and the participating school system school board reviewed and approved all protocols and procedures prior to the start of the study.

13.2. Stimuli

As in Experiment 1 and Experiment 2, the stimuli were 8 novel "stem" fact pairs, for a total of 16 facts (eight additional fact pairs were included as part of a larger study but not included in this analysis). The stimuli were updated from Experiment 2 to reflect the science curriculum of students in grades 4 and 5, changing some facts to which 5th grade students may already have been exposed. Graphic presentations consisted of a title and 3–4 labels.

Stimuli were pilot tested in fifth grade elementary classrooms in the community (the collaborating school for the main data collection was not included in pilot testing), in groups of five to eight, by three of the authors. Pilot testing was used to assess the novelty for children in the target age range. Pilot testing also determined that both stem facts were necessary to reliably produce or select the integration facts. That is, children in a control condition in which they only received one of the two stem facts did not reliably generate the integration facts or select the answers in forced-choice testing above chance performance. In contrast, children who received both of the to-be-integrated stem facts generated the integration facts more frequently and selected the correct answer in forced-choice testing with above chance performance.

13.3. Procedure

The testing procedure followed the three phases of Experiment 2 with two key changes designed to support and assess extraction of the target information from the graphic. First, after all graphic labels were presented, the researcher asked a question about the target information depicted in the graphic presentation. This question was the intervention to provide support for extracting and encoding the target fact, as per the goal for Experiment 3. A child was called on to answer the question and was corrected by another child if the question was answered incorrectly. If the class did not reach the correct answer (rare), the correct answer was provided by the researcher. This intervention was intended to bring the graphics to a similar level of transparency as the text facts which were presented on screen and read aloud. All stem facts were presented via Power Point® and prerecorded audio were played through speakers.

Second, after the conclusion of open-ended and forced-choice testing of integration facts in Phase 3, children were asked stem-fact recall questions in forced-choice format. Due to time constraints, children were asked 8 stem facts from the text-graphic pairs, representing half of the total stem facts. There were several constraints on counterbalancing that lead to a different number of stem fact questions being asked in text and graphic format across the four randomized orders of presentation (from 1 to 3 in graphic form; average of 2 across the four different presentation orders). The four presentation orders were used equally often across classrooms.

13.4. Scoring

Children were given 1 point for each integration-fact question answered correctly in each of open-ended and forced-choice testing. For each testing format, the maximum possible score was 4 in each of the text-text and text-graphic conditions. Children also were given 1 point for each stem-fact question answered correctly in forced-choice testing. Given the variable number of stem-fact questions representing facts in text and graphic format across conditions, scores were converted to proportions based on the number of questions in each category. Proportion correct ranged from 0 to 1.

14. Results

Descriptive statistics on children's performance in the text-text and text-graphic conditions, in open-ended and forced-choice testing for integration and stem fact questions, are provided in Table 1, Panel C. As in Experiment 2 and other previous classroom studies (e.g., Esposito & Bauer, 2017, 2019), open-ended performance was low and lacked the variability needed for further analysis. Thus, the remaining analyses proceed with forced-choice data only.

We began by analyzing the memory integration performance. Children's performance in forced-choice testing met assumptions of normality in terms of skewness and kurtosis (skew $\leq |.20|$ for all variables; kurtosis $\leq |.77|$ for all variables). In addition, both text-text ($n = 135$, $M = .40$, $SD = .27$) and text-graphic ($n = 135$, $M = .46$, $SD = .27$) force-choice performance was significantly above chance (.33), $t(134) = 2.91$, $p = .004$ and $t(134) = 5.73$, $p < .001$, respectively. A 2 (Grade: 4th, 5th) x 2 (Stimulus type: text-text, text-graphic) mixed factor ANOVA (with repeated measures on stimulus type) revealed a main effect of stimulus type $F(1,133) = 4.68$, $p = .03$, $\eta_p^2 = .03$ showing that text-graphic performance was significantly higher than text-text performance. The effect size was small, however. There was no main effect of grade $F(1,133) = .03$, $p = .87$, $\eta_p^2 < .001$, nor a Grade x Stimulus type interaction $F(1, 133) = 0.36$, $p = .55$, $\eta^2 = .003$.

We next analyzed memory for the individual stem facts with each pair and across format types (text and graphic). Both skew and kurtosis were again at acceptable levels (skew $= -.28$; kurtosis $= -.45$). Stem fact performance when the fact appeared in graphic presentation, in text presentation, and overall stem fact performance were all significantly above chance (.33), $t(138) = 8.79$, $p > .001$, $t(138) = 11.06$, $p > .001$, and $t(138) = 13.58$, $p > .001$, respectively. A 2 (Grade: 4^{th}, 5^{th}) x 2 (Stimulus type: text, graphic) mixed factor ANOVA (with repeated measures on stimulus type) showed a main effect of presentation format ($F(1,137) = 4.78$, $p = .03$, $\eta_p^2 = .03$) indicating performance was higher when the fact was presented as a graphic than when it was presented through text. As above, the effect size was small. There was also a main effect of grade ($F(1,137) = 6.34$, $p = .01$, $\eta_p^2 = .04$) indicating that across presentation formats, performance in Grade 4 ($M = .67$, SD $= .23$) was higher than Grade 5 ($M = .55$, SD $= .22$). There was no Grade x Stimulus type interaction ($F(1,137) = .05$, $p = .83$, $\eta_p^2 < .001$).

Finally, we note that the children who participated in the present Experiment 3 were older ($M = 10.7$ years) compared to those in Experiment 2 ($M = 9.34$). Self-derivation performance is better in older compared to younger children (Bauer & Larkina, 2017). To determine whether age differences across the experiments could explain the different patterns of performance, we examined the performance of the Grade 4 children in each experiment, thus equating the samples for age. The pattern of results held such that 4th graders in Experiment 2 had significantly lower performance on text-graphic pairs compared to text-text pairs in forced-choice testing, $t(26) \geq 2.10$, $p \leq .045$. Among the 4th graders in the present experiment the difference was reversed such that performance on text-graphic pairs was nominally higher than on text-text pairs in forced-choice ($t(49) = -1.16$, $p = .24$) testing.

15. Discussion

The goal of Experiment 3 was to explore whether the challenge of integrating text with graphic presentation across separate episodes of learning observed in Experiment 2 was due to difficulty extracting and encoding the target fact embedded within the graphic or difficulty integrating across two different mediums. Consistent with the first explanation, the support extracting the target fact from the graphic at encoding improved integration performance in the text-graphic condition to the extent that there was no longer a deficit in comparison to the text-text condition. As was observed in Experiment 2 and other classroom studies of memory integration (e.g., Esposito & Bauer, 2017, 2019), children's open-ended integration performance was notably low. However, when tested with forced-choice integration questions, children were successful at selecting the correct answer to the integration question. Most importantly, Experiment 3 demonstrates that children can successfully integrate across text and graphic stimuli across separate episodes of learning when their encoding of the graphic representation is supported.

The goal of Experiment 3 was to support the encoding of facts presented through graphics to the extent that they were encoded as well as the facts presented through text, eliminating difficulties in interpretation and extraction as a cause for lack of memory integration. Stem fact performance indicates that the supports provided were successful. Prompting children with a question to ensure accurate encoding of the target information from the graphic stimuli significantly increased stem fact encoding compared to stem facts presented through text. Sufficient encoding of the stem facts presented as graphics seemingly translated into better integration performance: in forced-choice testing, performance was significantly higher in the text-graphic condition compared to the text-text condition (the same pattern was observed in open-ended testing, but near-floor performance restricts interpretation). This pattern of results is the opposite of what was found in Experiment 2. Direct comparison of same-grade children in Experiments 2 and 3 indicated that age differences could not account for the pattern. Instead, this suggests that the challenge of integrating across these specific modalities was reduced when the demand to extract target information was supported. Importantly, the graphics included in this experiment were simplified compared to typical textbooks and the target information was brought specifically to the attention of the students. When this was not done, in Experiment 2, text-graphic integration performance suffered compared to text-text integration.

16. General discussion

In the present research, we examined memory integration of separate episodes of new learning as a function of differences in the presentation formats (text-text vs. text-graphic) between related facts across both laboratory and classroom contexts. The demand to integrate across text and graphic presentations increases as children move from "learning to read" to "reading to learn" in the middle-elementary grades. Yet, the difficulty in extracting information from graphic presentations and integrating across different presentation formats could pose a challenge to memory integration, indicating children also first need to learn to read *graphics* before reading *graphics* to learn. In three experiments, children were presented with pairs of facts that, when integrated with one another, could support successful responses to integration questions. Across all studies, we compared performance in a text-text condition with that in a text-graphic condition. In Experiment 1, children were tested individually in the laboratory and showed no difference in performance between the two conditions. In Experiment 2, children were tested in their classrooms. In the classroom context, the hypothesized costs of extracting information from a graphic and integrating across different presentation formats (text-graphic) emerged. Children had significantly higher performance in the text-text compared to the text-graphic condition that required interpreting a graphic and integrating across presentation formats. In Experiment 3, children were again tested in classrooms, but with support for extracting the target information from the graphic presentations. With the additional support at encoding, performance in the text-graphic condition was significantly better than the text-text condition. Direct comparison of same-grade children from Experiments 2 and 3 indicated that the different pattern of performance was not due to the overall older age of children in Experiment 3 relative to Experiment 2. Instead, the results of Experiment 3 indicate that with support at encoding, children are able to integrate across different presentation mediums and across separate episodes of learning.

We predicted the text-graphic presentation format would be challenging for two main reasons. First, children have difficulty understanding and extracting the intended meaning from graphic presentations (Arteaga et al., 2012; Boling et al., 2004). Second, integration across two different presentation formats (text and graphic) represents a low surface similarity condition that is associated with lower performance across productive processes (Bauer et al., 2012; Esposito & Bauer, 2019; Gentner, 1988). However, in Experiment 1, performance was generally high across both conditions in both stem fact recall (indicating the children extracted meaning from the graphics) and integration (indicating low costs to integrating across modalities). In the classroom context of Experiment 2, open-ended performance was low in both conditions; forced-choice performance was higher and the expected cost associated with navigating graphics and different presentation formats emerged. When extracting the target information from the graphic was supported in Experiment 3, stem facts presented in graphic form showed better selection than those presented in text form. Similarly, integration performance in the text-graphic condition was higher than the text-text condition. The added support at encoding effectively mitigated the challenges of integration across presentation modalities that was observed in Experiment 2.

It is possible that the classroom context exacerbates the initially predicted challenges of two different formats of presentation across lessons. Although the necessary differences in the protocols between the laboratory and the classroom preclude direct comparison, the pattern of higher performance in the laboratory compared to the classroom is also reflected in other memory integration work (e.g., Esposito & Bauer, 2017). There is likely more interference in the classroom where children have engaged in many hours of instruction compared to the laboratory setting, where they have spent limited time. The "cognitively crowded" conditions of the classroom make each episode less distinct. This could have added extra challenge to extracting the target information from a graphic, recognizing the relevance to a fact presented in text, and integrating the two to form new information. The classroom context could also increase the role of transfer appropriate processing (Morris, Bransford, & Franks, 1977). The integration questions were all in text format, which may have provided an advantage to text-text fact pairs, relative to text-graphic pairs. In Experiment 3, when encoding of the target information was supported, children were able to overcome the challenge of integrating across time and medium in the classroom as well as any additional challenge of transfer appropriate processing.

Another possible interpretation of the differences in the pattern of results between studies is the demographic distribution of the participants themselves. The participants in Experiment 1 were predominantly of middle and upper socio-economic demographics whereas the majority of participants in Experiments 2 and 3 were of relatively lower socioeconomic status (SES). SES is a predictor of vocabulary (e.g., Hoff, 2003), and vocabulary consistently emerges as a unique predictor of self-derivation through integration performance (Esposito & Bauer, 2018; Varga et al., 2019). Therefore, differences in verbal ability could contribute to differences in performance in the text-text condition across the studies. Students' performance in Experiments 2 and 3 was not unlike previous studies in similar classroom contexts with comparable demographic distributions (Esposito & Bauer, 2017, 2019; Varga et al., 2019). Thus, both the classroom context and participant demographics could have contributed to differing patterns of results between studies. The results underscore the need for laboratory studies to be tested for replication in ecologically valid contexts, such as the classroom.

The research informs our understanding of navigating changes in surface similarity in productive processes. Navigating low surface similarity conditions and integrating across texts and graphics has been considered challenging (e.g., Gentner, 1988; Mayer, 2005). Prior research on memory integration in the laboratory indicates that there is a cost to low surface similarity in features of related episodes of new learning (Bauer et al., 2012). This cost was not observed in Experiments 1 or 3, extending the boundary of memory integration to include integration across different presentation formats. These results require updating our understanding of the conditions in which low surface similarity hinders memory integration. In isolation, one could interpret the results of Experiment 1 as evidence that graphic presentations facilitate memory in such a way to overcome the low surface similarity conditions of cross-presentation formats. Certainly, there is ample evidence that pictures are more easily remembered than text (e.g., Glaser, 1992; Paivio, Rogers, & Smythe, 1968). However, this interpretation must be tempered in light of the results of Experiments 2 and 3. From the experiments in the classroom, we learn that graphic presentations must be carefully constructed and their interpretation supported in order to benefit integration.

16.1. Implications

Given the known challenge of extracting the intended meaning from graphic presentations, it may be that the children who participated in Experiment 2 did not encode the intended information. For example, they may have encoded that the apple has 4 parts, but not that the seeds are called pips (see Fig. 1). Children struggle even in text-only learning to encode the target information rather than peripheral information (for review, see Garner, Brown, Sanders, & Menke, 1992). The "seductive details" of graphic presentations could be even more salient, leading to lower encoding of target information. This indicates that children in classroom settings may require more support in extracting the target information from graphic presentations. The results of Experiment 3 provide evidence of this interpretation. Understanding the most successful types of supports is an interesting direction for future research. For example, perhaps repetition of material at encoding would suffice. It may also be that the teacher plays an important role underscoring the importance of target information.

In the reported studies, we made stimulus decisions based on ecological relevance. Graphics found in textbooks and other classroom materials are often complex with distracting elements (e.g., McTigue & Flowers, 2011). The graphics included in these experiments were simplified based on teacher preference compared to much of what is presented in textbooks. To understand how students learn through the integration of graphics and text, we ensured that the graphic stimuli contained labels typical of educational graphics. The target information was brought specifically to the attention of the students in Experiment 3. When this was not done (Exp. 2), integration performance suffered compared to text-text integration. Thus, the typical graphic presentation depicted in many textbooks may hinder integration, rather than facilitate it, without sufficient support at encoding.

The present research represents the first test of memory integration of separate episodes of new factual knowledge provided through different presentation formats. The results contribute generally to the conditions that support inferential processes and specifically to memory integration of science curriculum across text and graphic presentations. Overall, children were successful in integrating new information across presentation formats in both the laboratory and classroom contexts. In the classroom, memory integration of new factual knowledge was difficult. Performance in forced-choice testing was higher. The expected cost to memory integration of information presented across two different presentation formats was found when comparing text-graphic to text-text performance in the classroom context, but was not evident when support extracting the target graphic information was provided at encoding. Thus, under the conditions of the classroom context, integrating across two presentation formats presented an additional challenge that was mitigated with support of graphic comprehension. Outside the laboratory, learning episodes take place through multiple formats (books, television, lectures, exploration, etc.). The present research indicates that there are conditions that support memory integration across multiple presentation formats and extend the generalizability of integration to contexts and conditions that are ecologically relevant.

16.2. Limitations

The present research is not without limitations; we note three in particular. First, Experiment 1 could not be directly compared to Experiments 2 and 3 due necessary differences in protocols as well as demographic differences between the populations. However, the different contexts revealed important limits on the generalizability of findings from the laboratory to an important ecological niche of the child: the classroom. Second, time restrictions limited our ability to test stem-fact recall in Experiment 2. Thus, we were precluded from assessing whether lower levels of performance in Experiment 2 might be due to lower levels of stem-fact recall by the children in the second study. This left the question of whether performance was due to poor encoding of the graphic presentations or difficulty integrating across the presentation formats. This was, however, resolved in Experiment 3 when integration performance improved with the addition of encoding support for the graphic presentations. Third, the present research focused on elementary school aged children and the results do not necessarily generalize to younger or older age groups. These limitations provide suggestions for directions for future research.

16.3. Conclusions

In conclusion, the present research addressed the question of whether providing information across different presentation formats would pose a challenge to the productive process of memory integration across separate episodes of learning in laboratory and school contexts. We learned that children in the laboratory were successful in both a same-format and a different-format condition. Children in the classroom were also successful in both conditions, as tested in forced-choice selection. In Experiment 2, they showed a cost to integration across two different presentation formats compared to the same-format condition. This cost was mitigated in Experiment 3 when support was provided for extracting the target information from the graphic presentations. The findings simultaneously support the conclusion that high surface similarity across presentation formats can benefit integration of new factual knowledge while also providing evidence that surface similarity is but one contributor to performance.

Acknowledgements

Support for this research was provided by NICHDHD067359, by NSFBCS1528091, and by IESR305A150492 to Patricia J. Bauer, and Emory College of Arts and Sciences. The authors also thank the participating school system for Experiments 2 and 3, members of the *Memory at Emory* laboratory group, and the children and families who participated in this research.

References

Arteaga, P., Batanero, C., Contreras, J. M., & Cañadas, G. R. (2012). Understanding statistical graphs: A research survey. *Boletín de Estadística e Investigación Operativa, 28*(3), 261–277.

Bauer, P. J. (2012). The life I once remembered: The waxing and waning of early memories. In D. Bensten, & D. C. Rubin (Eds.), *Understanding autobiographical memory: Theories and approaches* (pp. 205–225). Cambridge, UK: Cambridge University Press.

Bauer, P. J., & Jackson, F. L. (2015). Semantic elaboration: ERPs reveal rapid transition from novel to known. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(1), 271–282. https://doi.org/10.1037/a0037405

Bauer, P. J., & Larkina, M. (2017). Realizing relevance: The influence of domain-specific information on generation of new knowledge through integration in 4- to 8-year-old children. *Child Development, 88*(1), 247–262. https://doi.org/10.1111/cdev.12584

Bauer, P. J., & San Souci, P. (2010). Going beyond the facts: Young children extend knowledge by integrating episodes. *Journal of Experimental Child Psychology, 107*(4), 452–465. https://doi.org/10.1016/j.jecp.2010.05.012

Bauer, P. J., King, J. E., Larkina, M., Varga, N. L., & White, E. A. (2012). Characters and clues: Factors affecting children's extension of knowledge through integration of separate episodes. *Journal of Experimental Child Psychology, 111*(4), 681–694. https://doi.org/10.1016/j.jecp.2011.10.005

Bauer, P. J., Blue, S. N., Xu, A., & Esposito, A. G. (2016). Productive extension of semantic memory in school-aged children: Relations with reading comprehension and deployment of cognitive resources. *Developmental Psychology, 52*(7), 1024–1037. https://doi.org/10.1037/dev0000130

Bauer, P. J., Cronin-Golomb, L. M., Porter, B. M., Jaganjac, A., & Miller, H. E. (2020). Integration of memory content in adults and children: Developmental differences in task conditions and functional consequences. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000996. Advance online publication.

Bauer, P. J., Esposito, A. G., & Daly, J. J. (2020). Self-derivation through memory integration: A model for accumulation of semantic knowledge. *Learning and Instruction, 66*, Article 101271. https://doi.org/10.1016/j.learninstruc.2019.101271

Benson, P. J. (1997). Problems in picturing text: A study of visual/verbal problem solving. *Technical Communication Quarterly, 6*(2), 141–160. https://doi.org/10.1207/s15427625tcq0602_2

Bergey, B. W., Cromley, J. G., & Newcombe, N. S. (2015). Teaching high school biology students to coordinate text and diagrams: Relations with transfer, effort, and spatial skill. *International Journal of Science Education, 37*(15), 2476–2502. https://doi.org/10.1080/09500693.2015.1082672

Boling, E., Eccarius, M., Smith, K., & Frick, T. (2004). Instructional illustrations: Intended meanings and learner interpretations. *Journal of Visual Literacy, 24*(2), 185–204. https://doi.org/10.1080/23796529.2004.11674612

Brown, A. L. (1982). Learning and development: The problems of compatibility, access and induction. *Human Development, 25*(2), 89–115. https://doi.org/10.1159/000272791

Bryant, P. E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature, 232*(5311), 456–458. https://doi.org/10.1038/232456a0

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction, 20*(2), 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review, 14*(1), 5–26. https://doi.org/10.1023/A:1013176309260

Coleman, J. M., & Dantzler, J. A. (2016). The frequency and type of graphical representations in science trade books for children. *Journal of Visual Literacy, 35*(1), 24–41. https://doi.org/10.1080/1051144x.2016.1198543

Cook, M. P. (2006). Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education, 90*(6), 1073–1091.

Cromley, J. G., Bergey, B. W., Fitzhugh, S., Newcombe, N., Wills, T. W., Shipley, T. F., et al. (2013). Effects of three diagram instruction methods on transfer of DIAGRAM comprehension skills: The critical role of inference while learning. *Learning and Instruction, 26*, 45–58. https://doi.org/10.1016/j.learninstruc.2013.01.003

Devetak, I., & Vogrinc, J. (2013). The criteria for evaluating the quality of the science textbooks. *Critical analysis of science textbooks* (pp. 3–15). New York, NY: Springer.

Esposito, A. G., & Bauer, P. J. (2017). Going beyond the lesson: Self-generating new factual knowledge in the classroom. *Journal of Experimental Child Psychology, 153*, 110–125. https://doi.org/10.1016/j.jecp.2016.09.003

Esposito, A. G., & Bauer, P. J. (2018). Building a knowledge base: Predicting self-derivation through integration in 6- to 10-year-olds. *Journal of Experimental Child Psychology, 176*, 55–72. https://doi.org/10.1016/j.jecp.2018.07.011

Esposito, A. G., & Bauer, P. J. (2019). Self-derivation through memory integration under low surface-similarity: The case of multiple languages. *Journal of Experimental Child Psychology, 187*, Article 104661. https://doi.org/10.1016/j.jecp.2019.07.001

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses USING G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction, 20*(3), 216–224. https://doi.org/10.1016/j.learninstruc.2009.02.021

Garner, R., Brown, R., Sanders, S., & Menke, D. J. (1992). Seductive details" and learning from text. In K. A. Renninger, S. Hidi, A. Krapp, & A. Renninger (Eds.), *The role of interest in learning and development* (pp. 239–254). New York, NY: Psychology Press.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy*. *Cognitive Science, 7*(2), 155–170. https://doi.org/10.1207/s15516709cog0702_3

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59*(1), 47–59. https://doi.org/10.2307/1130388

Gentner, D., & Smith, L. A. (2013). *Analogical learning and reasoning*. Oxford Handbooks Online. https://doi.org/10.1093/oxfordhb/9780195376746.013.0042

Glaser, W. R. (1992). Picture naming. *Cognition, 42*(1–3), 61–105. https://doi.org/10.1016/0010-0277(92)90040-O

Goswami, U. (2011). Inductive and deductive reasoning. In U. Goswami (Ed.), *Childhood cognitive development* (pp. 399–419). Oxford, UK: Wiley-Blackwell.

Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low-and high-ability children. *Contemporary Educational Psychology, 24*(2), 95–123.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*(6), 2797–2822. https://doi.org/10.1111/j.1467-8624.1996.tb01889.x

Lesnick, J., Goerge, R. M., Smithgall, C., & Gwynne, J. (2010). *Reading on grade level in third grade: How is it related to high school performance and college enrollment?* Chicago, IL: Chapin Hall at the University of Chicago.

Levin, J. R., Anglin, G. J., Carney, R. N., Willows, D. M., & Houghton, H. A. (1987). On empirically validating functions of pictures in prose. In D. M. Willows, & H. A. Houghton (Eds.), *The psychology of illustration* (pp. 51–86). New York, NY: Springer.

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*(4), 715–726. https://doi.org/10.1037/0022-0663.82.4.715

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19*(6), 668–682. https://doi.org/10.1016/s0022-5371(80)90355-2

McTigue, E. M., & Flowers, A. C. (2011). Science visual literacy: Learners' perceptions and knowledge of diagrams. *The Reading Teacher, 64*(8), 578–589.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*(5), 519–533.

Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science, 11*, 137–138.

Paoletti, G. (2005). Writing-to-learn and graph-drawing as aids of the integration of text and graphs. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Effective learning and teaching of writing* (pp. 587–597). Dordrecht: Springer.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., et al. (2007). *Organizing instruction and study to improve student learning. IES practice guide. NCER 2007-2004*. National Center for Education Research.

Preston, A. R., Shrager, Y., Dudukovic, N. M., & Gabrieli, J. D. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus, 14*(2), 148–152. https://doi.org/10.1002/hipo.20009

Roberts, K. L., Norman, R. R., & Cocco, J. (2015). Relationship between graphical device comprehension and overall text comprehension for third-grade children. *Reading Psychology, 36*(5), 389–420. https://doi.org/10.1080/02702711.2013.865693

Schnotz, W. (2014). *The Cambridge handbook of multimedia learning: Integrated model of text and picture comprehension*.

Siegler, R. S. (1989). Mechanisms of cognitive development. *Annual Review of Psychology, 40*(1), 353–379. https://doi.org/10.1146/annurev.ps.40.020189.002033

Varga, N. L., & Bauer, P. J. (2013). Effects of delays on 6-year-old children's self-generation and retention of knowledge through integration. *Journal of Experimental Child Psychology, 115*(2), 326–341. https://doi.org/10.1016/j.jecp.2013.01.008

Varga, N. L., & Bauer, P. J. (2017). Young adults self-derive and retain new factual knowledge through memory integration. *Memory & Cognition, 45*(6), 1014–1027. https://doi.org/10.3758/s13421-017-0711-6

Varga, N. L., Esposito, A. G., & Bauer, P. J. (2019). Cognitive correlates of memory integration across development: Explaining variability in an educationally relevant phenomenon. *Journal of Experimental Psychology: General, 148*(4), 739–762. https://doi.org/10.1037/xge0000581

Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental Psychology, 12*(4), 289–297. https://doi.org/10.1037/0012-1649.12.4.289